# The Translator's Tool Box

## A Computer Primer for Translators

### by Jost Zetzsche

**Version 9, December 2010**

# Table of Contents

# Introduction

As a technical translator and localization consultant, I've been continually surprised at the lack of technical expertise and knowledge about software tools among many translators and project managers. I've seen countless hours wasted on tasks that could have been done automatically or in a fraction of the time. And as an editor, I've often struggled to improve texts that were translated with an adequate level of linguistic or subject-matter expertise, but whose quality was sub-par because the translator didn't know how to use the necessary tools or formats.

At some point after it became common for translators to use computers for their work, it seems that many of us became convinced that we were really not smart (read: *technical*) enough to become proficient computer users. The irony is that many of us translate highly technical and complex subject matter every day. There is no lack of intelligence among us—merely a prevailing not-smart-enough-for-computers fallacy that we have bought into.

It is time to adopt a new paradigm for our profession: Not only is it acceptable to use computers well—it is critical to our success.

## The Purpose of This Book

This book attempts to bridge the gap between our technical paralysis and our potential. It provides translators with an overview of the technical issues that most of us are regularly confronted with. It starts with basic activities such as zipping and unzipping files and file management in *Windows*, and then moves on to more complex tasks such as working with desktop publishing applications and conversion issues.

If you're completely comfortable with your software equipment and your level of technical translation expertise, you probably won't need to read this book. But if you feel that you could use your computer time more efficiently, I encourage you to continue reading.

I have focused primarily on a *Windows*-based environment not because I think that *Windows* is better than any other operating system, but because it is the most commonly used environment among translators and certainly the best supported among translation-related tools.

*It is important to remember that Windows operating systems can run on either a traditional Windows PC or any other computer that supports Windows through a virtual environment, including Intel Macintosh computers.*

*Oh, and in case you wondered about the little fellow that accompanies all the notes and can be admired on the cover of this book. This is Jeromobot, the patron saint of the modern translator!*

The specific product names that I feature in the tutorials are not necessarily a reflection of any favorable judgment on these in comparison with other competing products. Instead, they represent either the most commonly used products or the ones that I am most familiar with.

## How to Read This Book

This book is not a tutorial on what to do or not to do with your computer. Instead, it is a list of suggestions based on what has worked well (or not so well) for me. Think of it as a tool box from which you can choose the tools that may be appropriate for you at any certain time.

The comprehensive index at the end of the book will help you to quickly find the information you need. To help you find some of the "tips and tricks" that I list throughout the book, I have preceded the alphabetical index with a "How to" section. Because you may not know exactly what you are looking for, I encourage you to actually read or at least scan through the book.

Finally, read with courage and creativity! Computers and the plethora of specialized software programs are powerful tools for translation, tools that are more accessible and affordable than ever before. And with this tool box at your disposal, the only limits to your craftsmanship as a translator are the boundaries you set for yourself.

## How to Read The Updated Version of This Book

If you've used this book before, you may be primarily interested in the many changes to this latest edition. I have posted an update log at www.internationalwriters.com/toolbox/corrections.txt that contains detailed information on what has changed between different versions.

Oh, and one more thing. This book is designed in a printer-friendly format, so you can certainly go ahead and do that. But please consider the environment before you contemplate printing the whole tome!

# Operating Systems

*The most important program that runs on a computer is the "operating system." Operating systems provide a software platform on top of which other programs, called "application programs," can run. Because the application programs must be written to run on top of a particular operating system, your choice of operating system largely determines the applications you can run.*

Software programs tend to inspire either a visceral love or hate response among their users, and this is probably nowhere more true than with operating systems. Generally, *Mac* and *Linux* users love their operating systems (and shake their heads about everyone else, especially *Windows* users). *Windows* users often dislike their platform, but they use it because many of the programs they want or have to use are not available for the other operating systems, or they would have to be run on a relatively expensive PC-emulator program on top of a *Mac* or *Linux* platform. With the arrival of *Windows 2000* and its successors, however, *Windows* users should not feel quite as unfortunate anymore. These versions are infinitely more stable than any of their predecessors, and they have many features built into their systems that make it easier for translators to use.

## The Benefits of Windows 2000 and Higher

These operating systems are essentially upgrades of the *Windows NT* series—the operating systems widely used by corporations with large networks—rather than the "home" series (*Windows 95*, *98*, and *ME*). For you as a user, this means first of all an increase in stability. More importantly, there is a significant difference in the way that the *NT* kernel (the central module that runs the operating systems) runs programs: in case of an application failure, the operating system remains stable, resulting in (believe it or not, dear *Windows 95/98/ME* user) no reboots after application crashes. While this is an improvement for all users, *Windows 2000*, *XP, Vista*, and *7*'s method of

handling Unicode and non-Western languages is of particular benefit for translators. It is possible to process non-Western text, including double-byte (Japanese, Chinese, Korean) or bi-directional text (Arabic, Hebrew), without any additional programs or utilities.

*Unicode is a standard for representing characters as integers. Because it is possible to represent all characters of all known languages with Unicode, there is no longer a need for different character sets (or "code pages") between languages. See also page 156.*

*Multilingual display thanks to Unicode:*

100 000+

Català · Deutsch · English · Español · Français · Italiano · Nederlands · 日本語 · Norsk (bokmål) · Polski · Português · Русский · Română · Suomi · Svenska · Volapük · 中文

10 000+

العربية · Asturianu · Azərbaycan / آذربایجان دیلی · বাংলা · বিষ্ণুপ্রিয়া মণিপুরী · Bosanski · Brezhoneg · Български · Česky · Cymraeg · Dansk · Eesti · Ελληνικά · Esperanto · Euskara · فارسی · Galego · 한국어 · हिन्दी · Hrvatski · Ido · Bahasa Indonesia · Íslenska · עברית · Basa Jawa · ქართული · Kurdî / كوردی · Latina · Lumbaart · Latviešu · Lëtzebuergesch · Lietuvių · Magyar · Македонски · मराठी · Bahasa Melayu · नेपाल भाषा · Norsk (nynorsk) · Nnapulitano · Occitan · Piemontèis · Plattdüütsch · Shqip · Sicilianu · Simple English · Sinugboanon · Slovenčina · Slovenščina · Српски · Srpskohrvatski / Српскохрватски · Basa Sunda · Tagalog · தமிழ் · తెలుగు · ไทย · Türkçe · Українська · Tiếng Việt

1 000+

Afrikaans · Alemannisch · ኦሮምኛ · Aragonés · Armãneashce · Arpitan · Kreyòl Ayisyen · Bân-lâm-gú · Basa Banyumasan · Беларуская · Беларуская (Акадэмічная · Тарашкевіца) · भोजपुरी · Boarisch · Corsu · Чăваш · Deitsch · ދިވެހި · Eald Englisc · Føroyskt · Frysk · Furlan · Gaeilge · Gàidhlig · 古文 / 文言文 · Հայերէն · Hornjoserbsce · Ilokano · Interlingua · Ирон æвзаг · ಕನ್ನಡ · Kapampangan · Kaszëbsczi · Kernewek · Ladino / לאדינו · Ligure · Limburgs · Lingála · മലയാളം · Malti · Māori · Mohron · Nāhuatlahtōlli · Nedersaksisch · नेपाली · Nouormand · Novial · O'zbek · पालि · Pangasinán · پنجابی · Қазақша · Ripoarisch · Rumantsch · Runa Simi · संस्कृतम् · Sámegiella · Scots · Kiswahili · Tarandíne · Tatarça · Точикй · Lea faka-Tonga · اردو · Vèneto · Võro · Walon · West-Vlams · Winaray · ייִדיש · 粵語 · Yorùbá · Zazaki · Žemaitėška

100+

ܐܪܡܝܐ · Avañe'ẽ · Авар · Aymara · Bamanankan · Башҡорт · Bikol Central · چىنگە · Chavacano de Zamboanga · Diné Bizaad · Dolnoserbski · Emigliàn–Rumagnòl · Eʋegbe · Gaelg · گیلکی · ગુજરાતી · ᎦᏬᏗ · Hak-kâ-fa / 客家話 · ʻŌlelo Hawaiʻi · Igbo · ᐃᓄᒃᑎᑐᑦ / Inuktitut · Interlingue · कॉशुर / كٲشُر · Kongo · Кыргызча · ພາສາລາວ · lojban · Malagasy · Māzǝrūni / مازرونی · Mìng-dĕ̤ng-ngṳ̄ · Молдовеняскэ · ߒߞߏ · Ekakairũ Naoero · Nēhiyawēwin / ᓀᐦᐃᔭᐍᐏᐣ · Norfuk / Pitkern · Нохчийн · ଓଡ଼ିଆ · Afaan Oromoo · অসমীয়া · ਪੰਜਾਬੀ / پنجابی · Papiamentu · Qırımtatarca · Romani / रोमानी · Kinyarwanda · Gagana Sāmoa · Sardu · Seeltersk · සිංහල · Словѣньскъ · Af Soomaali · Reo Tahiti · Taqbaylit · Tetun · ᏣᎳᎩ · Tok Pisin · ᎤᎭᏫ · Türkmen · Удмурт · ئۇيغۇرچە / Uyghur · Tshivenḓa · Wollof · 吴语 · Zeêuws · isiZulu

Other languages · Weitere Sprachen · 他の言語 · Kompletna lista języków · 其他语言 · Другие языки · Aliaj lingvoj · 다른 언어 · Ngôn ngữ khác

There are even multi-language versions of these operating systems that allow you to switch the user interface between 24 (*2000*), 33 (*XP*), 30 (*Vista*), and 36 (*7*) languages. While the multi-lingual versions for *Windows 2000* and *XP* were difficult to obtain for anyone outside the Microsoft Development Network (MSDN—see http://msdn.microsoft.com/en-us/goglobal/bb688178.aspx) with the versions before *Vista*, they are now offered as an integral part of *Windows Vista Ultimate* and *7*. To change the display language in *Windows Vista* or *7 Ultimate,* select **Start> Control Panel> Regional and Language Options** and make a selection or install an additional language under **Display language**.
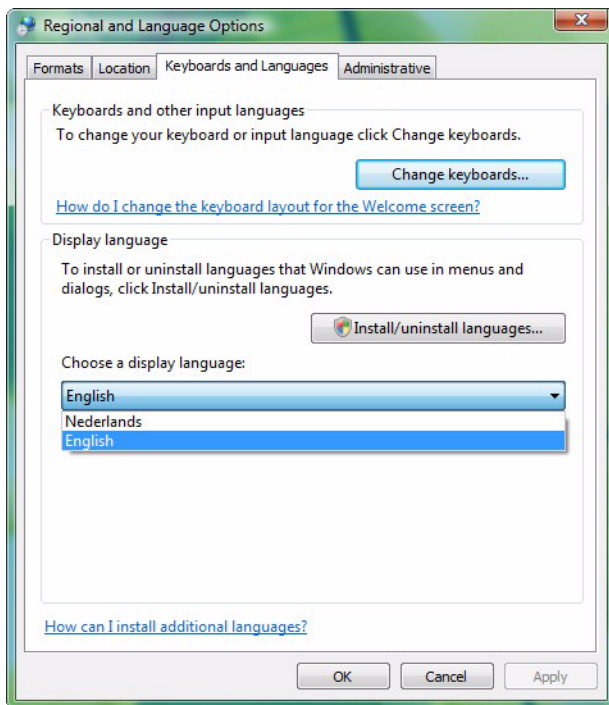


Figure 1: Switching the display language in Windows Vista

*For a language professional, in my opionion this feature would be the only reason to purchase the Ultimate version of Windows 7 (or Vista).*

If you prefer the "classic" Windows look with your version of *Windows XP* and higher, you can select this in three different areas.

You can change to the less task-oriented but more precise classic view of the Control Panel under **Start> Control Panel> Classic View**. (The instructions in this primer are all related to the *Classic View* of the Control Panel).
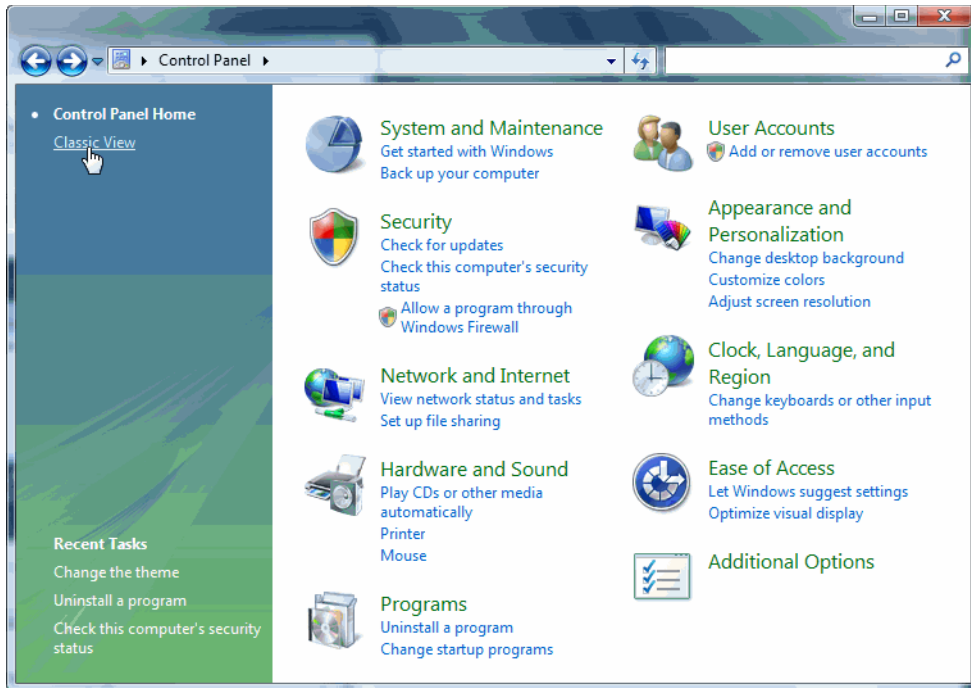


Figure 2: Switching into the "Classic View" of the **Control Panel** in *Windows Vista*

To change the "theme," i.e., the general appearance of dialogs, windows, and menus to the old look, select **Start> Control Panel> Display> Themes> Windows Classic** (*XP*) or **Start> Control Panel> Personalization> Theme> Windows Classic** (*Vista* and *7*).
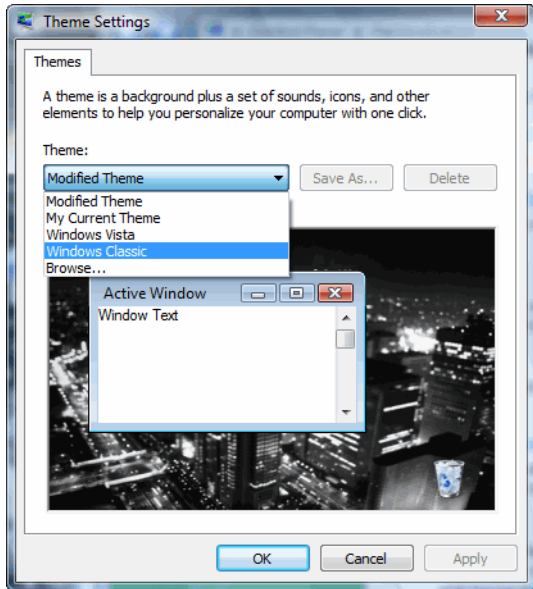


Figure 3: Switching into the "classic" view in *Windows Vista*

And lastly, the **Start** menu can be switched to the classic menu by selecting **Start> Control Panel> Taskbar and Start Menu> Start Menu> Classic Start Menu**.

# Windows Explorer

The most helpful and often-used *Windows* component on my computer is the *Windows Explorer*.

*In Windows 2000 and higher, Microsoft has tried to hide the Windows Explorer in relative obscurity by placing it within the **Accessories** folder (**Start> Programs> Accessories> Windows Explorer**). To more quickly open the Windows Explorer, right-click on any folder or the Windows **Start** menu and select **Explore**, or press the key combination WINKEY (the key with the Windows icon to the left of the spacebar on most keyboards)+E.*

The *Windows Explorer* is the application where most of my file management activities take place; where I can easily sort files by name, date, size, or type; and where I can customize the breadth of my searches for files and folders.

In most editions of Windows, clicking on the *Explorer* icon opens the *Explorer* in the `My Documents` folder. This is helpful if most of your work files are located there, but I always found this a nuisance when I wanted to access `Program Files`, `Windows`, or other main folders on my main C: drive.

Here's a great way to either have several shortcuts for different purposes or (my preference) one shortcut to the *Explorer* that simply opens right on top of the C: drive.

Right-click any blank part of the Desktop and select **New> Shortcut**. Type (or copy and paste):

> `explorer /e,/root,c:\program files` (to open Program Files—you can substitute any other path)
>
> —or—
>
> `explorer /e` (to open on top of the C: drive)

Click **Next**, give the shortcut an appropriate name, and then click **Finish**. That's it! Now you can drag that icon to your **Start** menu or your *Quick Launch* area (the icon bar to the right of the **Start** button).

There are, of course, extended third-party tools that act like a *Windows Explorer* on steroids. Many translators (and many computer users in general) say they "couldn't live without" applications such as *Total Commander* (see www.ghisler.com). Some of the helpful features of these tools include the ability to preview documents without actually having to open them or to add notes to files (which is really helpful when you have a bunch of files that need to be proofread; adding a little "done" to each finished file eliminates confusion.)

The *Windows Explorer* experience in *Windows Vista* and *7* is quite different and you might need some time to get used to it, but there are several advantages that you won't want to miss once you get used to them.

One is the preview feature which allows you to automatically preview HTML, *MS Office*, PDF, and text files within the *Explorer* (you need to enable this by selecting **Organize> Layout> Preview Pane**).

*In Windows 7 the previewed files also include graphics files and other formats. In addition, the Preview pane can be activated through selecting the Preview button on the toolbar.*

An element that might be confusing at first but may become one of your favorite features is the new address bar. The path to the selected file is no longer displayed in the usual manner with backslashes (such as `C:\Windows\Fonts`); instead, it is provided through a "breadcrumb trail." This

is an interactive address which, in case of very long addresses, is shortened and provided with little right-arrows between the different locations on the path. Clicking on any of those arrows displays all other possible branches that go off from that point so that you can quickly navigate there.
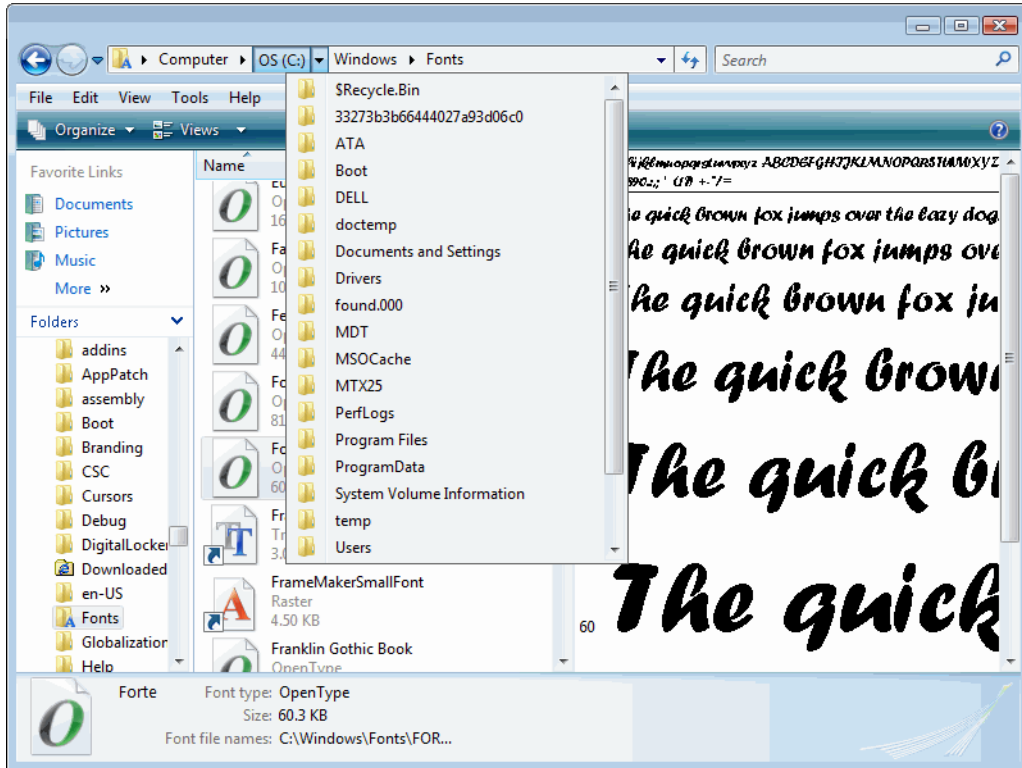


Figure 4: The *Explorer* in *Windows Vista* with the active breadcrumb and preview features

If you prefer the old-fashioned method of displaying addresses, you will only need to right-click the address bar and select **Edit Address**.
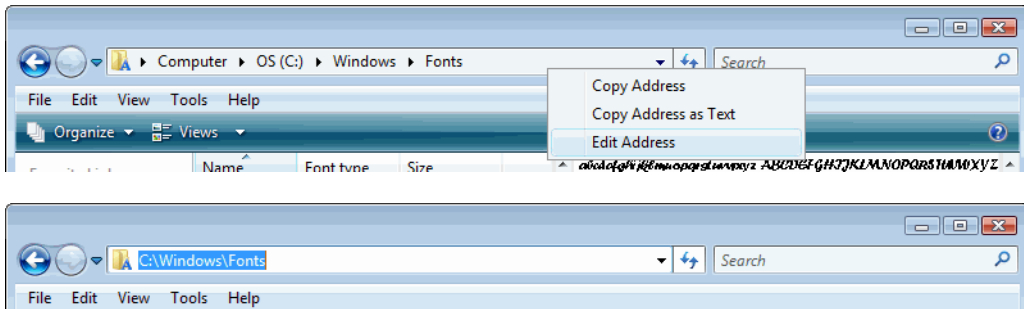


Figure 5: Switching the breadcrumb view to the traditional path view

# Tweaking the User Interface of Windows

Microsoft has been distributing a little application called *TweakUI* since sometime after the release of *Windows 95*. This was prompted by an ever-increasing number of user calls complaining about the lack of customizability of *Windows*.

*TweakUI for Windows versions prior to XP is accessed through the Control Panel, but in Windows XP it is run as a separate program, accessible from the Programs folder.*

*TweakUI is not available for Windows Vista and 7. There are however commercial third-party product— for instance Tweak VI and Tweak-7 (see www.totalidea.com)—that offer a large amount of customizability features.*
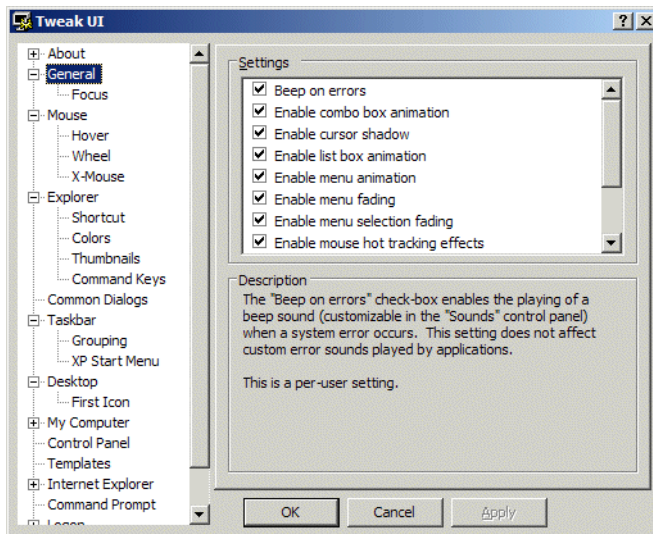


Figure 6: *TweakUI for Windows XP*

Some of the options provided to manipulate the way *Windows* displays data are more helpful than others, but looking through them may give you the opportunity to finally turn off the one feature in *Windows* that you have always hated. *TweakUI for Windows XP*, which is quite a bit more

comprehensive than its predecessors, can be downloaded at www.microsoft.com/windowsxp/Downloads/powertoys/Xppowertoys.mspx (for earlier versions of *Windows*, make a Google search for "TweakUI").

*On the same website you can also download a number of other tools, including the wonderful ClearType Tuning, a feature that makes fonts on your computer screen a lot more readable.*

*Though the basic functionality is part of the standard Windows XP, Vista, or 7 installation (you can activate it under **Start> Control Panel> Display> Appearance> Effects> Use the following method to smooth edges of screen fonts> ClearType** in XP and Vista and **Start> Control Panel> Display> Adjust ClearType text> Turn on ClearType** in Windows 7), this utility will allow you to adjust the display of fonts to your personal preference.*

## Helpful Shortcuts

### To Send Files to Other Drives or Programs

To send any file or folder to any drive (or any program)—including your floppy drive or CD writer—right-click on the file and folder and open the list under **Send to**.
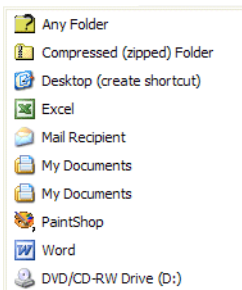


Figure 7: Right-click access to a variety of options

These shortcuts are stored in the `SendTo` folder (under `Windows` in *Windows 9x* and *ME,* `Documents and Settings/<user>` in *Windows 2000* and *XP*, and `Users/<user>/AppData/Roaming/Microsoft/Windows` in *Windows Vista* and *7*). You can delete any of the existing shortcuts or add links to any program you would like to have listed there. In the above example, I added shortcuts to *Word* and *Excel* in the `SendTo` folder so that I can open all possible files in these programs through a right-click.

*In Windows XP, Vista, and 7 there is a much quicker way to access complicated folder locations than to navigate to them step by step. You can use the Run command (**Start> Run** in Windows XP or the search bar in the Start menu in Windows Vista and 7, or, even easier, WINKEY+R in all versions) and enter shell: followed by the folder name. For instance, the command* `shell:sendto` *will for instance open the SendTo folder or* `shell:fonts` *the Fonts folder. This only works with commonly used folders within Windows.*

Another helpful way to open files quickly in many programs (especially *Office* and desktop publishing applications) is simply to drag the file into the open program while no other file is displayed. When the cursor with the file is located over the dark grey background, a plus symbol will be displayed. Releasing the mouse cursor will then open that file in the appropriate program.
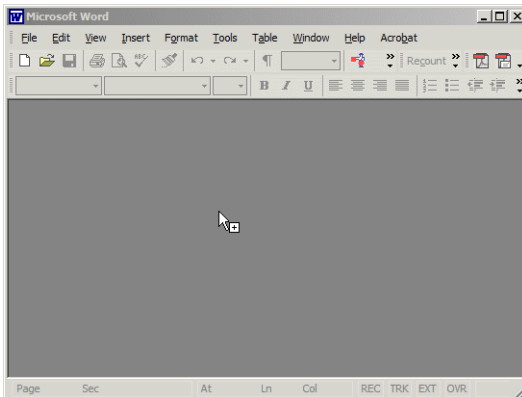


Figure 8: Opening a file by dragging it into an open application

## To Search with Wildcards

*A wildcard is a special symbol that stands for one (?) or more (*) characters. This means that a\*b could be any combination of characters starting with the letter a and ending with the letter b, whereas a?b can only be a three-character combination starting with a and ending with b.*

Wildcards in file searches are very powerful. Right-click on any folder, select **Search**, and (for instance) type a\*.exe to find any program file (.exe) that starts with an a.
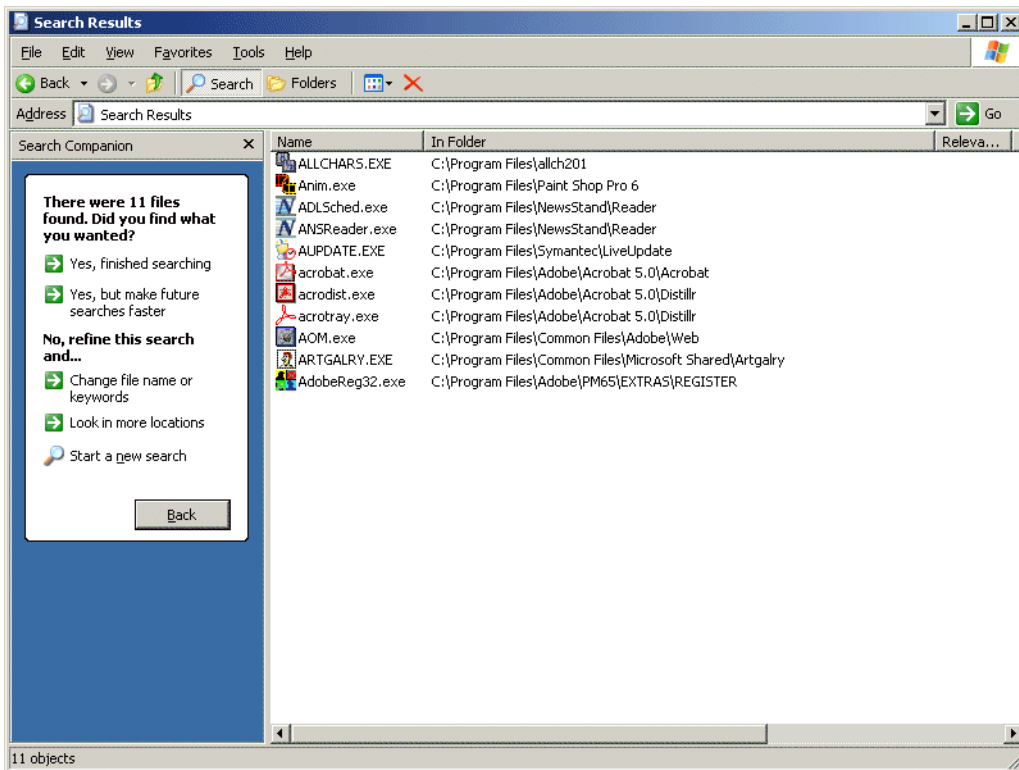


Figure 9: Searching with wildcards in *Windows XP*

The combination of this search feature and the method to open files (see page 16) also makes it possible to open many different files at the same time, even if they are located in different folders. Just use the search method described above, highlight all of the files in the *Search Results* dialog (press CTRL+A), and drag them into the application.
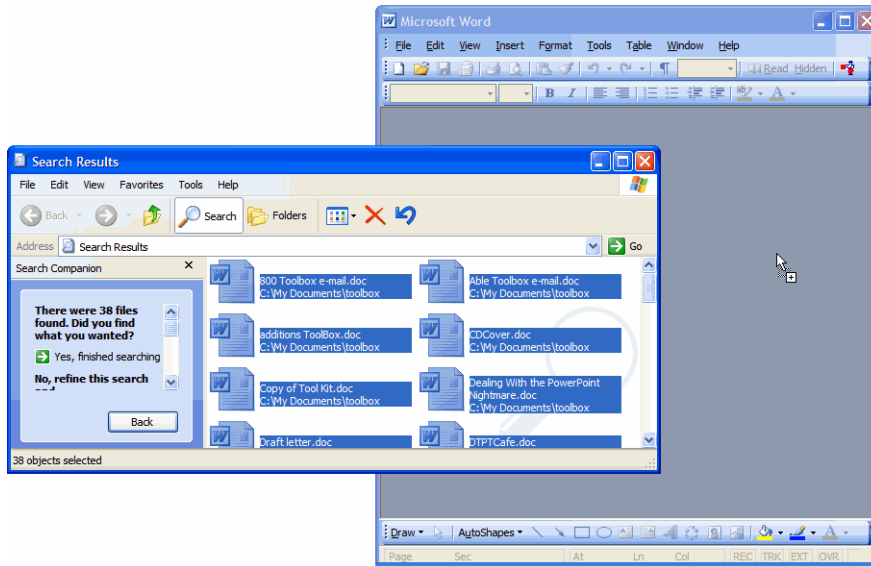


Figure 10: Opening multiple files simultaneously

*One of the main features of Windows Vista and even more so Windows 7 is accessibility and the ease of search. Unlike earlier versions of Windows where searches were rather slow and tedious, it is now possible to find data in file names or within files (the supported file types include HTML, text-based, XML, and all kinds of MS Office files) in a matter of seconds if the file is contained within one of the "indexed locations" on your computer. These typically include everything under the Start Menu and the* Users *subdirectory. However, you can change this under* **Start> Control Panel> Indexing Options***.*

There are also many third-party search tools available. One of the more outstanding ones has to be *Everything* (see www.voidtools.com). This tool is an extremely small utility (both in terms of its download size and the files it creates) that is able to index all names of files and folders on a computer—I had more than 200,000, and it needed less than three seconds. Once indexed,

the files and folders are all listed in *Everything*'s main application window; you can look for any part of the name by typing it into a search box (using wildcards or not) and the results are displayed instantaneously. You can search your complete hard drive or on any folder or folder group in *Windows Explorer*—just right-click and select **Search Everything**.

We all know that the search feature in *Windows XP* is painfully slow, so there is obviously not even the chance of comparing *Everything* to *XP*, but even against *Vista* and *7* it wins hands-down
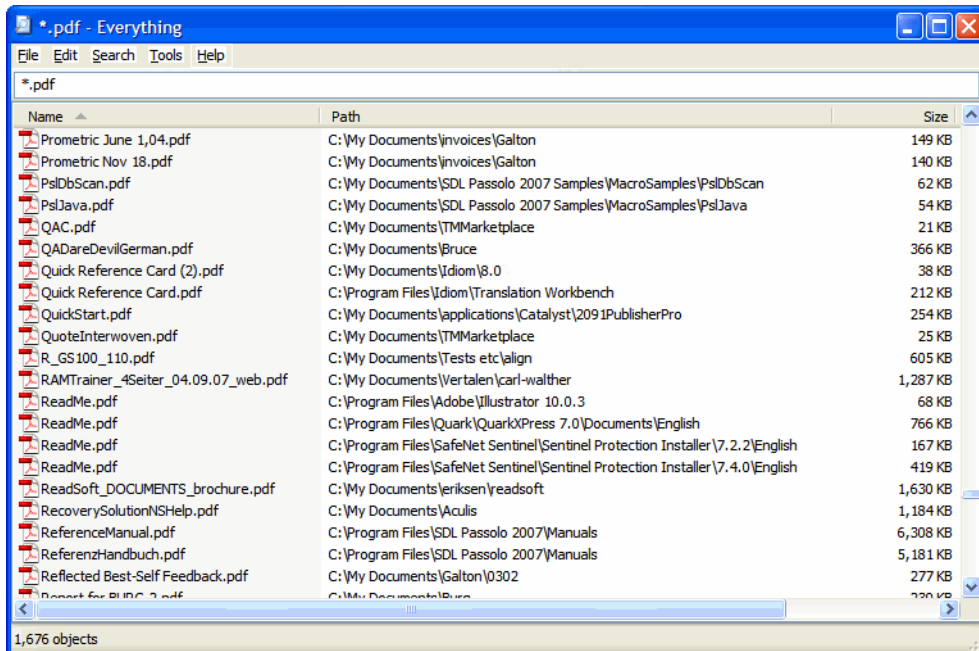


Figure 11: *Everything's* main window with a filter for PDF files

## To Copy Files or Folders

Holding the **Ctrl** key while you drag a file or folder to another place within the *Windows Explorer* makes a copy of the file or folder rather than moving it.
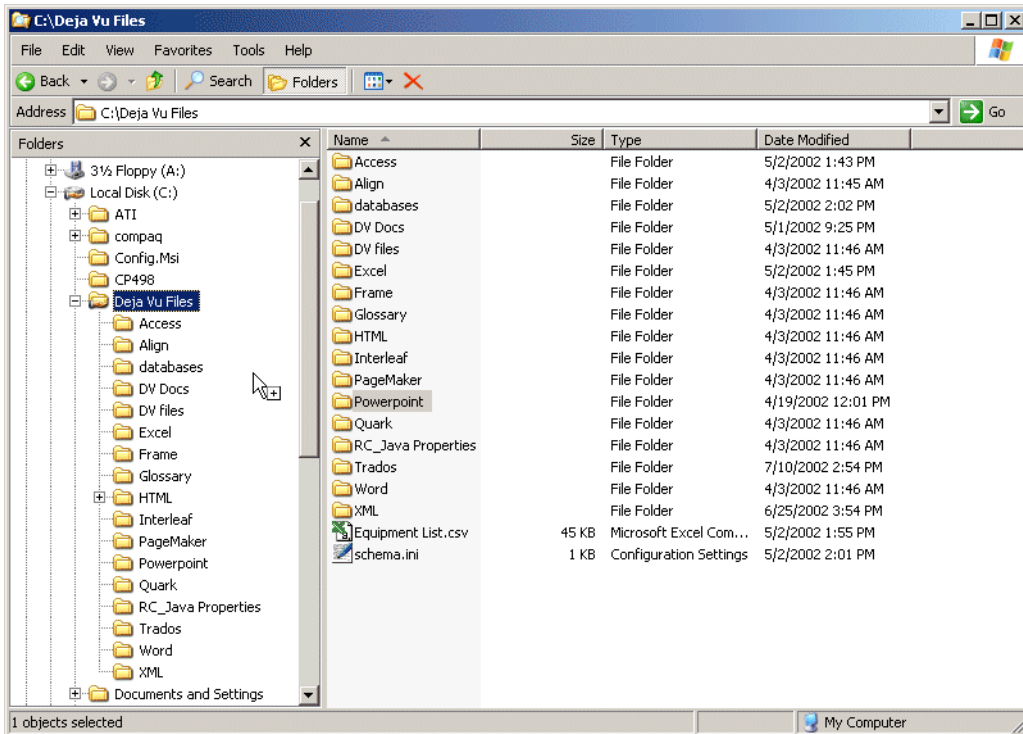


Figure 12: Dragging and copying a file in the *Windows Explorer*

Moving the file or folder within the same folder will make a copy of that file or folder and rename it to `Copy of <OldName>`.

*The same procedure (highlighting an item and dragging it while pressing the **Ctrl** key) also works within most Windows applications to duplicate the highlighted item. This procedure can be especially helpful when you work in a bilingual translation environment and you need to copy non-translatable items (e.g., product names or codes) from source to target.*

*If you press the **Shift** key rather than the **Ctrl** key while copying a file, a link to the file will be created in the same directory (in Vista and 7 it will be copied to the Desktop).*

### To Switch Between Applications

If you have several applications open, you can switch between those by selecting them on the *Windows* taskbar. In situations where the taskbar is not visible—for instance, when displaying a *PowerPoint* presentation—it is easier to do this by pressing the ALT+TAB key combination:
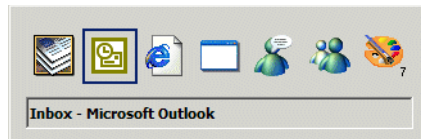


If you continue to press the ALT+TAB combination, you can rotate through the open applications. Releasing the keys will open the appropriate program.

For *Windows XP*, Microsoft also offers an enhanced version of the ALT+TAB feature, called *Alt-Tab Replacement*. It not only shows the program icons but also a thumbnail preview of each page. You can download this little utility at www.microsoft.com/windowsxp/downloads/powertoys/xppowertoys.mspx.
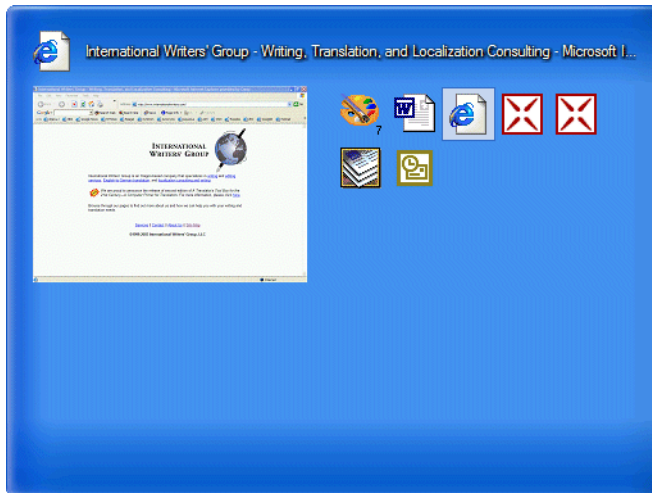


Figure 13: XP's Alt-Tab Replacement

In *Vista* this feature is the standard switch, called *Windows Flip*. *Vista* also offers the *3D Flip*, activated by pressing the WINKEY+TAB instead of ALT+TAB.
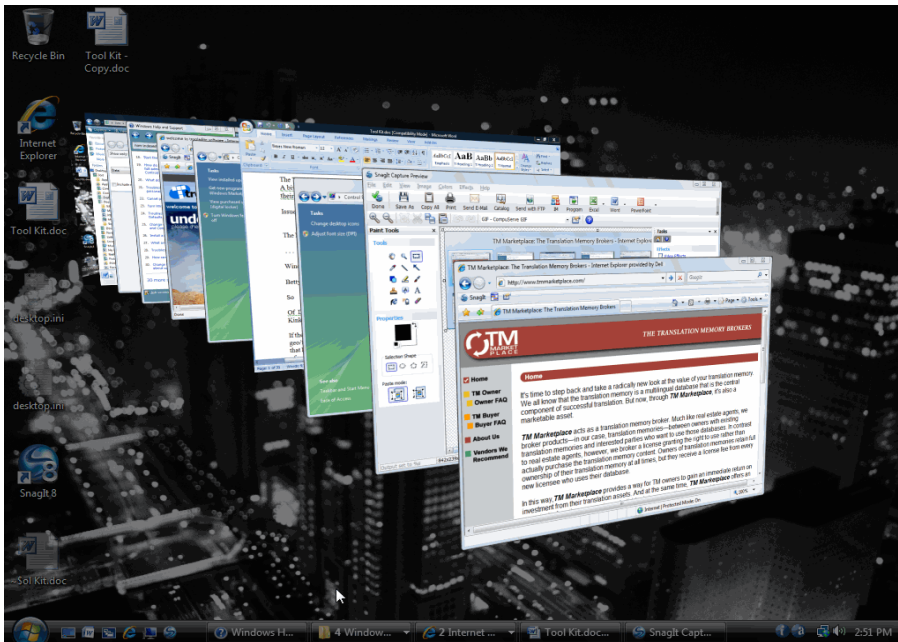


Figure 14: Vista's 3D Flip

In addition to the *3D Flip* feature, *Windows 7's* reworked taskbar (which has no distinction between the *Quick Launch* and the main area) adds a number of other features that allow for quick switching between tasks. If you have a number of instances of one application open, there is only one icon for all these instances on the taskbar with a stacked appearance. Hovering over that

icon displays a number of thumbnail images. If these thumbnail images are too small for you, you can put your mouse cursor on one of them to see a much larger representation. This should make it easier for you to decide which instance you want to work in. This feature is called *Aero Peek*.
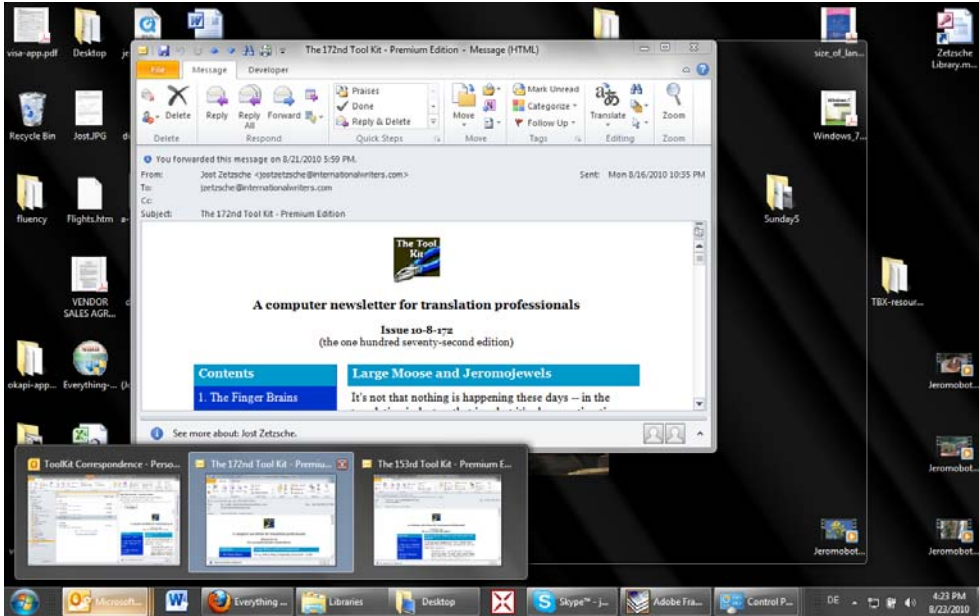


Figure 15: *Windows 7's Aero Peek*

### *Memorable Keyboard Shortcuts*

Most non-Asians who study East Asian languages find it much easier to remember characters of Chinese origin with the help of (real or imagined) pictographic aids. The same aid can be used with some well-chosen, fairly universal keyboard shortcuts.

The easiest to "see" this with is X (as in CTRL+X) for Cut (see the picture of scissors?). But how about CTRL+V for Paste? Can you see the proofreader's classic insert mark in the V? The same concept accounts for the Y in CTRL+Y for Redo, and CTRL+Z for Undo is a pictographic representation of a scribble-out.

Most other keyboard shortcuts are rather English-centric (because they are associated with the English word for the respective action: CTRL+O for Open, CTRL+N for New . . .); nevertheless, it is extremely helpful to learn this basic set of shortcuts because they are used across the majority of programs and languages.



*If you have too much time on your hands and would like to refresh your memory on all kinds of keyboard shortcuts for Windows products, here is a super-comprehensive list: www.microsoft.com/enable/products/ keyboard.aspx.*

## Folder and File Structure

The nature of our work as translators requires dealing with a large number of files, and often with numerous versions of the same file (original, translated, edited, proofread, etc.). Many clients, especially translation vendors, will require you to maintain a certain file and folder structure for their projects; others do not, but it will be advantageous to you if you maintain a uniform structure. Here is what I usually do: In a client folder (I use *My Documents* for this—or *Documents* according to *Vista* terminology), I maintain folders for each of my clients in which I create subfolders according to the date that I receive the job. I find this organization of folders helpful in many respects— end client names of projects often mean little or nothing to me a few months after the project is finished, plus I have a good indication of what should be archived and deleted from my hard drive (as in everything older than two years, or some other criteria).

The best way of doing this is by labeling each subfolder within a client's folder with year-month-day since this gives you the easiest way to sort. Now, it's possible to do this manually, but it's easier to add the date to the folder name automatically.

The following is a trick that works well in Windows XP and Vista, but I have not been able to make it work in *Windows 7*:

First you need to change the way that *Windows* displays dates. Select **Start> Settings> Regional and Language Options**. On the first tab click **Customize** (in *XP*) or **Customize this format** (in *Vista*). In **Customize Regional Options** select the **Date** tab and choose `yyyy-MM-dd` from the drop-down list. Click **OK** and **OK**.
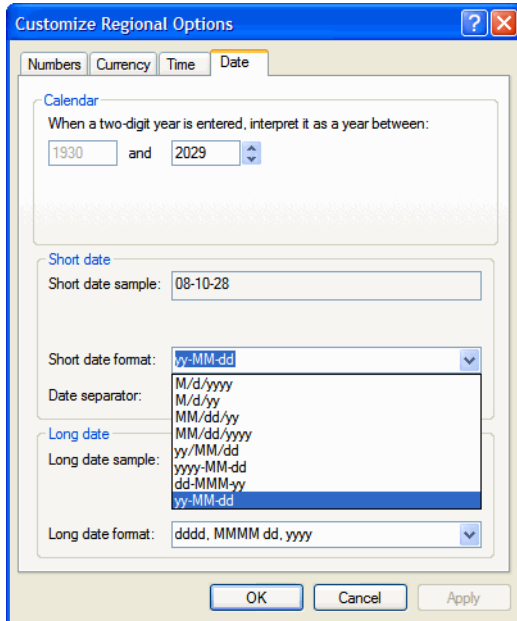


Figure 16: Customizing the date format

Now you need to make changes to the registry (for more information on the registry, including how to make a registry backup before doing this, see page 36).

Select **Start> Run** or press the WINKEY+R to open the **Run** dialog box. Enter `regedit` and press **Enter**. In the ensuing **Registry Editor** navigate to

`HKEY_CLASSES_ROOT\Directory\shell`

on the left-hand pane. Right-click the shell icon on the left-hand pane and select **New> Key**. Type `Date_Folder`, and press **Enter**. Select that folder and double-click on the **(Default)** icon in the left-hand pane. The **Value Data** box appears and you need to enter the command as you want it to appear on your context menu. For example, you can type

```
New folder with today's date
```

Click **OK**. Now right-click the **Date_Folder** key on the right-hand pane and select **New> Key**. Name it

```
command
```

and press **Enter**. Select the command icon on the left-hand pane and double-click the **(Default)** icon on the right-hand pane. Into the **Value data** box you need to type:

```
cmd.exe /c md "%1\%%DATE%%"
```

and click **OK**. Exit the **Registry Editor**.
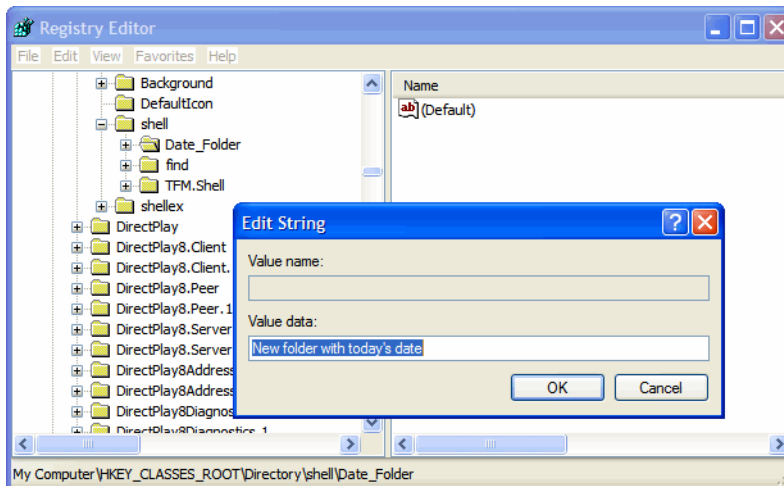


Figure 17: Editing the registry

Now when you right-click a folder icon in *Windows Explorer* or on the desktop, your new command will pop up. When you choose it, a **Command Prompt** window will briefly appear on the screen but then a new folder will come up inside the selected one.
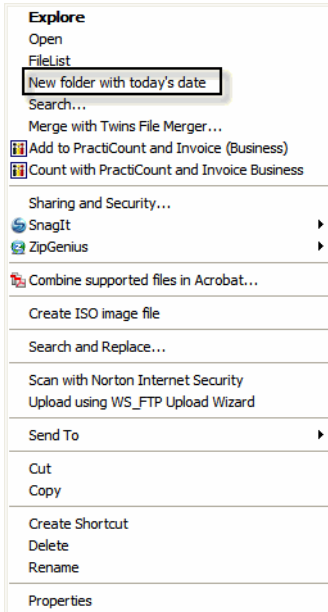


Figure 18: *Windows Explorer* right-click menu with option for new folder based on date

Naming conventions for files—if not prescribed by the client—should also have a certain logic, and it is generally helpful to have an indication in the file name of whether a file is an original, translated, or edited file (`filename_o.doc` vs. `filename_t.doc` vs. `filename_e.doc`). If you would like to batch rename a great number of files, you can find more information on page 121.
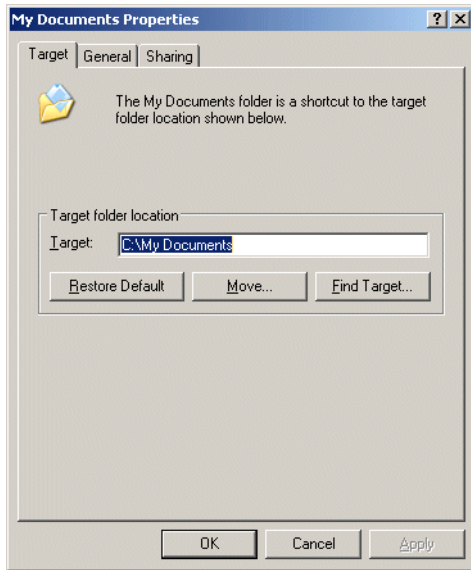


Figure 19: Changing the path of the *My Documents* folder

*A word about the My Documents or Documents folder: In Windows 2000 and higher this folder has been moved from the root of the C: drive to a user-specific folder. This is so you can log onto these operating systems with different user names, each of which can have a different **My Documents** (XP) / **Documents** (Vista, 7) folder. If you are the only user of your computer, you can move the (**My**) **Documents** folder back to the root of the C: drive. To let Windows know that you want this to be the default (**My**) **Documents** folder in which you want to save new documents, you will have to right-click that folder, select **Properties**, and then make the necessary changes under **Target** (2000 and XP) or **Location** (Vista and 7).*

# Controlling Which Programs Are Automatically Started

If your computer is very sluggish in starting up or you don't know what some of the icons in the taskbar stand for, it may be a good idea to control which programs are automatically launched when the computer starts up.

*If you are using Windows XP and you have more icons in the taskbar then you would like to have displayed, there is a helpful way to control their behavior. Select **Start> Control Panel> Taskbar and Start Menu> (Notifications Area**—Vista)**> Hide Inactive Icons**. To the right of that check box, select **Customize** to set the behavior of each of the icons that are presently being displayed or have been displayed in the past.*

There are two ways to control which programs are started up.

Any program that is listed under **Start> (All) Programs> Startup** will be launched automatically when you start *Windows*. To delete any association from that list, you can simply right-click it and select **Delete**.

*On the other hand, if you want to have your email program (or any other program) started every time you start Windows, you can also add a link to your **Startup** folder. To do this, right-click on the .exe file in its installation directory and select **Create Shortcut**. Once the shortcut file has been created, you can drag or copy it into the Startup folder.*

However, simply selecting **Delete** will not stop all automatic startup programs from running. To accomplish this, press WINKEY(+R) and type `msconfig`.

The *System Configuration Utility* will appear.
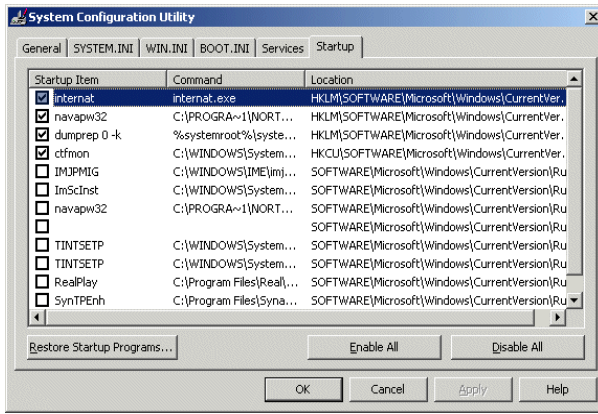


Figure 20: *System Configuration Utility* in *Windows XP*

All utilities and programs on the **Startup** tab are started automatically. You will need some of these programs to start up, but many can be unchecked (this depends on your computer configuration) to promote a faster startup and better performance.

You can find several lists on the Internet that will help you make an informed decision on which of these items should be started up and which not. Two of these lists include www.sysinfo.org/startuplist.php or www.answersthatwork.com/Tasklist_pages/tasklist.htm.

*For some odd reason, Microsoft decided to abolish the System Configuration Utility for Windows 2000 (only to reinstate it again for Windows XP and higher). If you have Windows 2000 and you would like to use this utility anyway, you can easily use the version that comes with Windows XP or higher. To retrieve it from Windows XP or higher, search for* msconfig.exe *in the* C:/WINDOWS *directory (for instructions on how to search, see page 17). Once you find it, copy it to the* C:/WINNT/system32 *directory on the Windows 2000 computer and it will run without a hitch.*

In the *Windows NT* line (*NT*, *2000*, *XP*, *Vista*, *7*), many programs are run as so-called services. These are listed on the **Services** tab of the *System Configuration Utility*; this is also where you can stop or start services. For a better description of each of the services and the ability to decide whether services should be started manually, automatically, or should be disabled, you can open the **Services** dialog under **Start> (Settings>) Control Panel> Administrative Tools> Services**.
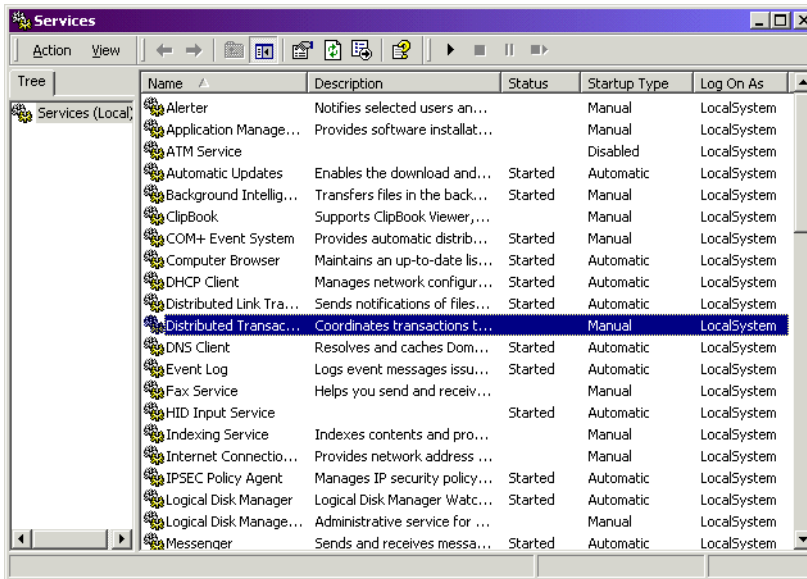


Figure 21: **Services** dialog in *Windows 2000*

Double-clicking on each of the services will open a dialog in which you can adjust the settings.
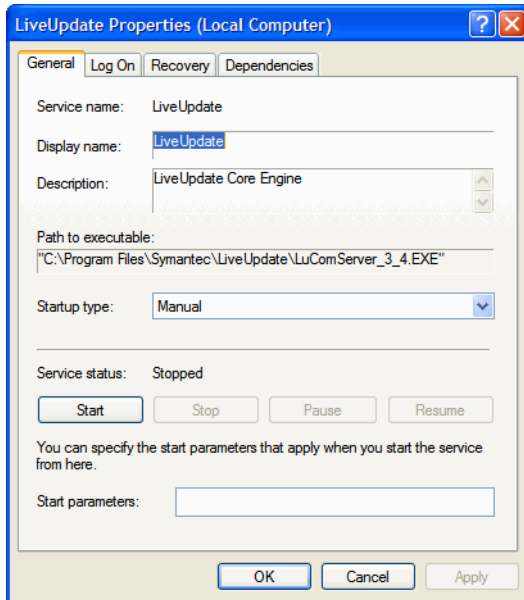


Figure 22: **Services Properties** dialog in *Windows XP*

## Avoiding the Animated Environment

Unless you own a much faster computer than you need, it may be a good idea to avoid some of the animated "Look & Feel" features that come with *Windows* or can be purchased with *Windows Plus!* for earlier versions of *Windows*. Many of these features are a heavy drain on the computer's memory.

Especially notorious in this respect are "themes" that can include animated mouse cursors, different sound schemes, and complex desktop or screensaver graphics. These can be disabled under **Start> Control Panel> Personalization> Theme** (*Vista*), **Start> (Settings>) Control Panel> Display> Themes** (*2000* and *XP*), or **Start> (Settings>) Control Panel> Display> Appearance> Schemes** (*98*). Other animated mouse cursors can be disabled under **Start> (Settings>) Control Panel> Mouse> Pointers> Scheme** (pre-*Vista*) or **Start> Control Panel> Personalization> Mouse Pointers> Scheme** (*Vista* and *7*).

For the *Aero* interface in *Windows Vista* and *7* your computer will also need extra resources, but the main resource hog in that interface is the transparency feature. You can disable this under **Start> Control Panel> Personalization> Window Color and Appearance> Enable Transparency**.

## Keeping the Computer Clean

A translator's computer, especially for those of us who work in software localization, often functions as a sort of public lab for all kinds of software applications that have to be installed for testing or to serve as reference material or as client-requested translation aids. This is not good for any computer. While I am not as radical as some who regularly re-install everything from scratch (I simply find it too time-consuming), I do try to keep my computer as clean as possible.

Most *Windows* users know that software cannot be uninstalled by deleting the corresponding folder under **Program Files**. Instead, it must be done through **Start> (Settings>) Control Panel> Add/Remove Programs** (pre-*Vista*) or **Start> Control Panel> Programs and Features** (*Vista* and *7*).
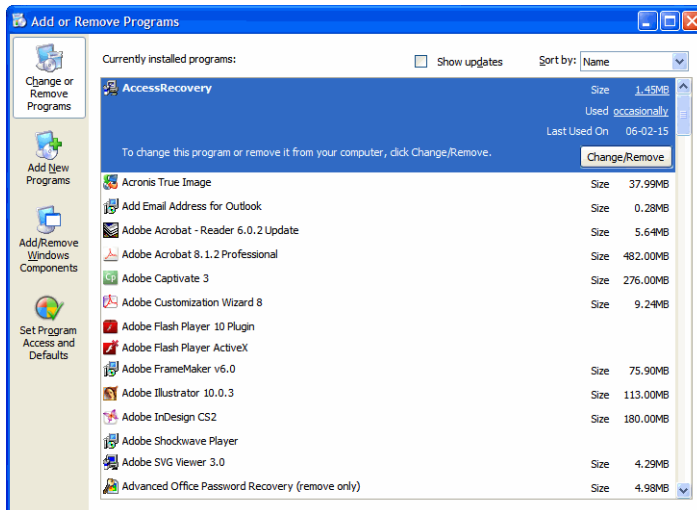


Figure 23: Removing programs

What many do not know is that many uninstallation programs are either not smart enough to find all the required files and references, or they are not even supposed to. Whenever you change anything in any of the files that were originally installed with that software, the file will not be uninstalled (this includes spell-checking dictionaries, for instance). The only way to uninstall those files is to actually go to their installation path (usually under `C:\Program Files`) and delete them manually in *Windows Explorer*.

*Rather than having to go through this long path of opening something in the Control Panel, it is also possible to drag some of the more often-used shortcuts directly to the Windows Start menu:*
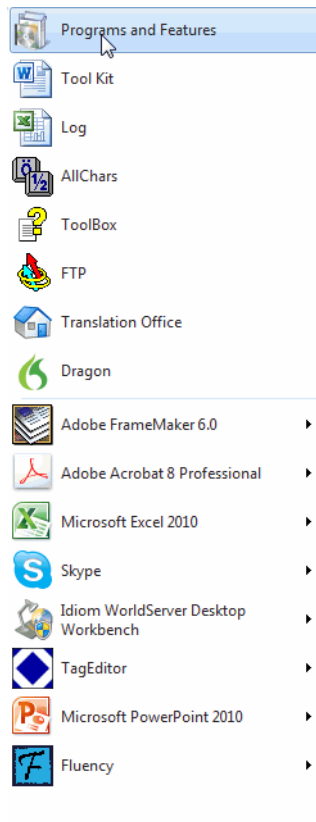
**The Registry**

Another sore spot in any *Windows* installation is the registry, a database used by *Windows* to store configuration information. The registry consists of information about your programs, operating system, all associated hardware and their drivers (little programs that make your hardware perform in the desired manner), and your personal settings for these programs. You can access the registry by pressing WINKEY(+R) and entering `regedit`, which will open a view of the registry that allows you to search for certain keys, values, or attributes and then edit them.
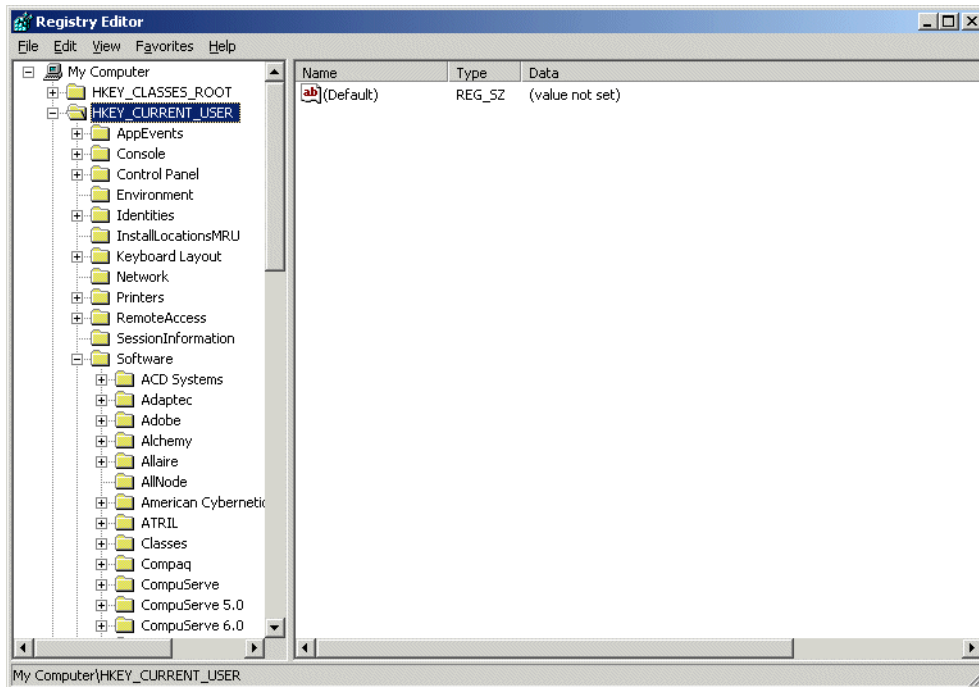


Figure 24: Registry Editor

But be forewarned: this is a very risky undertaking that could literally cripple your computer, so only do this if you have either very clear instructions on what to look for and edit.

*Before you use any of the registry cleaning utilities, make a web search for that application and see what other people have to say about it. Although this is not exactly a guarantee for success, it should give you a better idea of what (or what not) to expect.*

*And just to make sure, it's also a good idea to perform a backup of your registry under File> Export in the Registry Editor.*

## Disk Cleanup

It's an important practice to clean your computer of unnecessary files. There are numerous ways of doing this, including one pre-configured way within *Windows*. You can access the necessary dialog by right-clicking on the drive in question (usually the C: drive) in *Windows Explorer* or **My Computer** and selecting **Properties> General**. Clicking on the **Disk Cleanup** will make your computer look for unnecessary files in your computer, including temporary files.
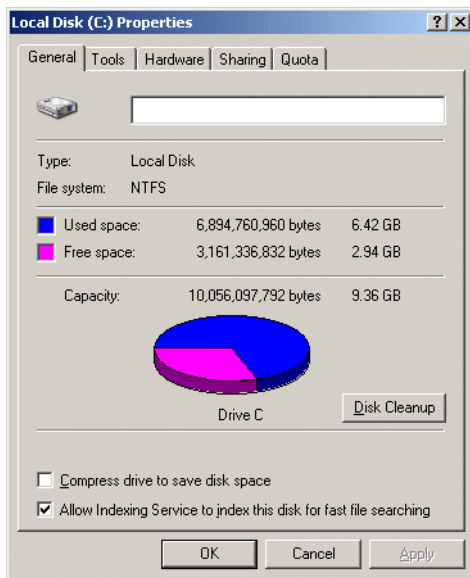


Figure 25: Accessing Disk Cleanup on *Windows 2000*

Once the computer determines which files could be deleted, it will display them divided by category and let you select which files you would like to delete.
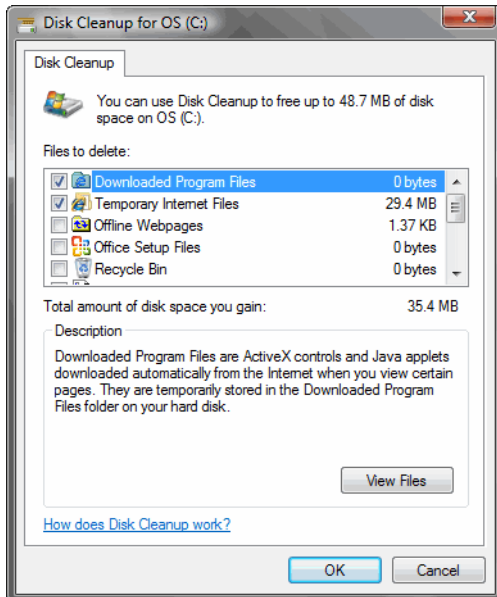


Figure 26: Selecting files for deletion on *Windows Vista*

*In the graphic above you can see that one of the items in the list is the "Recycle Bin." This is a Windows security mechanism by which it assures that files you delete from the hard drive will only be "truly deleted" once you empty the Recycle Bin.*

Deleting your temporary Internet files will only delete temporary files that have been collected with the *Internet Explorer.* It will not delete "cookies." Cookies are small text files that are stored on your computer by a web server so that you can be recognized when re-visiting its website.

If you need to delete cookies, you can do that under

• **Tools> Internet Options> General> Delete (Cookies)** (*Internet Explorer 6* and higher)

• **Tools> Options> Privacy> Cookies** (*Firefox*)

- **Tools> Clear Browsing Data> Delete Cookies** (*Google Chrome*)

- **Edit> Preferences> Security> Show Cookies> Delete all** (*Safari*)

- **Tools> Preferences> Cookies> Manage Cookies** (*Opera*)

Another way of automatically controlling how much space is used by temporary Internet files is to adjust settings in your browser. Depending on the speed of your Internet connection (the faster it is, the fewer files you should automatically store), you can reduce the amount of disk space, i.e., the number of files stored, in *Internet Explorer* under **Tools> Internet Options> General> (Browsing history)> Settings**.
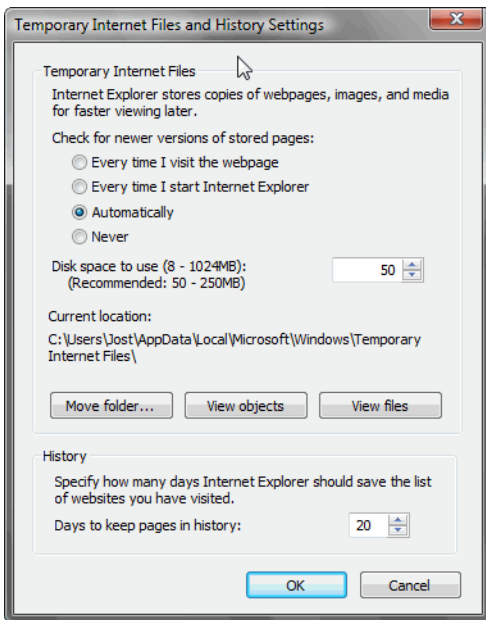


Figure 27: Setting the amount of disk space for temporary Internet files to 50 MB in IE 7 and 8

In *Opera* you can control this under **Tools> Preferences> Advanced> History**.
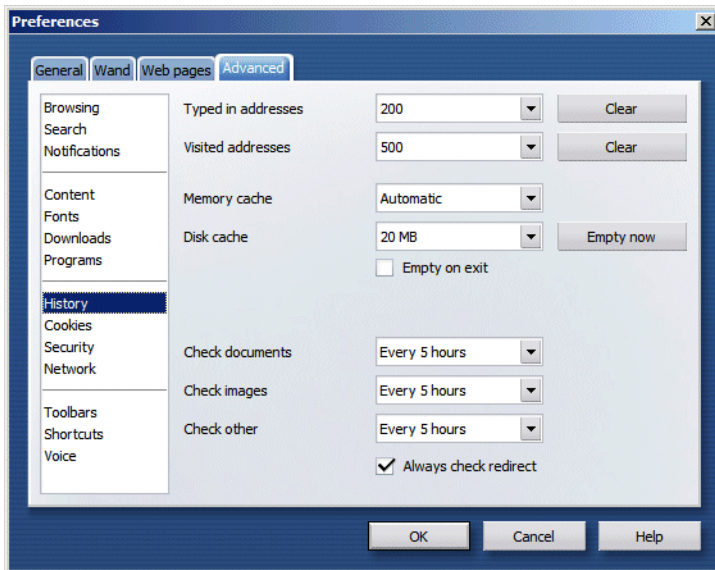


Figure 28: Setting the amount of disk space for temporary Internet files to 20 MB in *Opera*

In *Firefox* this option is available under **Tools> Option> Advanced> Network> Cache**.
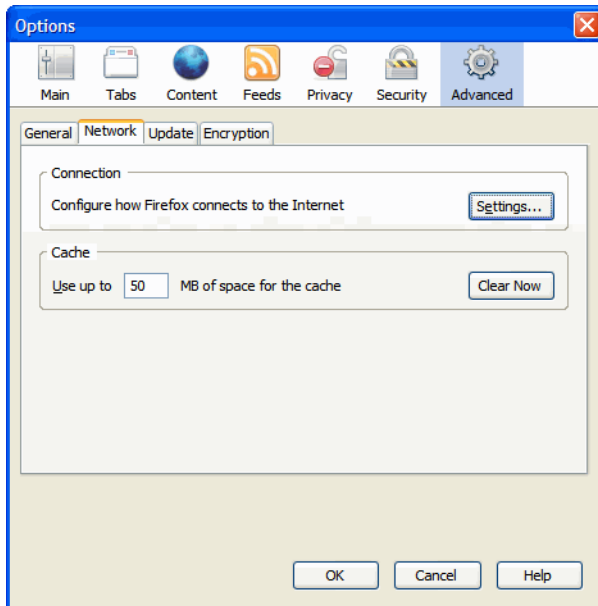


Figure 29: Setting the cache size in *Firefox*

In *Safari* you have access to this feature under **Edit> Preferences> Security**.
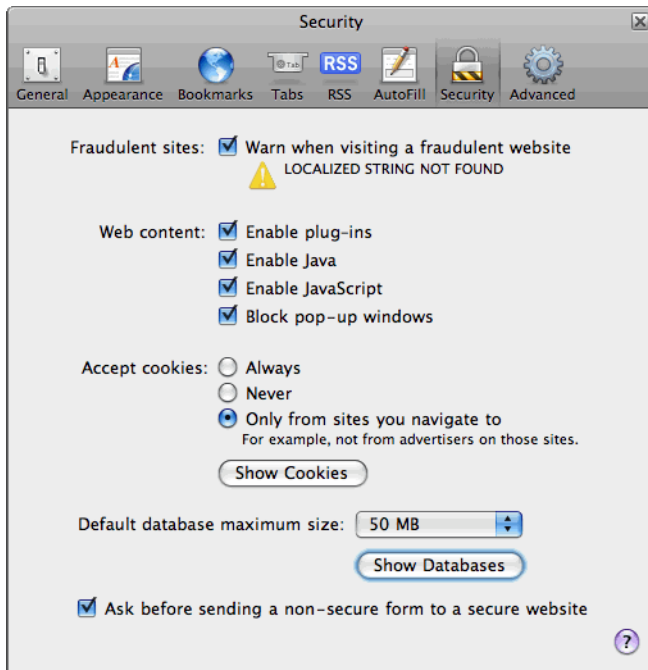


Figure 30: Setting the cache size in *Safari*

*Google Chrome* does not allow you to manage the amount of temporary files.

# Error Checking and Defragmenting Drives

Another way to speed up your computer is by defragmenting. It's a good idea to do this regularly, especially if you often install and uninstall programs. A file is fragmented when it is divided into pieces ("noncontiguous clusters") that are scattered around the disk. This is invisible to you, but it can slow down the speed at which your data is accessed because the disk drive must search through different parts of the disk to put together a single file.

However, before you defragment, it is usually a good idea to check your hard drive for any errors with the *ScanDisk* program. This can be done by right-clicking on the drive in question (usually the C: drive) in *Windows Explorer* or **My Computer** and selecting **Properties> Tools> Check Now**.
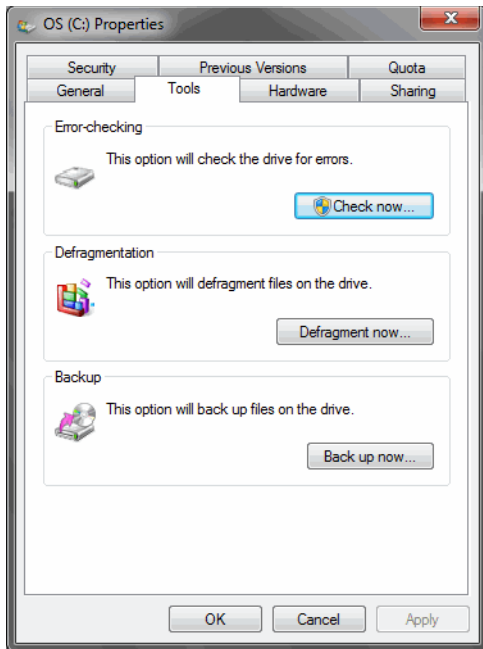
Figure 31: Error checking and defragmenting the hard drive

*Windows 2000* and higher will require a restart, during which the *ScanDisk* program is run.
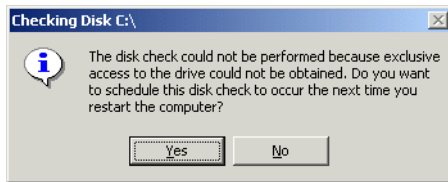


Figure 32: Prompt to restart for ScanDisk in Windows 2000

Though this may seem a little more tedious, in reality it will make your life much easier. The *ScanDisk* program needs exclusive access to the hard drive, so any additional access to the hard drive from any other process or program in *Windows 95* and *98* will make it stop and start from scratch. This will require you to identify the offending programs and terminate them before *ScanDisk* can finish.

Once *ScanDisk* has successfully finished checking the drive for errors, you can start the defragmentation. An easy way to access *Window*'s defragmentation program is by right-clicking on the drive in question (usually the C: drive) in *Windows Explorer* and selecting **Properties> Tools> Defragment Now**. While *Windows Vista* and *7* start the defragmentation process right away, *Windows 2000* and *XP* give you the option to analyze whether a defragment is necessary. Depending on the state of your drive, defragmentation can take several hours and is thus a process that should be done overnight (unless you share a bedroom with your computer—the chattering disk will keep you up all night).

*Windows Vista* and *7* also offer you the opportunity to schedule regular defragmentations.
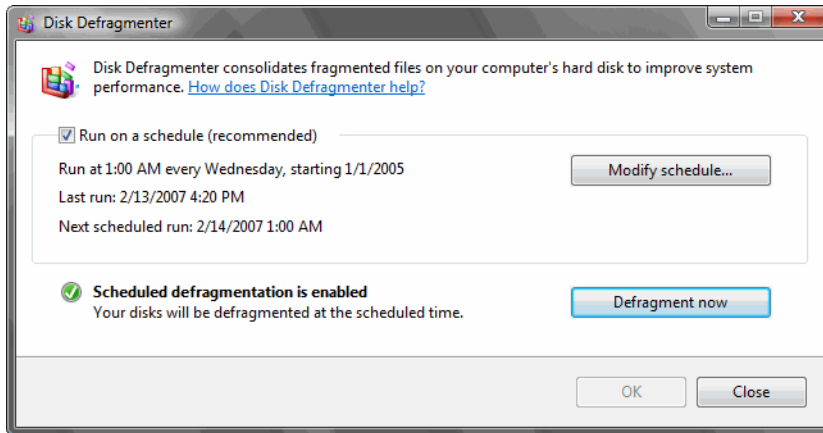


Figure 33: Setting up a schedule for defragmentation in *Vista*

## Starting the Computer in Safe Mode

There is a lot that can go wrong with a computer; ironically, the more you know about a computer and the more you experiment with it, the greater the danger that it will not even start up anymore at some point. There are a number of things you can do in a case like that, but my personal favorite (which has saved me a number of times) is starting in "*Safe Mode*."

In *Safe Mode*, the only programs that are loaded are the operating system and drivers for the mouse, keyboard, and standard display modes, greatly increasing your chances for successfully loading your computer. Once you are in *Safe Mode* you can undo what you messed up before and then reboot into *Normal Mode*. And sometimes problems even disappear once you have booted into *Safe Mode*.

To enter *Safe Mode*, continually press the F8 key as your computer starts up until you see a screen where you can select *Safe Mode* as your startup option. Once booted into *Safe Mode*, you can adjust your settings and simply restart. Your computer will then automatically boot into *Normal Mode*.
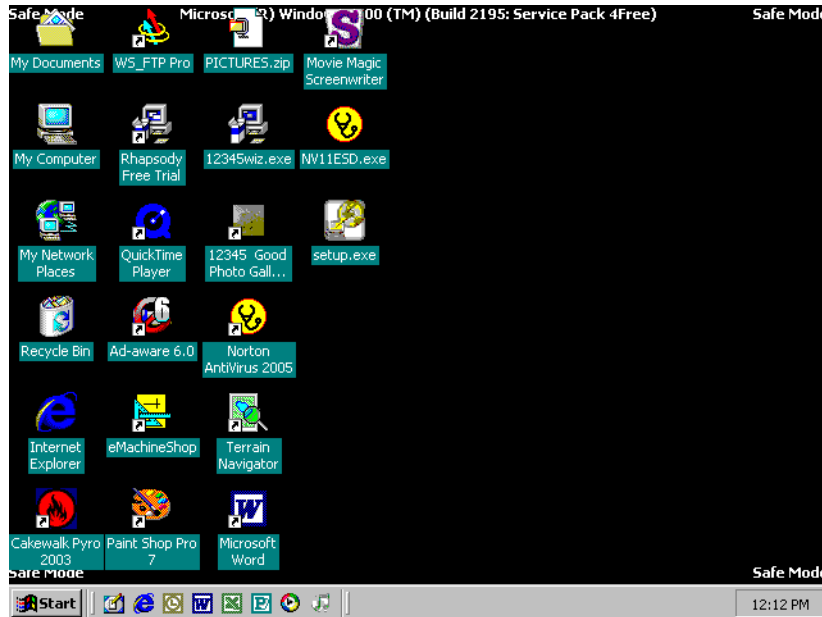


Figure 34: *Windows 2000* in *Safe Mode*

## Restoring Your Computer

*Windows ME, XP, Vista*, and *7* have a very helpful feature called *System Restore*, which allows you to reverse to an earlier point in the life of your computer. Every time a change in the system configuration is performed (and automatically every other day), *Windows* creates a "restore point" that allows you to jump back to that earlier point. If you perform something on your computer that you are not sure about, you can also create a restore point manually.

To access *System Restore*, select **Start> (All) Programs> Accessories> System Tools> System Restore**.
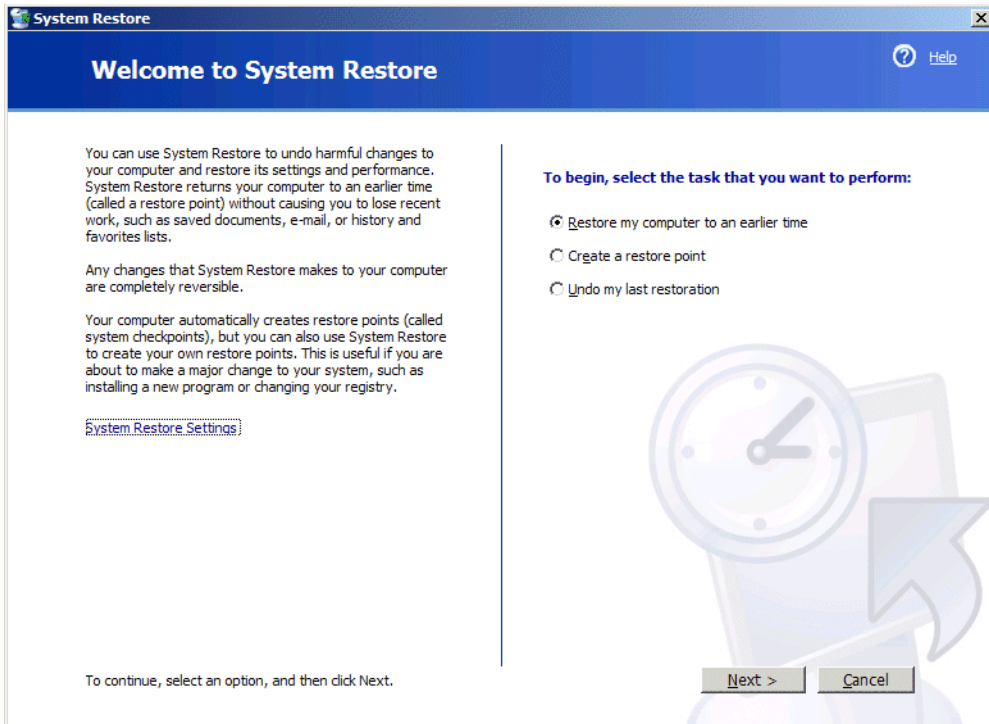


*Figure 35: Windows XP's System Restore*

Select whether you want to restore your computer or create a restore point and click **Next**.
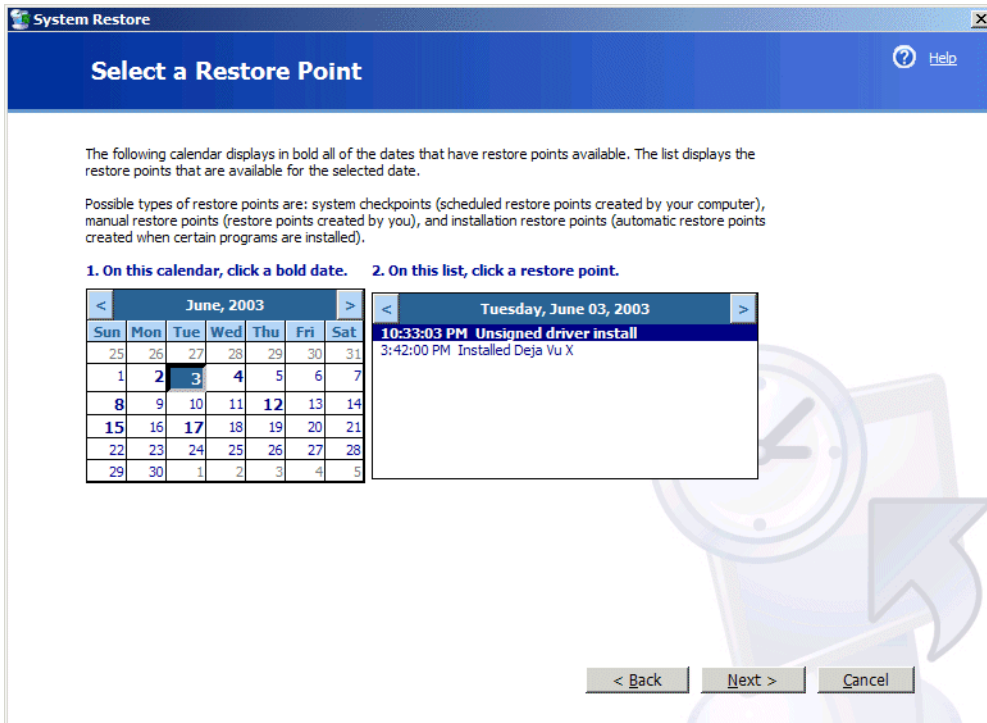


Figure 36: Selecting a restore point

If you chose to restore your computer, you can now select the date and the system change you would like to restore it to. Selecting **Next** will restart the computer to that point.

> *Any **programs** that have been uninstalled or installed during that time period will also be reversed. However, **documents** that you may have worked on will not be affected by this.*

# Backing Up Files

Though most of us had to say farewell to floppy disks as a backup method—files are usually too big for that and most computers don't have drives for 3½-inch disks anymore—we are once again in the golden age of inexpensive backups. With prices of CD-Rs and even DVD-Rs only slightly higher than floppy disks, and CD/DVD-RW drives as near-standard equipment on most computers, it is very easy and convenient to make regular backups of the projects you are working on.

Because it often takes a little more time to write very large files to a CD/DVD, I have found it most convenient to make a daily backup of my current projects on an external drive and a backup on a CD/DVD once the project is finished. The CDs/DVDs I usually store at a location separate from my computer.

# Backing Up and Restoring the System

While *Backing Up Files* describes how to back up single files or file folders you have worked on, there are also ways to back up the complete system. This process is called "ghost imaging" (you may also hear people refer to it as "imaging" or "ghosting"). It makes an exact image of a drive in your computer or the complete hard drive and allows you to burn that on several CDs/DVDs or an external hard drive. In the case of a disk failure, this will allow you to read that image back to the hard drive rather than having to install everything from scratch.

The two most well-known ghosting applications are *Acronis True Image* (see www.acronis.com) and *Norton Ghost* (see www.symantec.com/ghost).
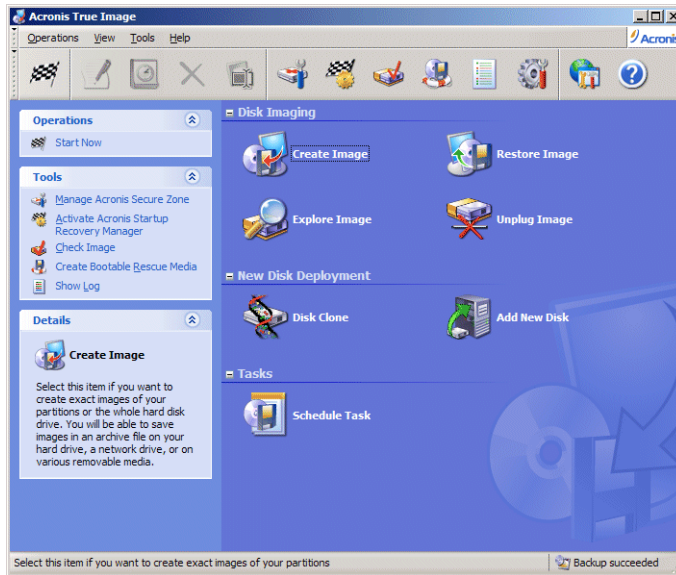


Figure 37: Image options with Acronis True Image
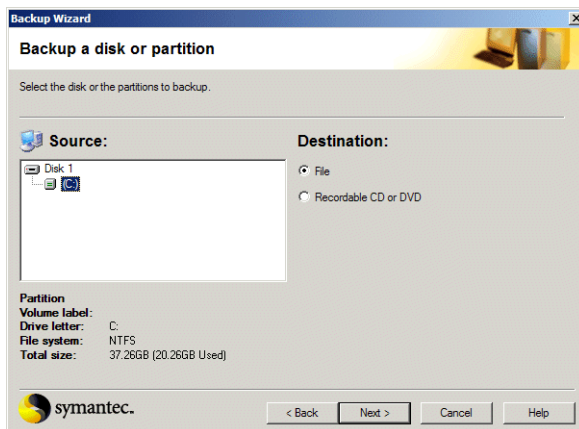


Figure 38: Creating an image with Norton Ghost

In the *Business* and *Ultimate* editions of *Windows Vista*, you can also find an integrated imaging tool under **Start> Accessories> System Tools> Backup Status and Configuration> Complete PC Backup**. While this is a great feature for which you don't need any additional product, the downside is that you cannot restore individual files or folders like you can in *True Image* or *Ghost*. In other words, this feature really is only for the unfortunate case of a complete disk failure where you have to restore everything from scratch.
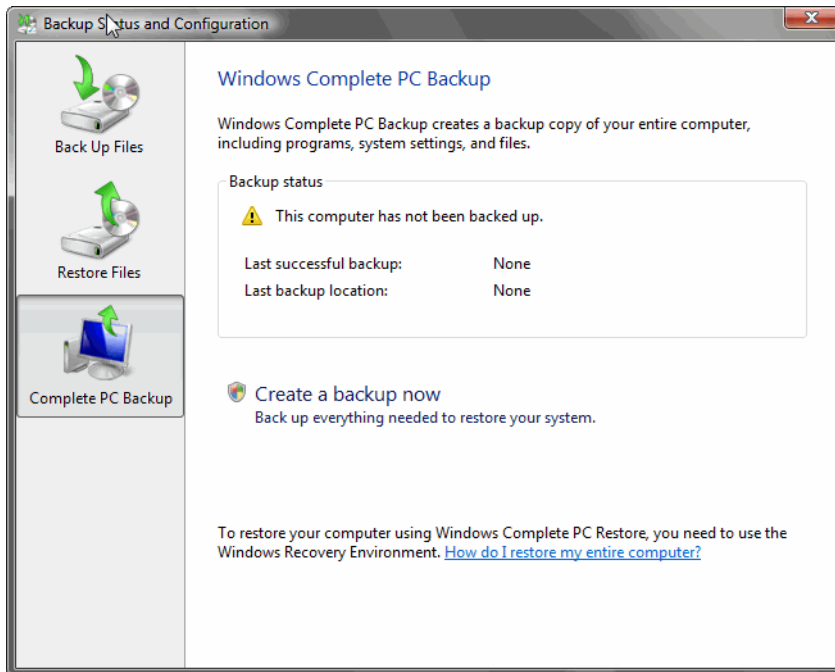


Figure 39: *Vista's Complete PC Backup* feature

Backup is one of the areas where Microsoft felt that it had offered a halfhearted solution for *Windows Vista* and tried to do the right thing for *Windows 7*. The first thing Microsoft fixed was the availability of the backup, which is now a feature in all versions of *Windows 7*. The next improvement is the increased flexibility in deciding what exactly you want to backup. As a result, *Windows 7* users might be the first in a long line of *Windows* users for whom a third-party backup program is not par for the course.

Both *Windows 7* and *Vista* offer yet another feature called "shadowing." (You have to admit that these terms—"imaging," "shadowing," and "ghosting"—suggest that some developers have a rather colorful imagination.) This tool restores files to a previous version even though you did not actually back the files up. It does this by "looking" into the *System Restore* archives to see whether there is an older version of the file and allows you to restore it to that version. And all you need to do to access the feature is to right-click on the file in question and select **Restore previous version**.
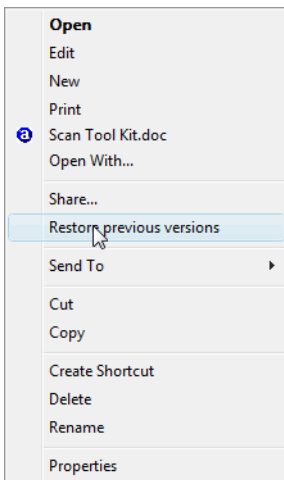
Figure 40: Selecting the **Restore previous versions** feature

Once you select the command, several earlier versions of the file might be displayed:
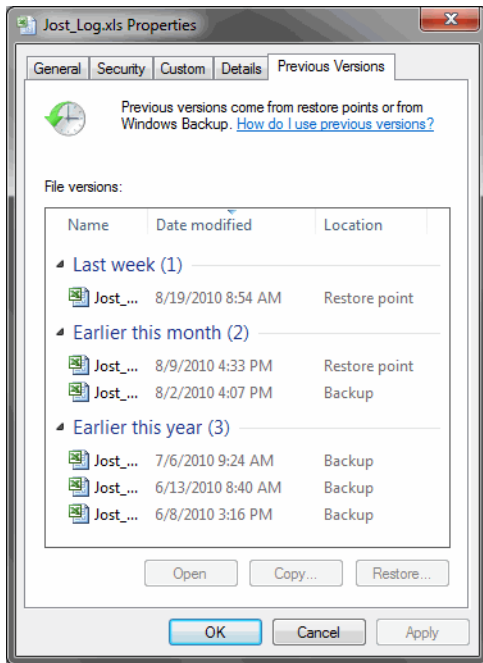


Figure 41: Previous versions of an *Excel* file

Now you can select the version of the file you need to restore.

*Once you do this, there is no undo: your current version of the file in question is gone and has been replaced. This might work out in almost all cases, but in the few cases where this makes things even worse (there is typically only one previous version per day so you might not get the version you want), there is also the option to highlight one of the files and select* **Open** *or* **Copy** *to check whether it's the correct version. Once you know it is, go ahead and save it over your existing file.*

## Taking Inventory of Your Computer

The little freeware download *Belarc Advisor* (see www.belarc.com) provides a helpful and detailed report of all the hardware and software that you have installed, including operating system, processor speed and memory, system serial number, printers, controllers, displays, and even passwords for installed programs applications that you might otherwise forget.



Figure 42: Analysis results in *Belarc Advisor*

## Keyboard Languages

It may sound strange in this age of unlimited choice, but there are times when it would be helpful if computers gave fewer choices for how to accomplish a certain task. (Needless to say, there are other times when just the opposite would be true!) One area where there are far too many choices is entering non-English characters in a *Windows* environment or within a tool like *Microsoft Word*.

Here are some of the choices for entering non-English characters with the facilities that *Windows* and/or *Word* offer:

* The archaic way: The *Character Map*. You can either start this under **Start> (All) Programs> Accessories> System Tools> Character Map**, or through a slightly modified version within *Word* under **Insert> Symbol (> More Symbols)**. Here you can find all the supported symbols and characters for each individual font to select and paste into your text. This is a great choice for the casual non-English user, but certainly not for the professional translator.

* The *Word*-centric way: *Word* provides a number of keyboard shortcuts to enter special characters, such as CTRL+SHIFT+@ followed by "a" for "å". You can find a full list of these under office.microsoft.com/en-us/ assistance/HP051865621033.aspx or in your *Word* help system. However, I have two gripes with this: you still need two keys to get to your special characters and, even worse, once you leave *Word* you're lost.

* The *Word*-centric way, part II: Customized shortcuts within *Word*. You can select a character in the *Word Character Map* (see above), click **Shortcut Key**, press the key combination you want to use (i.e., an ALT+ combination or a function key), and then click **Assign**. Not good either. Though you can get by with just one keystroke combination, you're still lost outside of *Word* or on any computer other than your own.

* The work-out way, aka the ANSI code: This poor but unbelievably popular way among translators consists or four (4!) keystrokes for one character. To activate this, make sure that you have your NUM LOCK key enabled (the small keypad on the right of your keyboard), and type the number of that character on the small keypad as you press the ALT key. The above-mentioned "å" has the key combination 0228. Phew! Like I said, a great way to train your memory to remember all kinds of code and exercise your finger muscles, but this certainly is not conducive to a productive work environment!

Clearly, things can't be as bad as these methods suggest, and most of you know that the best way by far for dealing with special international characters is by installing a language-specific or the *US-International* keyboard.

First things first, though. For the uninitiated, there is a distinction between a virtual and a physical keyboard. The physical keyboard is the hardware keyboard that you use to type and on which every key is labeled with a certain letter, number, or symbol. If you bought your computer in the U.S., chances are that you have a US-English QWERTY keyboard (representing the first six proper letters). If you bought your computer and/or keyboard in—let's say—Germany, you will probably have a German QWERTZ keyboard. The funny thing is that the labels are only meaningful if that physical keyboard matches the "virtual keyboard" —i.e., the way that your computer assigns the physical keys to the actual output on your screen. If they don't match, the virtual keyboard decides the output.

You are free to select as many virtual keyboards as your heart desires (if they are among the more than 100 different keyboards plus various other input systems supported by *Windows*), and in fact for many languages there is a good selection to choose from. For instance, one of the keyboards for U.S. English is the *US-International* keyboard, which is particularly interesting in our context because it provides ready access to a number of important international characters if you press the right ALT key.
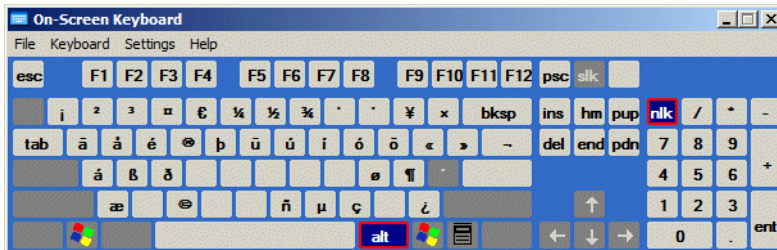


Figure 43: Characters on the US-International keyboard when pressing the right ALT key

*You can find the On-Screen Keyboard of the image above under* **Start> Programs> Accessories> Accessibility> On-Screen Keyboard** *(Vista and 7:* **Start> Programs> Accessories> Ease of Access> On-Screen Keyboard**.)

*Starting from Windows XP SP2, there is also a British equivalent, the "United Kingdom Extended" keyboard. This keyboard particularly supports languages like Welsh, replaces the apostrophe key as a dead key with the grave accent key, and introduces some other changes to the US-International keyboard.*

*In one additional change of versions after Windows XP SP2, the right Alt key plus c will generate ç.*

Aside from the keys that can be accessed like this, you can also "create" international characters with a combination of a "diacritical mark" (a so-called "dead key") followed by a letter:

- " + a = ä

- ' + a = á

- ' + c = ç

- ` + a = à

- ^ + a = â

- ~ + n = ñ

All this is great, but it also causes what many users consider to be the drawback of the *US-International* keyboard: the characters ", ', `, ^, and ~ are "dead keys," which means that they don't "type" if you use them in a normal text. Only when you type the next character will the system "know" whether you meant the character as a diacritical mark or a real character and output either one or two characters. If you are not used to this so-called "sequence checking" process, it can feel quite disconcerting, and, worse, some *Windows* installations tend to behave irregularly with printing or not printing the "dead keys."

Here is how to install an additional keyboard. Depending on the version of your *Windows* system, select **Start> Settings> Control Panel> Keyboard> Language** (*Windows 9x*), **Start> Settings> Control Panel> Input Locales> Change** (*Windows 2000*), **Start> Control Panel> Regional and Language Options> Languages> Details** (*Windows XP*), or **Start> Control Panel> Regional and Language Options> Keyboards and Languages> Change keyboards** (*Windows Vista* and *7*). This or a similar dialog will appear:
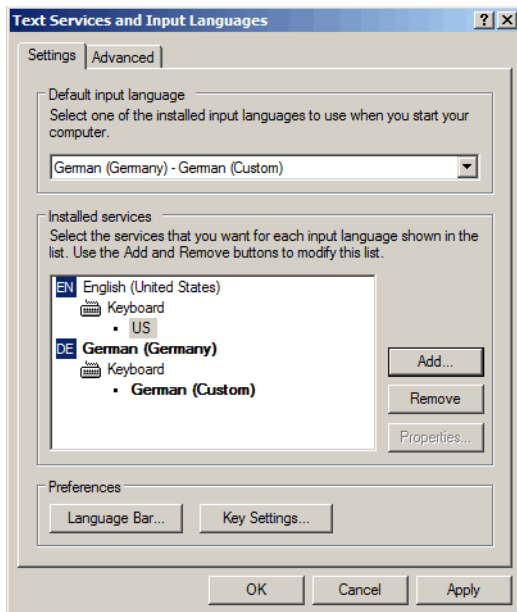


Figure 44: **Input Languages** dialog

Select **Add** and define which additional languages and/or keyboards you would like to have installed on your system. (On some installations, especially *Windows 95*, you will have to have the original media to install additional keyboards.)

When you select **OK**, the new keyboard will show up in the list of installed keyboards.
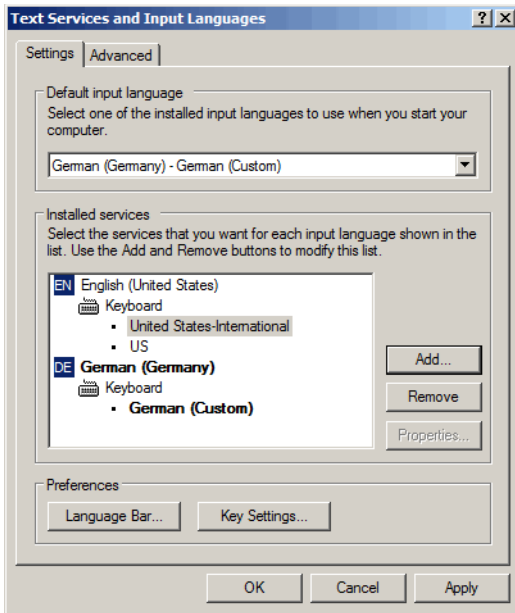


Figure 45: **Input Languages** dialog with a newly added keyboard

After you leave this dialog, you will have a little language icon displayed on your task bar.



Figure 46: Language icon on *Windows* taskbar

This icon displays your currently selected languages and allows you to switch between the different keyboards. Should you have more than one keyboard for one language installed (for instance, both the *US* and the *US-International* keyboards for English), a little keyboard is displayed to the right of the language icon. Clicking on that keyboard will allow you to select the specific keyboard you need.

*If you cannot see the keyboard, right-click on the language icon and select **Additional icons in task bar**. The same right-click command also gives you access to the **Restore** (or **Show**) **the Language bar** command that places a full language bar on the top of your screen, or the **Settings** command which displays the configuration dialog for the installation of a new keyboard without having to go through the ridiculous paths described above.*

This is all very easy. It becomes a little more hairy if you have to select languages that either don't deal (exclusively) with alphabets (such as Japanese, Chinese, or Korean) or use a completely different writing system, such as bi-directional languages (Hebrew, Arabic).

If you are still using *Windows 9x* or *NT*, you will either need a localized version of *Windows* in those languages or an additional program on top of *Windows* that will allow you to write. From *Windows ME/2000* on, these languages are supplied with the operating system if the appropriate "locales" are enabled. In

*Windows ME/2000*, select **Start> Settings> Control Panel> Input Locales** (see first graphic) and **Start> Control Panel> Regional and Language Options> Languages> Details** in *Windows XP* (see second graphic). Make sure to enable the appropriate languages or language groups.
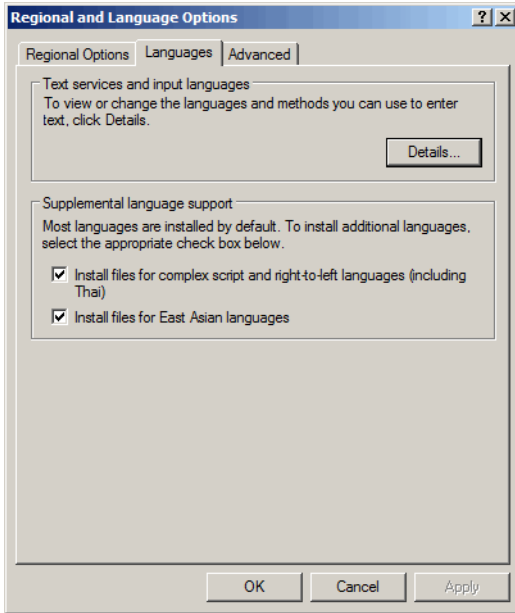


Figure 47: Enabling complex languages in *Windows XP*

In *Windows Vista* and *7*, all languages are enabled by default.

Once your locales are enabled, you can go back to the keyboard dialog and add keyboards (or: "Input Method Editors") for those languages—just as with any other keyboard under **Regional Options**.

Because many of the more complex writing systems offer a variety of options for their input systems, it is important to remember to activate **Additional icons in task bar** as described above. If you do not do that, you will not be able to use the keyboards properly.
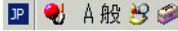


Figure 48: Example of Japanese keyboard with access to various features
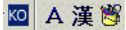


Figure 49: Example of Korean keyboard with access to various features



Figure 50: Example of Simplified Chinese keyboard with access to various features

Aside from the options that *Windows* offers you in the standard installation, there are many things that can be said about ways to change the mapping of your keyboard so that it works in one specific language or several, or performs certain processes ("macros") when pressing certain keys. A very powerful program which allows you to reassign keys is the *Microsoft Keyboard Layout Creator* or *MSKLC* (see www.microsoft.com/globaldev/tools/msklc.mspx). This allows you to take an existing language-specific keyboard, change some settings, and save that new keyboard as a customized keyboard for your language. Or you can create a new keyboard from scratch.

There are some drawbacks. *MSKLC* only works on *Windows 2000, XP*, or *Vista* (not *Windows 7*) and it's not particularly easy to use. Once you load an existing keyboard, you need to first make modifications and save the resulting file (those commands are all available in the **File** menu); only then can you build a project that will result in an installation program for the new keyboard (you can access those commands from the **Project** menu).

The great news is that it is free and the documentation is really pretty good. I used this program to swap the Y and Z keys on my German keyboard so they're in the same order as the English keyboard and I can avoid all those sillz tzpos.
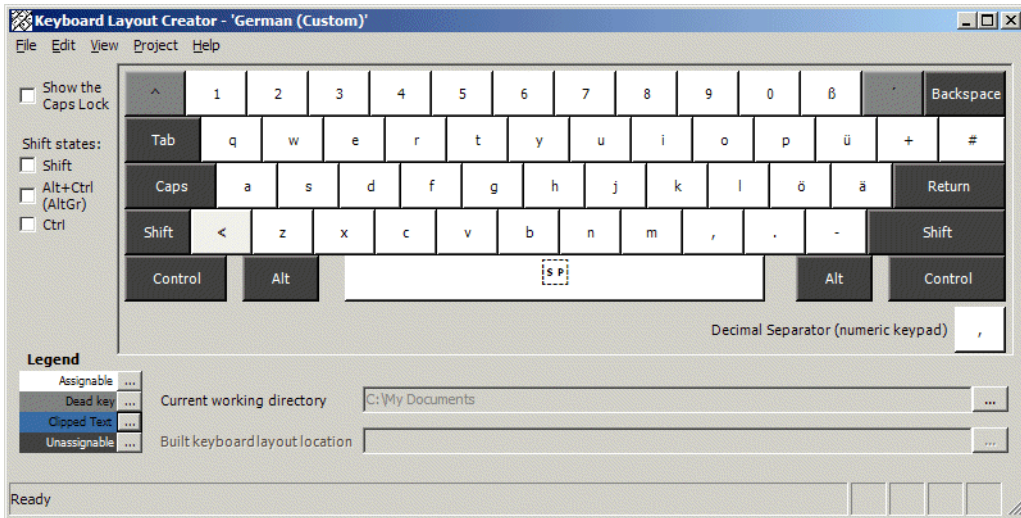


Figure 51: Custom German QWERTY keyboard layout with *MSKLC*

A little freeware utility by the name of *AllChars* (see sourceforge.net/projects/allchars/) not provides you with a large number of preconfigured keyboard shortcuts (that all follow the pattern CTRL (release!) + one key + second key):
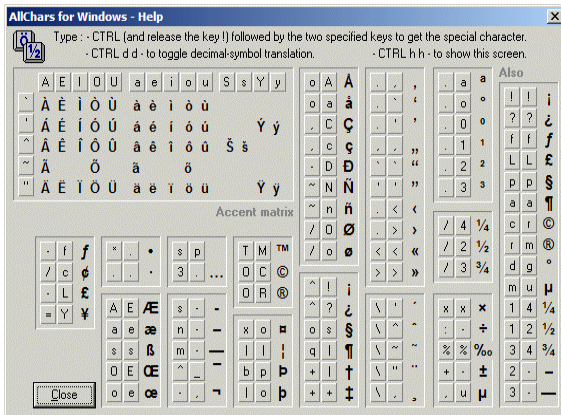


Figure 52: Pre-configured options in AllChars

and also allows you to configure your own combinations:
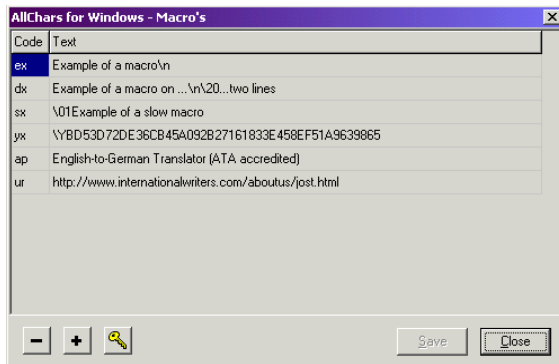


Figure 53: User-defined options for AllChars

This little utility works in almost any *Windows* application.

*My old version of AllChars (version 2.01) works much better for me than the presently available version 4, which interferes with the Alt key usage in some programs and has problems with user-defined macros. If you experience the same problem and cannot find an earlier version, I would be glad to send you a copy of the old program.*

One last thing about keyboards: There are very few things that I hate as much as when I hit the **Caps Lock** or the **Insert** key without knowing it and the following text is either in all caps or overwritten.

Fortunately, Windows allows you to have a little beep sound go off every time you hit the **Caps Lock**, **Num Lock**, or **Scroll Lock** keys. To activate this feature, select **Start> (Settings>) Control Panel> Accessibility Options** (Vista: **Start> Control Panel> Ease of Access Center> Make the keyboard easier to use**) and check **Use Toggle Keys**.

To make the **Insert** key beep every time you hit it, you can download the free and tiny *Insert ToggleKey* utility at www.mlin.net/misc.shtml.

# Windows Updates

A *Windows* feature that you should definitely be familiar with is the *Windows Update*, which gives you access to all security patches and other enhancements of your operating system as they become available.

In most installations of earlier versions of *Windows*, there is a link in the *Windows* **Start** menu that will take you to the appropriate website (windowsupdate.microsoft.com). In newer versions of *Windows* (*XP SP2* and *Vista*, and *7*) the update is performed without much intervention from your side.

A list will appear with a collection of customized updates for your computer (operating system, previous updates, version of *Internet Explorer*), categorized into critical updates, non-critical updates, and driver updates.
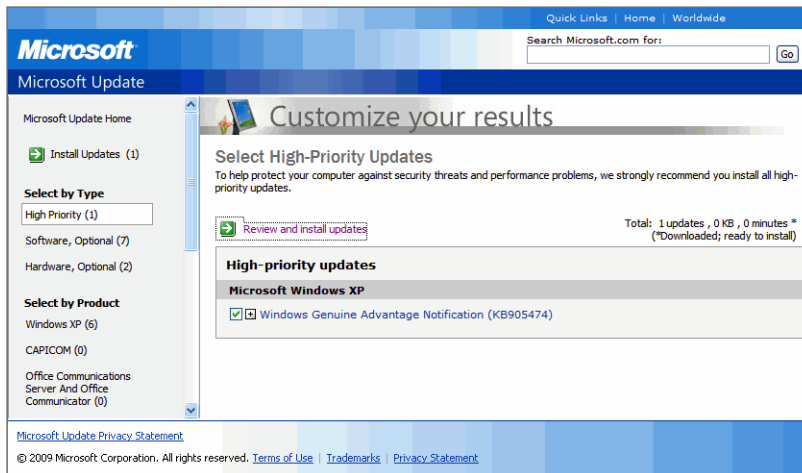


*Figure 54: Windows Update*

In general, you should make sure that you download the critical updates but carefully review whether you need any of the other updates.

In the fall of 2005 Microsoft introduced a new, more comprehensive service with *Microsoft Update*. *Microsoft Update* combines the *Windows Update* service with *Office Update* as well as several other similar services for other Microsoft products. It's quite easy to install—you can find a link on the present Windows update site.

Once this is installed and configured by you—you can choose to only be notified of available updates or you can have Microsoft download and install all updates for you—this service replaces the other services, i.e., if you try to go to the *Windows Update* site, it will reroute you to the *Microsoft Update* site.

Every few months Microsoft publishes a *Service Pack* (SP) for some of its *Windows* systems. (Or when an operating system disappoints completely: with the first release of *Windows 98*, Microsoft released a completely new *Second Edition* version, *Windows 98SE*.) The *Service Pack* can also be downloaded from the *Windows* or *Microsoft Update* site and is generally a bundle of all previously published security patches plus some other additions that Microsoft believes are enhancements.

# Web Browsers

Web browsers—those programs that help you locate and display web pages—are another of the rather emotional topics where everyone feels very strongly about the browser that he or she uses (especially if it is not *Internet Explorer*). I used to use *Internet Explorer* because it was one of the first to auto-detect different character sets and download the necessary fonts for it. However, since *Firefox* (see mozilla.com) now also does a good job with this, in addition to offering a huge number of add-ons to make it do virtually anything, I use *Firefox*.
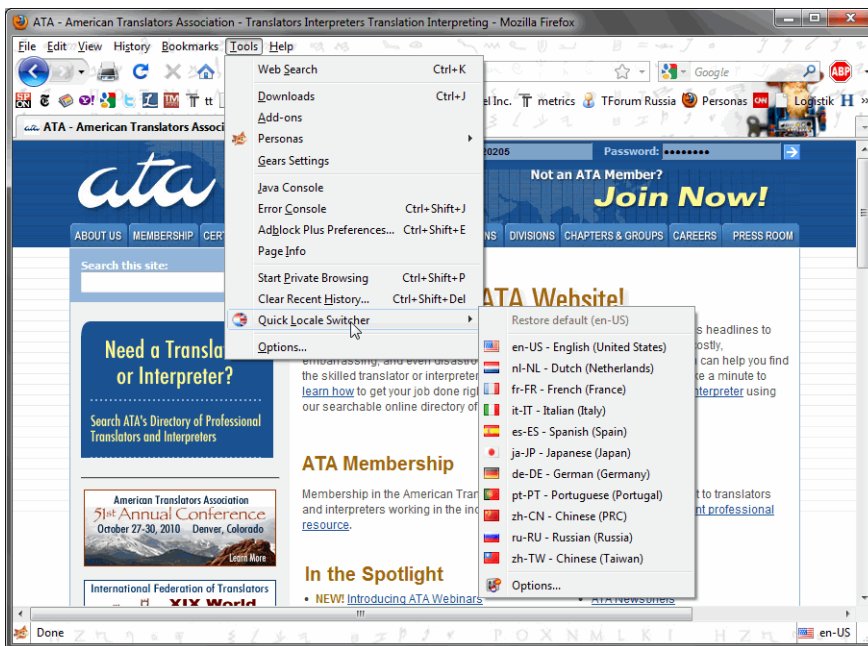


Figure 55: *Mozilla Firefox* with the *Quick Locale Switcher* add-on

*Plus, the Firefox browser is the only browser that allows me to have Jeromobot watch over everything I do. You can download the Jeromobot persona (skin) from the screenshot above under www.getpersonas.com/en-US/persona/242788.*

I also have *Opera*, *Safari*, and *Google Chrome* on my computer for testing my website and others.

## Browsing Tips

When I started translating professionally, the Internet was already a formidable resource that held a lot of translation-related information. But even I know the feeling of rummaging through books and other "hardware" material to find answers that I just couldn't find anywhere else.

While this will always remain so to a certain degree, here are some tricks that should make your Internet searches on *Google* and *Bing* just a little more focused.

Most everyone knows the use of quotation marks to find "just that specific expression," the + sign to force the search engine to include the following word in the search, or the - sign to specifically exclude sites that contain the succeeding word.

Here are some additional tricks:

- If you would just like to look at which web page titles contain the words "Chinese translation," type `intitle:"chinese translation"` (both *Google* and *Bing*).

- If you are interested in all web pages that have the word "translation" in their URL (website address), type `inurl:translation` (both *Google* and *Bing*).

- If you only want to look in the body text of websites (rather than the URL or the title), for instance to find out where your own web page is quoted, type `intext:www.<the name of your website>.com` (*Google* only).

If these tricks have not really impressed you, the next ones will:

- If you want to look for something in only a certain kind of document (such as a PDF file) and not in any other, type `filetype:pdf "translation memory"`. The result will be all PDFs that are registered with the search engines and contain the phrase "translation memory." If you would like to specifically exclude PDFs, you can type `-filetype:pdf "translation memory"` (both *Google* and *Bing*).

- If you want to return webpages for a specific language, you will just need to specify the language code directly after the keyword `language:`. For example, if you are searching for your name on Chinese-language websites, you will need to enter `John Doe language:zh` (*Bing* only).

- To return webpages from a specific country or region, you can specify the country or region code directly after the keyword `loc:` (or `location:`). You can even combine this with an OR search. For esample, to see webpages about machine translation from the U.S. or Great Britain, enter `"machine translation" (loc:US OR loc:GB)` (*Bing* only).

- Or how about this one: Unless you have one favorite online dictionary you always go to when you need a definition, you can also type `define:translation` (*Google* only).

## Using URLs to Find Translation Data

Sometimes it is very helpful  to interpret URLs (web addresses) from a language/translation point of view. There are numerous parts of a URL that could identify the language of the webpage that it displays. And if that is the case, chances are that the same webpage is also displayed in other languages (otherwise there is not much reason to note the language in the first place).

For instance, look at this long URL from the Microsoft help site:

http://windowshelp.microsoft.com/Windows/en-US/help/fe7ea80e-52a2-48d6-947a-05e02e78bc371033.mspx

This URL has two language identifiers. One is very obvious—en-US (in this case a mixture of the standards ISO 639-1 and ISO 3166). The other identifier may not be as obvious: the last four digits at the end are the widely used Microsoft Locale ID (see msdn.microsoft.com/en-us/goglobal/bb964664.aspx).

To change that page into, say, Japanese, you could just manually replace the URL in those two places with the appropriate codes:

http://windowshelp.microsoft.com/Windows/ja-JP/help/fe7ea80e-52a2-48d6-947a-05e02e78bc371041.mspx

and come out with the Japanese counterpart with all the Japanese terminology at your fingertips.

Or here is another one. There is a great English and German parallel SAP glossary at http://help.sap.com/saphelp_glossary/en/index.htm. To change the language in that case all you need to do is replace the "en" with "de".

But since this is a glossary within an HTML frame, it's not quite as easy to get to specific entries. If you click on any of the actual English entries in the above page, the URL does not seem to change. However, *Firefox* offers an easy way to let you open the page within the frame as a standalone page. Once you click on an entry and you have the English term and description displayed, right-click on that page and select **This Frame> Open Frame in New Tab**. This might open:

http://help.sap.com/saphelp_glossary/en/3b/
57a67b78608045852d629395c6844b/content.htm

And sure enough, just by changing it to:

http://help.sap.com/saphelp_glossary/de/3b/
57a67b78608045852d629395c6844b/content.htm

we get to the translated page.

While this particular example is only good for those who work in that language combination, there are many other cases where this can be adjusted easily to other websites and language combinations.

## File Transfer

If you send files by email, it's almost always a good idea to zip the files. Aside from reducing the upload and download time because of smaller file size, zipping adds an extra layer of protection to your files, does not write-protect your files, sends one file instead of many, and bypasses many virus protection applications that would otherwise block access to files with certain extensions.

*For information on zipping utilities, see Compression Utilities on page 127.*

For large files, you should not send attachments by email, but via the File Transfer Protocol (FTP), the same protocol that is used to upload files to websites. Although many browsers—especially *Internet Explorer*—have a decent or good FTP integration, I would always advise having an FTP program ("FTP client") aside from the main browser. The version that I use is *WS_FTP Pro* (see www.wsftp.com), but there are a number of other equally good applications out there including the open-source *FileZilla* (see www.filezilla-

project.org). My reason for suggesting the use of an FTP client is that I think it is easier to manage files (with a *Windows Explorer*-like interface) and it provides a helpful way to store a large number of FTP addresses for quick access.
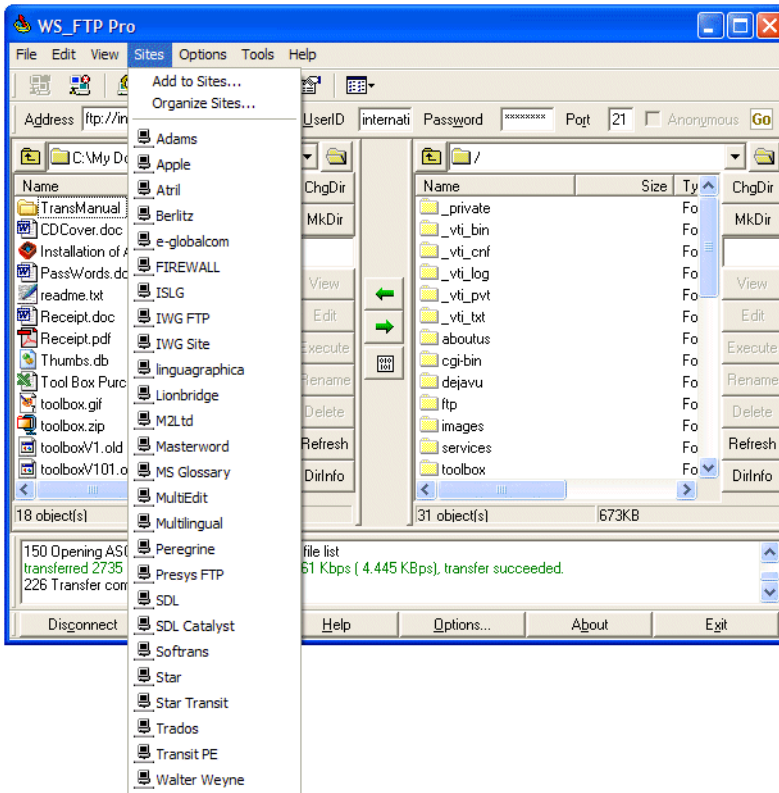


Figure 56: *WS_FTP* interface with direct access to a number of stored FTP sites

# Reducing Your Risk

This chapter is an introduction to the most common threats to computers and data. Like everything in the Tool Box, select just the tools that fit your particular need.

There is no reason to become paranoid. On the other hand, it's helpful to have an idea of what's out there so you can adjust your behavior and choose your defense mechanisms. After all, most—if not all—of us deal with sensitive data as translators.

Aside from the financial consequences that might result from unintended disclosures of the data you were entrusted with, a slip-up like that could also cause significant harm to your reputation.

I will begin by describing the most common threats to our computers and data, followed by a description of various tools that will let you reduce the risk posed by these threats. Finally, I'll include a few last words of advice on this subject.

## Common Threats

This section lays the groundwork by introducing some of the terminology you might encounter when reading about computer threats.

### Malware

Malware (**Mal**icious soft**ware**) is software that was designed to harm or enter a computer system without its owner's informed consent. The term refers to a variety of forms of software or program code that are hostile, intrusive, or annoying.

These are the most common types of malware:

#### Virus

Computer viruses have been around since the dawn of personal computers, so it's likely that you've already made a more than personal acquaintance with one or more of them.

A computer virus is a program that can copy itself and infect your computer without your permission or knowledge. Until a few years ago, the most common way for a computer virus to spread was by removable media, such as a floppy disk, CD, or a USB drive. With the increased use of the Internet, e-mail, cloud-based services and file sharing, these have become common vehicles for attacks, too.

Just as in real life, viruses usually require an infected host.

### Worm

A computer worm can spread itself to other computers without needing a host for the transfer.

The havoc that a worm can wreak is limited only by the author's imagination. The more common attacks focus on creating backdoors on computers or turning computers into "zombies" (see page 77). Often these "zombies" are combined into systems called "botnets" (see page 78).

### Trojan Horse

A Trojan horse is a computer program that disguises itself as something else while it installs malicious software on your computer.

Similar to its namesake, the Trojan horse of classical mythology, malicious code is hidden in an apparently useful, interesting, or at least harmless program or file. Once the unsuspecting user executes the program or file, the malicious code is also executed and installs a backdoor program (see page 75). This backdoor program either allows unauthorized remote access to your computer by other parties, or it installs a keylogger (see page 75) that captures data you enter—such as login and password information, account and PIN numbers—which then are transmitted to a third party.

Quite often, social engineering techniques (see page 76) are used to lure you into opening or executing these files and programs.

### Spyware

Spyware is software that is installed surreptitiously on your computer without your informed consent, intent on intercepting your interactions with the computer or taking over partial control.

Not only can spyware collect all kinds of information about you and your web-surfing habits, but it may also redirect your web browser activity without your knowledge.

### Backdoor

A backdoor to your computer is a program or method that bypasses normal authentication and protection, allowing for remote access to your computer.



*Although a hardware device might also be used for creating a backdoor, most of us do not have visitors or customers onsite at our offices, so the risk for a hardware-based backdoor is most likely negligible.*

### Keylogger

Widely available on the Internet, keyloggers can be either software- or hardware-based. Their purpose is to spy on the computer usage of others, for example, by obtaining usernames and passwords or by capturing authentication information of online banking users.

### Adware

Adware is software that has advertising functions integrated into it or bundled with it. It may automatically display or download ads to your computer after that software has been installed or while it's being used.

There are some comparatively benign specimens—like the old advertising-supported versions of the *Opera* web browser and the *Eudora* e-mail client—but you may also encounter some rather obnoxious specimens that generate pop-ups (see page 78) on your computer without any web browser open.

### Rogue Security Software

Rogue security software is computer malware that pretends to remove malware but does not do anything or itself installs other malware. Often it disguises itself under a name that is similar to well-known and legit programs. You can find a list of rogue security software applications on this Wikipedia page: http://en.wikipedia.org/wiki/List_of_rogue_security_software.

# Attacks

In addition to the malware mentioned above, there are also a few other types of attacks that may affect us.

Here's a brief description of each of these to help you understand them better.

## Phishing

Phishing is an attempt to acquire sensitive information—such as usernames, passwords, account information, or social security numbers—by masquerading as a trusted entity via electronic communications. To achieve this goal, phishing often employs "social engineering techniques" in an effort to fool users.
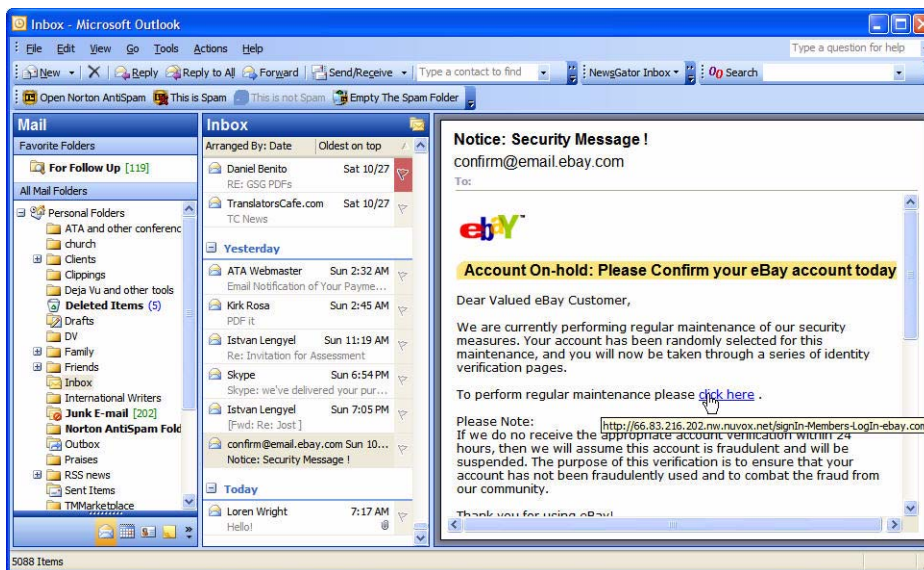


Figure 57: Phishing attempt with fake eBay email

This screenshot nicely illustrates one such phishing technique. At first glance this might look like an e-mail from eBay, but on closer inspection it becomes clear that this is an imposter's phishing e-mail. The first giveaway is that there is no specific recipient listed. More importantly, when you hover your cursor over the alleged eBay URL, a different URL is revealed as a tool tip.

Another form of phishing is advance-fee fraud, AKA "Nigerian scams." Any one of you will have received emails that promise great riches in exchange for an upfront payment, but recently there have been some of those emails in distribution that were specifically aimed at translators. Ted Wozniak, the owner of the translator payment watch list paymentpractices.com, has taken it upon himself to compile a list of those for reference purposes at www.paymentpractices.net/Scams.aspx.

## Drive-By Download

Sometimes also called a "drive-by installation," this term refers to a download and installation that occurs without your knowledge, and thereby without your consent, rather than just the mere download of some type of malware.

Such drive-by downloads can happen when you visit a website, view an e-mail message, or click on a pop-up window. While some drive-by downloads require a very limited amount of user interactions, such as a mouse click, others may exploit a vulnerability in the operating system or in an application, such as your e-mail client or your web browser.

## Denial-of-Service (DoS)

A denial-of-service attack (DoS attack) attempts to render a computer resource—such as an Internet site or a service—unavailable to its users.

Typically, DoS attacks attempt to consume the resources of a targeted computer so that it can no longer provide its services, or they interfere with the communication media between the target and the users so they can no longer communicate properly.

If multiple computers, quite often from different geographical areas, are involved in such an attack, this is referred to as a Distributed-Denial-of-Service (DDoS) attack, typically executed by zombie computers organized within a botnet (see page 78).

## Zombie

A zombie computer is a computer that is connected to the Internet and has been compromised by a virus, a Trojan horse, or a hacker to make it accessible to the people who "own" the compromised systems. The actual owner of the computer tends to be unaware that his or her system is being

used to send e-mail spam (see page 79), commit click fraud, conduct a denial-of-service attack (see page 77), or for other nefarious purposes. Typically, a compromised machine is just one of many under remote direction in a botnet.

## Botnet

A botnet is a collection of software robots, or bots running automatically and autonomously on zombie computers remotely controlled by crackers (**cr**iminal h**ackers**) via a common command and control infrastructure.

## Man-in-the-Middle (MITM)

A man-in-the-middle attack allows the attacker to read, insert, and modify messages between two parties without either party knowing that the link between them has been compromised.

The attacker may simply be eavesdropping to obtain victims' details during a phishing attack, or he or she might jam all communications to one party (denial-of-service) or use information gathered to execute a replay attack.

## Nuisances

In comparison to the malware and attacks that specifically target your computer, the following issues are mere nuisances, though they can be quite aggravating if you don't have the right tools at hand to repair their damage.

### Pop-up Ads

Pop-up advertisements or pop-ups are a form of Internet advertising aimed at driving traffic to a website or capturing e-mail addresses and other identifying information.

Pop-under ads are a variation of this technique that opens a new window underneath the active window rather than on top of it. Typically you don't see them until you close your current browser window, so it becomes a lot more difficult to determine which website originally opened this pop-under window with the ad in it.

**Spam**

The terms e-mail spam, bulk e-mail, junk e-mail, UCE (unsolicited commercial e-mail), and UBE (unsolicited bulk e-mail) all refer to nearly identical messages sent to a large number of recipients via e-mail.

No matter what they're called, we don't like them!

**Tracking**

Some websites try to track users' browsing habits by using cookies, small chunks of text sent to the browser by the server and then sent back unchanged upon each subsequent access to that server.

There are justified uses of cookies that are helpful to the user, such as authentication, configuration of site preferences, and electronic shopping carts. However, so-called tracking cookies, such as those third-party cookies catered by DoubleClick, are frowned upon, and more and more users have privacy concerns about the tracking of their browsing behavior.

Cookies are only data, not program code; therefore, they cannot delete information or read information from your computer. They do not generate pop-ups, nor are they used for spamming or for advertising.

It is also quite common for website operators and advertisers to embed tiny transparent (or colored) GIF images (typically a single pixel in size) on web pages as a means of tracking who accesses each page how often. These little critters are also referred to as web bugs.

In recent years, some companies have started to offer tracking of e-mails. This is usually achieved by placing a reference to a tiny transparent or inconspicuously colored image with a unique name in an HTML-based e-mail. This is unseen by the recipient, but it triggers an access log entry for the image when the e-mail is opened.

Not everybody likes the thought of inadvertently revealing to both a third party and the sender when they first opened their message and how often they looked at it.

# Tools for Reducing Risk

This section describes some of the countermeasures that are available and introduces you to a variety of tools that can reduce the risks posed by these threats.

While there are some hardware-based solutions available, most solutions consist of software. It's very important, however, not to forget about one of the most powerful defenses available to us, free of charge: common sense.

## Hardware

### Router

Most of us by now use some form of high-speed Internet connection, typically either a digital subscriber line (DSL) or a cable modem. Unfortunately, there are still a significant number of users who don't yet have the most basic means of protection in place that can render their computer virtually invisible to the Internet. The device I'm talking about is called a *router*.

Virtually all of the cable modem / DSL routers currently available block unsolicited traffic arriving from the Internet. Any "legitimate" network traffic from the Internet sent in response to a request from your computer—e.g., retrieving e-mail, browsing web pages, or downloading files—is let through and can get to your computer.

An additional benefit of these devices is that they are typically also equipped with four network ports, allowing you to set up a small network of computers at your workspace that can all share your high-speed connection.

I have been using *CISCO Linksys* routers for several years now and am quite happy with their performance and reliability as well as ease of administration. In the past few years, CISCO Linksys (see www.linksys.com) and other router manufacturers such as Netgear (see www.netgear.com) and D-Link (see www.dlink.com) have streamlined the initial setup to pretty much be plug-and-play. Just follow their install instructions and you'll be fine.

If you want to add an additional layer of protection, consider changing the default password when you are prompted for it during the install process. Otherwise, you can also do so via the CISCO Linksys router's administrative interface, which is accessible via the web browser of any computer attached to its LAN ports. In its default configuration, you simply have to point your browser to http://192.168.1.1.

*You can enter Netgear routers by entering 192.168.1.1 or 192.168.0.1 and D-Link routers by entering 192.168.0.1 into the address field of your browser. If you use another brand of router, please check your owner's manual for the address you need to access its administrative web interface*

On the **Authentication** screen, you will enter a password rather than a user name. If you changed the password, use whichever password you assigned to the router; if the router still has its default password, it would be `admin`.
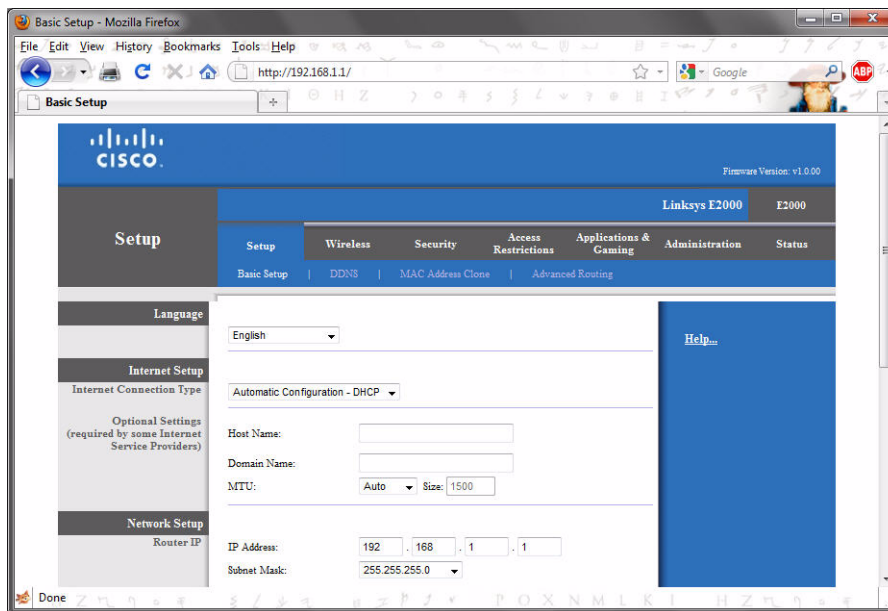


Figure 58: The router's setup interface after successful login

To change your password, click on **Administration**; this screen defaults to **Management** where you can change your router's password by entering it in the appropriate field and re-entering it to confirm. Don't forget to click on **Save Settings**.
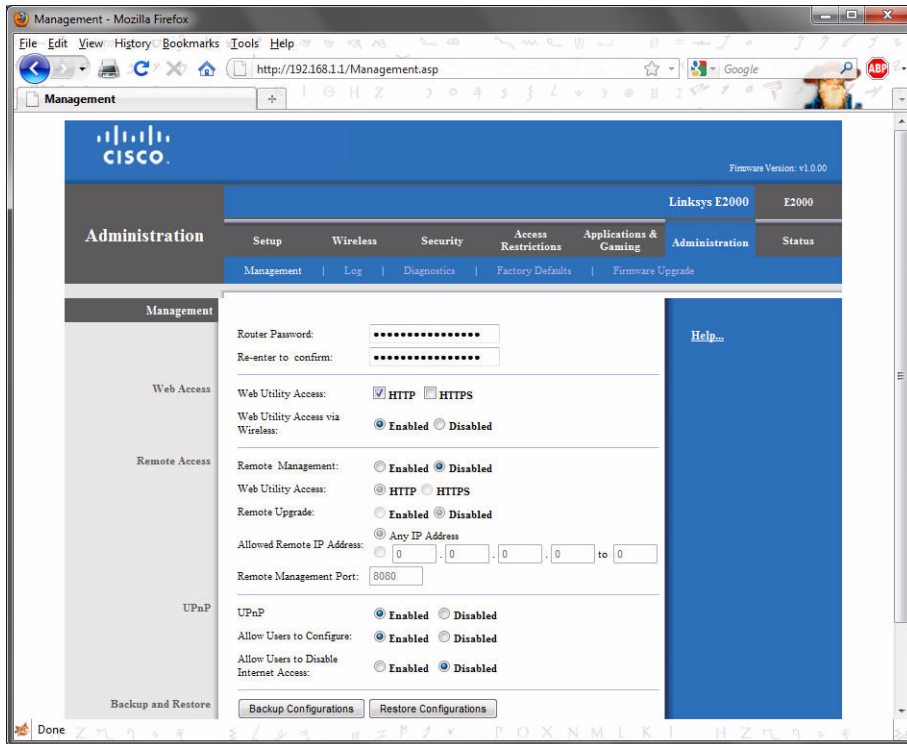


Figure 59: The management interface

If you opted to get a wireless router, make sure that it is configured to use encryption. If at all possible, elect not to use *Wireless B* (802.11b), since this rather dated technology only allows for the *Wireless Equivalency Protocol* (WEP) for encryption, which can be easily hacked into with freely available applications.

If you use *Wireless G* (802.11g) or *Wireless N* (802.11n), currently the most common types of wireless, make sure to configure your router's wireless network mode to not support *Wireless B*, otherwise you will not be able to use the much better protection afforded by *Wireless Protected Access* (*WPA* or, even better, *WPA2*) for encryption.

You can configure the network mode of your CISCO Linksys router under **Wireless> Basic Wireless Settings**.
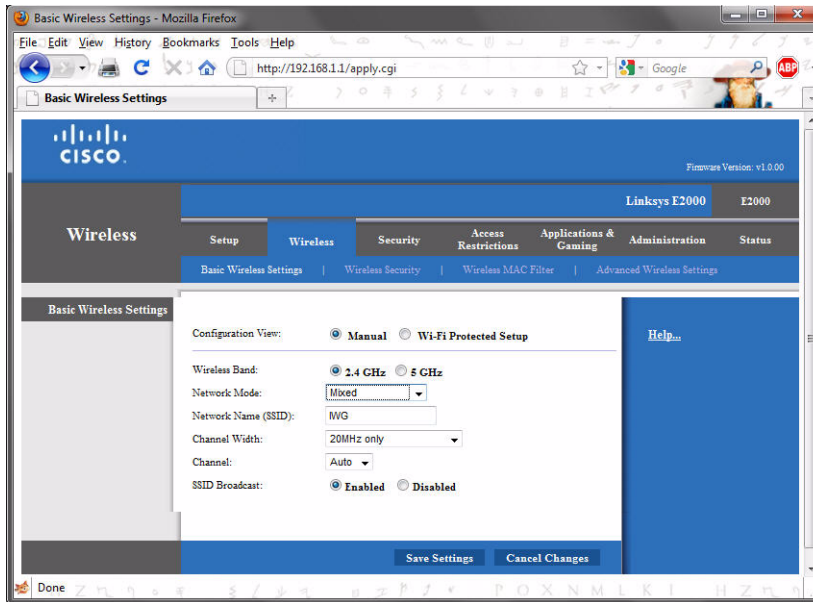


Figure 60: Basic wireless settings

Once again, don't forget to click on **Save Settings** when you are done with this step.

Now that you've restricted the type of networks you want your router to support, set up your wireless security to use **WPA Pre-Shared Key** so that only users who know the string you are using as your pre-shared key will be able to connect to your network.

You can configure the security mode of your CISCO Linksys router and enter the passphrase under **Wireless> Wireless Security**.
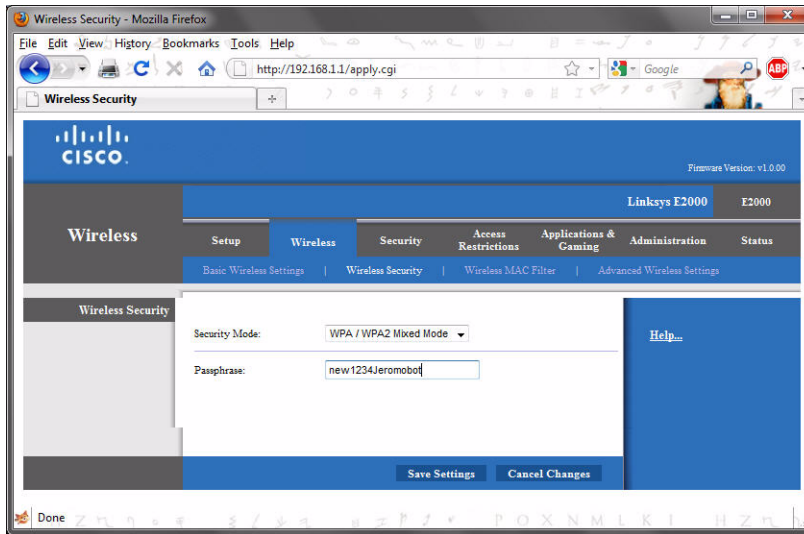


Figure 61: Entering a wireless key

As you can see from the above screenshots, there are a wide variety of additional features and settings available for those of us who want to venture into this terrain, but the steps outlined above are sufficient to provide a reasonable level of protection.

## Hard Disk with Built-in Encryption

If somebody were to steal your computer or even just your hard disk, they will be able access the data on the disk unless it is encrypted. Several vendors such as Seagate (see www.seagate.com) have now started to offer hard disks with built-in encryption.

*Alternatively, the whole disk can be encrypted using disk encryption software such as PGP Whole Disk Encryption (see www.pgp.com/ products/wholediskencryption). Once the whole disk has been encrypted, you are then prompted for your password / pass phrase each time your computer is booted. Your computer's disk thus is rendered inaccessible to anybody who does not know the necessary password or pass phrase.*

*If you don't see a need for whole disk encryption but think that encrypting a physical or virtual volume might sufficiently serve your needs, you might not have to look all that far from home, as long as your computer is running Windows XP Professional, Windows Vista (Enterprise or Ultimate), or Windows 7 (Professional, Ultimate, or Enterprise). These operating systems include EFS, the Encrypting File System, which optionally can be activated to encrypt files and directories.*

*Note that with EFS, plain text files can be easily recovered. To circumvent that you should configure encryption at the folder level. When you wish to encrypt individual files, copy them to an encrypted folder.*

## Privacy Filters

Privacy filters are thin sheets of plastic that are about 1 mm thick and can be mounted in front of the display, darkening the screen to anybody who is not sitting at close to a 90-degree angle in front of the screen. This can be helpful when you're working on confidential information in a public place, or if you simply don't want someone staring at your screen while you work.

## Cable Lock

Cable locks are available from a variety of vendors. They are primarily used for laptop computers, hooking into their so-called Kensington Security Slot (K-Slot) and allowing you to loop the cable around an immobile object, such as part of a desk, before securing it in the K-Slot. This way you don't have to worry about leaving your laptop at the hotel while you are shopping, meeting with friends, or having a drink.

# Software

To help address the various risks outlined at the beginning of this chapter, a wide variety of software is available to combat those risks.

There are a number of separate products that target each of these threats individually in a highly specialized fashion, but it's easier for most of us to deal with the bulk of these risks in an all-in-one fashion rather than with half a dozen or more separate products.

Another reason for choosing an all-in-one solution is that by now, virtually all providers of "free" security products stipulate that their "free" product is only free for personal, non-commercial use. Because we earn a living with the help of our computers, it may not be the most ethical thing to violate these licensing agreements by operating our business computers with a product that was only licensed for personal use.

And once you have to pay for a personal firewall, antivirus, anti-spam, or anti-phishing software, you might as well shell out a few dollars more and obtain one of the various "Internet Security" suites available from various companies.

First, I would like to spend a little time to briefly describe the type of products and the components of the suites that are available for risk reduction.

### Firewall Software

Although you (hopefully) already have a router in place (see page 80) to protect you from attempts to access your computer from the Internet, this protection is not available while you are on the road, nor does a router protect you from an infected computer located on your local area network (LAN)—for example, if another computer on your network gets infected or taken over by hackers. In addition, most routers for the small office and home office market consider all outgoing network traffic to be benign, even though this is not always the case.

This is where firewall software comes into play. If it is a full-fledged firewall, it will check both incoming and outgoing network connections.

*Up until Windows Vista, the Microsoft Firewall only inspected incoming traffic but did not look at outgoing traffic, thereby lulling its users into a false sense of security.*

In cases where no pre-defined rules are available, the firewall software typically generates a pop-up message that asks you for a decision. To allow you to make a qualified and informed decision, a "more info"-type link or button is generally provided which opens a browser window that gives additional details about the application. More often than not, it also offers a recommendation on how to handle it.
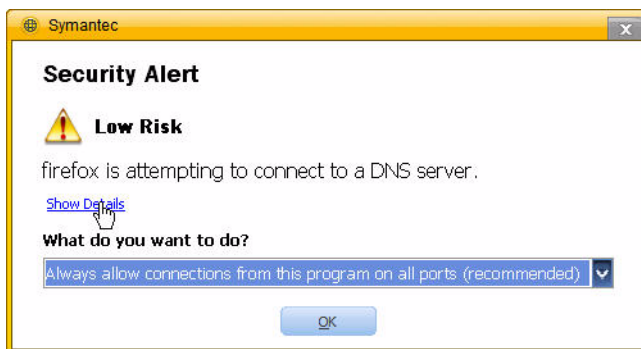


Figure 62: Entering a wireless key

If you elect to combine separate specialized applications and choose to buy a separate personal firewall product, the better-known personal firewalls are

- *ZoneAlarm* (see www.zonealarm.com),

- *Sunbelt* (formerly *Kerio*) *Personal Firewall* (see www.sunbelt-software.com), and

- Comodo Firewall (see personalfirewall.comodo.com).

Also, some companies have elected to offer only complete Internet security suites which contain a firewall component without also offering the firewall as a separate product. Among these contenders are:

- *F-Secure Internet Security* (see www.f-secure.com),

- *Kaspersky Internet Security* (see www.kaspersky.com),

- *Symantec's Norton Internet Security* (see www.symantec.com), and

- *McAfee Internet Security Suite* (see www.mcafee.com).

## Antivirus Software

Antivirus software attempts to identify, fend off, and remove computer viruses and some other malware.

Typically, this is done in a two-pronged approach:

- A memory-resident component monitors your computer for suspicious behavior and inspects all programs and files that are opened for certain "signatures" typical of malware. This also includes the content of web pages you are browsing.

- Regularly scheduled scans of all the files on your computer's hard disk(s) look for files containing the "signatures" that might indicate an infection.

While traditionally the primary focus of antivirus software has been on signature/definition-based detection, which is a reactive approach, information security experts in recent years have pushed more and more for a proactive approach in the form of behavior-based detection.

In recent years, *Norton Antivirus* and *McAfee Antivirus* have started to offer versions of their antivirus products which allow for installation on up to three computers. Among the other contenders for protection against computer viruses are *Panda* (see www.pandasecurity.com), *Kaspersky* (see www.kaspersky.com), and *Grisoft AVG* (see www.grisoft.com).

## Anti-Spyware

Anti-spyware software attempts to identify, block, and remove spyware and some other malware, such as adware.

This takes a similar approach to that of antivirus software: a memory-resident component plus regularly scheduled scans.

If you are using *Windows XP*, you should take advantage of the free *Windows Defender* (www.microsoft.com/athome/security/spyware/software/default.mspx), an anti-spyware acquired by Microsoft. *Windows Defender* is included with *Windows Vista* and *7*.

For most individual business users, Safer Networking's *Spybot – Search & Destroy* (see www.safer-networking.org/en/spybotsd) would be one of the few options that are available free of charge. The licensing terms of Lavasoft *Ad-aware*'s (see www.lavasoftusa.com) or SpywareBlaster's (see www.javacoolsoftware.com/spywareblaster.html) free version restrict it to personal, non-commercial use.



*With any of these products it is important to make sure that  program updates are regularly applied.*

In addition, virtually all Internet security suites now include an anti-spyware component.

## Anti-Phishing

Anti-phishing measures can be found embedded in most recent versions of the popular browsers, such as *Internet Explorer*, *Opera, Google Chrome* and *Mozilla Firefox*, and also as extensions and toolbars for browsers.

In *Internet Explorer 7* (no such thing in earlier versions of *IE*), the phishing filter is accessible under **Tools> Phishing Filter**, in *IE 8* under **Safety> SmartScreenFilter**.



Figure 63: Phishing prevention in *IE 8*

In *Mozilla Firefox*, you can find the phishing filter settings under **Tools> Options> Security**.
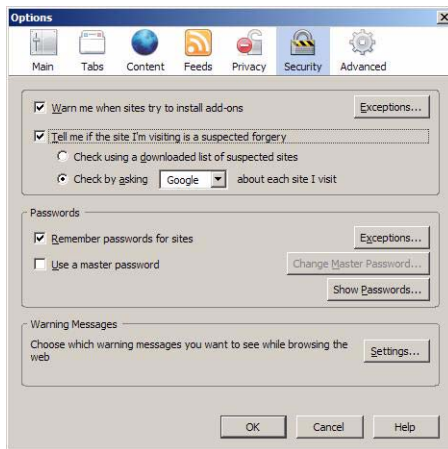


Figure 64: Phishing prevention in *Firefox*

In *Opera*, this feature is called "Fraud Protection" and can be found under **Tools> Preferences> Advanced> Security**.
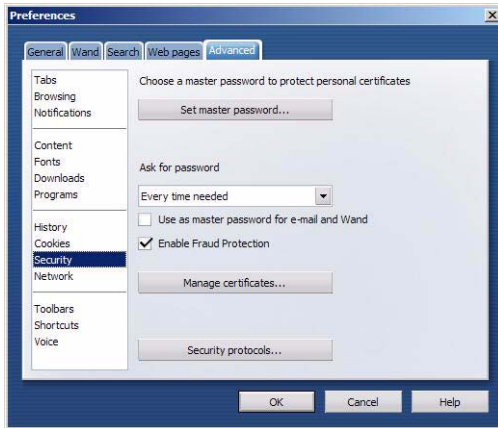


Figure 65: Phishing prevention in Opera

In *Chrome* you can set phishing protection under **Tools> Options> Under the Hood**.
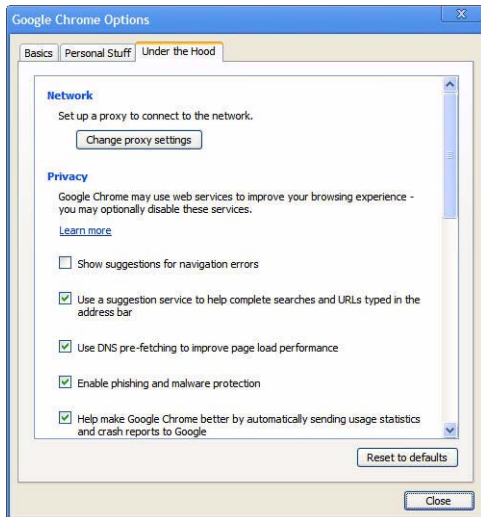


Figure 66: Phishing prevention in *Chrome*

One other way to prevent landing at undesired places is to use the *Firefox* add-on *LongURLPlease* (see addons.mozilla.org/en-US/firefox/addon/235815/ ). This add-on displays artifically shortened URLs (for instance, tiny.com or bit.ly) in their long form in a tool tip when you put your cursor on it. This is especially helpful when browsing through services like Twitter.

### Pop-Up Blockers

While there used to be a definite need for dedicated pop-up blockers to curb the flood of pesky pop-up or pop-under ads, basic pop-up blocking functionality these days can be found in all the major web browsers. Some of the newer browsers, such as *Google Chrome*, even have pop-up protection enabled by default.

In *Internet Explorer* 7 and *8*, you can exercise limited control over the pop-up blocker settings under **Tools> Pop-up Blocker Settings**.
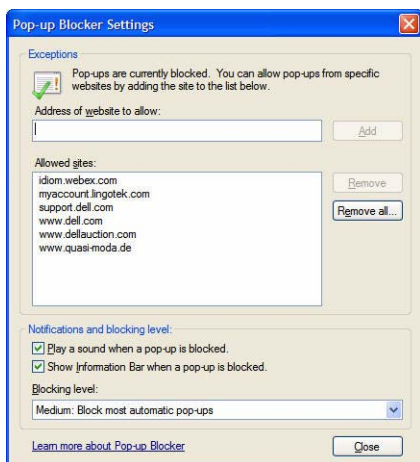


Figure 67: Stopping pesky pop-ups in *Internet Explorer*

In *Mozilla Firefox*, you can activate or deactivate pop-up blocking under **Tools> Options> Content**.
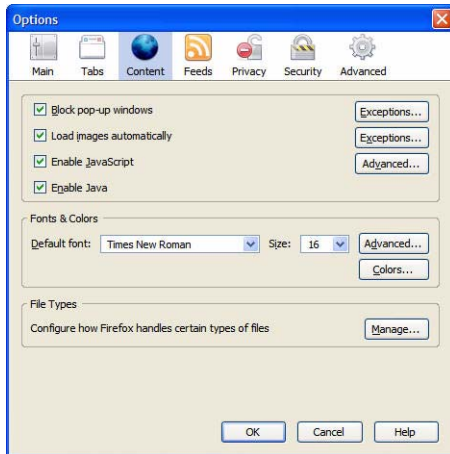


Figure 68: Stopping pop-ups in *Firefox*

In *Opera*, you can exercise a certain amount of control over pop-ups under **Tools> Preferences> General**.
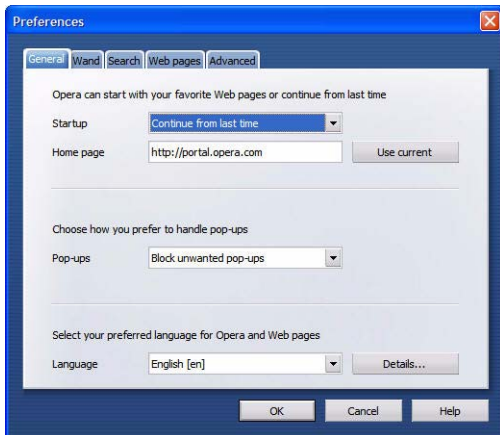


Figure 69: Stopping pop-ups in *Opera*

*For those who want to exercise even more control over the behavior of the various ads, there are also toolbars—such as the Google Toolbar (see toolbar.google.com) and Yahoo! Toolbar (see toolbar.yahoo.com)—which include pop-up blocking options, as well as third-party products—such as the Firefox add-on Adblock Plus (addons.mozilla.org/en-US/ firefox/addon/1865), and the browser-independent Privoxy (see www.privoxy.org), which goes one step further by also keeping most of the other ads away from your screen.*

### Anti-Spam

To curb the huge amounts of unsolicited commercial e-mail flooding your inbox, a variety of techniques and approaches are available.

One of the easiest ways of filtering out spam early on is to activate the spam filter(s) offered by your Internet Service Provider (ISP) or your mail service provider (MSP).

The catch with this approach is that you will regularly have to check your ISP's or MSP's filtered-out mail to make sure that your customers' e-mail or messages from potential new clients have not accidentally been misclassified as spam.

If you are using applications such as *Mozilla Thunderbird*, *Outlook 2003 and above*, or *Windows (Live) Mail*, you already have a pretty reliable and powerful spam filter built into your mail application.

In *Mozilla Thunderbird*, which refers to spam as "Junk" and "E-mail Scams," you can control the behavior of the filter under **Tools> Options> Privacy**.
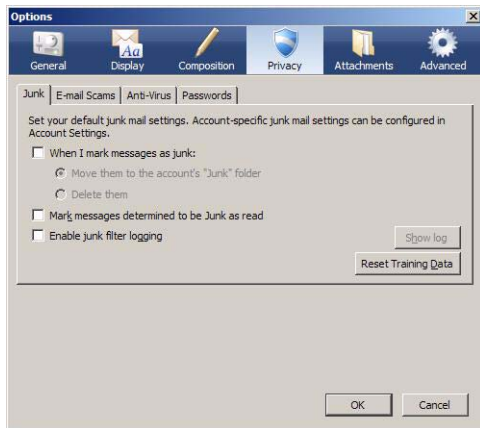


Figure 70: Spam options in *Thunderbird*

To configure "Junk Mail" handling in *Outlook 2003* and *2007*, go to **Tools> Options> Preferences> Junk Mail** (in Outlook 2010: **Home> Junk> Junk E-mail Options**).
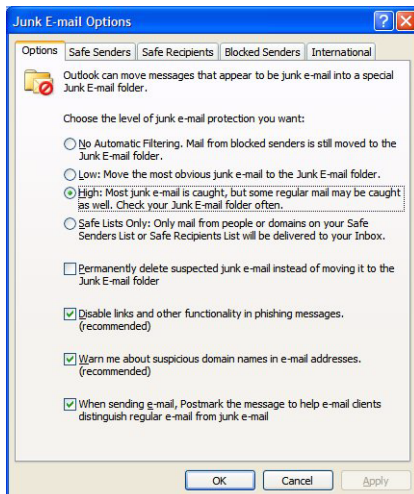


Figure 71: Spam options in *Outlook 2007*

In *Windows Live Mail* (the mail program that came as a downloadable option for later versions of *Vista* and *Windows 7*) the safety options are available under **Tools> Safety Options** or, if the menus are not enabled, under **Menus> Safety Options**.
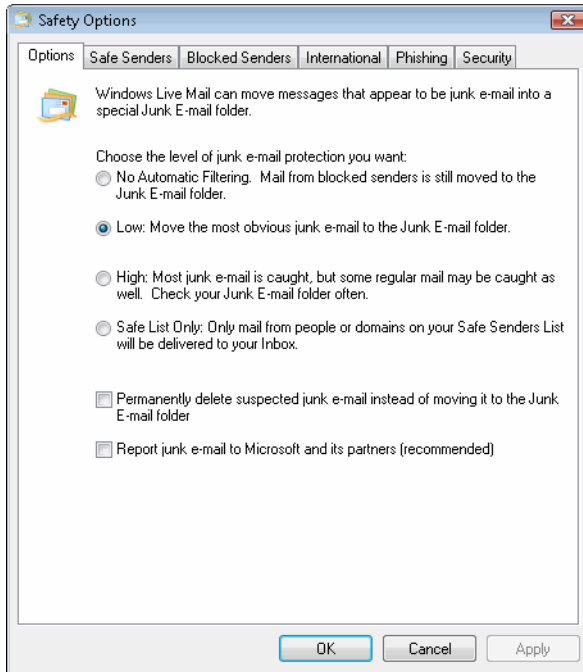
Figure 72: Spam options in *Windows Live Mail*

Virtually all the major Internet security suites include an anti-spam component which tends to be capable of identifying spam to a varying degree and which allows you to reclassify messages as spam or "ham" (non-spam).
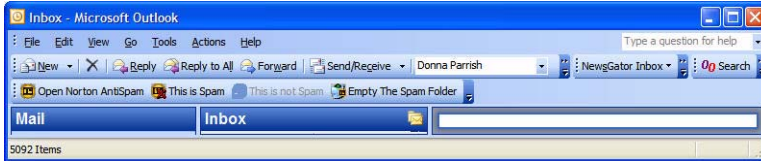


Figure 73: *Norton AntiSpam* filter in *Outlook*

*If you don't want to let spam get to your computer in the first place, or you don't want to depend on the Microsoft mail clients catered to by most of the current anti-spam products, you may want to look at a product like Mail Washer (see www.mailwasher.net). This product plugs in between the mail server and your mail client. The major catch with Mail Washer is that you can't retrieve your e-mail using your mail client; instead, you have to use Mail Washer, which then invokes your mail client. You must configure your mail client to not retrieve mail from your mail server itself.*

# A Few Final Words of Advice

There is no need to be scared, worried, or overly paranoid about securing your computer and protecting the information stored on it.

Simply install a router between your high-speed DSL or cable-modem connection and your computer or home/office network, put one of the various Internet security suites mentioned above on your computer, and you have already established a very good baseline of protection.

If you want to go the extra mile, consider whole-disk encryption or one of the hard disks with built-in encryption options.

But most important of all, apply common sense:

• You wouldn't want to fall prey to a scammer or non-reputable business, so apply due diligence and briefly google the name of new tools you are considering buying.

- Be cautious with the use of repair services for your computer and hard drive. If you have information on your hard drive that you would prefer not get into the wrong hands, consider having your computer looked at with its hard disk removed.

- Be cautious about selling your old computer or hard drive. If you have information on your hard drive that you would prefer not get into the wrong hands, consider removing your hard disks and physically destroying them, for example, with a sledge hammer. If this step strikes you as too violent, consider using the secure shredder function provided by various security products, including *Spybot – Search & Destroy* and *PGP Desktop*.

- Stay up-to-date on firewall, antivirus, and anti-spyware signatures, so that you always have the most current protection.

- Stay up-to-date on security-related patches for both your operating system (see *Windows Updates* on page 64) and any applications you are running.



*A very cool program to do this with is Secunia Personal Software Inspector (see secunia.com/vulnerability_scanning/personal/)—it gives you a list of all necessary updates and links to those in one fell swoop. And whatever Secunia does not find, FileHippo will (www.filehippo.com/ updatechecker).*

# Text and HTML Editors

## Text Editors

*"Text editors" are programs that allow you to create and edit text files, often with advanced editing features for software source code. The term "word processor" typically refers to programs that support many additional formatting features, such as Word or WordPerfect.*

Every version of *Windows* comes with two text editors: *Notepad* and *WordPad*.

*Notepad* (accessible under **Start> Programs> Accessories> Notepad**) is a very limited (and sluggish) text editor that is the *Windows* default program for any kind of text file. It can be useful if you want to look at the underlying code of any file (it will always open compiled formats, including *Word* documents, in the source code view).



Figure 74: Translation memory exchange (TMX) file in *Notepad*

*WordPad* (accessible under **Start> Programs> Accessories> WordPad**) is a text editor with some more functionality, including some formatting capabilities and the ability to open *Word* files in the intended formatting (without actually opening the memory hog *Word* itself).
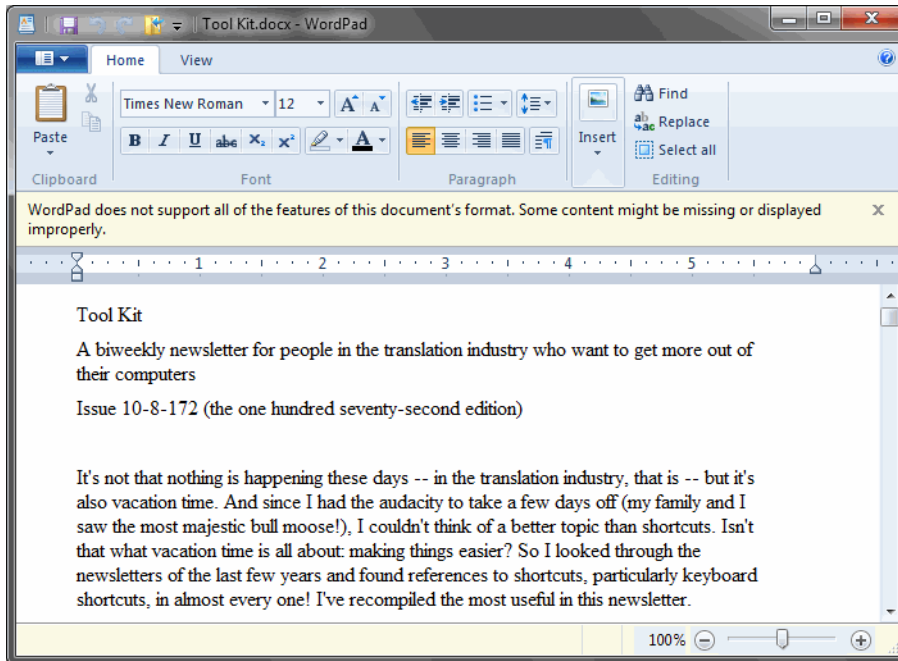


Figure 75: *WordPad* in *Windows 7* with an open *Word 2010* .docx document

There are, however, much more powerful (and fairly inexpensive) programs out there. Most of these have been developed by developers for developers, so we as translators usually only see the very surface of what these programs can do, but this is usually enough to be duly impressed. The most commonly known ones for *Windows* are probably *TextPad* (see www.textpad.com), *UltraEdit* (see www.ultraedit.com), and *EmEditor* (see www.emeditor.com). I use *UltraEdit* for all Western languages and *EmEdit* for East-Asian languages. But regardless of what I use, all of these programs do wonderful things and are typically very sufficient for what we need as translators.

Here are some examples of what these programs do:

Have you ever tried to open a 20- or 30-MB text file in *Word*? Depending on the speed of your computer, this can take up to a few minutes as *Word* tries to load the whole thing all at once. In contrast, these little text editor programs can open any size text file in just a few seconds. For instance, many of you have worked with the so-called Microsoft glossaries. Searching through these glossaries with *Excel* or *Word* renders them practically unusable because they are so large and response time is much too slow.

*The original "Microsoft glossaries" are not really glossaries; instead they're large translation memories with the translation data of the user interface for many of Microsoft's software products. From 1994 through the summer of 2006 they were available for free on one of Microsoft's FTP sites. In 2006 these files were replaced with a multilingual glossary, which in 2009 was replaced by the Microsoft Language Portal at www.microsoft.com/language. The portal offers an online search interface to Microsoft glossaries and translation memories, access to the Microsoft style guides for many languages and the ability to download extensive glossaries in the termbase exchange (TBX) format (see page 270). Many translation environment tools now offer support for TBX so that you can import the files right into your termbase. And for users of tools that do not yet offer TBX support, you can use the free and powerful ApSIC Xbench (see page 272) to convert the TBX files into something more palatable.*

*The translation memory files that were previously available on the FTP site can still be accessed on MSDN (msdn.microsoft.com/subscriptions) or TechNet (technet.microsoft.com/en-us/subscriptions). In order to access the files (which are in the same CSV format as before), you will have to pay an annual fee for the various available levels. The MSDN Operating Systems subscription costs US$699 and the TechNet subscription is available for US$349 (in August 2010).*

*To search these files from Microsoft you can also use the above-mentioned ApSIC Xbench.*

In the example below, a search for `directory structure` in all of the old German "Microsoft glossaries" (a total of 61 files and 98 MB) took *UltraEdit* less than ten seconds and resulted in a list of all the occurrences in all the files and simultaneous links to the exact occurrence.
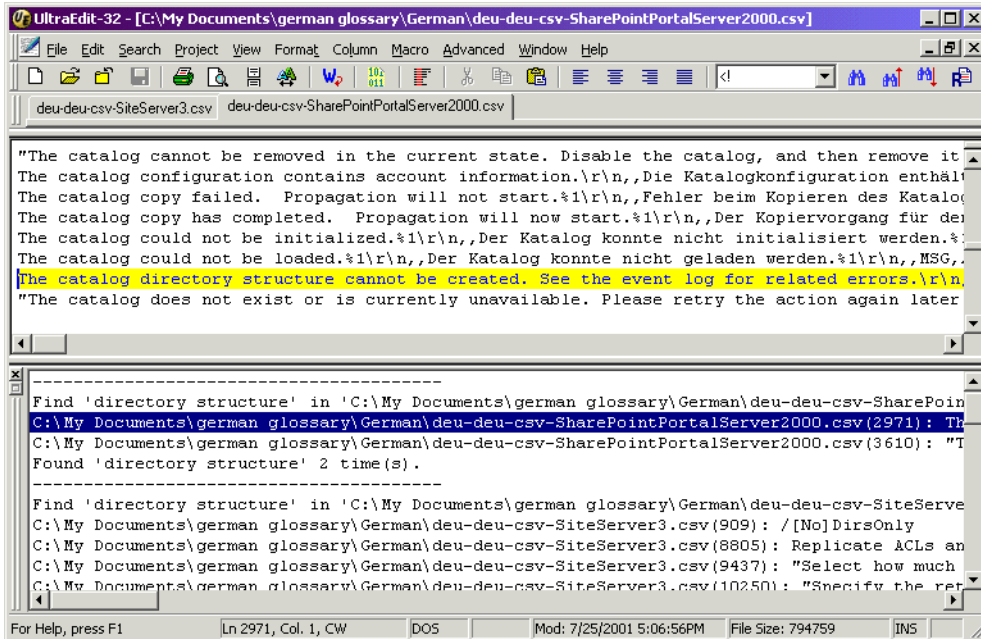


Figure 76: Batch search in *UltraEdit*

*Any of the searches can be made with wildcards. For wildcards, see To Search with Wildcards on page 17.*

Another helpful feature compares different versions of files. In the example below you can see that *UltraEdit* highlights the differences in two versions of an HTML file.
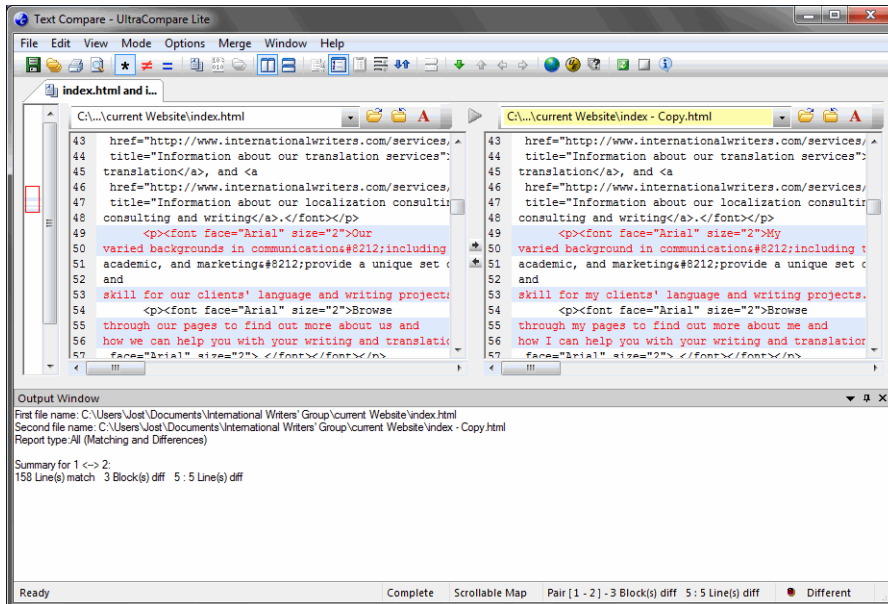


Figure 77: File comparison in UltraEdit

*If file comparisons are very important to you, you can also use a specialized tool such as ExamDiff (see www.prestosoft.com/examdiff) or Araxis Merge (see www.araxis.com). These give you additional reporting options as well as the ability to compare directories.*

These text editors are also able to automatically recognize typical formats such as HTML or XML and show them in color coding to ease the editing process.
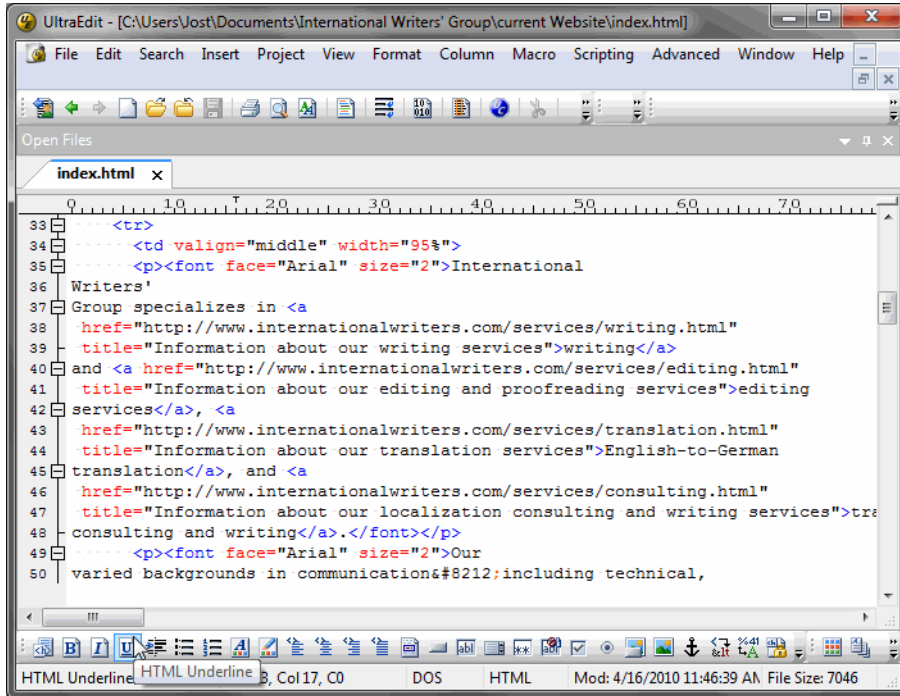


Figure 78: Color-coded HTML file in *UltraEdit* with automatically opened HTML tools at the bottom of the window

In the following example I instructed *EmEditor* to mark the occurrence of a certain word in a file, which *EmEditor* recognized as a resource file and color coded as such (marking the translatable text in pink).



Figure 79: RC file view in *EmEditor*

One important issue is the change of code pages (between different forms of Unicode, ASCII, DOS, or Mac formats under **File> Conversions** in *UltraEdit* and between language-specific code pages under **File> Save As** in *EmEditor*):
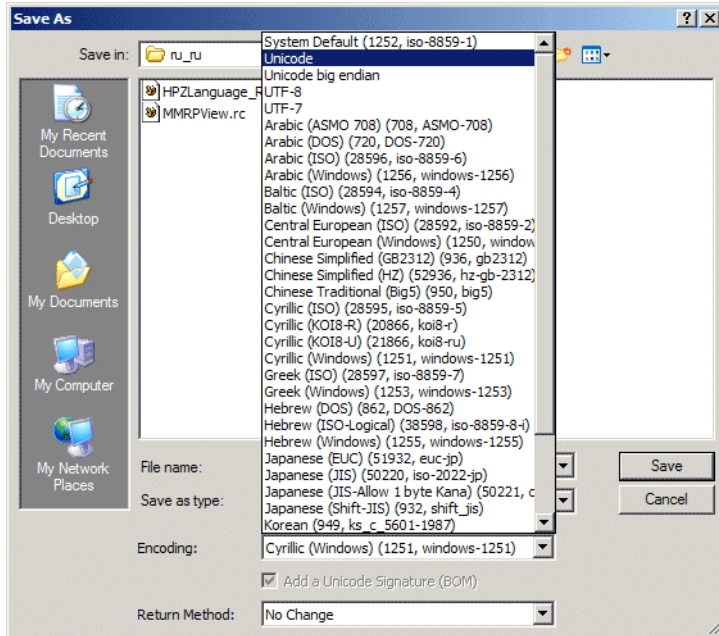


Figure 80: Possible conversions in *EmEditor*

Here is another use of text editors (though you can admittedly use *Notepad* for this also):

Most of us have received files of indeterminate type—we either don't know the extension (and even sites like www.filext.com can't help) or the extension is gone or corrupt.

Here is a quick and dirty way to help with that. You can open the file in question in a text editor. If everything is "humanly" readable, the file is in a text-based format and can be translated in an appropriate text, HTML, or XML editor. If the file opens with a lot of strange characters, it is some kind of binary file (a file that can be read only by computers) that cannot be edited (or saved!) in a text editor.

The good thing is that many of these files have a "magic number" (that's what it's really called!), i.e., a clue to their identity in the first line (the "header line"). Here is a list of the more common ones:

- GIF image files have "GIF89a" or "GIF87a" in their header.

- JPEG image files begin with "0xFFD8" or with "JFIF" or "Exif".

- PNG images contain "PNG" or "P N G" in their header.

- TIF/TIFF graphic files begin with either "II" (II for Intel, or little-endian) or "MM" (MM for Motorola, or big-endian).

- PDF files start with "%PDF".

- PostScript files and programs start with "%!".

- Many EXE or DLL files start with "MZ" or "ZM" (after the developer Mark Zbikowski).

- MIDI music files have "MThd".

- ZIP files begin with "PK" (for Phil Katz, author of the compression utility PKZIP).
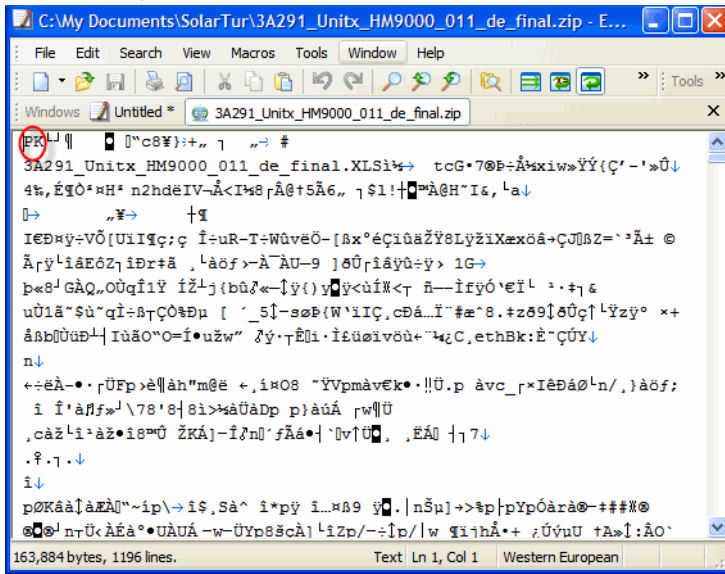


Figure 81: A "cracked" zip file with the magic number "PK" in *EmEditor*

This last magic number is particularly helpful because there are many file formats that pretend to be something very fancy and unique when in reality they are "only" ZIP files with a new and different extension. Finding out that these are ZIP files allows you to change the extension to .zip and unzip them with the compression utility of your choice. This in turn can sometimes give you access to data you would not have access to because you might not have the appropriate application.

What's extremely important to remember is this: NEVER save non-text-based files in a text editor. If you need to give it a different extension, do that in Windows Explorer.

I could write a lot more about the benefits of these programs, but here's what I've found to be true: If I can imagine some kind of logical operation within a text-based file, chances are that it can be done with one of these programs.

# HTML Editors

*HTML is an abbreviation for HyperText Markup Language, the authoring language used to create documents on the World Wide Web. HTML defines the structure and layout of a web document by using a variety of tags and attributes. Tags and attributes are enclosed with < and>. Translatable text typically includes all text between tags (the part that is displayed by the browser) as well as some attributes within tags (for instance, the "alt" text that pops up when you move your mouse over a graphic).*

Although the afore-mentioned text editors have HTML capabilities (see page 103), many users prefer to use specialized HTML editors for HTML files. Again, there are many different approaches out there, from the high-powered flagship products such as *Microsoft Expression* (see www.microsoft.com/Expression) or *Adobe Dreamweaver* (see www.adobe.com/products/dreamweaver) to (also high-powered but much more affordable) hands-on products like (the free and powerful) *Nvu* (see net2.com/nvu/).

If your intention is to use an HTML editor for translation or editing purposes, I would strongly recommend using a text editor or one of the hands-on tools because they tend to be "clean" tools, i.e., they do not add any additional and unwanted coding that your client most likely abhors.

Applications such as *FrontPage* used to have a bad reputation for changing code. They have gotten much better about protecting the original code in recent releases, but you still run the risk of changing the code if you do not adjust the settings accordingly. In the case of *FrontPage*, make sure that you select **Tools> Page Options> HTML source> Preserve existing HTML** (and in the case of *KompoZer* select **Tools> Preferences> General> Retain original source formatting**).

```html
<!DOCTYPE HTML PUBLIC "-//W3C//DTD HTML 4.0 Transitional//EN">
<HTML><HEAD><TITLE>International Writers' Group - Writing, Translation, and Localization
Consulting</TITLE>
<META content="MSHTML 6.00.2719.2200" name=GENERATOR>
<META http-equiv=Content-Type content="text/html; charset=iso-8859-1">
<META
content="copywriting, copy writing, translation, localization, editing, proofreading,
localization consulting, localisation consulting, localisation, German proofreading,
German editing, German translation, writing services, editing services, International
Writers' Group, International Writer's Group, Oregon coast"
name=keywords>
<META
content="The International Writers' Group offers a combination of services, including
copywriting, translation, editing and proofreading, and localization consulting"
name=description></HEAD>
<BODY bgColor=#ffffff>
<CENTER>
<TABLE cellSpacing=0 cellPadding=2 width=593 border=0>
  <TBODY>
  <TR>
    <TD align=left>
      <P> </P>
      <P> </P>
      <P> </P></TD>
    <TD align=right><IMG alt="International Writers' Group"
    src="indexHomeSite_files/IWGlogo.jpg" border=0></TD></TR></TBODY></TABLE>
<TABLE cellSpacing=0 cellPadding=1 width=592 border=0>
  <TBODY>
  <TR>
```

Figure 82: HTML page after being saved in non-obtrusive editor (same as original)

```
<!doctype HTML PUBLIC "-//W3C//DTD HTML 4.0 Transitional//EN">
<html>

<head>
<title>International Writers' Group - Editing, Translation, and Localization
Consulting</title>
<meta content="Microsoft FrontPage 5.0" name="GENERATOR">
<meta name="ProgId" content="FrontPage.Editor.Document">
<meta content="text/html; charset=iso-8859-1" http-equiv="Content-Type">
<meta content="copywriting, copy writing, translation, localization, editing, proofreading,
localization consulting, localisation consulting, localisation, German proofreading, German
editing, German translation, writing services, editing services, International Writers'
Group, International Writer's Group, Oregon coast" name="keywords">
<meta content="The International Writers' Group offers a combination of services, including
copywriting, translation, editing and proofreading, and localization consulting"
name="description">
</head>

<body bgcolor="#ffffff">

<body bgcolor="#ffffff">

<center>
<table border="0" cellPadding="2" cellSpacing="0" width="593">
  <tr>
    <td align="left"> <p> </p>
    <p> </td>
    <td align="right">
    <img alt="International Writers' Group" border="0" src="services/IWGlogo.jpg"></td>
```

Figure 83: Same HTML page after being saved in *FrontPage* without **Preserve existing HTML** selected

Never work in HTML files in *Word* unless you are specifically instructed to do so (see image below).

```
<!DOCTYPE HTML PUBLIC "-//W3C//DTD HTML 4.0 Transitional//EN">
<html xmlns:v="urn:schemas-microsoft-com:vml"
xmlns:o="urn:schemas-microsoft-com:office:office"
xmlns:w="urn:schemas-microsoft-com:office:word"
xmlns="http://www.w3.org/TR/REC-html40">

<head>
<meta http-equiv=Content-Type content="text/html; charset=iso-8859-1">
<meta name=ProgId content=Word.Document>
<meta name=Generator content="Microsoft Word 9">
<meta name=Originator content="Microsoft Word 9">
<link rel=File-List href="./indexWord_files/filelist.xml">
<link rel=Edit-Time-Data href="./indexWord_files/editdata.mso">
<!--[if !mso]>
<style>
v\:* {behavior:url(#default#VML);}
o\:* {behavior:url(#default#VML);}
w\:* {behavior:url(#default#VML);}
.shape {behavior:url(#default#VML);}
</style>
<![endif]-->
<title>International Writers' Group - Writing, Translation, and Localization
Consulting</title>
<!--[if gte mso 9]><xml>
 <o:DocumentProperties>
  <o:Author>Jost Zetzsche</o:Author>
  <o:LastAuthor>Jost Zetzsche</o:LastAuthor>
  <o:Revision>2</o:Revision>
  <o:TotalTime>1</o:TotalTime>
```

Figure 84: Same HTML page after being saved in *Word*

*To be fair, Word 2003 and above has an option under **File> Save As** that is called **Web Page, Filtered**. Though this eliminates some of the additional coding, Word remains an unfortunate choice for an HTML editor.*

If you must use an HTML editor but are insecure about what kind you should use, a good guideline is to either use non-intrusive programs such as the ones mentioned above, or the same program that the webpages were originally designed in.

*If they were designed in programs such as FrontPage or Dreamweaver, you can find that information in one of the meta tags in the HTML code: `<META NAME="generator" CONTENT="Adobe Dreamweaver XX">`.*

*If you are working in double-byte languages (Chinese, Japanese, or Korean) or bi-directional languages (Arabic or Hebrew), make sure that the HTML editor of your choice supports your language.*

## Saving HTML Files from the Internet

As a general rule of thumb, you never want to translate files that you save from a website. There are two reasons for this: First, most browsers add or change some code when saving files from the Internet. Second, the vast majority of webpages are dynamic, which means that they retrieve their content from databases; by saving a single HTML page you will see little or no database content.

For quoting purposes, however, it could be a good idea to save some files from the Internet (if you keep the above limitations in mind).

It is easy to save single web pages. In *Internet Explorer* and *Firefox*, open the web page, select **File> Save as**, and make sure that *Webpage, complete* or *Webpage, HTML only* is selected under **Save as Type**.

Typically, you do not want to quote on a single web page but on a complete website instead. So-called spiders such as *Tenmax Teleport Pro* (see www.tenmax.com/teleport) allow you to download a complete (static) website at one time so that you can browse it offline or analyze it for quoting or other purposes.



Figure 85: Downloaded website in *Teleport Pro*

*If you need to download websites with a safe encryption (starting with https://), you will need to use the slightly more expensive Teleport Ultra.*

*Other comparable tools include the freeware program HTTrack (www.httrack.com); Quadsucker (www.quadsucker.com/quadweb) is a commercial counterpart to this.*

A much more complex translator-specific spider program is *WebBudget* (see www.webbudget.com).

*WebBudget* is not "just" a web spider (it calls that function **Map a site**). It is also a translation memory tool specifically geared toward HTML and other tagged formats. Unlike most other web spiders, none of the downloaded HTML files is modified during the download process. And while its translation memory component does not have the sophistication of the larger translation environment suites (see *Translation Environment Tools* on page 190), the program is very simple to use. I especially like the fact that there are a number of HTML-specific settings on the **Options** tab (such as what kind of tags are translatable or how words are to be counted) that are both easy to understand and to modify. Loading an HTML file for translation is as easy as opening it in any HTML editor; and by simply clicking the **Extract** button, the translatable text will be displayed to you in the lower translation pane.

The **Quote images** feature is also a unique and helpful way to quickly view which images are part of a website or project, and which of the images have to be translated and charged for.
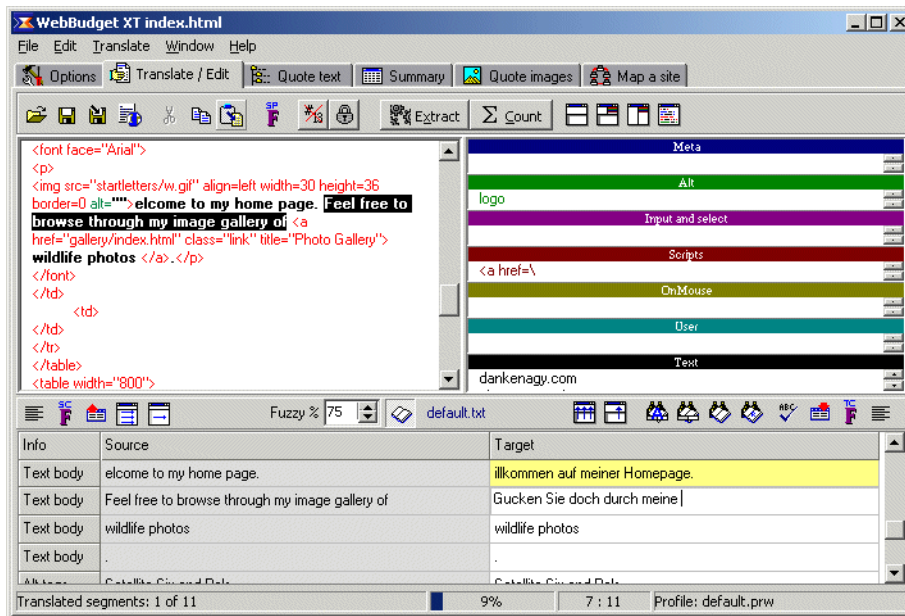


Figure 86: Translation view in *WebBudget*

# Utilities

Before discussing the "real (and expensive!) programs" such as office suites, computer-assisted translation tools, desktop publishing, or graphics applications, here is an overview of smaller, inexpensive (or even free) programs—called utilities—some of which have very powerful capabilities that can make many things easier for you. Though there are many thousands of these little applications available, and I am sure that there are many that are just as useful as—or even more useful than—the ones I am describing here, I have limited myself to the ones that I use on an almost daily basis in my work as a translator.



*There are also a number of software libraries which are good places to look for utilities. My favorites are www.tucows.com, www.download.com, www.majorgeeks.com, and www.snapfiles.com with a huge number of freeware (software that is given away for free by the author, who retains the copyright for it) and shareware programs (software that you can download and try for free, and then pay a licensing fee if you like it and use it).*

# Graphic Management Utilities

One of the helpful features of *Windows ME* and *XP* is the possibility to view graphic files in the so-called *Filmstrip* view, allowing for a quick and fairly detailed view of graphic files.
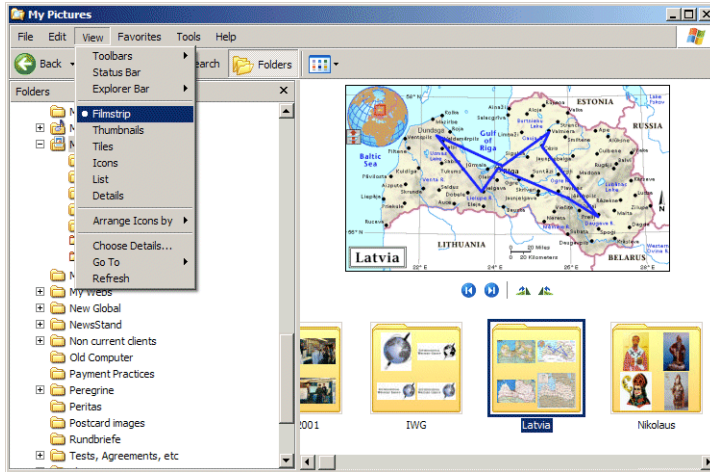


Figure 87: Filmstrip view in Windows XP

*You can activate the Filmstrip view by going to the My Pictures folder (inside the My Documents folder) and selecting* **View> Filmstrip** *in the Windows Explorer.*

*Vista abandoned the* **Filmstrip** *view and replaced it with the Windows Photo Gallery view that displays the pictures full-screen. Microsoft must have realized that this was not a move toward the better, because in Windows 7 it now offers the downloadable Windows Live Photo Gallery, an enhanced version of XP's Filmstrip. (You can also download this utility for other versions of Windows under download.live.com.)*

This feature is similar to other existing programs that are specifically geared toward managing image files, such as the freeware program *XnView* (www.xnview.com). *XnView* lets you view, convert, copy, sort, and edit any image file.
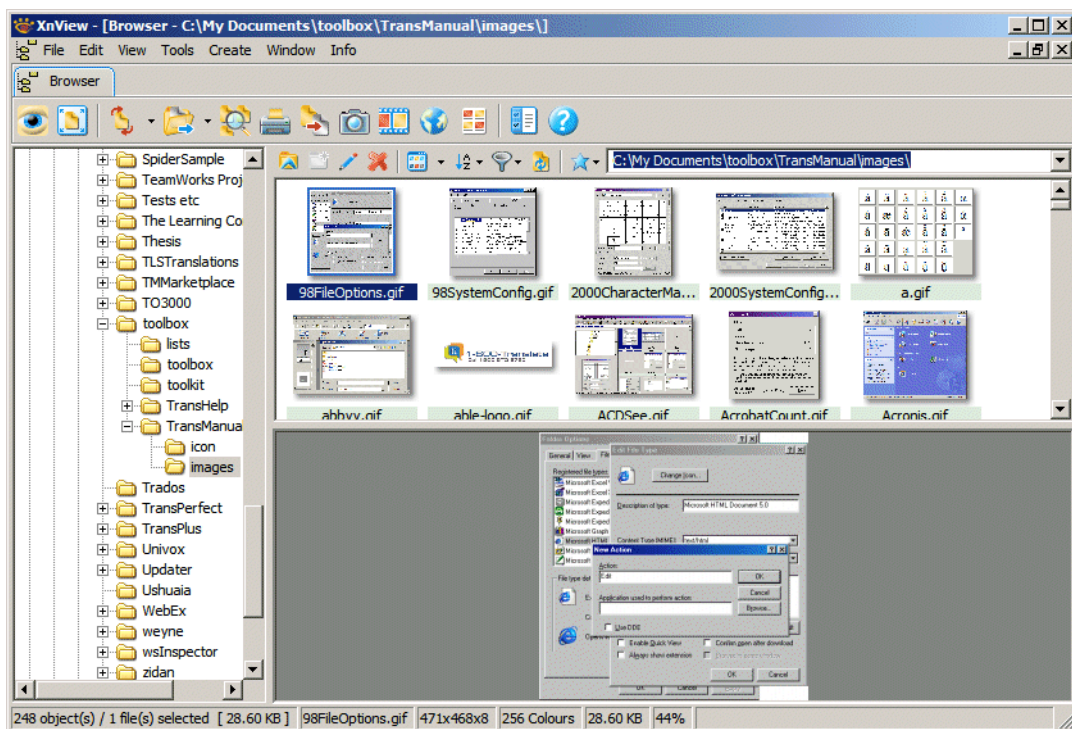


Figure 88: Image management in *XnView*

This program is particularly useful when dealing with a large number of image files such as you might have in a manual or help system. It quickly lets you view the individual images, decide which images need to be translated or generated again, and open these images in the graphic editor of your choice.

# Renaming Utilities

A small but useful tool is the freeware *Rname-it* (see www.brothersoft.com/ rname-it-4690.html) that allows you to batch rename a large number of files. This can be helpful when you need to change an extension (for instance, from `.HTM` to `.HTML`) or when you need to change the actual file name of a large number of files (for instance `filename.doc` to `filename_edited.doc`).
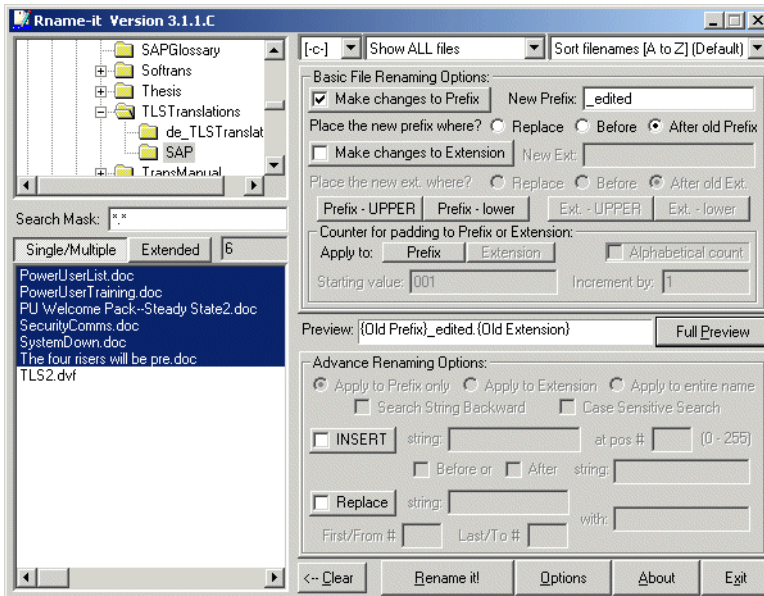


Figure 89: Renaming files in *Rname-it*

*Another useful function of this utility is the ability to change the time and date stamp of any file. This comes in particularly handy if you have worked until 5 am and prefer your project manager not to see that . . . .*

Another highly acclaimed renaming tool is *1-4a Rename* (see www.1-4a.com/
rename). This tool not only offers a larger range of functions than *Rname-it*
(including the ability to replace names in subdirectories!) but also shows that
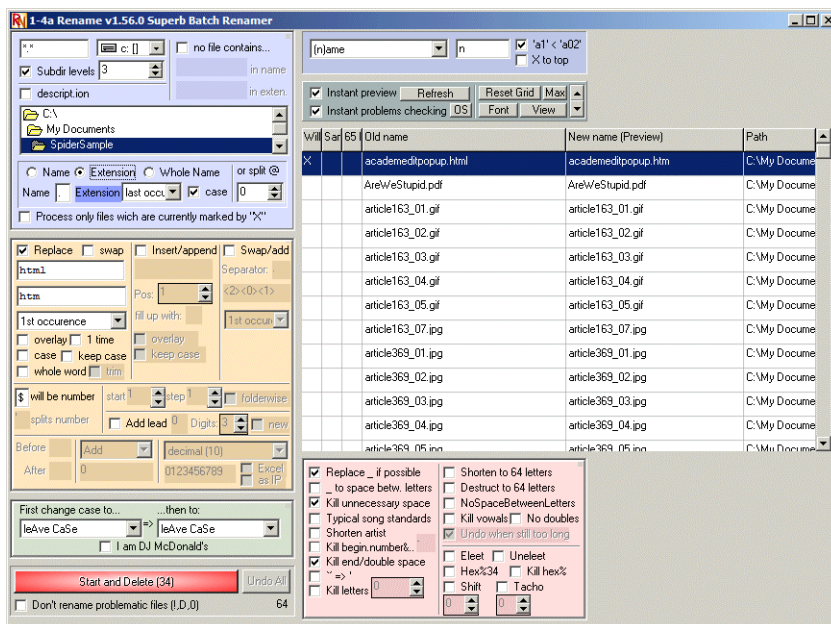developers can't resist giving renaming tools ridiculous names.



Figure 90: 1-4a Rename's interface

# Search Utilities

While the text editors described above have powerful search and replace functions, some utilities specialize in that and offer some additional features that can prove to be very handy.

*Search and Replace* (see www.funduc.com/search_replace.htm) can be opened on any individual file or folder from the *Windows Explorer* or any other folder view through the right-click menu.

In the resulting dialog, you can select a mask (i.e., a filter) for your file type (for instance, `*.txt` for all text files, or `a*.*` for all files starting with `a`) and a search string, as well as the string to replace it with.



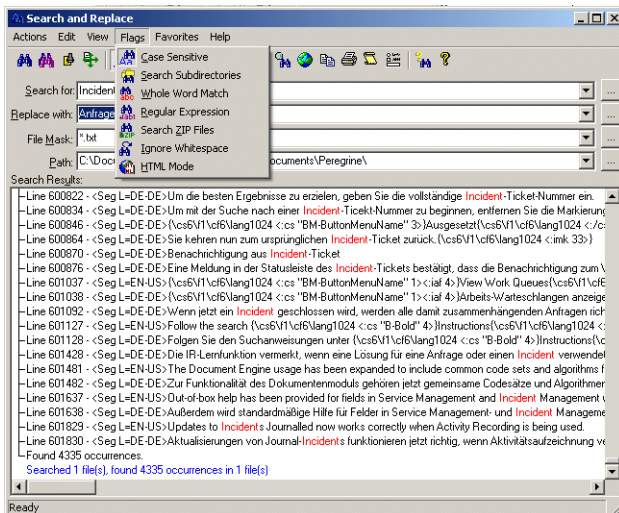Figure 91: *Search & Replace* with various options

*For information on wildcards, see To Search with Wildcards on page 17.*

*If you need to search with very complicated expressions, a separate Search & Replace Regular Expression Wizard is also available from www.funduc.com.*

When *Search and Replace* has shown you how many of the desired strings you have in your file(s), you can either decide to batch replace or on a case-by-case basis.

Search and text retrieval programs that approach searches differently are called indexing tools, such as *dtSearch* (www.dtsearch.com) or *Copernic* (www.copernic.com).

Unlike *Search and Replace*, they don't perform any searches in the actual files but rather in indexes that are linked to the files. Admittedly, this sounds kind of confusing, but the principle is this: If you have a large amount of data (let's say all your email, including attachments, of the last three years), this may be sorted by name or date but not by the actual data. So for any program to find a certain word or phrase within these humongously large files, it would actually have to go through every line of data that is contained in these files. If, however, you had a preconfigured index containing information on all the words contained in these files, together with information on where to find them, these programs could access that information virtually instantaneously.

Starting with *Windows Vista* and continuing with *Windows 7*, one of the operating system's main emphases has been accessibility and ease of search, so it has integrated a new search mechanism. If files are located inside one of the "indexed locations" on your computer, it's just a matter of seconds to find text within the file or the file name itself. You can change the settings in *Vista* and *Windows 7* under **Start> Control Panel> Indexing Options**.

A search tool for translation-related things outside and on your own computer is the wonderful *IntelliWebSearch* (see www.intelliwebsearch.com). This free little application copies highlighted text from any *Windows* program with a number of user-definable shortcut keys, opens your default browser or dictionary, and sends the copied text to up to 50 customizable search engines, on-line dictionaries, or dictionaries stored on your hard drive.
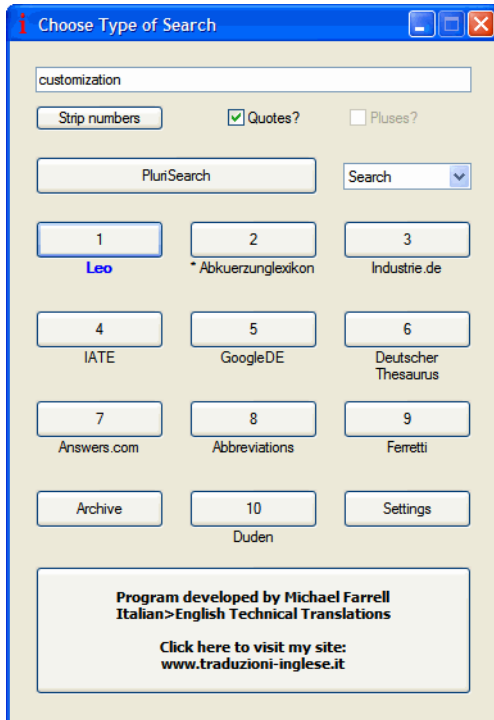


Figure 92: *IntelliWebSearch* with customized search options

It takes a little fiddling to find the right coding for the various search options, but the tool comes with a good number of preloaded searches and with a little bit of patience and the help provided under www.intelliwebsearch.com/gb/help.html it's quite easy to develop your own searches.



Figure 93: *IntelliWebSearch* with customized search for EN>DE in the IATE



Figure 94: *IntelliWebSearch* with customized search for computer-based dictionary

# CD Emulators

CD emulators are not essential equipment but they can be very convenient programs. They allow you to create up to 23 virtual CD drives on your hard drive, make an image (a complete copy) of the content of a CD and thus enable you to play several CDs at a time, all at higher speeds than through the original CD-ROM drive.

For us as translators this can be particularly interesting if we need to work with several CD-based dictionaries at a time and don't want that "disc-jockey" feeling of having to change CDs all the time (or if we want to use a dictionary and listen to a CD or watch a DVD at the same time. . .).

Again, there are a number of products out there: FarStone's *Virtual Drive* (see www.farstone.com/home/en/shtml/vdpoverview.shtml), ZTekware's *Original Virtual CD* (see www.ztekware.com), or *Virtual CD* (see www.virtualcd-online.com).

I'm familiar with *Virtual CD*, which has a very intuitive and easy interface. It allows access to all features through a right-click on an icon in the taskbar.
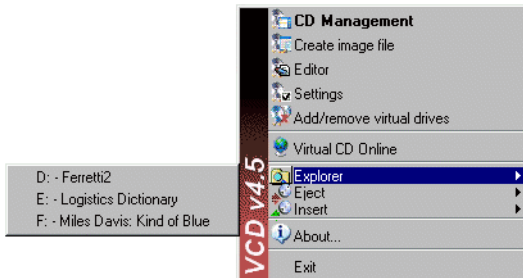


Figure 95: Right-click access to all functions of *Virtual CD* and three emulated CD drives.

Once the emulated CD drives are created, they appear "as equals" to any "real" CD-ROM drive under **My Computer**:
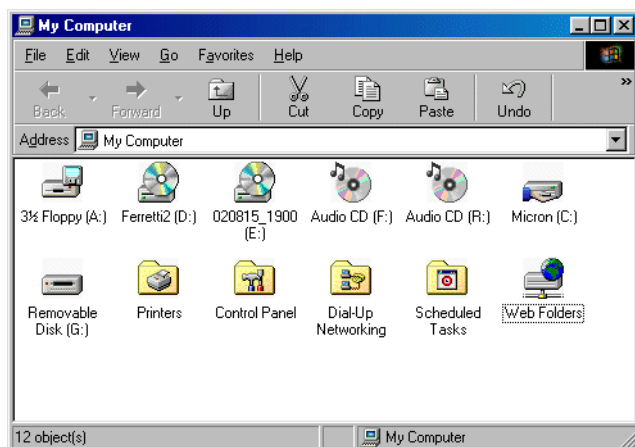


Figure 96: **My Computer** with several emulated CD drives.

So, what's the catch? Well, aside from the fact that some DVDs and CDs have a copy protection that does not allow them to be copied, there is none. That is, if you have a very large hard drive that can handle a lot of data. Even though these programs all have some kind of compression capability, it's not uncommon to have 600 MB of data stored on just one single CD.

# Compression Utilities

Arguably the most important utility you need to be able to receive and send files properly is a compression program. Nothing can frustrate a client or customer more than receiving a file of several megabytes that would have been maybe a tenth of the size or even less if it had been sent in compressed format (see *File Transfer* on page 71).

*Some file formats, such as .rtf or .bmp, are particularly well suited for compression because they can be minimized significantly; others, such as .jpg, .gif, or .pdf, often shrink very little when being compressed because they are compressed in themselves to start with.*

Other important reasons for using compression programs are that they allow you to send one file instead of many, and compressed files can also be sent as password-protected files for safety reasons.

A search on the Internet reveals that there are probably as many different programs out there as you could come up with word combinations containing the word "zip"—*ZipMagic*, *PowerZip*, *Quick Zip*, *ZipGenius*, *BitZipper*, *ALZip*, and *TurboZip* form only the tip of the iceberg—and of course *PKZIP* (see www.pkware.com) from the "inventor" of the zip format, and the market leader *WinZip* (see www.winzip.com), which is now owned by Corel. And, yes, there are a great number of compression programs that do not contain the word "zip". . ..



*Windows ME, XP, Vista, and 7 actually contain a limited zip program, as do other file management systems such as the above-mentioned "Explorers on Steroids" (see page 11).*

My long-time personal choice of *WinZip* was recently replaced by the Italian *ZipGenius* (see www.zipgenius.it) which works with a much greater number of compressed file formats and has other advanced features that I like (oh, and it's free). As with most other programs of its kind, it is closely integrated with *Windows Explorer*, i.e., a right-click on any file, group of files, or folder(s) gives you access to the program.

And while in earlier versions of compression utilities, the context menu tended to be rather cluttered with a number of options, newer versions typically put an end to this mess by giving only one option. This provides access to a whole new sub-menu with the various old and new zipping options (which, by the way, are configurable), including the ability to directly email the newly created zip file (which saves space on your hard drive and means one less step in your workflow).
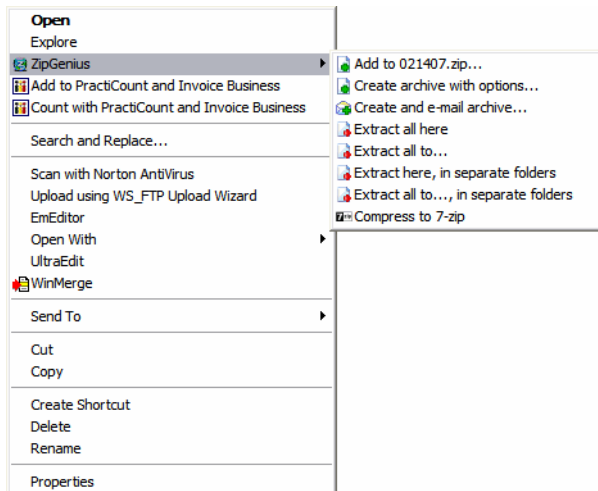
Figure 97: Right-click access to *ZipGenius* options

For the above-mentioned password protection ability, select the **Create archive with options** command, which will open a dialog that gives you access to the password option.
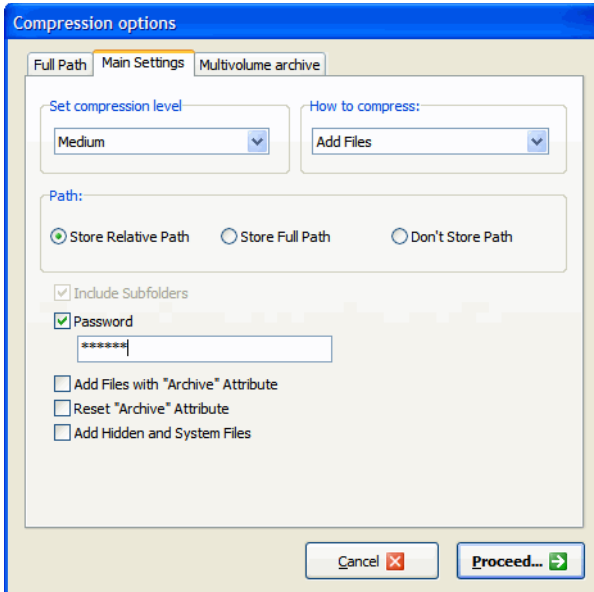


Figure 98: Access to *ZipGenius'* password feature

*If you are really paranoid about the safety of your files, you will want to give a password in which you mix numbers and lower-case and upper-case letters. The cracking utilities that are available to crack these passwords (see Password Cracking Utilities on page 131) take significantly longer to break these mixed passwords.*

Another feature that most zip tools offer is the ability to split files into smaller chunks so that they fit into an email or on a CD. Once you want to use the file(s), the tools allow you to reassemble them into one large zip file again.

*If you use the split feature frequently, you may also benefit from tools like FileSplit that specialize in this particular feature while also allowing you to zip files (see www.partridgesoft.com/filesplit).*

And here's what I just discovered recently. Often I receive five different zip files for a project. It has always annoyed me to have to right-click on each of them individually and select the appropriate unzip command so that the files will be unzipped into a folder that carries the name of the zip file. Then I discovered that you can also select several zip files at a time (by holding the **Ctrl** key and clicking on each of them). You can then select a command to extract them "**. . . in separate folders**," which creates as many folders as there are zip files.

## Password Cracking Utilities

Another helpful tool to at least know about (this one might indeed not have to be in your tool box until you really need it) is a password cracking utility. Though this may sound rather ominous, these tools can be used legally and are even rather necessary at times. The three kinds of password encryptions that arguably give translators the most headaches are those for *Office* files— particularly *Word* files—zip files and PDF files. Now, some of those files may be encrypted for good reason, but for others, especially those you are asked to translate, you'd better know the tool you'll need to open them.

As with so many other tools, there are a great variety of tools out there that allow you to find the magic word, but the tools that I have been using successfully come from ElcomSoft (see www.elcomsoft.com)—unfortunately in different (paid) versions for the different products. Now, the magic word is not magic for nothing, and it isn't easy to find, even for the smartest software. Plus, there is also a reason for the different levels of "password strength" that you are asked for on the various websites and programs you may be choosing a password for.

Like any of their competing tools, ElcomSoft tools essentially have two strategies. The fast and quick way is the dictionary attack. This is for simple-minded folks like me who know that they would forget their password if it were not the dog's name or something like that. This attack only takes a few seconds and all it does is to run a large list of terms against the actual password until the correct one is found. If that method is not successful, a second method is applied, the so-called brute force method.



Figure 99: Selecting the right kind of attack in an effort to crack a zip file

Typically you can tell the program certain parameters (like only lower- or upper-case letters, with or without numbers/special characters, or the presumed length of the password) and depending on how complex and accurate these are, a successful attack can take minutes or a whole day.
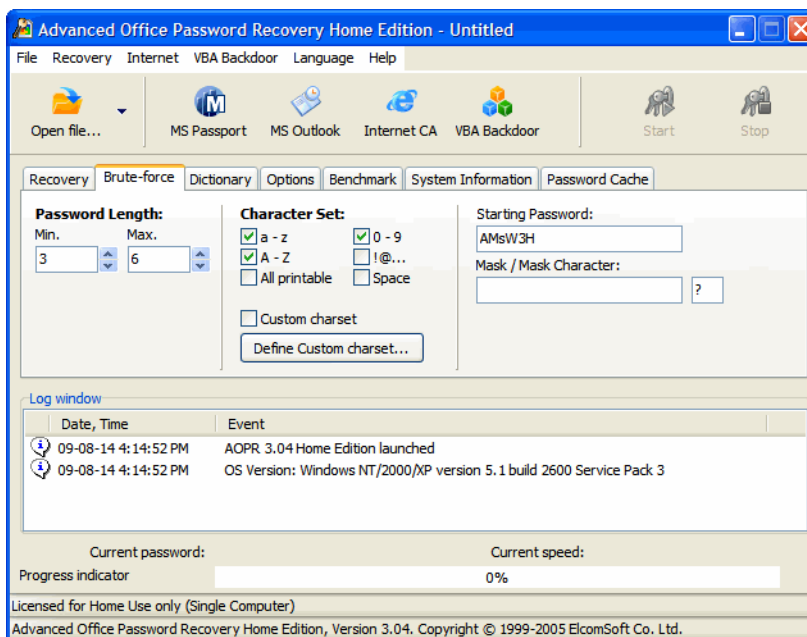


Figure 100: Selecting specific options for a brute force attack for cracking *Office* files

# Measurement Conversion Utilities

While some software programs, including some translation environment tools, provide for automatic conversions between certain measurements, there are naturally restrictions to their abilities.

And while there are many websites that perform all kinds of conversions, it's still helpful to have a little freeware utility like *Convert* (see www.joshmadison.com/software). *Convert* not only allows you to convert between a multitude of measurements, but even lets you define your own parameters.
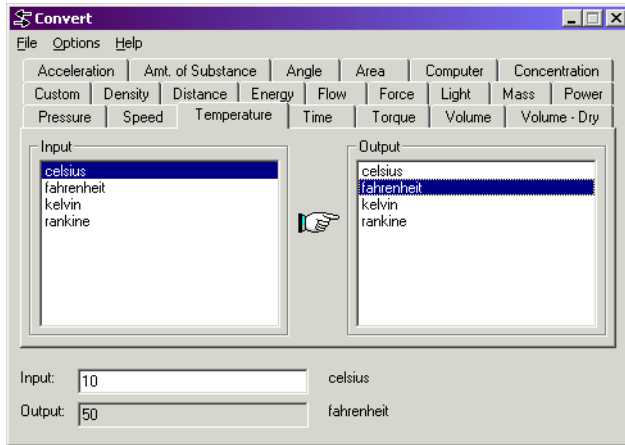


Figure 101: *Convert's* interface—the **Custom** tab gives you access to customizable conversions.

Unfortunately, *Convert* does not perform conversions directly within documents, which is something that *T4T-CNV* (see www.itla.ch/t4t/tools.php) does (within *Microsoft Word 2007* and below—*Word 2010* is not supported). An added benefit to this latter program is that you can choose whether you simply want to have your data converted (i.e., `1"` to `2.54 cm`) or appended (i.e., `1"` to `1" (2.54 cm)`).

# Word Count Utilities

Word counts are without a doubt one of a translator's main nightmares. Many programs do not provide for any word count mechanism (for instance, many DTP applications or HTML editors). If they do, they're only very basic ones (such as *FrameMaker*, which successfully hides this function in a submenu and does not count any index markers), and no program counts words in graphics. Even for tools that do word counts or are actually specialized in it, the results differ greatly.

A word count of our home page, www.internationalwriters.com/index.html, produced these results with the following tools:

• *Word*: 83 words

• *PractiCount & Invoice*: 154 words

• *Déjà Vu*: 173 words

• *Trados*: 184 words

• *UltraEdit*: 273 words

These tools applied three different theories to the count of these files:

• *Word* merely counts the words displayed in a browser, omitting all hidden text, such as keywords or pop-up texts for graphics, etc.

• *PractiCount*, *Déjà Vu*, and *Trados* are specifically geared toward counting or translating HTML files so they have to be able to reveal and count all translatable words.

• *UltraEdit* (see *Text Editors* on page 99) counts all words, including a lot of non-translatable coding information. It's not a very useful number to present to your client in an invoice (unless you are not interested in keeping the client, that is. . .).

Interestingly, even within these groups there are fairly significant differences. These are due to differences in the counting parameters. These include questions of delimiters (how are \ and – counted and how many words is `C:\Program Files\Netscape` or `format-specific`?) and numbers (are those to be counted and, if so, how many words is `255.255.255.0`?). And it becomes very hairy, of course, when it comes to non-alphabet-based languages or languages without spaces between words.

It seems that there are two main strategies for dealing with these problems. You can avoid word counts altogether and either go with an hourly rate or a character count (such as the 55 characters per line that many European translators do business by), or you can make a special point with your client to agree on a certain program for the word count.

For most text- or word-processor-based documents (.doc, .rtf, .wpd, or .txt), *Word*'s word counts have become a de-facto standard. This means that if you plan to count your words in any of these documents using other applications, you had better inform your client about it. The same accounts for most other formats that can be easily pasted into *Word* or converted to a *Word*-readable format, such as *Excel* and *Access*, or a number of DTP applications. If clients perform word counts on these files, chances are that they will use *Word* for that.
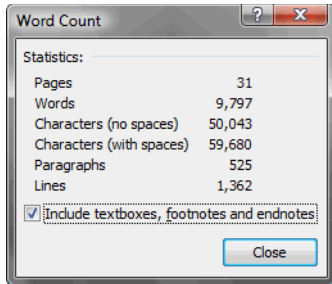


Figure 102: *Word 2007's* word count features

*Word's word count is available at **Tools> Word Count** (Word 2007: **Review> Proofing> Word Count**). In Word 2007/2010 and on the Mac versions of Word (Word 2001 and higher), the word count is displayed on the status bar by default.*

You will need to be cautious with *Word*'s word counts because *Word* is famous for skipping texts in comments, text boxes (before *Word 2007*), WordArt, headers, and footers. At www.shaunakelly.com/word/CompleteWordCount you can find a free and more "accurate" word count macro than the feature provided by *Word* itself.
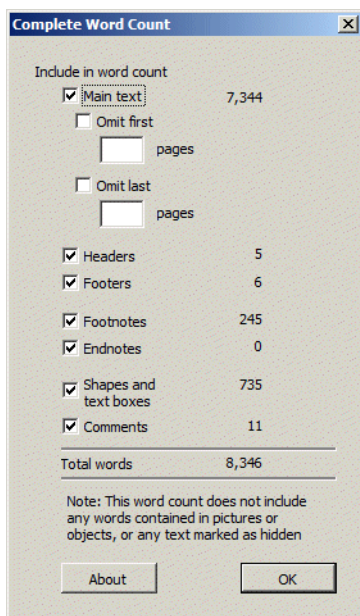


Figure 103: *Complete Word Count's* word count features

What these word counts do not address is batch counts of several documents at a time. I usually choose to do that (and word counts for many other supported formats) with translation environment tools (such as *Déjà Vu* or *Trados*), but there are also many specialized programs such as *FreeBudget* (see www.webbudget.com/fb4dload.htm) for text- or word-processor-based documents or *PractiCount & Invoice* (see www.practiline.com), the word count tool with the largest available feature set to date. It not only supports all of the file formats above and more (text- or word-processor-based, *Excel*, HTML, *PowerPoint*, XML, and PDF), but it also offers a variety of reporting options for

direct use in invoices, and it has a very customizable set of word count options. You can choose to use *Word*'s own word count module for most of the supported formats, or you can customize the rules by defining your own delimiters.
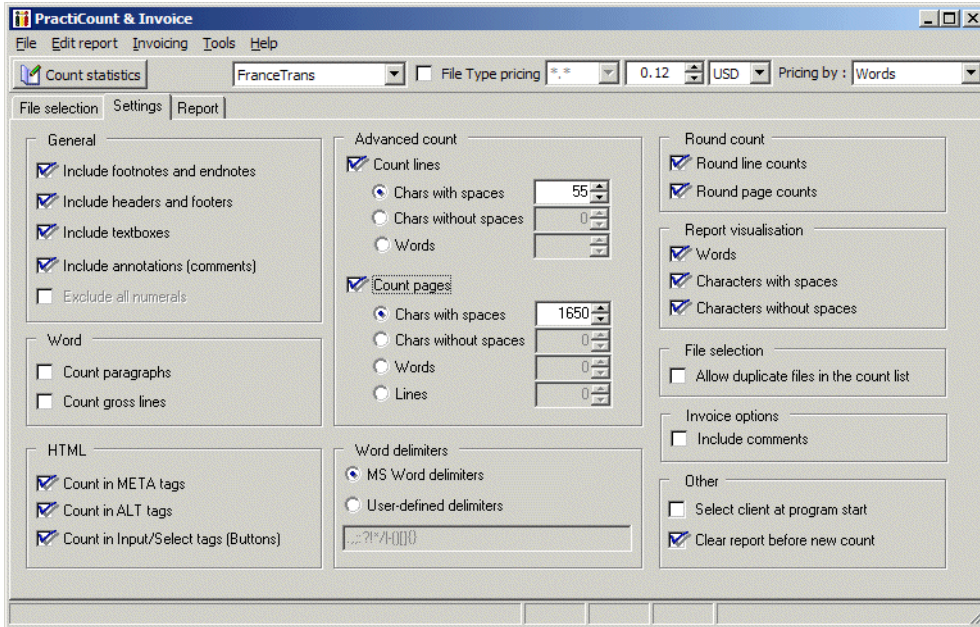


Figure 104: *PractiCount & Invoice's* settings

You can also count words in embedded text-based objects, and you can count editing time in *Word* and *PowerPoint* documents (the numbers are based on information that you can access under **File> Properties> Statistics** in *Word* and *PowerPoint*.)

Different file types can be mixed for different batch counts:



Figure 105: *Word* count summary view with easy options to save directly in various file formats

*PowerPoint files are counted by PractiCount with PowerPoint's own word count module (within PowerPoint, this is accessible through **File> Properties> Statistics**, **Office** button**> Prepare> Properties> Document Properties> Advanced Properties> Statistics** in PowerPoint 2007 and **File> Info** in PowerPoint 2010). Of course, this does not include text on any embedded objects (see page 180), and you should also be aware that the word count module in PowerPoint XP and above differs from earlier versions (e.g., hyphenated words are no longer counted as two words).*

*The count of PDF files should be taken with caution, as much of the text could be contained in embedded graphics.*

# Time Tracking

There is no need to explain why it is important for translators to have a good mechanism to track time. Some programs show you how much time you have spent working on them (for instance, *Microsoft Word* or *OpenOffice* under **File> Properties**) but that usually includes all the time you had the document open (while you had lunch, went to the bathroom, or took a nap).

The most common way to log the time we spend on an individual task is probably in an *Excel* spreadsheet. Two keyboard shortcuts have made it easier for me to keep track of my time in *Excel*:

- CTRL+; to enter the current date and

- CTRL+: to enter the current time.

A spreadsheet with a third field for the total time (formula: `=SUM(<end time field>-<start time field>`) can then ensure that your time is being calculated accurately.
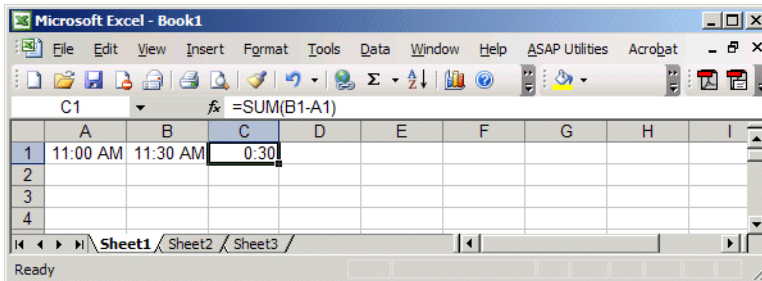


Figure 106: Sample of an *Excel* spreadsheet with formula as described above

While it is possible to record your time in this manner, there are some little programs available that make it a lot easier. *Time Stamp* (see www.syntap.com) is a free program (supported by optional donations) which allows you to track the start and end time for projects you are currently working on with a click on a button in your task bar. It's even possible to have several instances of the program running simultaneously so you can switch back and forth between different projects that you're working on. When you

are completely finished, all the time that was spent on each project is summed up and can either be printed out or saved as a text file. This is a nifty little program which requires neither a lot of computer resources nor a lot of time to learn.
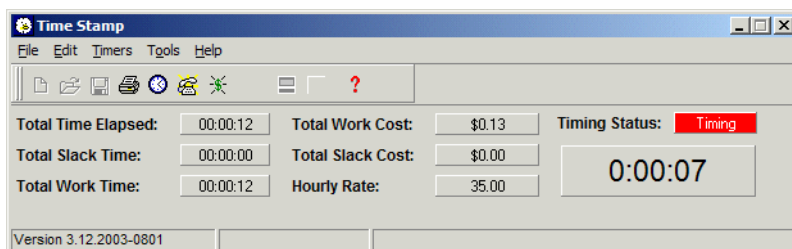


Figure 107: *Time Stamp's* user interface

The Ukrainian software maker AIT released a time tracking tool specifically geared toward language professionals. Similar to its generic counterparts, *ExactSpent* (see www.exactspent.com) tracks time for multiple jobs and/or clients simultaneously and even has a little (configurable) feature that reminds you if you have not touched your keyboard for some time. It leaves very little footprint on your computer and minimizes itself to the system tray, where it can easily be accessed and controlled with a mouse click or configurable keyboard shortcuts.
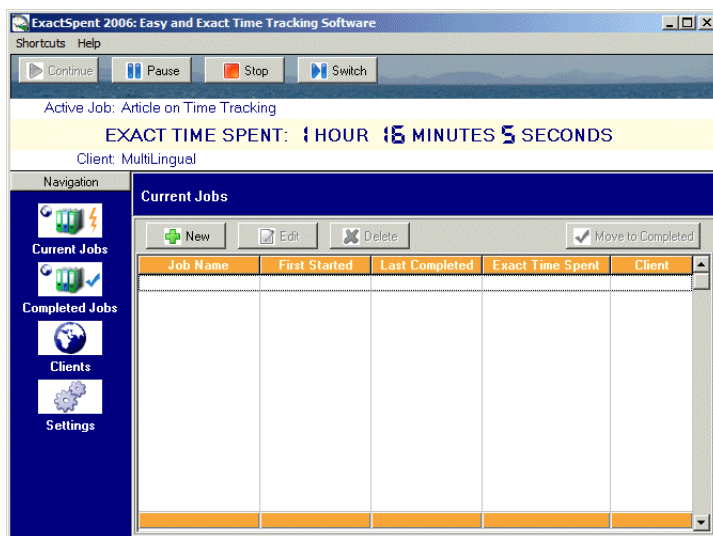


Figure 108: Tracking time with *ExactSpent*

# Clipboard Management Utilities

Have you ever wanted to be able to use your clipboard (i.e., the place where *Windows* stores copied information) more extensively?

Recent versions of *Microsoft Office* include the *Office Clipboard* (in *Office 2000* applications this is available as a toolbar under **View> Toolbars> Clipboard**, and in *Office XP* and above a slightly more usable application available under **Edit (Home)> (Office) Clipboard**), which allows you to collect up to 12 (*2000*) or 24 (*XP* and above) different clipboard items from anywhere on your computer and paste them individually or all at once into any *Office* document.
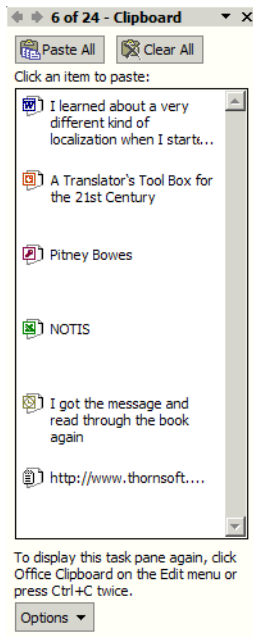


Figure 109: *Office XP Clipboard* with copied content from several *Office* applications

If you don't limit yourself to *Microsoft Office* programs, though, this is not very helpful, and has very limited functionality.

What I really needed, especially for editing tasks, was a program that would not only collect clipboard entries between all possible programs, but would also make it possible to directly print from my clipboard or store clipboard entries between different computer sessions (i.e., after I switched the computer on and off). I finally found this little helper in *ClipMate* (see www.thornsoft.com).

*ClipMate* is a little program that you can configure to start automatically every time you start *Windows* (**Config> User Preferences> General> Run at Windows Startup**). It collects an unlimited amount of clipboard content containing anything from text to graphics to complete files or folders. It is displayed as a little icon on your task bar and you can open the *ClipMate Explorer* by simple double clicking that icon.
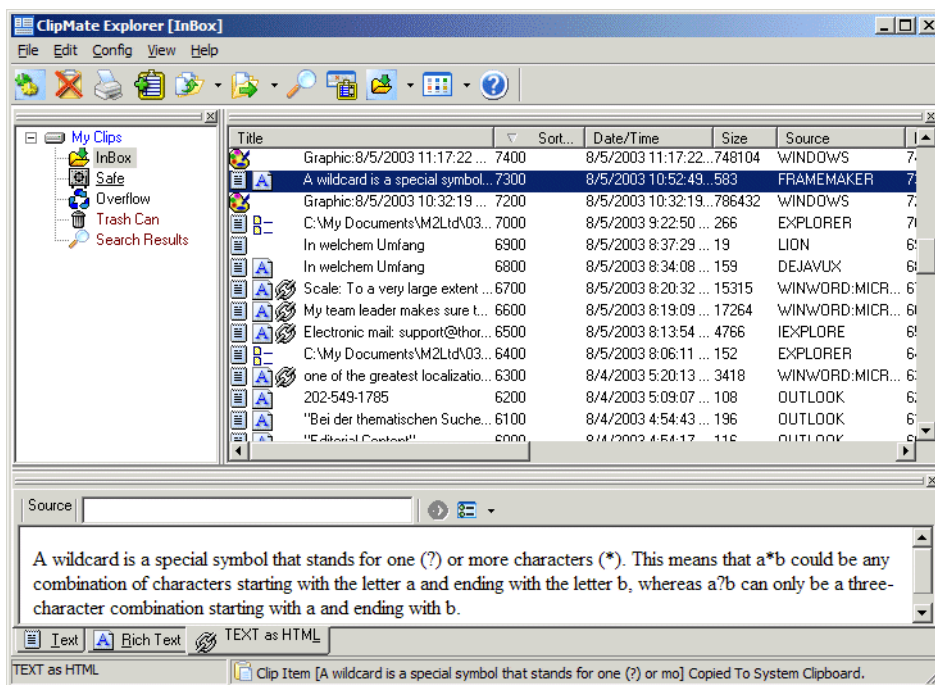


Figure 110: *ClipMate Explorer* with a preview pane (bottom) and a collection view (upper right)

*ClipMate* can accomplish all the tasks I need it for, and it's even possible to edit the clipboard content once it is stored in *ClipMate*.

*Before you purchase a tool such as ClipMate, you will need to run some tests to determine whether its functionality will interfere with other applications that make use of the clipboard. Trados 2007 and earlier (see Translation Environment Tools on page 190), for instance, uses the clipboard extensively and the combination of these two tools does not work well.*

# Collaboration Tools

No other medium has facilitated real-time collaboration between translators and editors more than the Internet. While the "traditional" method of online communication—e-mail—still plays an important role, there are an increasing number of other tools and communication channels that have certain benefits over e-mail.

One group of tools that I had long pooh-poohed because I assumed it was mainly a tool for adolescent puberty talks are the various instant messenger programs that are available. But I recently discovered that they can provide enormous productivity boosts if you work in a clearly defined work group.

*The most well-known messenger tools are the AOL Messenger (see www.aim.com), ICQ (www.icq.com/download), Yahoo! Messenger (messenger.yahoo.com), and the Windows or MSN Messenger (messenger.msn.com). It is important to check which IMs the people in your workgroup already use before you make your choice. Most networks are not compatible with each other, i.e., you cannot communicate with someone who has AOL Messenger if you yourself use the MSN Messenger.*

*An exception to that rule is Miranda (see www.miranda-im.org), an open-source Instant Messenger that allows you to connect to most other Messenger programs.*

I like these programs because I can see whether any of my "buddies" is online, allowing me to ask something and receive an answer immediately. Some *IMs* also allow for file sharing or computer sharing.



Figure 111: *Miranda* with "buddy list" of users from various other tools

The one protocol that's only indirectly supported by *Miranda* is *Skype* (see www.skype.com), and this isn't too surprising because *Skype* is something completely different. While *Skype* also offers text messaging comparable to any of the other providers, it is primarily a VoIP (Voice-over-Internet-Protocol) service that allows you to make completely free calls from computer to computer if the person you are calling also has *Skype* installed. Even if the other person doesn't have *Skype*, you can still make ridiculously cheap calls when the recipient uses a normal telephone.

The caveat? Well, it's not the price (the program is completely free). I suppose the danger is that it's just too easy (literally just one click) to call someone and chat, keeping you away from your work and wasting a lot of time in the process. If wisely used, however, *Skype* can be a great tool for virtual workgroups, especially because you can set up conference calls with five participants in a matter of five clicks (and at no cost if everyone has *Skype* installed).

Another feature that *Skype* offers is direct file transfer without any limit on the size of the file. You may or may not like this feature depending on your network or firewall situation.
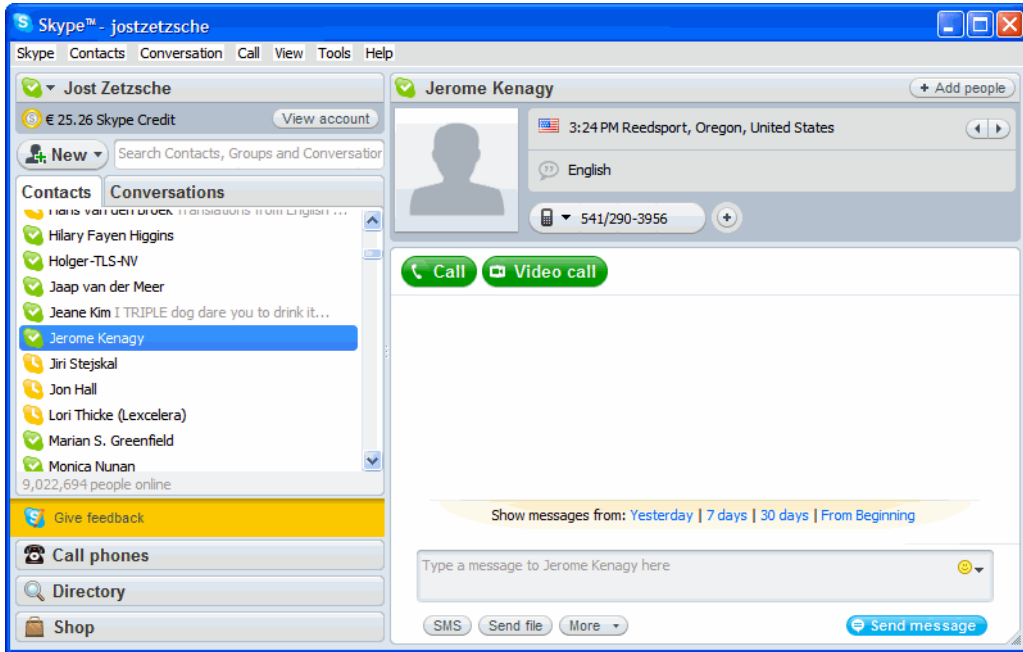


Figure 112: Skype's user interface

# Merging Files

Have you ever had a lot of files from one subject or client that would have been so much easier to handle if they could have been merged into one large file, for instance for alignment purposes?

Though it is often possible to copy and paste into one large master file, it can be tedious and frustrating if the original files are extremely large. *Twins File Merger* (see www.twins-software.com) is inexpensive shareware that allows you to merge as many MP3, MPEG, text, and *Word* files as you would like (the unlicensed copy allows you to only merge two files at a time). Like most of these specialized utilities, the use of this tool is very self-explanatory and the effect that it has on the performance of your computer system is very small.

Figure 113: Merging a large number of files in *File Merger*

# Office Suites

Nearly every new *Windows* computer comes with some kind of *office suite*, a software bundle that includes word processing, spreadsheet, and presentation applications, and, depending on the package, various other programs.

The most common suites include

- The open-source *OpenOffice* (see message URL www.openoffice.org) and its commercial and slightly extended version in Sun's *StarOffice Office Suite* (see message URL www.sun.com/software/star) and of course

- *Microsoft Office* (see office.microsoft.com/products).

All these suites include comparable combinations of programs and are certainly adequate as office suites for personal or professional use. The reason I use *Microsoft Office* is because so many other programs, especially translation environment tools, rely on *Word* in some way or other, and because it is, quite simply, the standard. Though all the other application suites stress their compatibility with *Microsoft Office*, in reality conversion problems to and from *Microsoft Office* remain, and I find it hard to imagine that clients who take it for granted that translators are equipped with *Word*, *Excel*, and *PowerPoint* have the patience to deal with conversion issues.

This certainly does not mean that you could not have some of the other suites as well, or that some of the other suites are less productive and/or more of a headache (after all, I "grew up" with a *DOS* version of *WordPerfect*, and it took me a long time to get used to *Word*). What I find most exciting about this

competitive situation is that at least *StarOffice/OpenOffice* (and now *Microsoft Office 2007/2010*) are XML-based, which points toward a future when files from all these suites and applications will indeed be truly exchangeable and conversion will no longer even be an issue.

*XML—eXtensible Markup Language—was developed by the World Wide Web Consortium as an alternative to HTML (see page 109)—which was not sophisticated enough for the creation of interactive content—and SGML—which proved to be too complex for web applications. The most important aspect of XML may be that it allows for the definition, transmission, validation, and interpretation of data between applications and between organizations.*

Still, I believe it's more important to consider which line of *Microsoft Office* should be used and how often you should upgrade than which of the office suites a translator should use.

## Release Versions of Microsoft Office

To find out which release version of *Microsoft Office* you should use, first look at all the other programs you use that in some way interact with programs of *Office*, most notably *PowerPoint* and *Word*. Many of the translation environment programs use *Word* plug-ins that often are not updated to the latest release of *Word* until months after Microsoft has published it, while other computer-assisted translation tools require a minimum version. Other programs that use *Word* plug-ins include *Adobe Acrobat*, virus check programs, and voice recognition programs.

If you are sure that you will not have any conflict with any other program, you can start to look into upgrading. In general, the best advice for upgrading *Microsoft Office* may be to wait until you buy a new computer (assuming that your new computer comes pre-installed with *Office*). The changes between the different versions often make very little difference in our work as translators, so it may be hard to justify the fairly significant expense.

However, there are some exceptions of *Office* upgrades that make a significant difference.

Depending on the language version of *Office* you are using, there will be a limited number of spell checkers pre-installed. You can buy a full add-on version with 51 languages (*Office 2003*), 48 languages (*Office XP*), or 30 languages (*Office 2000*) on Microsoft's or other websites.

For *Office 2007*, this system has changed quite a bit. Here you can buy the additional spell-checkers only as part of the *Multi-Language Pack* which also includes the user interface in the respective languages. You can either choose to buy all 37 covered languages or individual products for each of the languages. There is no compatibility between *Office 2007's* additional spell-checkers and those of earlier versions.

For *Word 2010*, things again have changed. There is now unfortunately no Multi-Language Pack for the non-corporate user and, of course, there is no compatibility between the language packs for *Word 2007* and *2010*. Each version of *Word* contains spell-checkers for between 3 and 7 languages (you can find a list under http://tinyurl.com/28bjpne). For additional languages you will have to purchase individual language packs. These also come with three or so spelling languages but only one additional user interface language.

*If you are a user of a "minor" language, you might just be lucky enough to be able to download an LIP (Language Interface Pack) for your language that includes the ability to run Office in that language, the spelling checker of that language, and sometimes even a help systems and templates in that language. Under http://tinyurl.com/2g6y39l you can find out right here which approximately 60 languages with what features are supported for that program.*

On the bright side, it is much easier to find, purchase, and install the language packs in *Office*. You will just need to select **File> Options> Language** in an *Office* program and there you'll find the respective links.

What you'll also find in that dialog is a link for a free download to have the *ScreenTips* (previously called *QuickInfo*—the tidbits of information that you get when you put your mouse cursor on any item in the user interface) in any language. Depending on your perspective, having this feature can fall anywhere on the spectrum between helpful and fun—but either way I recommend that you download it in a language that is not covered by the user interface.
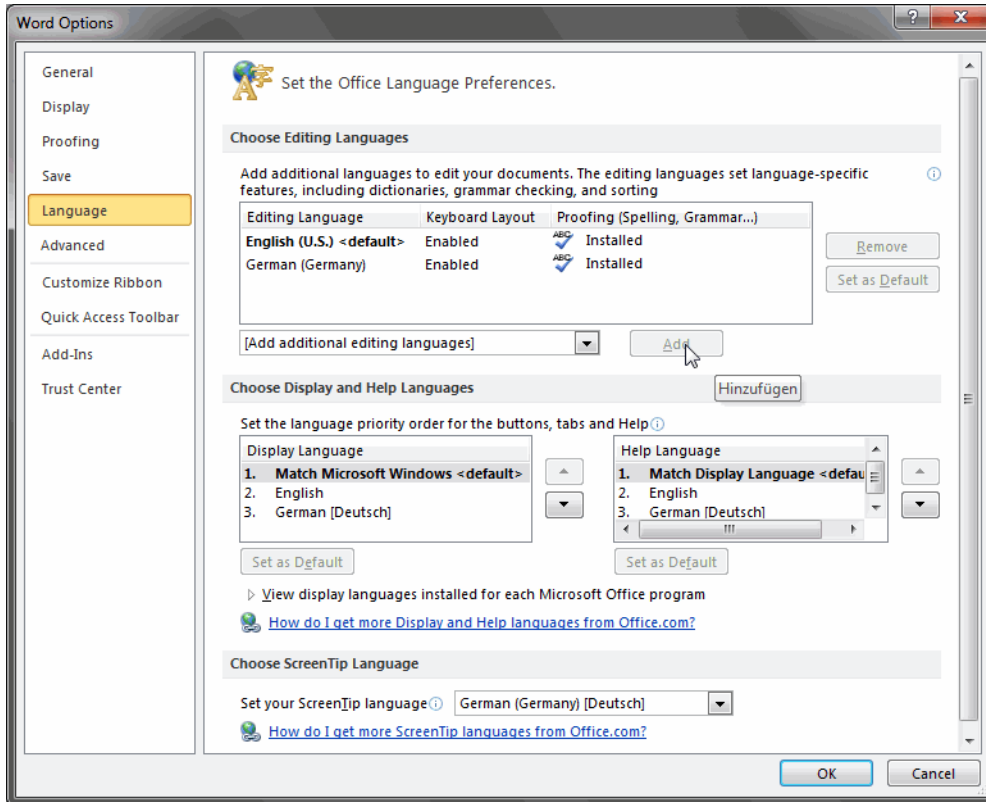


Figure 114: Language options in *Word 2010* with the **ScreenTip** option in different language (see ScreenTip for **Add**)

## Different Versions of Excel

*Excel 97* put an end to *Excel 95*'s limitation of 255 characters per cell (and upped that to 32,000) and 16,384 rows in a worksheet (extended to more than 65,000; *Excel 2007/2010*: more than 1,000,000). This was a significant improvement for translators who work with *Excel* to store or exchange glossaries and/or databases, which often go beyond either of those original limitations.

There is little difference for translators between *Excel 2000*, *XP*, *2003*, and *2007/2010* with the exception of a couple of functions:

• *Excel XP* and above have more advanced support for non-Western languages than previous versions.

- *Excel XP* and above provide a helpful integration into *Internet Explorer*. If you have *Excel XP* installed, you can right-click on any web page that contains a table (most web pages do) and select **Export to Microsoft Excel** from the shortcut menu. The text of this web page will automatically be copied into a newly created *Excel* spreadsheet. I find this helpful for copying some online glossaries.

Figure 115: Exporting glossaries from the web to *Excel*

- *Excel XP* and above offer a new way of listing search results with accompanying hyperlinks in the **Find and Replace** dialog (under **Edit> Find**; *2007*: **Home> Editing> Find & Select**). This makes it very easy to search glossaries in *Excel*.



Figure 116: Search results in *Excel 2003*

### Different Versions of *Word*

*Excel 2000, Word 2000, and Access 2000* added the very important Unicode support. *Word 2000* also allows you to save a text in a different code page, a feature that comes in very handy in many situations.

*You can access this feature under **File> Save as> Encoded Text** in Word 2000, **File> Save as> Plain Text** in Word XP and Word 2003, and Office button/**File> Save as> Plain Text** in Word 2007/2010.*

*A code page is a set of characters used to represent the characters of a particular language or several languages.The original (DOS) ASCII character set with 256 characters was sufficient for English and some Western languages, but not for many other languages, including the "double-byte" languages (Chinese, Korean, and Japanese) and the right-to-left languages (Hebrew and Arabic). These languages have their own code page. To unify all these efforts, Unicode was developed to include most (and eventually all) characters of all languages.*



Figure 117: Code page conversion in *Word 2010*

An interesting new feature for translators was the introduction of automatic language detection, i.e., *Word*'s ability to detect different languages automatically and activate the appropriate spelling and grammar checkers (if installed). This option is accessible under **Tools> Language> Set Language** (*Word 2007/2010*: **Review> Set Language**).



Figure 118: Automatic language detection in *Word 2007*

The drawback: *Word* documents with this feature enabled are significantly larger, and the (invisible) tags that *Word* places around special characters to detect the different languages tend to interfere with other programs in which you may process the *Word* document. Turning off the automatic detection will not delete the tags. To delete those, you will have to save the document to an earlier version in which this feature was not supported.

There were few new features in *Word XP* that would make a great difference to translators—even considering the new **Translation** option in *Word XP*—with the possible exception of a completely new view of tracked changes and comments (in any of the versions, this is a very important feature for anyone who does editing or proofreading work) and the ability to select several places in one document to apply formatting simultaneously (by pressing the **Ctrl** key as you make your selections).



Figure 119: **Track Changes** view and multiple selections in *Word XP*

*The Track Changes functions are accessible under **Tools> Track Changes** (up to Word 2003) or **Review> Track Changes** (in Word 2007/2010) (you can also activate them by double-clicking the little TRK/ Track Changes button in the status bar of Word).*

When you use the **Track Changes** function in Microsoft Word, be aware that there are important pitfalls to avoid.

Some clients like you to use the **Track Changes** feature so that they can get an impression about the quality of the original translator (or about how much you may over-edit a text…), while other clients want a clean text that has all editing marks removed and that can be finalized without further ado.

For a client of the second category it is not sufficient to simply hide edits from the screen view (by selecting **Tools> Track Changes> Highlight Changes> Hide from Screen** before Word XP, by selecting the appropriate **Show** command on the **Reviewing** toolbar in Word XP and 2003, or by selecting the appropriate command under **Review> Tracking** in Word 2007/2010). To make sure that you have deleted all edits in a document, select **Tools> Track Changes> Accept or Reject Changes> Accept All** (before Word XP); in Word XP and 2003, select **View> Toolbars> Reviewing**, click on the little down-arrow beside the **Accept Change** button, and select **Accept All Changes in Document**; and in Word 2007/2010 select **Review> Accept> Accept all Changes in Document**. Of course, with any version of Word you can also make case-by-case decisions to accept or not to accept an edit (the easiest way is to right-click the edits and select the appropriate commands).

*Word 2003* had even fewer new features that are relevant to translation or—for that matter—new features of any kind. . . . One thing that may prove to be helpful for you as you edit or proofread is the **Compare/View Side by Side** feature (accessible through the **Windows** menu in *Word 2003* and under **View> Windows** in *Word 2007/2010*) that allows you to simultaneously scroll through two texts at the same time.
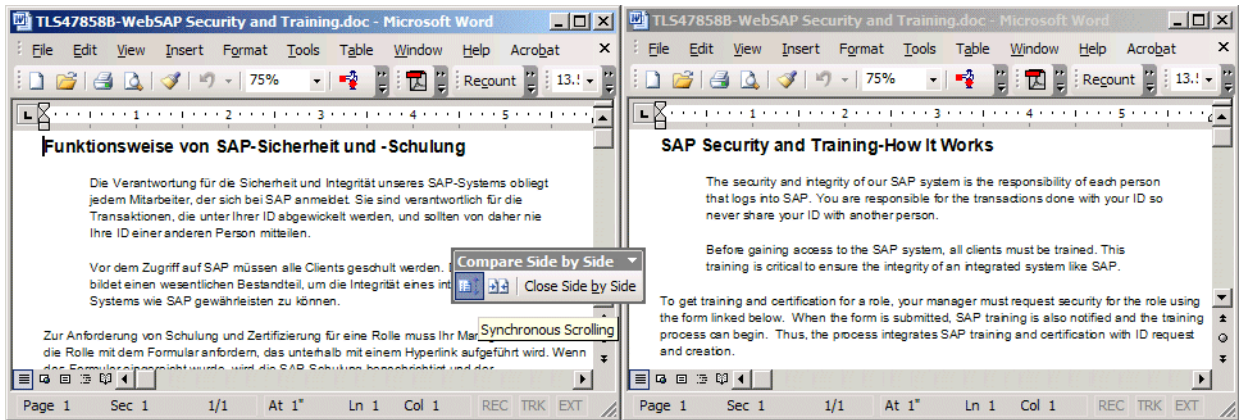


Figure 120: Simultaneous scrolling in *Word 2003*



*This feature is also available in Excel.*

*Word 2010* offers one new feature that many of you will appreciate: the **Navigation** bar, a three-tabbed pane to the left of your open document (of course, you can move it anywhere else inside or outside the Word window) that appears when selecting *View> Navigation Pane* or by hitting CTRL+F. Yes, you read that right: CTRL+F, the one shortcut that we all know to search for

something within a document. The reason for this is that one of the tabs, and possibly the most helpful, is indeed a search tab in which you can search for text or other items. All occurrences of the searched item will be listed, and you can then jump to it by clicking on the listings.



Figure 121: Navigation pane in *Word 2010* with list of occurrences of searched term

*If you need to do some "advanced searching" with the traditional Find dialog, you can select **Home> Find> Advanced Find** or press C<small>TRL</small>+H to open the **Replace** dialog. The other two tabs on the Navigation pane are the Heading and the Page view. The **Heading** view is kind of a fancy **Outline View** of earlier versions of* Word *where you can see the structure of your document by its headings and then rearrange or apply properties for whole sections. The **Page** view is very much like the **Pages** view in Acrobat or Adobe Reader where you have thumbnail views of each page that you can jump to by clicking on it.*

Another concept that Microsoft uses in its last few editions of *Office* is "Research." Most *Office* applications now have a **Research** command (available on the **Tools** menu or with the clever key/mouse combination A<small>LT</small>+C<small>LICK</small>) that allows you to automatically search a number of associated

dictionaries, thesauri, and other sources of information. The fact that the information provided differs radically between different languages shows that this is a concept which still needs some maturing; still, it can be helpful in some cases.



Figure 122: Doing research on an English term

**Different Versions of PowerPoint**

Anything below *PowerPoint 97* is usually not supported by translation environment programs, so it is advisable to have at least *PowerPoint 97* if you plan to translate *PowerPoint* files in computer-assisted translation applications.

**Different Versions of Outlook**

In my opinion, *Outlook 97* was completely useless as a standalone email client because it was primarily geared to work with *Exchange* servers. Microsoft apparently shared this opinion, leading to the release of *Outlook 98,* which was shipped with later versions of *Office 97*. Due to many of the anti-virus security measures that Microsoft tried to implement in the later editions of *Outlook*, many features that I thought were very useful in *Outlook 98* and *2000* are gone in *Outlook XP* and above (such as the feature to automatically add contacts to your list of contacts when you reply to an email).

Aside from the continued security limitations, *Outlook 2003* and above are still new programs in comparison to their predecessors. Although these aren't translator-specific issues, it made a huge difference to my daily workload not to have to deal with spam anymore because of *Outlook 2003*'s filtering capability. It also helped me manage my correspondence by sorting my email into virtual favorite folders that show me the messages I have not read or that I need to follow up on.

A welcome addition to *Outlook 2007/2010* is the ability to subscribe to RSS feeds—little snippets of information from blogs, news sources, etc.

*RSS stands for a lot of things, but the easiest is apparently "Rich Site Summary." This is a technology that allows you to be very specific about what kind of information you would like to have sent to you. As translators, for instance, we regularly visit a variety of websites: newspapers in our source and target languages, translation-specific newsgroups, discussion groups on translation tools and various other translation-related topics (many of which are located on groups.yahoo.com), and whatever else we desire for our non-translation lives (let's hope there is such a thing).*

*Many of the websites and other online resources mentioned above allow you now to subscribe to what are called "RSS feeds." RSS feeds consist of XML-based data that looks really ugly if you view it as a text file or right in your browser, but it looks quite proper if you display it in a specific RSS reader. So, because I'm interested in international news from the New York Times, and because I also know that the NYT publishes very well-defined RSS feeds, I now subscribe to those (see www.nytimes.com/services/xml/rss). Instead of having to go to www.nytimes.com every half hour, the NYT now sends me data blurbs in real time with links for more information. And the same applies to all of the other sources I mentioned above along with many, many others.*

*To collect these RSS feeds you can use standalone desktop programs, web-based programs, plug-ins for browsers, or e-mail programs. Almost every browser and/or email application now supports them directly without the aid of a third-party tool.*

Another helpful feature of *Outlook 2007/2010* are the integrated search features that previously were only available through third-party add-ons.



Figure 123: *Outlook 2010* with an applied search and various RSS feeds

## Other Office Applications

For any of the other programs that may or may not be directly part of the shrink-wrap version of *Office* but are considered by Microsoft to be part of the *Office Suite*—including *Publisher*, *Visio*, or *Project*—it probably does not make a large difference to any typical translation task which version you have (if any).

### Compatibility

There is the issue of compatibility, of course. Before *Office 2007* all files of all versions from *Office 97* on were compatible with each other (with the exception of *Access* and *Publisher*). Most newer *Office* files cannot be opened in versions of *Office 95* and earlier.



Figure 124: Warnings for loss of features when down-saving *Word* document

The different applications of *Office 2007/2010* use a different file structure (and in fact, even a different set of extensions) and are not compatible with earlier versions. However, it is possible to down-save any file within *Office 2007/2010* applications to an earlier format (select the Office button (**File**)> **Save as**). Alternatively, Microsoft has also released a *Compatibility Pack for Word, Excel and PowerPoint* at http://tinyurl.com/y5a879.

So, to come back to the question of the upgrade: Should you upgrade? It probably would be a <u>very</u> good idea to have at least *Office 2000* because of its Unicode support. Otherwise you should also upgrade if you think you would lose work and/or productivity otherwise. If none of the above applies, I would advise waiting for your next computer.

# Working with Office

Aside from *Outlook*, which I use as my email client, I rely relatively little on *Office* applications, because hardly any of my translation or editing work is done directly in any of its programs. I do translate many *Word*, *Excel*, and *PowerPoint* files, but I typically use a translation environment tool to retrieve the text from the original document (for which I typically do need the *Office* program installed on my computer). I then translate or edit it in that environment rather than directly in the *Office* applications.

## Working with Word

Because *Microsoft Word* is such a highly customizable program, there is a plethora of tips and tricks for working with *Word*.

*There are many, many resources for Word out on the Internet. The most comprehensive that I am aware of is the Word Most Valuable Professional site at word.mvps.org.*

Here I am specifically concentrating only on tasks that are important to us as translators, and/or actions that are not easily obtainable through *Word*'s menu commands. If you use *Word* a lot, you should familiarize yourself with the following options:

- **Tools> Options** (before *Word 2007*)—This is kind of like the "command center" for many functions that can be modified in *Word*. It's a good idea to learn about the commands on the different tabs. I have found that the

**View** tab is an especially good place to look when you're not happy with the way certain features are displayed in *Word*.

Figure 125: The **View** tab on the **Options** dialog in *Word*

In *Word 2007*, you can access the **Options** dialog by selecting the **Office** button and clicking on **Word Options** in the lower right-hand corner, in *Word 2010* you access it by selecting **File> Options**.



Figure 126: The **Advanced** tab on the **Options** dialog in *Word 2007*

• **Tools> Macro> Record New Macro** (*Word 2007/2010*: **View> Macros> Record New Macro**)—This is a very easy way to "record" macros (i.e., a series of commands that can be initiated by a keystroke or toolbar button) for recurring tasks.

- **Tools> Customize**—This allows you to modify *Word*'s toolbars to your heart's content.



Figure 127: Dragging a new button to the *Word* toolbar

*If you want to get back to Word's original toolbar configuration, you can simply delete the normal.dot (Word 2007: normal.dotm) template in which all these settings are stored (up to Windows XP it is typically installed under* `C:\Documents and Settings\<user>\Application Data\Microsoft\Templates,` *in Windows Vista and 7 it is to be found under* `C:\Users\<user>\AppData\Roaming\Microsoft\Templates`*).*

In *Word 2007*, it is unfortunately not possible to modify the toolbars (ribbons). It is only possible to modify the *Quick Access Toolbar* on the very top of the Word window (by clicking on the down-arrow or by right-clicking on any button and selecting **Add to Quick Access Toolbar**) or the status bar at the bottom of the window (by right-clicking on it).

The developers of Word 2010 have fortunately listened to the outcry of the user community and made the ribbons customizable again (under **File> Options> Customize Ribbon**).

- You're probably familiar with *Word*'s **Format Painter**, the icon with the paintbrush. You can click on or select any text in your document, select the **Format Painter**, and then copy the formatting of the selected text by highlighting another block of text. What you may not know is that you can also use the same procedure and double-click the **Format Painter** icon. After being double-clicked, the icon remains activated and the desired format is available to you until you press the **ESC** key.

> *Speaking of formatting: If you need to get rid of formatting for a specific block of text, you can highlight that text and press* C*TRL+*S*PACE.*

- Everyone knows that you can highlight a word or a section of a document and drag it to another part of this or any other currently open document. What many of you may not know is that you can also drag any fragment onto your desktop. There it will become a "Scrap" file which will open on double-click as a new *Word* document or which you can drag into any other *Word* document at any other time. Furthermore, if you drag that fragment to your desktop while pressing C*TRL+*S*HIFT*, *Windows* creates a little scrap link that opens your document to the very place in the original text if you click it. This is a great trick if you're tired after a long day of translating or editing and want to jump right back the next morning with fresher eyes to the place where you left off.

> *It is unfortunately no longer possible to create a scrap on your desktop with Word 2007/2010.*

- Opening *Word*: By default, *Word* opens a new blank document when you open the program. Sometimes this is helpful, but often it is not. To avoid this, right-click on the icon that you use to open *Word* with, select **Properties**, and add " /n" to the **Target** line. Now *Word* will open without a blank document.



Figure 128: Changing the open properties for *Word*

*This is especially helpful when you need to open a document that is not configured to be opened in Word. Instead of having to go through the rather tedious process of **File> Open**, etc., you can now simply drag the document from Windows Explorer into Word, where it will open automatically (see page 15).*

If you need to open a new file in *Word*, you can do that either by pressing CTRL+N, which will open an empty blank document, or you can select **File** (*Word 2007*: **Office** button)**> New**. The menu command will not open an empty file but instead give you access to your templates. (In *Word XP* and *2003* a side bar will be displayed. Clicking on the **General Templates** option will display the following dialog.)



Figure 129: Templates dialog in *Word 2007*

• Before processing *Word* documents, it may be advisable to disable all language-specific functions such as optional hyphenation and spell checking (accessible through **Tools> Language**; in *Word 2007/2010* hyphenation is accessible under **Page Layout> Page Setup> Hyphenation** and spell checking under **Review> Proofing**). You should also make sure that there are no tracked changes and/or comments in the *Word* file; these typically should be reviewed and accepted and—in the case of comments—deleted.


*These options are accessible through **Tools> Track changes** (Word 2007/2010: **Review> Changes**). In Word 2000 and below you will have to manually locate the individual comments, right-click them, and select **Delete**.*

• If you work with programs that automatically try to run templates in *Word*—as a translator you're likely to have a translation environment tool, voice recognition program, *Acrobat*, or one of the other programs that do this—you might quickly get annoyed with the long start-up time that *Word* requires when it has to load all these templates. Or, even worse, when it crashes because some of the templates conflict with each other.

The easiest way would be to just delete the templates. But in certain situations they do offer functionality that you want to use.

Here's what you can do: Move them out of a startup folder and into a folder where they can be started manually instead of automatically.

### *To Move Templates*

**1** Select **Tools> Templates and Add-Ins** within *Word* (*Word 2007/2010*: Select the Office button (**File**)**> Options> Add-Ins> Word Add-ins** under **Manage> Go**).

**2** The **Templates and Add-Ins** dialog appears. The templates with a checkmark are activated.

Figure 130: Selecting and deselecting templates in *Word*

**3** Though it is possible to uncheck these templates and disable them for this session, they will be started again the next time you open *Word* if they are located in a startup folder (see the **Full Path** on the bottom of the dialog).

**4** To change the location, close this dialog and the instance of *Word* and go to the *Windows Explorer* (or any other folder view).

**5** There are two different locations where *Word* uses startup folders (if you have used the default installation path):

`C:\Documents and Settings\<user>\Application Data\Microsoft\Word\STARTUP` (in *Windows NT*, *2000*, and *XP*), `C:\Users\<user>\AppData\Roaming\Microsoft\Word\STARTUP` (in *Windows Vista* and *7*)

and

`C:\Program Files\Microsoft Office\Office<X>\Startup` (all *Windows* version).



*If you are not able to find your AutoStart templates in these folders, right-click on C:, select* **Search***, and make a search for the name of the template (see Helpful Shortcuts on page 15).*

**6** Cut the templates out of these folders (CTRL+X) and paste (CTRL+V) them into:

`C:\Documents and Settings\<user>\Application Data\Microsoft\Templates` (in *Windows NT*, *2000*, and *XP*), `C:\Users\<user>\AppData\Roaming\Microsoft\Templates` (in *Windows Vista* and *7*)

or

`C:\Program Files\Microsoft Office\Templates` (Windows 95, 98, ME).



*You can also save them at a different location, but it may be helpful to have most of your templates stored in one location.*

**7** The next time you start Word, the templates will not be loaded automatically, but you can load them manually by selecting **Tools> Templates and Add-Ins**, adding the templates in question, and activating them.

## Working with Excel

For a long time, *Excel* files were almost abandoned by the translation industry. Only one computer-assisted translation tool—*Star Transit* (see *Translation Environment Tools* on page 190)—supported their translation through its own environment. Only in the last few years have most of the other major translation environment tools also started to support *Excel*. From my perspective as a translator, this early abandonment stands in awkward contrast to the relatively large share of *Excel* files that I translate.

*Many of the more general things that have been said about Word in the previous section are the same or similar for Excel, including the use of macros (see page 169), customization of toolbars (see page 170), the* **Format Painter** *(see page 171), or opening properties (see page 172).*

As far as translation-specific tasks go, a problem that I have encountered a number of times concerns the very large size of some *Excel* files. In the case of embedded objects, such as graphics or flowcharts, or a very large amount of data, this may be legitimate, but what also may cause the large file size is sloppy formatting. Every worksheet from *Excel 97* through *Excel 2003* on contains more than 65,000 rows (and *Excel 2007/2010* contains more than 1,000,000; see page 153). If only a few of these rows are used, this large number of potential rows does not contribute to the overall file size. However, if there is any kind of entry—such as a formatting mark—in any of the otherwise empty remaining rows, all of them may contribute to the file size (and it becomes a nightmare to work with these files, especially in a translation environment tool). The easiest way to fix this (after you have verified that this is indeed only unintended, sloppy formatting!) is to delete the unwanted cells. Highlight the first unused row, press CTRL+SHIFT+DOWNARROW, right-click on the selection, and choose **Delete**.

One more thing that may be important when using *Excel* is to understand the difference between comma-separated (.csv), tab-separated (.txt), and *Excel* (.xls) files.

*Excel* files are complex files that can contain formatting, embedded objects, formulas, and numerous worksheets. In comparison to that, comma-separated and tab-separated files are very simple text files that are built according to this pattern (for tab-delimited files, replace the comma with a tab):

```
"first record in first row","second record in first row","third record in first
row"
"first record in second row","second record in second row","third record in
second row"
```

If you open this file in *Excel*, it will be displayed just like an *Excel* spreadsheet; in fact, in many cases, the file will automatically open in *Excel* when you double-click on it. The reason why these files are so often used is that these formats provide for generally accepted ways of exchanging data between all kinds of databases. The Microsoft glossaries (see page 101) that are delivered as .csv files provide the best case in point.

One of the most exciting *Excel* add-ons that makes many of the text-related (and other) tasks in *Excel* a lot easier is *ASAP Utilities* (see message URL www.asap-utilities.com). This free collection of programs contains more than 300 (!) different utilities to streamline working in *Excel*.

Some of the functions that I really like include the ability to count characters in individual cells (a command in the **Information** submenu), helpful formatting and selection functions, and the ability to write numbers with a leading zero (it was always a pet peeve of mine that you couldn't do this.) During installation, you will be asked whether you would like to have it started

every time you start *Excel* (I chose "**Yes**"). *ASAP Utilities* shows up as a separate menu in *Excel* (or as a toolbar—the applicable command is available in the **System** submenu). If nothing else, you'll enjoy seeing what some of the other 95% of *Excel*'s unused features are . . . .



Figure 131: *ASAP Utilities* menu in *Excel*

## Working with PowerPoint

For translation tasks, *PowerPoint* is certainly the most tedious of the *Office* programs. This is mainly because of the people who primarily use it—marketing people—and their lack of understanding of how to properly format a document. For instance, in almost every *PowerPoint* presentation you will be presented with issues such as hard returns for line breaks. Before processing a *PowerPoint* presentation in a translation environment tool, it is always a good idea to spend a few minutes going through the document and cleaning up its gross formatting sins.

Before quoting on a *PowerPoint* project, always make sure that all text is actually translatable and not an embedded object such as a graphic. You can check this by right-clicking on the slide. If picture-related commands show up (see graphic below) or the picture toolbar appears, you are dealing with a graphic rather than text.



Figure 132: Picture-related commands on *PowerPoint* context menu



*For word counts in PowerPoint, see page 139.*

Before you start to work on translating a *PowerPoint* document, check whether you have installed all fonts that are used in the document. You can do this by opening the document and selecting **Format> Replace Fonts** (*PowerPoint 2007/2010*: **Home> Editing> Replace Fonts** under **Replace**). If any of the fonts under **Replace** displays a question mark, contact your client to either send you the font or to tell you to change it to another font.



Figure 133: The **Replace Font** dialog in *PowerPoint*

### Changing spelling languages in *PowerPoint*

To say that it is annoying to change spell-checking languages in *PowerPoint* is an understatement. It's easy to change the spell-checking language for individual text boxes, but not for a complete presentation.

But do not despair: here is way to do it once and for all. Though this may seem a little technical, it is not nearly as bad as it first seems. Let's first cover *PowerPoint 2003* and below and then continue with the different process of *PowerPoint 2007/2010*.

### *PowerPoint 2003* and below

On the Microsoft website under support.microsoft.com/ default.aspx?scid=kb;en-us;245468, you can find a macro that allows you to change the spelling language of your *PowerPoint* presentation into the language of your choice. Within the macro you will find the line

```
shp.TextFrame.TextRange.LanguageID = msoLanguageIDEnglishUS
```

To change this declaration into the language of your choice, you'll need to change the `msoLanguageID` at the end of that line. You can find your language (as well as many other languages you most likely have never heard of) at msdn.microsoft.com/en-us/library/aa432635.aspx.

To insert a macro into a **PowerPoint** presentation, select **Tools> Macro> Macros** (or select **Alt+F8**). Type in the name *Lingo* and select **Create**. Delete everything in the right-hand window (that starts with "Sub Lingo()"), paste the code from the Microsoft website, and change the `msoLanguageID`.



Figure 134: Adding the macro in *PowerPoint's Visual Basic Editor*

Close the Visual Basic editor and you will now have a macro in your *PowerPoint* presentation.

To run the macro, you need to once again select **Tools> Macro> Macros** (or select Alt+F8), highlight the *Lingo* macro, and select **Run**. The spelling language in all text boxes should now be changed.

Since you don't want to recreate this process for every file you are working on, you now need to find a way to make this macro available for all files (unlike in *Word*, macros in *PowerPoint* are stored on a per file basis).

The first thing step is to verify the security level for your *PowerPoint* installation. Different versions of *PowerPoint* contain different levels of security measures, some of which prevent running of "Visual Basic projects," i.e., macros from other files. In *PowerPoint XP* and *2003*, you need to have **Trust access to Visual Basic project** enabled under **Tools> Security> Trusted Publishers** (*PowerPoint XP*: **Trusted Sources**), and the security on the **Security Level** tab needs to be set to **Medium**. In **PowerPoint 2000** there are no limitations and you don't need to do anything.

When these settings are secured, you can use the macro from another file in your current file. There are two ways to do this. You can have both files open, select **Tools> Macro> Macros** (or select **Alt+F8**), and select **All open presentations** under **Macro in**. This will display the macro in question that you can now run. Or, much easier, you can create a button that contains the macro. Open the file with the macro, select **Tools> Customize> Commands**, and scroll down the list of **Categories** until you get to **Macros**. On the right-hand side you will see *Lingo*. Drag *Lingo* to your toolbar and that new button will now give you access to the ability to change your language for any **PowerPoint** presentation, PROVIDED that you do not delete the original .ppt file that contains the macro.



Figure 135: *PowerPoint's* with a newly created Lingo button

## *PowerPoint 2007/2010*

Unfortunately things are not made easier for the newer versions. Still, it's possible to change the spelling language and here is how you would do it:

Open a text file and copy the following content into that file:

```
Sub Lingo()

   Dim sld As Slide
   Dim shp As Shape

   For Each sld In ActivePresentation.Slides

      For Each shp In sld.Shapes

         If shp.Type = msoTextBox Or msoPlaceholder Then
         If shp.HasTextFrame Then

            shp.TextFrame.TextRange.LanguageID = msoLanguageIDEnglishUS

         End If
         End If
      Next
   Next

End Sub
```

Just like for the earlier versions of PowerPoint, the line

```
shp.TextFrame.TextRange.LanguageID = msoLanguageIDEnglishUS
```

contains within the `msoLanguageID` at the end of the line the declaration into which language the file needs to be changed. This needs to be changed to your desired language. You can find your language at msdn.microsoft.com/en-us/library/aa432635.aspx.

Make the change and save the file as a *.bas file.



Figure 136: Saving the macro as a *.bas file in *Notepad*. Note that placing the file name in quotation marks prevents the addition of .txt at the end of the file name

For the next step, you'll have to do some preparation (you'll need to do this only once). If the Developer ribbon is not displayed, select

- **File> Options> Customize Ribbon** and in the list of main tabs, select **Developer** for *PowerPoint 2010* or

- Office button> **PowerPoint Options> Popular> Show Developer tab in the Ribbon** for *PowerPoint 2007*.

To adjust the security settings, select Office button (**File**)> **(PowerPoint) Options> Trust Center> Trust Center Settings> Macro Settings> Trust access to the VBA project object model** and select **Disable all macros except digitally signed macros**.

Once that is done select the **Developer** ribbon and click on **Visual Basic** (if you don't want to deal with the **Developer** ribbon you can also press ALT+F11).

In the now opened Visual Basic Editor select File> Import and browse to the location where you saved the *.bas file and open that file.

Once that is done, all you need to do is to press F5 or click the Run button and the language in your PowerPoint file will be changed.



Figure 137: Visual Basic editor ready to run the macro

OK, I admit. This sounds all a little complicated, but here is the bright side: most of the steps only have to be done once per computer (or PowerPoint installation). Once the *.bas file is created, the Developer ribbon ready and the security settings selected, all you need to (for every PowerPoint presentation) is to

• Press ALT+F11 or select **Visual Basic** on the **Developer** ribbon

• Open the *.bas file

• Press F5 or press **Run**.

Not too bad!

Here is the caveat, though: for none of the *PowerPoint* versions, the *Notes* text is changed by any of these acrobatics. That will have to be done manually one by one. Oh well.

# Computer-Assisted Translation Tools

I am a passionate user of CAT tools and have many of the commercially available tools installed on my computers, but the point of this chapter is not to promote one tool over another. Instead, I want to convince you that you can reach new levels of productivity and quality if you harness some of these tools described here. For that reason, I will do my best in this important section to give you a comprehensive overview of what tools are available.

First of all, the category of CAT tools encompasses much more than "translation memory tools." By definition, any tool that is specifically designed to aid the translator in the translation process falls under the category of CAT tools. In the following pages, I will focus on *translation environment tools*, but will also talk about other kinds of tools.

*Here's why I use the phrase "translation environment tool" (TEnT):*

*Ironically, language professionals have created very poor names for a number of things in the translation industry. One of my pet peeves is "localization" (or, worse, "l10n")—a term that seems to invite everyone to come up with his or her own definition, creating great confusion in the process. Though I've given up on changing that, I would like to change the way we use "CAT" and "translation memory tool."*

*CAT, or computer-assisted translation tools, is a great term for describing the numerous families of software tools that translators use for their work (see above). Unfortunately, we often use "CAT" as a synonym for so-called "translation memory tools," when the latter is really only a sub-category of the former.*

*I also think that the term "translation memory tool" is not serving us well. "Translation memory" is just one feature included in tools like Trados, memoQ, Déjà Vu, Transit, etc.—albeit an important one. Terminology management, analysis, code protection, project management, batch processing, spell checking, code page conversion, and many more features are also part of these tools. In fact, some of these features, especially terminology management, are or should be central to how we work with these tools. But by naming this category of tools "translation memory tools," we focus almost exclusively on this one feature, and in turn overuse it disproportionally.*

*The phrase "translation environment tool" (TEnT) describes much more accurately the various ways that we should use these tools in our translation work.*

I like categorizations because they sometimes help to convey the big picture more clearly, so I have created three different categories of CAT tools:

**1** tools that independently provide specific functions for the translator.

**2** tools that provide functions which enhance the use of TEnTs (Translation Environment Tools).

**3** tools that provide a comprehensive environment for a large variety of translation-related features, i.e., TEnTs.

There certainly is a lot of overlap between these categories. Many TEnTs, for instance, also provide many of the features that the more specialized tools provide. However, it's sort of like *MS Word*: it does pretty much everything, but the more specialized tasks (such as word counts, working in text-based or HTML or XML files, switching code pages, etc.) are performed much better and more efficiently by the more specialized tools. And that's not too surprising—highly charged, passionate folks are investing all their energy in doing one thing right, so it would be a shame if they could not get that done extremely well.

Here are the functions of the **first main category** and examples of the tools that cover those functions:

- **Resource lookup**—These include tools such as *Wordfinder* (see www.wordfinder.com), which assembles dictionary resources and provides lookup in them from any *Windows*-based tool, and tools like *IntelliWebSearch*, which allows for lookup in online and offline resources (see page 124).

- **Terminology management**—Tools such as *Lingo* and *AnyLexic* specifically allow you to import term lists or termbases, add to these from any *Windows*-based tool, and provide lookup in these termbases (see *Standalone Terminology Tools* on page 269). Another kind of terminology management tool would be something like *ApSIC XBench*, which allows you to easily manage, convert, and search large existing term lists like the ones from *Microsoft* or *Apple* (see page 101 and page 272).

- **Project management**—There are a whole slew of tools that provide project management functions for companies, starting with tools like *Worx*, *Plunet*, *OTM*, and *]project-open[*, and continuing to some of the

workflow tools that also provide the features of TEnTs, such as *Star James* or the corporate solutions from SDL. And then there are also tools that are primarily geared toward the freelance translator. You can find much more information on these tools on page 282.

- **Word counts**—These are explained in more on page 134. What is important to realize, of course, is that these tools don't count on the basis of repetition or memory matches—that is the forte of TEnTs.

- **Software localization**—Tools that provide these functions are really in a class of their own. Their core function is to provide an environment for both the translator and software engineer that allows them to translate, test, and fix the localized software before deployment. You can find more information on these tools on page 276.

The **second category** is made up of tools that cater to the needs of TEnTs— either by making them better in a specific area or even giving them additional abilities that they flat-out don't have.

- **Term extraction**—This is a really useful but woefully underused feature that allows you to extract term lists from translation memories or translated and aligned documents. It's extremely helpful because it allows you to build up the core of your terminology database for your TEnT almost on the fly, thereby foregoing the often tedious and painful way of doing everything manually. You can find more information on page 265.

- **Text extraction**—In a really interesting (and smart) move, some TEnT makers have stopped building solutions of their own for very complex formats—notably for *QuarkXPress*. Instead they are promoting the products of other companies that allow you to extract text from complex formats and then process it in a regular TEnT. You can find information on a number of these tools under *Desktop Publishing Formats* on page 287.

- **Conversion and Maintenance**—This refers to the conversion that is specifically geared toward translation format. The open-source *Olifant* offers some conversion and many maintenance features for translation memory files, as does the above-mentioned *XBench*. You can find more information on these tools under *Translation Memory Quality Maintenance* on page 265.

- **Alignment**—For the longest time this was just an awful after-thought component of TEnTs, but there are now some very specialized products out there and they make all the difference as far as I am concerned! You can find more information on the process and the tools that support it under *An Excursion: A Word of Caution About Alignment* on page 234.

- **Quality assurance**—Tools that offer only quality assurance—or maybe it should be called quality control (they are typically used as a last check)— are intriguing on two levels: They provide high-level quality control on translation files and translation memories, but they have also forced TEnT vendors to respond and include QA components on their own. Find more information under *Quality Assurance Tools* on page 255.

And lastly there are, of course, the comprehensive TEnTs.

# Translation Environment Tools

## What Are Translation Environment Tools (TEnTs)?

In a traditional sense, these are applications that extend the memory of the translator by allowing him or her to build up databases of translated material and leverage that against newly translatable content.

In reality, they're usually more than that because most of them

- also allow the user to build up terminology databases that complement and extend the functionality of the translation memories, and

- allow translators to work in very complicated file formats that they may not understand or otherwise be able to support by hiding or protecting the code and displaying only translatable content.

Furthermore, many of the tools provide methods for analysis, quality assurance, and productivity.

## A Little Bit of History

TEnTs have been around for long enough that they sport a history of their own.

*Trados*, today's market-leading tool, was originally developed by the Trados translation company. In 1990 it released its first commercial product *MultiTerm* (Trados' terminology management component), and in 1992 *Workbench* (Trados' translation memory application) for DOS was released. In 1994, Trados released a *Windows* version with a *Word* interface (see www.lim.nl/monitor/trados-2.html).

In that same era, several other translation environment tools also entered the public arena.

The translation agency Star released a product that was originally designed for in-house use: *Star Transit*, with its terminology component *TermStar*. IBM released its *Translation Manager* (*TM/2*) product in 1992 (which it buried in 2002 and revived once again in 2010 as the open-source product OpenTM2— see page 224).


*Curiously, these three tools all were initially developed in the small German town of Böblingen.*

As the first *Windows*-based commercial product, Atril's *Déjà Vu* was released in 1993.

The last few years have seen a number of new translation environment tools enter the market (see *Categories of Translation Environment Tools* on page 192) and a number mergers and acquisitions of translation environment tool vendors as demonstrated by the acquisition of *Trados* by SDL in June of 2005 and of *Idiom* in 2008, or the acquisition of the German *Logoport* by Lionbridge in early 2005.

In 2009, long after IBM had decided to withdraw from the translation tool market, another truly big-time player, Google, entered the fray with the release of the *Google Translator Toolkit* (see page 227).


*Old tools are discontinued at nearly the same pace, such as Alpnet's (now SDL) TSS/Joust, SDL's Amptran, Quintillian, Clear-CAT, SDLX, and Cypresoft's Trans Suite 2000.*

There is a very impressive list of all kinds of translation software programs (in particular machine translation) published by John Hutchins, the great chronicler of translation software. His *Compendium of Translation Software— directory of commercial machine translation systems and computer-aided translation support tools* (see www.hutchinsweb.me.uk/Compendium.htm) really is a very interesting document, if only to see how much software there actually is to support our work. One very practical application of the document is the index of language pairs for machine translation in the very back of the manual. I often receive questions about whether certain language pairs are supported by a particular system. Well, here are the many answers.

## Who Should Use Translation Environment Tools?

The most obvious users would be translators with repetitive texts. In my opinion, translators who work in technical, medical, or legal fields waste a lot of time (and money!) if they do not make an initial investment in a translation environment tool. But it doesn't stop there. Because most of the available packages include decent or good terminology management tools (see *Terminology Handling* on page 244), any translator who has to control terminology can greatly benefit from these applications.

Also, if you work in more complex file formats than *Word* documents or you do not want to worry about formatting, TEnTs separate translatable from non-translatable content and will help you tremendously.

Or if you would like to use more advanced quality assurance features than just spell-checkers, you should also look at TEnTs.

Or if you ever need on-the-fly access to previous translation—TEnTs can do that for you.

Are those enough good reasons?

## Categories of Translation Environment Tools

For organizational purposes I have always tried to classify the available tools into work interface categories. While I will still use this classification for now, it is on the way to becoming obsolete since *Microsoft Word*, originally the preferred interface, has lost its importance and is increasingly becoming an interface for entry-level TEnTs.

Still, these are the main categories:

- tools that perform all or most of their work through macros in *Microsoft Word* that allow an association with translation memory(s) and terminology database(s)

- tools that let the translator work in an independent, mostly tabular environment

- tools that strictly use an online interface

In the following sections, I will introduce the different tools within their categories, briefly describe the one or two outstanding features of the different tools, and eventually spend more time with three veterans—*Trados*, *Transit*, and *Déjà Vu*—to describe the typical features of a TEnT in more detail.



*Tools that I will not include are those that support only one or two language pairs, such as the Danish-English(-German) WebWordSystem (see www.webwordsystem.com) or the Russian-English MT2007 (see mt2007-cat.ru) or tools that support only one file format, such as Webbudget (see page 114).*

*I also will not discuss translation management systems including proprietary systems, such as translation.com's GlobalLink, commercial tools like Andrä's ontram, and open-source tools like Globalsight Ambassador. These tools are becoming increasingly important for our industry, but there are a number of distinctions that made me exclude them. While most of them support exchange standards, their workflow does not allow for third-party tools to participate. This means that if your client uses one of the above tools, chances are that you will have to use the translation editor that comes with the tool. The good news is that these editors are typically free; the bad news is that you have to get used to a new work environment and are often not able to use your own resources (translation memories, terminology databases, etc.). Also, a purchase or an implementation of these tools, if at all possible, is only feasible for the very large language providers or the translation buyer.*

## Tools That Use Microsoft Word as Their Main Translation Interface

The most well-known application that formerly made heavy use of the *Word* plug-in was the market-leader *Trados* (see www.trados.com), but with its latest release—*Studio 2009*—it has finally completely abandoned *Word* as a translation interface. Other tools that still support the *Word* (or, in the case of *Anaphraseus*, *OpenOffice*) interface as their only or their most important interface include

- the open-source *Wordfast Classic* pendant *Anaphraseus* (see anaphraseus.sourceforge.net),
- *AnyMem* (see www.anymem.com),
- *JiveFUSION* (see www.jivefusiontech.com),
- *MetaTexis* (see www.metatexis.com),
- *MultiTrans* (see www.multicorpora.com),
- *Similis* (see www.lingua-et-machina.com),
- *Snowball* (see www.trmem.com),
- Lionbridge's *Translation Workspace* (see www.geoworkz.com), and
- *Wordfast Classic* (see www.wordfast.net),

This group can be further categorized by its range of applications:

- *Wordfast* covers a relatively large range of applications and formats because it essentially comes in two different versions, *Wordfast Classic* (formerly *Wordfast 5.5* and below) with its *Word* interface and *Wordfast Pro* with a *Java*-based interface that allows for the processing of many other file types (including *InDesign*, *FrameMaker*, XML, and other *Office* formats aside from *Word*).

- Both *MultiTrans* and *Translation Workspace* offer an *XLIFF Editor* for any kind of tagged formats (*MultiTrans* also comes with additional interfaces for *WordPerfect* and *PowerPoint*).

- The other tools are primarily focused on native *Word* files or other formats that can be accessed through *Microsoft Word* (either through a tagging mechanism or by "calling" into other applications).

### Wordfast Classic

The most successful tool in this group presently is *Wordfast Classic*, a tool developed by Yves Champollion. Yves is related by name and blood to Jean-François Champollion, the fellow who translated the Rosetta Stone. The history of the product itself is a little more mundane but still rather interesting. Released as a freeware product in 2000, *Wordfast* stunned the market—particularly *Trados*, to which *Wordfast* at first sight looked very similar. In August 2001, Champollion and the Italian translation agency giant Logos formed a joint venture but continued to give the program away for free. The partnership ended about a year later, and in October of 2002 *Wordfast* became commercial. In 2007, *Wordfast* again allied itself with a translation giant, this time translations.com.

*Wordfast Classic* has garnered a significant following among freelance translators, and its reputation typically is that of an inexpensive and easy-to-learn tool.

One clear advantage of *Wordfast* is that it runs on any platform supported by *MS Word*, including *Macintosh.* (It does not on *Word 2008* for Mac, but it does work again on *Word 2011* for Mac). A drawback is naturally that the amount of supported file formats is limited. While it is possible to tag and process HTML and XML files through the free *PlusTools* application (which can also be downloaded on *Wordfast*'s website) or to "call" content from other MS Office formats into Word and translate it there, *Wordfast Classic*'s main format is *Word* files.

*To mitigate this limitation, a completely revamped version of Wordfast— Wordfast Pro—was released in 2009. This version supports both Word and other file formats (see page 213).*

One outstanding feature is VLTM. VLTM stands for *Very Large Translation Memories*, and these are TMs that are hosted on a *Wordfast* server in 10 different language pairs (most of the European languages, but also languages such as Japanese, Arabic, and Korean, all of these into or out of English). The contents come from a variety of non-copyrighted sources that were aligned and are provided free of charge to registered users of *Wordfast*. All you need to do is install a little plug-in that you can download on the *Wordfast* website, enable the VLTM search in *Wordfast* (click the little *f* icon on the toolbar and select the **VLTM** tab under **Translation Memory**), and you automatically connect to the language combination on the *Wordfast* server that matches the combination of your main TM. The search is very fast and the number of hits

quite impressive, but it's disappointing that there is no indication of source or subject area, so I'm not sure how helpful it would be as a very reliable translation memory; however, as a general resource VLTM is certainly valuable.



Figure 138: *Wordfast* translation with match from VLTM

*While Wordfast was by far the earliest version that directly integrated data from online resources, a number of competitors now connect to resources such as MyMemory—mymemory.translated.net—(Trados Studio, memoQ, and MultiTrans) or the TAUS Data Association's memory data—www.tausdata.org—(Translation Workspace, MultiTrans, GlobalSight).*

### JiveFUSION

*JiveFUSION* (formerly *Fusion*) has been primarily active on the Canadian market so far. *JiveFUSION* is a tool that strongly emphasizes collaboration—it was one of the first tools that offered server-based translation memories.



Figure 139: *JiveFUSION's* translation view

### Similis

*Similis* differentiates itself from most other tools because it comes with a very high-level linguistic "knowledge" in seven EU languages (English, Dutch, German, Spanish, Italian, Portuguese, and French) which it derives from a powerful engine that was originally developed by Xerox for its *XTS* tool (see page 267). This engine gives *Similis* the analytical power to apply linguistic rules to a number of processes, including alignment and automatic extraction of terms and phrases from translation memory content. In both of these processes *Similis* offers extremely high accuracy (of course, only in the language pairings mentioned above). For the actual translation process it offers two different environments: a hybrid *Word/Similis* environment for the translation of all files directly compatible with *Word* (*Word*, RTF, text files,

etc.) and a separate environment for HTML and XML files. What makes the translation memory matches remarkable is the existence of "chunks," fragments of translation memory matches that the program was able to automatically extract from larger matches with the help of the *XTS* engine.



Figure 140: *Similis'* translation view with access to internal dictionaries

*In late 2010, the makers of Similis announced that the fully functional Freelance edition would be a free download, effective immediately!*

### MetaTexis

*MetaTexis* is a German tool with a surprisingly powerful and highly configurable database engine behind it. While currently all files are still processed in *MS Word*, a new independent interface is being developed.



Figure 141: *MetaTexis'* translation view within *MS Word 2007*

### MultiTrans

*MultiTrans* does not completely fit into this category. In fact, it is not a "traditional translation memory" tool to start with, but a "bi-text" or "corpus" tool, or, according to the tool's latest terminology preference, a "text-base translation memory" tool. Rather than matching on a sentence-by-sentence level, *MultiTrans'* corpora are full source and target texts with an approximate matching capacity that allows alignment to be done virtually on the fly. What also distinguishes corpora from traditional translation memories is the display of all the context of the original text.

*MultiTrans* was originally designed to cater to the needs of the Canadian government, whose millions of pages translated from and to French and English made it unreasonable to go through a manual alignment process.



Figure 142: *MultiTrans'* translation view with *MS Word* on the bottom and *Translation Agent* on top

> *Aside from the Word interface, MultiTrans also offers the translation of files in a PowerPoint and Word Perfect interface as well as a completely independent XLIFF Editor (which needs to be purchased as an add-on) for the translation of tagged file formats, including HTML, XML, InDesign, and of course XLIFF.*

## Translation Workspace

*Translation Workspace* is based on *Logoport*, the product that Lionbridge purchased from a German company and has used since 2006 for most of its own translation projects.

In 2010 Lionbridge decided to start offering this as a SaaS—Software as a Service—product to everyone else. The means that your actual usage and makes you pay accordingly (with a price that is determined by a mixture of users and words). The exception is that if you work for Lionbridge, you "only" have to pay a base fee and no actual usage fee (this last point has raised considerable controversy).

*One interesting aspect of this is that you can upgrade or downgrade the contract at any point, therefore making it possible to dynamically respond to business fluctuations.*

The system itself is a hybrid system. While all the work is done on your computer, with all the documents that you are translating on your local machine, the supporting data (TM, glossary—it really is not a full-fledged termbase—and all administrative controls) are based on Lionbridge's servers. The interface in which you translate is either within *Word* (including *Word 2007*, though *Word 2010* is not supported yet) or an independent tool somewhat reminiscent of *Trados TagEditor*, called *XLIFF Editor*.



Figure 143: *Translation Workspace's Word* interface with TM and terminology matches and a preview feature

The *XLIFF Editor* is able to translate *Office 2007* files, *Trados* TTX files, *FrameMaker* files, and XML- and HTML-based formats. The *Word* interface can access any *Word* or RTF-based file.



Figure 144: *Translation Workspace's XLIFF Editor* with TM and terminology matches

During the translation, you'll have to connect to a "tenant," which is essentially a space on the server that can hold translation memories, glossaries, and review packages and that can be shared with others.

One feature that is unique is its approach to the review process. This takes place in a separate, completely web-based, tabular interface with error-tracking, version control, etc. Though you will have to expend some extra effort to create the review packages (upload the translated, bilingual files), you'll have the benefit that the very last version of your translated and edited files ends up in the translation memory.

### Snowball

*Snowball* is a new kind of translation memory tool (I'm actually not sure that it is a "true" TEnT since most of the translation environment it provides is indeed the translation memory—they use the phrase *Integrated Translation Environment*). The proposition of this tool is to be as quiet and unobtrusive

and easy as it possibly can be. The 2MB installation allows you to create bilingual TMs that you can use while translating *Word* documents. It does not have a separate terminology database, but you are encouraged to enter term pairs alongside complete sentence pairs when you use the "word for word" rather than the "whole sentence" mode. It's an inexpensive tool for the very occasional translator, someone who is not willing to learn anything more than a few shortcuts (which truly is all you need to operate it) and who has tons of perfect matches—there is no fuzzy matching yet.



Figure 145: Translation in *Word* with an attached *Snowball* TM with two perfect matches

## AnyMem

Like *Snowball*, *AnyMem* is a low-priced tool that runs exclusively within *Word*. While it offers a term search within the translation memory, it does not have an independent terminology component. It uses the same system to process text within bi-lingual *Word* that *Trados 2007* and earlier and *Wordfast Classic* use; as a result the files are interchangeable.



Figure 146: Translation in *Word* with an attached *AnyMem* TM and a fuzzy match

## Anaphraseus

*Anaphraseus*, formerly *OpenWordfast*, is a tool with an infrastructure and display very reminiscent of *Wordfast Classic*. The only major difference is that, unlike its commercial sibling, it does not run in the *MS Word* interface but only within *OpenOffice* for *OpenOffice* files or *Word* files that have been converted to the *OpenOffice* format. (And of course it is not nearly as versatile

or stable as *Wordfast*.) I am not sure that this tool would be a good choice as a beginner's TEnT since it is still (and has been for a long time) in beta stage, but it's something to keep in mind to look at once it's more stable (or, since it is an open-source tool, to help make it more stable).



Figure 147: Translation with *Anaphraseus* within *OpenOffice*

## Tools That Use an Independent Desktop-Based Translation Interface

The second group—the category that presents all files in a uniform independent format through a desktop application—includes:

- *Across* (see www.across.net),

- *Alchemy Publisher* (see www.alchemysoftware.ie/products/ publisher.html),

- *Déjà Vu* (see www.atril.com),

- *Fluency* (see products.westernstandard.com/fluency.html)

- *Heartsome* (see www.heartsome.net),

- *MadCap Lingo* (see www.madcapsoftware.com/products/lingo)

- *memoQ* (see www.memoqtm.com),

- *SDL Trados* (see www.trados.com),

- *SDLX* (no longer marketed but included in the *Trados Studio package)*,

- *Star Transit* (see www.star-transit.com),

- *Swordfish* (see www.maxprograms.com), and

- *Wordfast Pro* (see www.wordfast.com),

as well as the open-source tools

- *ForeignDesk* (see www.foreigndesk.net)—see page 225,

- *OmegaT* (see www.omegat.org),

- *Open Language Tools* (see open-language-tools.dev.java.net), and

- OpenTM2 (see www.opentm2.org).

### *Across*

The German tool *Across* heavily emphasizes workflow aspects of translation while equipping the translation professional with all the tools (translation memory, terminology database, QA tools, etc.) needed to translate successfully. Of course, the first thing that differentiates *Across* is the price tag. It's free for freelance translators! And while the free freelance edition seamlessly hooks into the workflow created by the corporate editions, it is completely functional on its own.

The underlying database system is an *SQL Server* system (very powerful but also very resource-heavy) in which all TM and terminology entries are stored simultaneously. One differentiator is that *Across* offers morphological recognition in its term searches in various European languages, including English, Danish, Dutch, Finnish, German, Norwegian, Russian, Swedish, French, Italian, Portuguese, and Spanish. Of course, all other languages work with perfect term recognition.



Figure 148: *Across'* translation view

## SDLX

*SDLX* has really seen very little development over the last few years, and with the release of *Trados Studio 2009* it is not being marketed anymore by its parent company SDL. *SDLX* was one of those tools that users either loathed or loved with its idiosyncracies like the famous "format painting." The desktop version is/was in many ways a toned-down version of *Déjà Vu*, but unlike *Déjà Vu* it had a fully functional and high-powered enterprise version as well.



Figure 149: SDLX' translation view

## memoQ

The Hungarian *memoQ* is a very process-oriented tool that makes the general workflow user-friendly even for a novice to TEnTs. The first screen that you see when you open the program gives you all the different options of what you can do; the creation of a translation project, a translation memory, or a termbase can be done without having to enter complicated information. The actual translation interface again looks very similar to the veteran *Déjà Vu* (see page 241), where the translatable text is presented in a table format, the

source on the left and target on the right, and matches from termbase and translation memory are displayed on the side. The import and export of files goes blazingly fast, and this is true for translation files as well as when you import TMX into a translation memory. The supported file formats include HTML, Microsoft *Office* formats, RTF, *FrameMaker*, *InDesign*, text, TTX, RESX files and XML. The latest format it supports are *Transit* project files.

Some of *memoQ's* features are unique:

* "translation memory driven segmentation" matches the segmentation (how a translation file is separated into segments) to the associated translation memory. This is especially helpful if you have a translation memory from a different tool that may have used a different set of segmentation rules. By employing this feature, the new translation file will match its segmentation to translation units in the TM.

* One other unique feature is the oddly named "longest substring concordance" (or, just as odd: LSC). This is an attempt to automate concordance searches according to user-defined parameters and without interrupting the workflow. In the **Translation results** pane there are now not only translation memory matches, terminology matches, and assembled rows but also ominous-looking matches of subsets of the string that needs to be translated with empty targets. These are the LSC matches. Clicking on them will produce the **Concordance** dialog, which will list all (or a predefined number of) the appearances of that particular substring in the translation memory within the context of the complete strings in the TM.

* *LiveDocs* is a concept that was introduced with version 4.5 as an attempt to quickly access earlier, TM-external translation without the painful alignment process. Rather than taking one source and one target file, matching them up, and then fixing it manually, you can now take any number of file pairs, align them on the fly, and keep them as matched up file pairs for reference purposes. Aside from that you can also use bilingual files (such as XLIFF files) and monolingual files (for reference purposes). In any given project, matches will show up just like TM matches, with the difference that you can see that they come from a LiveDoc rather than a TM.

Also, for HTML and *MS Office* files there is an integrated preview feature right in the translation window.



Figure 150: *memoQ's* translation view with the preview feature at the bottom and LSC hits under **Translation Results**

## Alchemy Publisher

*Publisher* is a little different from most other tools in how it extracts text from the originating documents, in particular when it comes to *FrameMaker* and *Word* files. Rather than converting the files into an interim format (.rtf in the case of *Word* and .mif in the case of *FrameMaker*), it communicates directly with the application and extracts text on an object basis. This means that

even within the *Publisher* interface, it is apparent where a specific piece of text originated—whether from a text box, a heading, an index maker, etc. The translation memory and terminology databases are simple text-based files. Direct access to *Trados* TMs and termbases is also possible.

In addition to *FrameMaker* and *Word* as well as HTML, XML, HTMLHelp, and software-related development files, *InDesign* and *Quark* are supported (the latter two only after a conversion through the third-party plug-in *CopyFlow*—see page 301).



Figure 151: *Alchemy Publisher's* translation view

*A clear drawback of Publisher's approach to Word and FrameMaker files is that you will need to have Word or FrameMaker installed on your computer to process the files. While this is typically no problem with Word, FrameMaker is not an application that is to be found on everyone's computer.*

### *Wordfast Pro*

*Wordfast Pro*, which originally was marketed as the successor of *Wordfast 5.5* (today: *Wordfast Classic*, see page 195), has now instead become a a parallel version. Until further notice *Wordfast Pro* and *Classic* are supported equally.

The concept of *Wordfast Pro* is very different than the *Classic* version. Rather than using a third-party interface (*MS Word*) for its translation, it comes with its own refreshingly simple and well-organized interface. The tool is *Java*-based so it runs on *Linux*, *Mac*, and *Windows* and all files, independent of type, can be viewed the same way and in the same interface.

The interim format into which files are converted for translation purposes is a custom XML format (TXML), and the supported translation file formats include *MS Office* formats, HTML, *FrameMaker*, PDF (through an external but integrated plugin), and *InDesign*.

Overall, the response of users of the *Classic* version to the new tool with its new interface has been cautious (thus the parallel maintenance of two separate versions) and it remains to be seen what the future holds. On the positive side, the underlying TXML file format is the same that translation.com's translation management system (TMS) *GlobalLink* uses so that it's likely that *Wordfast Pro* will see a push from that side.



Figure 152: *Wordfast Pro's* translation interface

### Heartsome and Swordfish

*Heartsome* and *Swordfish* are both *Java*-based tools that run on *Mac*, *Linux*, and *Windows*. In this same spirit of supporting exchange formats, they not only support the translation of XLIFF files (see page 278), they actually convert a number of formats (including RTF, *Office 2007*, *FrameMaker*, HTML, *OpenOffice*, *InDesign*, and a variety of software development formats) to XLIFF, provide for their translation within that format, and then convert them back into their original format.

While there will most certainly be fundamental differences between the two different tools, as of now (December 2010) they are quite similar. This is not too surprising since the main developer who originally worked for *Heartsome* has now branched out on his own with *Swordfish*. The first few differences can be found in the support of some varying formats (*Swordfish*, for instance, supports the new *Wordfast Pro* format TXML) and in their differing licensing schemes.

*The "original" Heartsome also is a split-off from its original company. In 2002 Klemens Waldhör, an employee of Alpnet, co-founded Heartsome Singapore and Heartsome Europe together with a fellow Alpnetian. This lasted until 2004 when the companies split into two independent entities. Heartsome Singapore started to market the Heartsome tool suite, and Heartsome Europe's product was called Araya (see www.heartsome.de). Araya was customized for corporate customers with XML- and server-based processes, and the Heartsome tool suite was and is marketed to the whole range of actors in the translation industry.*

*Waldhör is now in the employ of FOLT, the German industry association that develops the open-source translation management system OpenTMS (see www.folt.org), so we can expect to see some of Araya's design in that system in the near future.*

Along with the actual translation editor, both come with a variety of other products (one for XLIFF, one for TMX, one for the termbase exchange format TBX, and various conversion utilities).



Figure 153: *Heartsome's Translation Editor* on a *Windows* computer

Figure 154: *Swordfish's Translation Editor* on a *Windows* computer

## *Fluency*

*Fluency's* translation interface is a multi-pane interface with access to a number of the linguistic resources mentioned above, including a pane for translation memory matches, a source and a target, and a target preview pane. However, the interface is not in the traditional table-like format with

source on the left and target on the right; instead, you'll see a source pane that displays the text with all its context (sort of) and the target pane that displays segment by segment independently (the current segment is highlighted in the source pane.



Figure 155: *Fluency's Translation Editor*

The supported file formats include all the expected formats, but also some surprising ones, including *Microsoft Publisher* files and, as mentioned above, PDF.

*I can see some of you cringe when you read "MS Publisher"—yes, I know, it may have the reputation of a desktop-publishing program for dummies, but who wants to say no to a well-paying client with Publisher files to translate? And to my knowledge, Fluency is the only tool on the market that supports Publisher files.*

*Fluency's* developers have collected all kinds of processes and third-party utilities that they felt would be helpful in the process of translation, integrated them into their tool and its workflow, and left it up to us whether to use them. These features include PDF conversion (from and to PDF), optical character recognition (OCR), an interface for manual transcription of items that can't be processed with optical character recognition, extensive language-specific glossaries, direct links to any number of web-based resources, and audio recognition.

It's important to remember that most of these are language-specific. So you will have to verify that your language or language combination is supported before you do some extensive testing. In fact, when you test or purchase *Fluency*, you do that in language-combination-specific editions (or in a more expensive edition for all supported languages).

Other unique features include "Track Term" which allows you to determine a translation for a certain term and force the program to warn you every time you choose not to use that and various panes to access language-specific Internet resources.

### MadCap Lingo

*MadCap* is the company that split off from Adobe after some of *MadCap's* current owners felt that Adobe was treating the help authoring product *RoboHelp*, which it purchased as part of a larger acquisition, too shabbily. So they started their own company and have since given Adobe a run for its money. (Once they were gone, of course, Adobe resumed work on *RoboHelp*.) *MadCap's* main product is the help authoring product *Flare*.

Early on, the people in charge at *MadCap* recognized that there was a strong link between the language and technical writing industries. This has finally resulted in the release of *MadCap Lingo*, a translation environment tool that easily integrates into the authoring/translation environment of *Flare* but can also be used as a standalone TEnT for file formats such as *MS Word* and *PowerPoint* (all versions), HTML, XML, DITA, and .resx files.



Figure 156: *MadCap Lingo's* translation interface with an HTML file

*MadCap Lingo* is a solid and user-friendly tool that performs very well, certainly and particularly with *Flare* projects, but may not have some of the bells and whistles of its more well-known competitors.

Interestingly you can either purchase a moderately priced license outright or you can buy a time-limited license for 3, 6, or 12 months. This might make a lot of sense for projects in which you'll have to temporarily work on *Flare* files but otherwise choose to use another tool.

*Note that MadCap Lingo does not work with bi-directional languages.*

### Text United

As of this writing (December 2010), *Text United* is still a tool in flux, and yet it has a very interesting concept behind it. Though it is a desktop-based tool, all its resources—including the translation files, the TM ("repository"), and the termbase ("glossary")—are server-based. And similar to Lionbridge's *Translation Workspace* (see page 201), this is sold as a SaaS, "Software as a Service," so you'll have to pay a monthly fee for its usage. In this case, however, the fee doesn't depend on the volume of words you translate but on the number of users.

The translation interface looks very modern with an *Outlook*-like **Home** screen, ribbons instead of menus, and a very lightweight application.



Figure 157: *Text United's* translation interface

Processes are role-based, i.e., a project is assigned to a translator (who can be identical with the project owner). The translator only has as many rights to the project and its resources as the owner assigns to him or her.

When importing a file (supported file types presently include HTML, *MS Word*, XML, and *InDesign*), a terminology extract of the file's content is performed that can serve as the base glossary for the translation of the project (for terminology extraction, see also page 265).

The makers behind *Text United* envision this tool as a stepping stone toward a future in the translation industry where it will be easier for translation buyers to directly contract with individual translators—through the database of users and the interface of the tool.

Clearly this is an interesting proposition: one that has been tried in the past but might just work in this case—after all, "disintermediation" is one of the hot topics these days in our industry .

### *Open-Source Tools*

As operating system-independent as *Heartsome* and *Swordfish* are two open-source tools, *Open Language Tools* and *OmegaT*. Both of them are *Java*-based and are thus supported on *Linux*, *Macintosh*, and *Windows*.

- *Open Language Tools* converts SGML, XML, HTML, *OpenOffice* and a number of software formats to XLIFF, and translates those within its own translation memory environment.



Figure 158: *Translation Editor's* translation view on a *Windows* computer

- *OmegaT* (see www.omegat.org) translates HTML, OpenXML (for *MS Office* formats), XLIFF, PO, text, LaTeX, various software development, and *OpenOffice* files and is the chosen tool for a very passionate and

knowledgeable crowd of open-source-loving translators. It may be the tool with the fastest pace of development.



Figure 159: *OmegaT's* translation editor on a *Windows* computer

- In 2010 LISA, the Localization Industry Standards Association, partnered with IBM, Welocalize, Cisco, and Linux Solution Group to resurrect some components of the IBM *TM/2* product and release it to the open-source community as *OpenTM2*. LISA's expressed goal with this tool is to establish it "as the reference implementation of existing localization industry standards, such as TMX." IBM's goal is to widen the pool of translators familiar with a tool that originates from the tool that they still use in-house. So far not so good. The tool is used by very few (if any) translators, partly due to the fact that it supports only HTML, *Java* .properties, *OpenOffice*, XLIFF, and generic XML files, but also because it's a very

outdated tool that you would have to definitely feel passionate about to work with effectively. It also does not feature a terminology component.



*Though neither of these tools supports the wide range of formats that translators typically have to work in, they certainly point to a possible future of translation environment tools: independent open-source applications.*

*Interestingly, it was translation giant Lionbridge that first released open-source software in 2001 with ForeignDesk, but it never really caught on, partly because of its interdependence with many of the Trados-tagged formats.*

### Tools That Use an Online Interface

After Uniscape's failed attempt to provide an online interface to translation memories (the company was swallowed by *Trados* in 2002), in 2006 *Lingotek* (see www.lingotek.com) was the first to come out with a new attempt at this kind of interface and approach to a translation environment tool. The distinguishing factors of these tools are that a) they are completely online-based so there is no need to install any software on your computer (no worries about updates, etc.); b) your data (translation files, translation memories, termbases, etc.) is stored not on your computer but on a remote server; c) the tools are typically offered through a SaaS framework, meaning you have to pay a monthly or annual licensing fee rather than buy a software license with (quasi) perpetual validity; and d) you will have to have an online connection to work.

Other companies that offer a similar interface include:

- *Google Translator Toolkit* (see translate.google.com/toolkit),

- *Wordbee* (see www.wordbee.com),

- *XTM Cloud* (see www.xtm-intl.com/xtmcloud), and

- *Wordfast Anywhere* (see www.wordfast.net/anywhere)

### *Lingotek*

*Lingotek* does not actually store translation memory-like data; instead, it indexes the data and allows you to access the index (which it calls "vault"). What this means in practice is that the only data sitting in the indices is source and target data without any descriptive data, user information, etc. These vaults are searched not just on a fuzzy and perfect match level like traditional translation environment tools, but they are also searched at a sub-segment level, meaning that theoretically every word or combination of words in the source text is searched in every combination (you can adjust a threshold for how many words in sequence you will want to have searched).

When you create a project, you can choose between loading data into a public vault that can be accessed by anyone, or into a private vault that only you and other parties that you determine can use.



Figure 160: *Lingotek Workbench's* translation interface with term match from the public TM and machine translation by *Google Translate*

The makers of *Lingotek Workbench* have recently been trying to newly brand their tool as a "crowdsourcing" tool, i.e., a tool that allows simultaneous access by many who can evaluate each other's translations.



*One area of concern with Lingotek is the fact that internal formatting of segments is handled rather crudely: the application of formatting is counter-intuitive and formatting markers are not stored in the vaults.*

### Google Translator Toolkit

In the summer of 2009, Google finally released its long-awaited translation environment tool: *Google Translation Toolkit* (*GTT*).

Unlike what was expected by many, this tool contains virtually no translation/ project management features, but instead a rather well-designed frontend that allows you to do one of two things: you can either upload a file (in HTML, word processing [not .docx!], or text files) or you can specify a URL and the corresponding HTML page will be uploaded.

In the process of setting up your project, you will need to choose whether you are using a "shared" translation—i.e., whether you are using and contributing to a large anonymous translation memory—or whether you would like to upload your own translation memory in TMX format. Here you can also upload or define a glossary (in a very strictly defined CSV format).

While you are working on your translation (or before, or after), you can also invite others to participate in your translation/editing efforts by selecting **Share> Invite people**. All they need is some kind of *Google* account.



Figure 161: *Google Translator Toolkit* with access to machine translation matches

So, why does *Google* offer the service? While it does state in the FAQ section that it may at some point charge power users, its main objective is not to generate cash but instead to generate high-quality data. *Google*'s machine translation engine *Google Translate* uses a statistical machine translation engine that relies on good bilingual data—lots of good bilingual data. So, any translation that you add (or improve) contributes to the quality of the machine translation performed by *Google Translate*. While there may be nothing wrong with that in exchange for a free tool, it is also possible to upload existing (TMX) translation memories. You can specify whether these are for your use or for your use and that of others, but they are always used by Google. Even after you "delete" them from the system. Many will not like that.

## Wordbee

*Wordbee* is a completely web-based translation environment tool that attempts to combine a translation environment with some relatively intuitive project management capabilities.

Since *Wordbee* is a web-based tool, you don't purchase a license but you pay a usage fee according to the SaaS—Software as a Service—concept. If you want to use it in a team environment, you can purchase any number of floating licenses that you can pass among your team members. A freelance license gives you space for 250 MB for documents and 250,000 segments in the TM; if you go beyond that, you can buy more space or download the data to your desktop.

Despite being browser-based, *Wordbee's* translation user interface is very simple and intuitive: keyboard shortcuts can be assigned and customized; right-click shortcut menus can be accessed in different areas of the program; and the translation view (in tabular format) along with the TM, dictionary (many preconfigured online dictionaries), preview, comments, and revision history (!) panes are also freely configurable.



Figure 162: *Wordbee's* translation interface

The list of supported file formats is solid (*MS Office*, *InDesign*, *OpenOffice*, RTF, XML, HTML-based formats, as well as various software development formats).

One area the developers were intrigued with is crowdsourcing (thus "Wordbee"). But rather than looking at crowdsourcing as a process with a large undefined mass (à la Facebook or Twitter), it uses the principles of crowdsourcing in a professional environment: in-house translators and project managers subscribe to RSS feeds where new projects are announced and then locked once a translator claims them.

### *XTM Cloud*

*XTM Cloud* is a product that is geared toward translation agencies and freelance translators alike. So, like *Wordbee*, it combines some project management features with translation features.

One of its most striking features is the Spartan and highly functional interface. While it is not always completely intuitive, it's well organized once you get the hang of it.

*XTM Cloud* is a SaaS-based offering, a Software as a Service. There is no download or installation; everything, including the files that need to be translated as well as the translation memory and a (full-fledged) terminology database, sits on a server.

Every translation file (the supported formats include *MS Office*, XML, *Visio* (!), *InDesign*, HTML, *FrameMaker*, PDF [converted internally to text], *Trados* .ttx files, and others) is internally converted to XLIFF. At any stage of the translation process it's possible to export it out of the system, process it on another XLIFF-supporting tool, and bring it back into the *XTM Cloud* system. This enables you to continue to work offline in case you have no Internet connection.

The prices are structured remarkably similar to Lionbridge's offering, i.e., according to the number of users and translatable words (see page 201), but are designed to beat out *Translation Workspace's* offering by between 20 and 50 percent. Also like Lionbridge, you can upgrade and downgrade your plan on a month-by-month basis (after an initial three-month commitment).

The translation interface is extremely uncluttered, and it allows you to start working with essentially no learning curve. Translation memory matches, (optional) *Google Translate* machine translation, terminology matches, version control data, and various levels of warnings are clearly displayed and highly accessible, and you can enter terminology entries to your terminology database seamlessly and without much effort.



Figure 163: *XTM Cloud's* translation interface

### *Wordfast Anywhere*

*Wordfast Anywhere* is Wordfast's answer to online translation environment tools. Unlike some of the competitors, it's probably not meant to be the one and only tool for the professional translator, but most likely a companion to the other two Wordfast products: *Wordfast Pro* (see page 213) and *Wordfast Classic* (see page 195). Its looks, though browser-based, are very much styled after *Wordfast Classic's Word*-based interface, but even if you have never worked with that tool, you should be able to start working with *Wordfast Anywhere* almost immediately.

You can either paste text from your clipboard into the translation pane or upload documents in various formats (*MS Office*, HTML, text, *FrameMaker*, *InDesign*, or the *Wordfast Pro* format TXML). Once your document is ready for translation, you can set up whether you want to use your own TM (and whether you want to keep that to yourself or share it with everyone else), use the large, public VLTM database (see page 196), and/or MT through *Google Translate*. (To set up all these settings, select **Tools> Open Setup Dialog Box**—the language of the user interface is a little erratic.)



Figure 164: *Wordfast Anywhere's* translation interface

Every user has their own workspace in which they can store up to ten documents. Once it's full they can either delete or download the documents (same with the translation memory and the glossary: they also can be downloaded at any time.) The size limitations clearly excludes *Wordfast Anywhere* as your primary tool, but it just might be the tool to use when no other tool is at hand. Particularly because it's free.

# Comparing TEnTs in Detail

In the sections below, I will concentrate my introduction primarily on the veterans: *Trados*, *Transit*, and *Déjà Vu*. I am doing this not to endorse these over any of the others but because these are good examples of full-fledged tools with very different product philosophies. Also, I will be talking about two different versions of *Trados* (*2007* and *Studio 2009*). While you might feel this is a little bit of a *Trados* overkill, the *Studio 2009* version is not only wildly different from the previous version of *Trados 2007* but it's also far from decided whether it is going to be the successor of the market-leading tool *Trados 2007*.

## The Similarities

In some areas, the major tools have become increasingly similar (aside from their inherent similarity that they all have translation memories and display translatable text):

- All of them have a freely configurable terminology component that plays a fairly major role in the translation workflow.

- All of them process texts in Unicode, thus giving access to all languages that are supported by the *Windows* operating system.

- All of them have a feature that allows "alignment" (see *An Excursion: A Word of Caution About Alignment* on page 234).

- All of them provide support for TMX, the XML-based translation memory exchange format, and other formats to exchange terminology data.

- All of them support a similarly broad range of file formats (including desktop publishing, word processing, and software development formats) and tagging standards (HTML and XML).

- All of them allow concordance searches (i.e., searches for one single word or expression within larger segments) in their translation memories.

- All of them have some general word processing features, such as spell-checking or search & replace.

- All of them provide fairly detailed analysis features.

This is more or less where the similarities end.

### An Excursion: A Word of Caution About Alignment

In the context of most TEnTs, alignment refers to the process of selecting file pairs in the source and target language that were translated outside of a translation memory environment, matching all the segments (sentences, headings, etc.), and creating a translation memory database from those matches. The resulting translation memory can then be applied to translate similar or identical texts. *Trados*, *Star Transit*, and *Déjà Vu* all contain alignment modules in some or all configurations. At first glance, alignment seems like a great process that anyone starting to use a translation environment tool should do to build up a nice translation memory database.

And while it's true that alignment is indeed a helpful process, it's often misused. I've encountered many situations where new users (both freelance and corporate) became enamored with the idea of using alignment to "magically" turn their existing translation materials into one large translation memory. They spent days or weeks devoting their time to this task, and in the process they became so frustrated with the use of their new tool that they laid it aside completely. The reason that alignment is often (and correctly) perceived as a tedious process is its manual nature. Although each of the alignment modules in the above-mentioned tools applies well-chosen parameters to the alignment "suggestions," they all have to be verified, and—as anyone knows who has done alignment before—often repaired. The parameters are typically punctuation and paragraph markers, repetitions, and non-linguistic matches such as numbers and abbreviations. This can go a long way toward making correct matches, but it often requires user intervention.

Typical cases where manual changes are required are differences in sentence delimitation (one sentence in the source becomes several in the target or the other way around), shifts in the order of segments, different use and/or placement of footnotes, and index markers.





Figure 165: Alignment views in *Déjà Vu* and *Trados* (note that in both cases the programs split the first sentence incorrectly in the Spanish target).

With all these difficulties, why would alignment still be a helpful process? Alignment can be a very powerful tool if you have specific sets of already-translated documents that correspond closely to new documents that now have to be translated. The amount of time you can save and the level of

consistency and quality you can achieve by aligning the existing documents and using that as the basis for your translation can be immense, and there's simply no reason not to go that route. But for other documents, unless you can hire someone else to do mass alignment of existing materials (someone with the odd combination of being both cheap and well-qualified . . .), I would strongly advise you to build up your translation memory database by simply performing translation in the tool of your choice and adding material to your translation memory segment by segment.

Now . . . there is also a less "traditional" form of alignment, which in most situations will be more successful.

Terminotix's *AlignFactory* (see www.terminotix.com) offers an uncommonly high accuracy of alignment because a) it uses a highly sophisticated alignment engine and b) it uses a number of filters that filter out any unlikely match (for instance, based on differing lengths of segments).



Figure 166: *AlignFactory's* filter settings for alignment

Furthermore, with *AlignFactory* you can also select thousands of file pairs (including PDF files), have them matched up (they have to follow certain naming conventions such as a language identifier), and then have them aligned in one big swoosh. And it really is one big swoosh: the speed of the

alignment is mind-boggling. In fact, it's so fast that I have repeatedly thought that something had gone wrong only to find that it had already successfully completed the alignment. While it's not perfect, it certainly has brought alignment to a different level.



Figure 167: *AlignFactory's* alignment results—note the correct alignment of the first segment.



*YouAlign (see www.youalign.com) is a free service that uses the same alignment engine. While there are limitations to file size and number of files you can align it should give you a goof idea of what AlignFactory can do.*

Another Canadian project is NoBabel's *AutoAligner* (see www.nobabel.com). Unlike the other tools, *NoBabel* is a SaaS (Software as a Service), meaning you don't have to install any software on your computer. Instead, you upload files that need to be aligned to an online server, and the software on the server does the alignment for you. In fact, you don't even have to "tell" the *AutoAligner* anything about the files, not even what language they are in. The files are analyzed on the server, the language is correctly recognized, and the

file pairs are matched up on the basis of the actual content (rather than the file name, the method by which all other tools operate). Once the files are matched up, the actual segment matching is performed on the basis of a linguistic analysis. This means that not every possible pairing will end up in the translation memory, but only those that are deemed appropriate (typically up to 5% are ignored). This also means that you can upload a list of terms in two different languages, both sorted according to the language-specific sort order, and you will still have the correct alignment.

The originating file types that are supported are text, *Word*, *WordPerfect*, and *HTML*. The number of files you can upload is unlimited and the languages that are presently supported are English paired with French, Italian, German, Spanish, Dutch, Portuguese, Russian, Polish, Chinese, Japanese, Korean, and Arabic. While the limited number of language combinations is a drawback, it makes sense because of the underlying linguistic data.

Once the resulting TMX or *Trados Workbench* text format files are generated and ready to download, you'll need to make a purchase decision with a price of US$.02 per translation unit (as of November 2009).



Figure 168: Download page for *AutoAligner's* aligned translation memories.

The drawbacks of this system is the absence of any context for segments and the fact that you will not get every single possible match (since it's only matching up pairs that it can "guarantee"). So if you have to translate a manual of version 1.1 of a product for which you already have version 1.0, it might make sense to align it the good old-fashioned tedious way.

And lastly, there is *Similis* (see page 267) which, much like *AutoAligner*, aligns at a much greater accuracy than traditional tools because of its underlying data. Unfortunately, *Similis* only works in seven European languages and the alignment process is excruciatingly slow.

## The Differences

### The Translation Memories

*Transit* is fundamentally different from *Trados* and *Déjà Vu* in the sense that rather than using an external translation memory that stores previously translated information, it provides for a "virtual" translation memory—*Transit* refers to it as "reference material"—by associating already-translated files.

The benefit of this is that there is no need for an "additional" database that may tie up computer or server space. The drawback lies in the large amount of translated file pairs that have to be retained to provide for the necessary "reference material."

A major difference between the translation memory databases of *Trados* and *Déjà Vu* lies in their formats. While *Trados 2007* uses a proprietary database format, *Déjà Vu* and *Trados Studio 2009* use generic industry-standard formats.

### The Work Environment

This is not only the most obvious difference, but it has a variety of repercussions on the actual translation work.

*Transit* and *Déjà Vu* each display translations in independent interfaces that are entirely different from one another.

*Transit* provides an open, semi-WYSIWYG ("What-You-See-Is-What-You-Get") interface.



Figure 169: *Transit's* bilingual view (note the partial formatting of source and target)

Like *Transit*, *Trados Studio 2009* also displays it in a semi-WYSIWYG view:



Figure 170: *Trados Studio's* bilingual view

*Déjà Vu* displays everything in a non-WYSIWYG, strictly tabular format:



Figure 171: *Déjà Vu's* bilingual view

Users of these tools describe the benefits of handling text in independent applications, such as in *Transit, Trados Studio 2009* and *Déjà Vu*, as follows:

- Regardless of the source text and file format, the translation environment always stays the same.

- No "file conversion" is necessary to make texts display in an environment such as *Word.*

- The program is independent of any third party (i.e., an upgrade to *Word* does not have any effect on the translation environment tool).

On the other hand, *Trados 2007*'s traditional method of displaying translatable text is in *Microsoft Word*:



Figure 172: *Trados 2007's* bilingual view in *Microsoft Word*

Users who like this interface point out that, when translating DOC or RTF files, it is helpful to have a full WYSIWYG interface and to be able to work in a familiar environment.

*To be able to translate other formats that could not be handled in MS Word, Trados 2007 and earlier also offered TagEditor, an editor which at first was only used for tagged formats but was eventually able to support all formats for which Trados 2007 offered support.*

## Code Handling

With the exception of plain text files, every file type contains some kind of coding information. This coding information can serve a variety of purposes, including formatting, programming information, and hyperlinks.

Translation environment tools have to deal with two different kinds of codes:

* Code that appears only outside (or between) segments

* Code that appears within segments

For instance, if you have the following text in an HTML file:

```
<FONT face=Arial size=2><P>The <B>cat</B> is black.</P></FONT>
```

the different tools display it in the following ways:



Figure 173: From top to bottom: *Trados 2007* (*TagEditor*), *Trados Studio 2009*, *Transit*, *Déjà Vu*

* *TagEditor* in *Trados 2007* displays all tags in the same form as the original tags and write-protects them. The bolded `cat` is displayed in a WYSIWYG style.

* *Trados Studio 2009* displays the text in WYSIWYG mode (the supported formats for the WYSIWYG mode include common formatting such as bold, italic, underline, etc.)

- *Transit* encodes and displays the outside coding information, hides the inner-segment-coding, and displays the formatted `cat` in WYSIWYG style (as in *Trados 2009*, the most common formats are supported for the WYSIWYG display).

- *Déjà Vu*, on the other hand, hides everything outside the actual segment, does not display any of the formatting in WYSIWYG style, and encodes the inner-segment coding with write-protected numeric codes.

It is easy to see that there are completely different philosophies behind the way text is handled, and different users accordingly prefer different ways of encoding.

One important aspect of this is how these segments and their coding information are stored in the translation memories. All of the tools store only the text and the inner-segment codes. The difference lies in how *Trados 2007* and *Transit* on one hand and *Déjà Vu* and *Trados 2009* on the other store that information. While *Trados 2007* and *Transit* store data about the actual tagging information (`<B>` and `</B>`) in source and target, *Déjà Vu* and *Trados 2009* only store placeholders for these codes. The benefit of the former method is that you could store a change of formatting (let's say from **bold** to *italics*) between source and target languages, whereas the benefit of the placeholders is that it is 100% compatible across file formats and formatting information.

### Terminology Handling

What are terminology databases good for if all the relevant material is already located in the translation memories? For new users of translation environment programs, the use of the terminology databases often seems superfluous if not downright confusing.

This has several reasons:

- The often-used name "translation memory" program seems to suggest that the emphasis is on the translation memory.

- There is a more immediate gain through perfect and fuzzy matches on a sentence-by-sentence basis than there is with terminology databases.

- Translation memories can be relatively quickly built up by aligning existing translated file pairs and/or automatically as you translate new texts.

- The construction of terminology databases is a comparatively tedious process: terms have to be individually highlighted in the translation or even entered into the terminology management application, and additional information has to be entered.

If it is indeed so tedious to build up and use terminology databases, what makes them so important?

As every experienced translator knows, translation is much more than the mere exchange of translated segments across situations and contexts. Though most translators have one or several fields of specialization within their language combination(s), very few, if any, work exclusively in a field in which language is so controlled that there is no need for additional information on individual terms and phrases except their one-to-one translations. We all know that semantic fields of words and phrases across languages do not match each other 100% (where one word or phrase in Language A would always match one word or phrase in Language B; simultaneously, that word or phrase in Language B will always match only that word and no other in Language A).

Obviously, if things were that simple, there would be no need for translators in the first place—machine translation would have long taken over our profession as translators!

The terminology database is the place where you can invest effort into defining your words and phrases grammatically, contextually, or even by contrast. If this is very helpful for you as a single translator, how much more would it be in a virtual translators' workgroup! Of course, none of this is news to anyone: any good dictionary offers the same concept. What makes these "dictionaries" (if you will) much more exciting is that you can build them up the way you want them. Furthermore, they are "living dictionaries" that present their findings for each of the segments you are currently translating without you having to do anything (if you have previously given them the data that they now share with you).

Why then is it helpful to have numerous different translations for—let's say—"cat" ("feline animal," "computer assisted translation," "Caterpillar," etc.) come up when I translate a text?

Because of the close association of the terminology databases with your translation project, and because of all the information that you or someone else has fed into the terminology database as you entered the terms, the application will actually recognize which of these terms is more relevant than another. Depending, for instance, on whether you are translating a text of the subject area "Flora and Fauna," "Translation Technology," or "Heavy Machinery" (to stay with our silly examples above), the application will make the more likely choice for you (while still allowing you to access the other ones).

In addition, some applications not only display this data to you from the terminology database but even try to assemble it for you—i.e., piece it together—which should convince you that it makes sense to spend some time building up these databases. If you are not a translator of extremely repetitive materials, this might also convince you that these tools may have a definite benefit even for you (see page 192).

*Trados* and *Transit* have for the most part marketed and sold their terminology databases either as part of a combination with their translation memory solution or as a standalone solution (see page 276).

*Transit* is probably the tool that spent the most early effort developing a sophisticated terminology tool, *TermStar*. In the screenshot below you can see some good examples of what kind of information can be entered into a terminology database: client, date, definition(s), homonyms, and of course translation.



Figure 174: View of open dictionary (terminology database) in *TermStar*

While this tool can be opened and used as an individual application, it becomes part of the workflow and interface when used in actual translation work. The terms that are displayed in the **Terminology** window can be entered with the associated keyboard shortcut, and new terms can be entered by highlighting source and target and entering them in the attached dictionary by choosing the **Rapid Entry** button on the **Quick Access** toolbar.



Figure 175: *Transit* translation project with dictionary access (lower right pane)

*Trados'* original terminology program *MultiTerm* (5.5 and earlier) was relatively unpopular with translators—you had to type new terms into the program and it was very difficult to import and export in and out of other formats and programs. In 2002 it released a completely revamped *MultiTerm* which now has become the standard terminology component.

The advantages of the new version are that it is based on standard XML rather than a proprietary database format; it exports into XML, HTML, and RTF; term entry is made less cumbersome (you can now highlight the source term and only have to type the target term); and remote applications of the program have become easier.



Figure 176: Sample view of *MultiTerm* termbase

*MultiTerm* provides the found term as a reference to aid in fixing a found fuzzy match or translating from scratch within *Trados 2007*.



Figure 177: View of the *Trados Workbench* translation memory with a fuzzy match and reference from *MultiTerm* (right pane)

For *Trados Studio*, this procedure has been changed. Here the terminology matches are displayed in a separate **Term Recognition** pane (upper right) as well as in an automatically displayed match proposition which can be entered by pressing the Enter key.



Figure 178: View of *Trados Studio* with an automatically offered terminology match

*Déjà Vu* has gone a slightly different route with its handling of terminology. The terminology component has always been closely integrated into the translation workflow—the terminology component was never sold and marketed apart from the main application. But while earlier versions still used a separate application—*Terminology Maintenance*—to perform any major changes to the databases, the current version's database files are accessed and maintained directly from the main interface. Other major changes between the old and most recent versions are that while *Déjà Vu 2* and *3* were mainly term-based and not freely configurable, *Déjà Vu X*—like *Trados* and *Transit*—is concept-based (as the graphic below illustrates) and completely configurable.



Figure 179: Sample view of *Déjà Vu X* terminology database

Although there were clear shortcomings in the terminology handling of earlier versions of *Déjà Vu*, it was and is generally well liked by its users. The reason for this is a feature called *Assemble* that otherwise only *memoQ* uses in a similar fashion. This feature provides the possibility of piecing together segments that cannot be found in the translation memories, or fixing fuzzy matches from the translation memories and turning them into perfect matches.



Figure 180: Example of a fuzzy match (segment with "producer - producteur") turned into a perfect match by using information from the terminology database ("manufacturer - fabricant")

Of course, there are also standalone terminology tools. You can find more information on those on page 269.

## Making the Move. . .

### Getting to Know the Different Tools

Most of the larger and better-known translation environment tools are not exactly cheap (that is, if you look at the purchase price only and not at the return on investment), and it seems quite unreasonable to purchase tools until you find the one that fits your needs. The good news is that you don't have to. Most of the tool vendors offer freely downloadable trial versions of their programs, some of which have a limited time use and others a limited functionality.

Currently the best way to compare the different TEnTs is TranslatorsTraining.com, a Flash video-based online resource to compare the 16 leading TEnTs. I started this website in cooperation with Intrawelt, an Italian translation company that specializes in online training, and (yes, I know I'm biased, but I can't help myself) it serves as an incredible resource that gives you a powerful comparison of the main features of the different tools.

### How to Purchase a Translation Environment Tool

Once you finally decide to buy one of the tools, I would advise you either to see whether any of the vendors is currently offering a special (some tool vendors offer discounts of up to 20% off the normal price at TranslatorsTraining.com), or to wait for the next translation-related event in your area. Many of the vendors offer special discounts during those events.

*If you talk to translators who have been in the business for many years, some will tell you that they were actually given a copy of Trados or Star Transit by one of their clients. However, I am afraid that this is truly a thing of the past now. . . .*

## The Future of TEnTs

The TEnT market is competitive. This is a positive situation for us as users, not only because of pricing but also because of new and better features. And, not surprisingly, once a tool introduces a fundamental feature change, other tool vendors feel that they need to follow suit.

The last few waves of fundamental changes included

- moves away from the *Word* interface by some of the more established tools (*Trados* and *Wordfast*—other tool vendors, including *Metatexis* and to some degree *MultiTrans*, are still working on it),

- the implementation of more advanced quality assurance features (see *Integrated Quality Assurance Features* on page 255),

- partnerships with other vendors or tools to offer new and richer feature sets including partnerships with project management applications vendors,

- online availability of translation memory, terminology, and, in some cases, the translation data itself,

- the inclusion of features to support crowdsourcing, and

- the inclusion of machine translation (MT) features into the translation interface of the translation environment tools.

This last point—inclusion of MT features—is sure to become more prominent, and in much more subtle ways than just by offering a gist or possible translation alternative. The next big step in translation technology will be a much deeper merging of TM and MT for some industries (primarily the software industry, but also other industries, such as automobile, customer service, and social networks).

Machine translation is not easily embraced as a positive development by most translators, but here is a sobering quote by Jaap van der Meer (in *MultiLingual* 71, 2005) on machine translation and its relation to translation memory:

> Disdain on the side of the professional translators for the hilarious and stupid MT mistakes gave birth to a new variant of MT called translation memory (TM). TM started off as a lower-level feature of commercial MT systems (…). But the success of TM came with dedicated products such as *IBM TM/2* and *Trados*. The marketing message was tuned in to what the professional translation industry wanted to hear: "Forget about MT; it doesn't work well. Instead, use our TM product because it leaves you in full control of the process.
>
> The message worked well. Within a period of 10 to 15 years, TM products have found their way to the workstations of more than 50,000 translators in the world. But the message had also caused a "cognitive disorder" in the translation industry, namely that TM is good and MT is evil, foregoing the fact that TM is just a new

variant of MT (…). The damage is done, however, and it will take years to convince the community of business translators that post-editing fuzzy matches from TM databases is, in fact, not different from post-editing fuzzy matches from any other MT system.

# Quality Assurance Tools

## Integrated Quality Assurance Features

We're at a point in the development of translation environment tools when it's generally understood that translation memory tools are not primarily time-saving applications; first and foremost they are programs that allow translators to increase the quality and consistency of their work. During the translation process this happens (or should happen) with the use of terminology databases and translation memories that give on-the-fly access to previously used translation or to terminology that the client requires for the current project.

Beyond that, virtually all translation environment tools offer quality assurance features such as spell-checks or checks for formatting integrity. What tool vendors have recognized only rather recently is that there is demand for more far-reaching quality assurance features.

Though many think of *Wordfast Classic* as a rather light weight tool, it offered advanced quality assurance features long before most other tools did. *Wordfast Classic* also pioneered the separation of quality assurance features ("quality check" or "QC" in *Wordfast* lingo) into two areas: The first was terminology checks based on a terminology database, i.e., a verification process to determine whether the same translated term or phrase was used in the terminology database as in the translated file.



Figure 181: Setting which terminology database is to be used for *Wordfast's* terminology check

*It's important to realize that this feature is not equally useful in all languages. Terminology checks in languages with heavy conjugation or declination, or agglutinative languages such as Turkish or Finnish, will typically find a lot of "translation errors" that are really just different forms of the correct term. A strategy to counter that is to enter various term pairs to cover the different word forms.*

In a second component of the program, *Wordfast Classic* offer various other areas of quality assurance, including punctuation, extraneous spaces, and identical source and target segments.



Figure 182: Quality check setting in *Wordfast*

*SDLX* offered its quality assurance checks as the major improvement when it released its version 2005. At that point it was probably the most comprehensive solution.



Figure 183: The *QA Check* module in *SDLX 2006*: the **Inconsistencies** tab

However, *Trados* (whose owner also owns *SDLX*) versions 7.1 and above includes a larger set of QA features than any of its competitors.

Figure 184: The *QA Checker* module in *Trados*: **Segment Verification** (*Trados Studio* - above) and **Punctuation** (*Trados 2007* - below)

One particularly helpful aspect of the *Trados QA Checker* is the fact that you can load and save a profile (under **QA Check Profiles**), enabling all members of one translation team to use the same QA procedures.

In a sense it is not a surprise that *Trados* offers such encompassing QA features. For a number of years, several tool vendors have been offering a variety of quality assurance tools that specifically provide quality assurance for *Trados* files. So when *Trados* got ready to offer it themselves, they could just pick and choose the most helpful features.

The only tool to rival *Trados*' QA features at the moment is *memoQ*:



Figure 185: A sampling of memoQ's quality checks

## Third-Party Quality Assurance Tools

Two tools that offer approximately the same range of features are *QA Distiller* from Yamagata Europe (see www.qa-distiller.com) and *Error Spy* from D.O.G. (see www.dog-gmbh.de).

Both of these tools work on *Trados* and TMX files, as well as some other formats (*Error Spy* also works on *Transit* and XLIFF files). The criteria that are checked include terminology, consistency, numbers, completeness, tags, acronyms, typography, and missing translations.



Figure 186: *QA Distiller* displaying errors

A third, free (**!**) tool, *Xbench* (see www.apsic.com/en/downloads.aspx), also offers some QA functionality on bilingual translation environment files. And while the QA features that this tool offers (for other *Xbench* features, see page 272) are more limited—see the image below—the range of bilingual files formats that it supports is greater than its competitors with *Transit*, *SDLX*, *IBM TM*, or *Trados* files.



Figure 187: QA features in *Xbench*

## Source Document Quality Assurance

Quality assurance tools that don't rely on translation memory technology include *Acrocheck* (see www.acrolinx.com) and *FormatCheckers for Word and FrameMaker* from Star (see www.star-group.net/ENU/formatchecker/ formatchecker.html).

*CrossCheck* compares hundreds of features between source and target in *InDesign* and *FrameMaker* documents, including missing links, untranslated text, number mismatching, terminology, text overflows, font size, missing symbols, and wrong number format.

*Acrocheck* is a corporate tool to check terminology, style, and grammar of documents in German, English, or French. That doesn't sound too fancy, but I was really impressed at the level of "intelligence" of the program. For instance, you can load customized style sheets or modify the pre-loaded ones to make the program steer authors to stay within the desired parameters. One of the goals of this is to optimize your return from translation memory, because the more unified the source text, the better by far the return from TM.

Unfortunately, there is no desktop version available. However, the makers of *Acrocheck* are looking at marketing a desktop version in the mid-term future.

A similar tool, though not quite as comprehensive (but available in a desktop version), is Star's *FormatCheckers* for *Word* and *FrameMaker*. This tool checks about 50 different potential errors in Word or *FrameMaker* documents, ranging from typographical errors to duplicated spaces, paragraph marks, manual references, and many others. Much like *Acrocheck*, the intention is to create well-formed documents before the translation even starts, thus hoping for a better return on translation memory matches and/or better entry of data into the translation memory.



Figure 188: Error checking in Star's *FormatChecker*

A slightly different and yet powerful way of also maximizing translation memory content is to author—i.e., write—the source document on the basis of the translation memory. There are several tools that offer this feature, including *Congree*, based on a partnership between the TEnT vendor Across and the Society for the Promotion of Applied Information Sciences at the Saarland University (see www.congree.com), SDL's *AuthorAssistant* (see www.sdl.com/en/products/global-ams/authorassistant.asp), Sajan's *Authoring Memory Tool* (see www.sajan.com/software/authoring-memory-tool.html), or *Star MindReader* (see www.star-group.net/ENU/mindreader/mindreader.html).

Though these do not strictly fall into the category of quality assurance, this is a particularly exciting new family of tools. Tools that allow authoring on the basis of a translation memory not only extend the use of the translation memory—it is obvious that you will have a huge amount of matches in the translation portion of a project if you adjust your writing to the source part of the translation memory in the first place—but it also offers a whole new world of opportunities to language providers! All of a sudden, authoring may become a much easier new service portfolio item for individuals or companies who have so far specialized in translation only.



*Quality assurance tools of a very different nature are the SDL products HtmlQA and HelpQA (for both, see page 331). These products allow you to compare and check translated and original versions of HTMLHelp and WinHelp for any errors or omissions and fix them before you compile them into the final product (see Help Systems on page 326).*

## Translation Memory Quality Maintenance

Some of the above-mentioned tools provide quality assurance features for existing translation memories, but none of them allows you to maintain translation memories on a large scale, and while some TEnTs offer decent features to do that, many do not. *Olifant* (see okapi.sourceforge.net/Release/Olifant/Help/) allows you to open (and save) TMX, *Trados*, and *Wordfast* translation memories and filter and manipulate them to your heart's content, all in a clean, multi-column interface.



Figure 189: Open TMX file in *Olifant* with the available commands in the **View** menu

*Apsic Xbench* (see page 272) also allows for the conversion of translation memory formats or other database exchange formats, and so do some of the tools that are provided along with *Heartsome* or *Swordfish* (see page 215).

## Terminology Mining

SDL offers separate programs (*MultiTerm Extract* and *SDL PhraseFinder*) to "mine" bilingual terminology for the creation of project-specific glossaries.

These programs offer the possibility of mining existing terminology and building up terminology databases or glossaries by taking existing pairs of source and target documents or bilingual translation memories, analyzing them, and presenting you with a proposed translated terminology list. Once this list is generated, it can be used as either a primary glossary for a project (or to send to the client), or as a common glossary that can be shared among multiple translators working on this project.

*MultiTerm Extract*, the tool that originally comes from the *Trados* side of things, works on a purely mathematical level ("if word A always appears in sentences for which word B always appears in the translated sentence, then these words must form a word pair"). This means it supports all *Windows*-based languages. *PhraseFinder*, the former *SDLX* companion, works on a language-based level for English, French, German, Spanish, Dutch, and Portuguese. This means that overall all languages are supported, but users of the languages that are supported by *PhraseFinder* have drawn the longer straw since the recognition will be more accurate (on the other hand, the *PhraseFinder* process is very resource-intensive and not particularly fond of large amounts of data).



Figure 190: *Trados' MultiTerm Extract* (formerly *ExtraTerm*)

The most powerful application in the field of term extraction used to be the *Xerox Terminology Suite* (*XTS*), which was designed for the deep pockets of corporate users and was very powerful because it was based on preconfigured linguistic data in various languages. Today the suite is owned by TEMIS (see www.temis.com), where development has virtually come to a halt.

However, the translation environment tool *Similis* (see page 198) has integrated the *XTS* engine and therefore comes with a very high-level linguistic "knowledge" in seven EU languages (English, Dutch, German, Spanish, Italian, Portuguese, and French). *Similis* is able to apply a combination of linguistic and statistical rules to a number of processes, including automatic extraction of terms and phrases from translation memory content with extremely high accuracy—but unfortunately only in a handful of languages.



Figure 191: *Similis* with automatically extracted terminology

Another terminology extraction tool is *SynchroTerm* (see www.terminotix.com). *SynchroTerm* aligns texts in their various formats to bitexts, extracts terminology, and presents long lists of terminology with reference information that can then be verified, annotated, and exported to a number of formats, including *Logiterm*, the two *SDL MultiTerm* formats, *Excel*, and the machine translation tool *PROMT*.

Aside from extracting terminology from aligned documents, it is also possible to import TMX translation memories and have terminology extracted. There are a good number of settings that allow you to govern the extraction, including many fields you can automatically add to each term pair (so that your TEnT will be able to take that into consideration when processing the data and make suggestions to you).

Theoretically all languages are supported with the tool; however, practically speaking there are different tiers of language support. In general, *SynchroTerm* relies on mathematical calculations to extract terminology pairs. For a great number of Western languages it also uses long lists of stop words to filter those out automatically, and for English and French it also makes use

of stemming rules, further improving the accuracy in those languages. All other languages are not supported right out of the box, but can be added manually by adding stop lists for those languages to the file `StopLists.txt`, which can be found in the installation directory.



Figure 192: Extracted terminology in *SynchroTerm* from a TMX file

# Standalone Terminology Tools

I have tried to make a strong case for using the integrated terminology components of a TEnT (see *Terminology Handling* on page 244). In my opinion this is the preferred way of dealing with terminology since it is directly integrated into the workflow and completely interactive, meaning that matches are displayed to the translator automatically. Still, direct integration into the translation workflow does not necessarily mean that the terminology applications cannot also be used as full-fledged programs that can and are in many cases used on their own—particularly in the case of *Trados/SDLX* with *MultiTerm* and *Star Transit* with *TermStar*. But besides those tools, there are also a number of other independent standalone terminology tools.

One of the first standalone terminology tools was developed by Alan Melby in 1982 and made commercially available in 1987. *MTX* (see www.linguatech.com/mtx.htm) enabled translators to compile their own glossaries as a separate task or while working in documents. It provided macros for *Word* and *WordPerfect* so that, with just the help of a keyboard shortcut, a search for an entry in the termbases could be launched. The exchange format of *MTX* is called MicroMATER (this was later developed into MARTIF, which in turn provides the basis for today's exchange standard TBX).

*TBX, the TermBase eXchange standard, is an XML standard developed by LISA (see www.lisa.org/TBX) that allows for the interchange of terminology data including detailed lexical information. At its core it is based on MARTIF. The adoption of this standard has gone very slowly, but at this point tools like Across, Heartsome, Swordfish, XML-Intl, Trados, and Star TermStar support it.*

While this tool can still be downloaded today, it is unsupported and has been replaced with a newer generation of tools that may be easier to use and provide support for a larger number of import file formats (such as *Excel*, CSV, or TMX—but typically not TBX).



Figure 193: *MTX's* interface with the open **File** menu

*Lingo* (see www.lexicool.com) is one of those modern standalone terminology management tools. It is very straightforward to use, supports all Unicode-compliant languages, works alongside any *Windows*-based application, and allows you to export and import glossary files in the various formats mentioned above.

For translators who feel that the jump to using a TEnT is too big or who are unhappy with the terminology management in their existing TEnT, this might be a good and inexpensive solution. Like its competitors, it allows you to perform a search without ever leaving the application you are working in; simply hit a keyboard shortcut, which then calls up the application with the search results window.



Figure 194: Search results in *Lingo*

A tool with a very similar set of features is *AnyLexic* (see www.anylexic.com) from the Ukrainian CAT developer AIT. While it supports neither TMX nor TBX imports (it supports only *Excel* and CSV imports), it comes in a standalone as well as in a server-based version allowing for multi-user access.



Figure 195: Search results in *AnyLexic*

A slightly different tool is *ApSIC Xbench* (see www.apsic.com/en/ products_xbench.html). Xbench is a free tool that performs quality assurance checks on a large number of TEnT formats (see page 272) but it also imports a huge number of bilingual file formats, including

- tab-delimited text files,

- TMX and TBX (!) files,

- *Trados* and *Wordfast* text translation memories,

- *Trados MultiTerm* XML files,

- *Trados* .ttx and bilingual *Word* files,

- *SDLX* .itd files,

- *IBM Translation Manager* files,

- *Star Transit* files,

- XLIFF files, and

- the Mac OS X glossaries (these are accessible at ftp.apple.com/developer/Tool_Chest/Localization_Tools/Mac_OS_X_Glossaries/).

It then indexes them and gives you near-instantaneous access to the content of these files. It's a very powerful tool that really stands alone in its class; the only drawback is that it requires a rather large amount of computer resources to run.



Figure 196: Search results in *ApSIC Xbench*

Lastly there is Terminotix's *LogiTerm* (see www.terminotix.com), a very powerful terminology reference tool. *LogiTerm* allows you to do a quick (and sometimes coarse) alignment of translated files (including *Word*, *WordPerfect*, *Excel*, *PowerPoint*, and HTML files) and store the aligned results as bilingual HTML files.



Figure 197: Bilingual HTML text created by *LogiTerm*

Once you have created your bilingual files, you can search any number of them with a search feature for any term (just enter the term under **Main Field**). *LogiTerm* will display all matches, including fuzzy matches (matches that are not 100% identical).



Figure 198: *LogiTerm's* bitext search feature

*Logiterm works in Western languages only.*

# Software Localization Tools

## Who Needs Software Localization Tools?

Depends on who's asking, I guess. Does the average translator who every once in a while translates software need a software localization tool? Not necessarily. Does the translation agency that offers "localization services" need one? Well, they should at least be aware of the tools out there and be ready to make the investment when the time comes. Does the end client who translates his software product into one or several languages need it? In my opinion, either he or his language provider absolutely does!

All this information doesn't help much if you don't know what software localization tools actually do. So, here goes:

Ten or fifteen years ago (and in many cases even today), software was translated by

- finding out which files contained translatables (in the case of most Win32 applications, the translatable strings were typically located in binary .exe or .dll files, i.e., files that cannot be opened with a text editor),

- breaking (=decompiling) these files apart with a development tool into numerous individual "flat" files (such as text-based .rc files),

- translating the .rc files in a text editor or in a translation environment tool,

- combining (=compiling) these files back into the original .exe or .dll files,

- testing these files extensively for cut-off text due to text expansion or any other errors that may have been introduced, and

- starting the process from scratch if any change in text occurred during the development cycle or any other editing had to be done.

```
/////////////////////////////////////////////////////////////////////////
//
// Dialog
//

IDD_WEBWIZ_SIGNONDLG DIALOGEX 0, 0, 174, 103
STYLE DS_MODALFRAME | DS_CENTER | WS_POPUP | WS_CAPTION | WS_SYSMENU
EXSTYLE WS_EX_CONTEXTHELP
CAPTION "WebWizard Admin Signon"
FONT 8, "MS Sans Serif", 0, 0, 0x1
BEGIN
    RTEXT           "Enter &Userid:",IDC_STATIC,15,14,54,8,0,WS_EX_RIGHT
    EDITTEXT        IDC_USERID_EDIT,77,13,63,14,ES_UPPERCASE |
                    ES_AUTOHSCROLL
    RTEXT           "&Password:",IDC_STATIC,15,34,54,8,0,WS_EX_RIGHT
    EDITTEXT        IDC_PASSWORD_EDIT,77,33,63,14,ES_UPPERCASE | ES_PASSWORD |
                    ES_AUTOHSCROLL
    RTEXT           "&New Password:",IDC_STATIC,15,54,50,8,0,WS_EX_RIGHT
    EDITTEXT        IDC_NEWPASSWORD_EDIT,77,52,63,14,ES_UPPERCASE |
                    ES_PASSWORD | ES_AUTOHSCROLL
    DEFPUSHBUTTON   "&Sign On",IDOK,29,78,50,14
    PUSHBUTTON      "&Cancel",IDCANCEL,92,78,50,14
END
```

Figure 199: View of an RC file in a text editor

Sound complicated and convoluted? Well, it was (and is).

Some of the more sophisticated companies such as Microsoft and Corel started to look for other solutions. They developed internal tools that performed the following functions:

• allowed developers to test the software before translation to determine whether a translation could indeed be performed into more complex languages such as Arabic/Hebrew or Chinese/Japanese/Korean

• allowed translators to go directly into the binary software files to translate those and build a glossary during the translation that functioned as a translation memory

• allowed translators to see everything in a WYSIWYG ("what you see is what you get") format and resize text fields that needed to accommodate text expansion or contraction

• allowed automatic testing for various errors

• eliminated the need for the various compilation procedures and at the same time streamlined updates to the software (like for a new release or bug fix), because the old glossaries could be applied and only new text needed to be translated

This, in a nutshell, is what software localization tools still do today, though they have included other development formats, most recently the new Microsoft development standard .NET or the translation exchange format XLIFF.



*According to the Sun website, "XLIFF is an XML-based format that enables translators to concentrate on the text to be translated. Likewise, since it's a standard, manipulating XLIFF files makes localization engineering easier: once you have converters written for your source file formats, you can simply write new tools to deal with XLIFF and not worry about the original file format. It also supports a full localization process by providing tags and attributes for review comments, the translation status of individual strings, and metrics such as word counts of the source sentences."*

*What that means is that instead of exchanging data after the translation (through TMX or the termbase standard TBX), XLIFF allows for the exchange of data during the translation process itself. Once my HTML, RTF, or FrameMaker file has been converted into the XLIFF format, I can use any tool that supports that format to translate it, without needing the original application. This applies to any computer-assisted translation tool as well as any original authoring tool (such as FrameMaker or others). What this in turn may mean is that eventually we all will work almost exclusively in XLIFF files and won't have to worry about ever-expanding tool sets.*

While Microsoft decided to keep its tool, *LocStudio*, internal, Corel decided to market its tool, *Catalyst*, to the rest of the translation and software development community. *Catalyst* is now the market leader in a field with numerous other players, many of which have remarkably similar feature sets.

Here is a list of the involved players:

- *Catalyst* (see www.alchemysoftware.ie): Supports Win32, .NET, Java, XML, HTML, XLIFF, HTML help files, and numerous database types and offers an interface to Trados applications.



Figure 200: *Catalyst's* translation environment

- *SDL Passolo* (see www.passolo.com): The second-largest player in the market and the most formidable contender to *Catalyst*. Originally developed for medical applications, *Passolo* today supports Win32, Delphi, Java, XML, XLIFF, HTML help files and offers an interface to *Trados* and *Star Transit* memories and terminology databases. It also supports all current versions of .NET, possibly in a more sophisticated manner than its competitors.



Figure 201: *Passolo's* translation environment

*In 2008, Passolo was purchased by SDL and Catalyst by translations.com, putting an effective end to the age of independent localization tools.*

*As an attempt to tie Passolo closer into the Trados product suite, Trados 2007 and Trados Studio 2009 now both include a limited version of Passolo.*

- *RC-WinTrans* (see www.schaudin.com): Originally a tool specialized in the translation of RC files (thus the name). Today it supports Win32, .NET, and Java Properties.

- *Multilizer* (see www.multilizer.com): Finnish tool that originally was designed for Delphi. Today it also supports Win32, .NET, Java, XML, and database contents.

 *Other smaller players include Language Studio (see ls.atia.com), Lingobit Localizer (see www.lingobit.com), Sisulizer (see www.sisulizer.com), and Visual Localize (see www.visloc.com).*

All of the tools come in several editions that have radically different price tags, and many of the above-listed abilities are sold as separate plug-ins. Typically there is a translator edition that excludes some of the more development-oriented functionality, and a developer or localizer edition that contains all the functionality. *Passolo* and *Catalyst* also come in editions that allow the developer to create files which can be worked on in a freely downloadable edition for the translator.

To come back to our original question, when these tools were first released, software developers across the board became nervous. They were afraid that a new development-oriented tool would likely cause problems—as most of us know, developers feel quite protective of their "baby," the software. At this point, however, it's clear that these fears are completely unwarranted. Unless software does not follow any of the supported development standards it's not only safe to use a software localization tool, it's silly not to—and a great waste of money, time, and energy to boot.

 *A number of "mainstream" translation environment tools, among them Transit NXT and Across, also support the direct editing of binary file formats.*

# Management Tools

There's a problem most translators face with project management tools: when business is finally good enough to justify implementing a management tool (both because of the purchase price and the volume of business that needs to be managed), their management workflows are so entrenched that it's hard to change. And old habits die hard. . . .

Some tools that have been mentioned in previous sections can take care of certain aspects of your translation work, including:

- *Outlook* (or any other applicable email and scheduling application) for managing schedules, due dates, and reminders (see page 163).

- Applications such as *PractiCount & Invoice* for writing invoices and/or reports (see page 137).

- Spreadsheet applications such as *Excel* to track hours and pricing or to use with a good template for invoices (see page 140).

Many of us use a combination of these and others to manage our businesses. Of course, there are a number of tools out there that very specifically help to manage workflow.

My favorite tool as a project manager (so many years ago. . .) was *Microsoft Project*. This program is impressive for its ability to track very effectively projects in an almost unlimited number of ways and save the results in a great variety of formats, including HTML, that can be shared with anyone. Awhile back I purchased a recent copy of *Project* for our small company and never really used it; for our small business it seemed like overkill to use such a "heavy" application to track projects.

I chalked that up to one of my few software investments that didn't pay off.

But there is another group of tools that have come of age: accounting and project management tools that are specifically created for the translation industry. The concept of these tools is to automate and organize repetitive tasks that are associated with your translation projects, including:

- generating quotes

- scheduling tasks

- managing all contacts with all relevant information

- managing your price lists broken down into certain tasks and clients

- tracking all activities with a certain client

- tracking the progress of projects and the completion of individual tasks

- writing customized invoices

- generating comparative overviews about the profitability or the importance of certain jobs and clients

Agencies or anyone who outsources some jobs have additional needs, including:

- vendor management

- a networked or web-based approach

You won't be surprised to hear that all this makes for a number of different categories of tools. The first category is the kind of tool that gears toward the management of jobs, invoicing, and vendors for agencies. These are the tools that I am aware of which do this:

- *LTC Worx* (see www.langtech.co.uk/content/view/19/23/)—*Worx* is the successor of *LTC Organiser*, the first commercial project management program for the language industry.

- *]project-open[* (see www.project-open.com)

- *Beetext Flow* (see www.beetext.com)

- *Projetex* (see www.projetex.com)

- *Plunet* (see www.plunet.de)

- *T.O.M.* (see www.jovo-soft.de)

- *XTRF* (see www.xtrf.eu)

If you look at the different websites of these vendors, you will quickly recognize different levels of professionalism, price, and approach. For instance, *]project-open[* is an open-source tool that allows a great deal of customizability; it also offers a number of additional paid modules that you can but don't have to integrate. *Worx* and *Plunet* are completely hosted online, and *T.O.M.* is really more geared toward smaller companies or freelancers.

Because of the set workflow paths that well-established language agencies already employ, the implementation of an outside system is sometimes associated with significant effort to adjust either the workflow or the system, with the latter often being the more difficult. For this reason, a number of language agencies have decided either to build a system for themselves (with a typical price tag of $100,000) that will integrate with an existing accounting system, or to use generic applications (project management apps like the *MS Project*, customer relationship management tools like *GoldMine* or *Salesforce.com*, or ERP products like *mySAP*) that they adjust to their specific needs.



*There are also tools that combine some workflow aspects (in particular vendor and file management) with translation memory technology, including SDL TeamWorks and SDL Synergy (see www.lspzone.com/en/ products), STAR James (see www.star-group.net/star-www/description/ star_james/star-group/eng/star.html) and XTRF-TM (see www.xml-intl.com/web/guest/products/xtrf-tm).*

And then, of course, there are also tools made for freelance translators. Some tools, such as *Customer Pro-File* (see www.linguabase.com/cpf.asp) or the above-mentioned *T.O.M.* and *PUDD*, are designed for smaller companies or freelancers. However, the tool that is probably the leading contender in this group is the little sister of *Projetex*, *Translation Office 3000* (see www.to3000.com), a no-nonsense database-based solution with little footprint that can significantly minimize your accounting time as a freelancer.



Figure 202: The **Invoice** window in *Translation Office 3000* with easy access to all other modules

I'm pleased to admit that I have finally given up my old entrenched ways and adopted *Translation Office 3000* for my management and accounting. I've thrown out my general accounting software (*Quicken*), adjusted the look of the customizable invoice templates in *Translation Office* to the look of my old ones, and figured out that, after a bit of setup, I'm much faster this way.

You'll have to try it for yourself and see whether the same is true for you.

# Translating Complex File Formats

Once upon a time the overwhelming majority of computerized translation was done in word processing or text formats. If there was any question about format it was maybe whether the source was in *Microsoft Word* or *WordPerfect* format (with the latter most often being the case!). Today, things clearly have changed. There are a host of formats out there, and fewer and fewer translators, especially those who work in more or less technical fields, get by with just using word processing application(s).

While I have dealt with office formats in earlier sections (see *Office Suites* on page 149), in this section I have attempted to categorize the most commonly required more advanced file formats. You will find descriptions of the programs for which these are written, how to distinguish between the translatable vs. untranslatable parts, and how these formats are supported by computer-assisted translation tools.

The categories of file formats are the following:

- Desktop publishing formats (formats for programs such as *QuarkXPress*, *PageMaker*, *InDesign*, *FrameMaker*, etc.)

- Graphic formats (pixel-based: .jpg, .gif, .bmp, .tiff, etc.; and vector-based: .eps, .ai, etc.)

- Tagged files (.html, .xml, .sgml, etc.)

- Software development formats (binary formats: .dll, .exe, .ocx, etc.; and text-based formats: .rc, .properties, .resx, etc.)

- Help systems (*WinHelp* and *HTMLHelp*)

- Database-based data

## Desktop Publishing Formats

The common denominators of desktop publishing programs seem to be these: they are expensive, they're not very intuitive for users, and they present considerable obstacles for translators.

This doesn't sound good, but here are the brighter aspects: Yes, they are expensive, but you may not even need to have them installed on your computer when you translate them. They are very difficult to learn on a real expert level—after all, graphic designers, desktop publishers, and prepress specialists are well-paid professionals—but as translators we only need to translate the files, not design them. And, yes, there are obstacles, but, fortunately, there are workarounds as well.

Generally, DTP programs can be categorized into two groups: those created for design-oriented publications and those intended for content-oriented publications.

In the first group are programs such as:

- *QuarkXPress* (see www.quark.com),

- *Adobe InDesign* (see www.adobe.com/products/indesign), and increasingly frequently

- *Adobe PageMaker* (see www.adobe.com/products/pagemaker).

The second group is home to applications such as:

- *Adobe FrameMaker* (see www.adobe.com/products/framemaker), and also increasingly frequently

- *Corel Ventura* (see www.corel.com/ventura).

The design-oriented programs provide superior graphic and font management and processing. Text is handled in individual and independent text boxes that can be placed anywhere in the application. They are typically closely integrated with graphic applications, and they offer advanced methods of prepress setup.

While the content-heavy applications also offer good graphics and prepress management (albeit not as advanced as the design-oriented programs), their main focus is on the processing of text, which shows in the advanced TOC (Table of Contents) and index generation, cross-references, page break management (widow and orphan rules), an independent character and paragraph setup, and the ability to output documents in a huge variety of formats. The latter is increasingly done through a tight integration into XML (see page 150).

Though working in these programs with translation environment tools is not as straightforward as it is with *Microsoft Office* formats, the benefit is that you do not have to have any of these programs installed to perform the actual translation as long as you work with a translation environment tool that supports these formats (most of the more advanced ones do).

## Content-Oriented DTP Programs

Any of the content-oriented formats—*FrameMaker* or *Ventura*— offers a fairly painless way of saving the original compiled format in an interchange format (MIF for *FrameMaker* and a tagged text format for *Ventura*) that can be easily processed. Because the emphasis for these files is on text and not on graphics, text is represented in one flow, and can be saved in a simple **Save as** process for each file (which is typically synonymous for one chapter). If you do not own these programs, it's no problem to ask the client to do this for you.

The very concept of these programs is that there will be as much automation in the layout as possible. This is achieved, for instance, through fairly sophisticated widow and orphan rules so that there will only be a small amount of additional pagination.

In general, these programs are very well suited for translation. There is no problem with non-Western languages even in Western versions of the system (provided that your operating system supports it), and the latest versions of *FrameMaker* now also fully support Unicode. The size of the files tends to be relatively small because graphics are usually linked and not inserted, and all of these programs are exceptional in the ways they publish and re-publish text in a great variety of formats, including HTML, XML, PDF, and RTF.

*Especially for FrameMaker, there are a great number of plug-ins that are very useful, in many instances even for translators.*

*The default track changes feature in FrameMaker (accessible under* **Format> Document> Change Bars***) is not nearly as advanced as you would hope for; it only marks entire paragraphs that have been changed at some place. Fortunately these have been replaced by a more comprehensive tracking feature in FrameMaker 8 (***Special> Track Text Edits***). If you still work in an earlier version of FrameMaker you might want to look at the TrackChanges plug-in (see www.intech.com/solutions/trackchanges) which allows you to track your changes in FrameMaker very much like it's done in MS Word.*

*The famously bad column and page breaks in indices in FrameMaker can be solved with the Index Tools (see www.siliconprairiesoftware.com), and the amazing Mif2Go (see www.omsys.com) does all kinds of magic when converting FrameMaker books to help files, MIF files, RTF files, etc.*

## Preparing FrameMaker Files

*FrameMaker* files typically come as part of a project organized within a .book file. The easiest and safest way to open them is by opening the .book file first and then opening the .fm files from within there.



*If the .fm files are displayed with an icon in the form of a question mark, you need to delete them from the book with the appropriate command from the menu and then re-add them from within the* **Add** *menu. Once the files are added, you can easily change the order of the files by simply dragging them within the .book interface.*

You will need to save the compiled .fm format within *FrameMaker* by selecting **File> Save as** and selecting the text-based .mif format. To avoid the individual opening and saving of each file, you can use the well-liked *MifSave* (see home.comcast.net/~bruce.foster/MifSave.htm) to do this as a batch process for a whole book. (By the way, it's totally okay to ask your client to do this for you if you do not have *FrameMaker* on your computer.)

Once all your files are preprocessed, they are supported in any of the larger translation environment tools (*Trados*, *SDLX*, *Déjà Vu*, *Transit*, *Across*, *memoQ*, *Wordfast Pro*, *Publisher*), most of whose representatives will tell you that their *FrameMaker* processing is one of their strongest features—which only goes to show that *FrameMaker* is a very translator-friendly format.

There are slight differences in the way that the different tools process the .mif files. While some tools (*Déjà Vu*, *Transit*, *Across*, *memoQ*, *Wordfast Pro*, *Trados Studio*) process the .mif files directly like any other file type, in *Trados 2007* and earlier you need to convert the MIF files into RTF files (so-called "STF" files) with a separate program that is part of the *Trados* suite of tools, the so-called *S-Tagger for FrameMaker* (usually located under **Start> Programs> Trados XX> Filters**), before you can translate them in either *Word* or *TagEditor*. The process of converting the files is slightly confusing if you do it for the first time, but the principle is this:

You will need to create two different directories, one of which will contain a set of files that will only serve as reference files so that the *FrameMaker* .mif files can be reassembled once they are translated in RTF format. The other will contain the set of files that are actually prepared for translation. Keep that in mind in both conversion processes (into RTF and back into MIF) so that you select the correct directories either way.



Figure 203: *Trados S-Tagger for FrameMaker* with tabs to convert files and verify tags

*Many translators do not like to work in TagEditor, but in many ways it is much more suited for the translation of FrameMaker files than Word. For one thing, TagEditor offers a much better tag protection than Word, and it also allows for an immediate code integrity check once the translation is done (you can select the code check under **Tools> Verify** in TagEditor). This makes the sluggish tag checking within the S-Tagger application obsolete.*

Both *Trados* (*2007* and before) and *SDLX* create additional files (the *Trados* version is called ancillary.rtf; the SDLX files are SDLX_ix.mif, SDLX_xr.mif, and SDLX_vr.mif). These contain background information such as index markers (*SDLX* only), cross-references, and information from the master page.



*Note that Trados Studio supports FrameMaker versions 8 and 9 only.*

Most other tools process the .mif files directly and translate all the background information individually for each file.

One exception to this is *Alchemy Publisher*, which allows the direct translation of *FrameMaker* .fm files. Clearly this is a tremendous time saver, but there is one striking disadvantage. Since *Publisher* uses *FrameMaker* in the background to process the files, you must have *FrameMaker* installed on the machine on which you translate .fm files. So, if you already have *FrameMaker*, *Publisher* might be a good option. If not, it's important to consider the additional cost.

### Preparing Ventura Files

*Trados* (2007 and before, but not the Studio edition) is the only TEnT that supports the *Ventura* format—but don't worry, there are very few translation projects in that format.

The process for translating *Ventura* within *Trados* files is very simple: You will need to export the content of the original .vp files to text files (**File> Export Text> ANSI text**), translate those in *TagEditor*, and reimport the translated text at the place where you want the text to be inserted (**File> Import Text**).

## Design-Oriented DTP Programs

With the other programs—at least as far as *QuarkXPress* and *PageMaker* are concerned—the process is very different.

In these formats, each text block, called a story, is saved in individual text boxes from which the text has to be manually exported into a tagged text format and re-imported if you want to process them in a translation environment program. While this is theoretically not an issue, it is very (!) time-consuming when you have to do this for tens or even hundreds of stories in one document.



Figure 204: Exporting a story from *PageMaker*

Fortunately, there are some applications available for these programs that allow for the batch export and import of these stories into one text file per original file (*CopyFlow* at www.napsys.com/cflow.html for *Quark* and *StoryCollector* as part of *Trados* [version 2007 and before] for earlier versions of *Quark*, *PageMaker*, and *InDesign*).



Figure 205: Importing a batch-exported story into *Quark* with *StoryCollector*

An issue with any of these programs is that there is often a fair amount of post-translation layout due to text expansion, etc. The text boxes in which the stories are located do not automatically expand, and often have to be manually resized once the translation is finished.

It's one thing to consider purchasing (and learning!) any or all of these tools, but a consideration that is just as important is the price you will have to ask for to translate a document in *InDesign*, *PageMaker*, or *Quark* in comparison to a document in *Word*. Are your clients able and willing to reimburse you for the larger amount of time that you are spending with these files?

In my experience it may not necessarily pay out to do projects in *PageMaker*, *Quark*, or *InDesign*, but it is one important way to distinguish yourself from your competition. (According to the 2003 American Translators Association's *Translation and Interpreting Compensation Survey*, only 18% of all U.S.

respondents use any kind of desktop publishing program!) Once you have mastered a job or two for a client in *InDesign*, or once you get your client's attention because you own these programs, it often paves the way for other lucrative non-desktop-publishing jobs.



*One way of being able to buy desktop programs relatively cheaply is through competitive upgrade programs. With a copy of Microsoft Publisher (that was part of an Office package), I was able to purchase a greatly reduced copy of PageMaker, which in turn allowed me to buy a competitive upgrade of both Quark and InDesign!*

This is not where the problems stop, though. Especially *PageMaker* (and *QuarkXPress* up to version 6.5) is still very "last century" when it comes to processing multilingual text. Though Unicode (see page 6) is a widely accepted standard that makes it easy to mix and match different writing systems on web pages and all kinds of other documents, some DTP programs are not up to par on this. Even though *Quark* does now support Unicode with its latest versions, *PageMaker* most likely will never do that because the folks at Adobe have a better choice when it comes to processing Unicode: *InDesign*.

## InDesign

After a fairly unsuccessful version 1, *InDesign* really gained traction beginning with version 2. Presently you will encounter *InDesign* files that are created in versions 2, and CS (3) through CS5 (7). To translate efficiently in *InDesign* you will need a program that exports all the stories (the above-mentioned text boxes) into one large file which can be processed in a translation environment tool. (Of course, it is possible to translate directly within *InDesign*, but the emphasis was on "efficient.")

## InDesign 2 and CS Support

*Trados* (versions *2007* and before; *InDesign 2* and *CS* are not supported by *Trados Studio*) offers little plug-ins as part of all its versions of the *Workbench* product that support *InDesign* versions *2* or *CS* (the plug-ins are stored under `C:\Program Files\SDL International\Txxxx_xx\FI\IND`—follow the instructions in the help file on how to install the plug-ins). Once you have installed the plug-in and opened the *InDesign* file, you will see a new *Trados* menu with all the necessary commands to export and re-import your file.

```
<STORY NAME="119" LOCATION="P16"><UNICODE-WIN>
<Version:2.000000><FeatureSet:InDesign-Roman><ColorTable:=<Black:COLOR:CMYK:Proc
ess:0.000000,0.000000,0.000000,1.000000>>
<DefineCharStyle:Bold Oblique=<Nextstyle:Bold
Oblique><cColor:Black><cTypeface:67 Bold Condensed
Oblique><cSize:12.000000><cHorizontalScale:1.000000><cLigatures:0><cTracking:0><
cBaselineShift:0.000000><cCase:Normal><cStrokeColor:None><cVerticalScale:1.00000
0><cNoBreak:0><cUnderline:0><cFont:Univers><cPosition:Normal><cStrikethru:0><cCo
lorTint:-1.000000>>
<DefineParaStyle:Normal=<Nextstyle:Normal><cTypeface:Medium><cLigatures:0><cBase
lineShift:0.000000><pHyphenationLadderLimit:0><pAutoLeadPercent:1.204987><pMinCh
arBeforeHyphen:3><pHyphenateCapitals:0><pShortestWordHyphenated:6><pHyphenationZ
one:0.000000><cFont:Helvetica><pDesiredWordSpace:1.100006><pMaxWordSpace:2.50000
0><pMinWordSpace:0.850006><pMaxLetterspace:0.039993><cHang:Baseline><pRuleAboveC
olor:Black><pRuleAboveTint:100.000000><pRuleBelowColor:Black><pRuleBelowTint:100
.000000>>
<ParaStyle:Normal><pBodyAlignment:Justify><CharStyle:Bold Oblique>
Note•<CharStyle:><cTypeface:47 Light Condensed
Oblique><cFont:Univers>:Throughout your workout you will hear a series of
"beeps" from your console. This is perfectly normal. These beeps signify various
changes taking place in the unit's programming. For instance, you will hear one
beep every time you press a key on the keypad, and one beep to engage higher
function keys such as X-Mode, and two beeps when they have selected a program
for you. If you've<cTypeface:><cFont:><cTypeface:47 Light
Condensed><cFont:Univers> <cTypeface:><cFont:><cTypeface:47 Light Condensed
Oblique><cFont:Univers>decided to end your workout and hit the Cool Down button
you will hear three. Stepping away or stopping completely will engage the Pause
mode which will signal with three beeps that you've stopped your workout before
it was finished and will begin a ten minute countdown before exiting the program
you were using. You will also hear beeps as you move from the warm-up into each
program.<cTypeface:><cFont:><cTypeface:47 Light Condensed><cFont:Univers>
<cTypeface:><cFont:><pBodyAlignment:></STORY>
```

Figure 206: View of Trados-exported InDesign text file in a text editor

As you can see in the above illustration, the text file is not just a "normal" text file; instead, it is a "tagged" text file where only the smallest part is actually translatable (essentially everything that in not enclosed by `<tag markers>`) and all the other data stores information about details such as formatting, placement, etc. While it theoretically would be possible to translate this within *Microsoft Word* or a text editor, it would be foolish to even try—chances are that you would break the code or overlook text.

*Trados* and *Déjà Vu* recognize these files as *InDesign* files, protect all coding information, and display only translatable text.



Figure 207: View of *Trados*-exported *InDesign 2* text file in *Trados TagEditor*



*Transit XV and NXT also support InDesign CS files through a specially purchased add-in.*

*Additionally, Transit NXT supports InDesign CS2-CS5 with either Windows or Mac plugins.*

## Support of InDesign CS2 and above

With the release of *InDesign CS2*, the accessibility of *InDesign* files became much more feasible for translation environment tools because it was now possible to save files into the XML-based .inx format. This format is supported by the vast majority of TEnTs. Note that you will have to have a copy of *InDesign* on your computer to save the file as an .inx file (or you can ask the client to do it for you).

It is also advisable to check what version of *CS2* through *CS5* your tool officially supports since there are fairly major differences between the XML structure of the different versions of *InDesign*. Since *InDesign* has become the quasi-standard desktop-publishing format, you should be able to expect your TEnT vendor to update quickly to the latest format of *InDesign*.

## Third-Party Support of InDesign

There are a number of third-party tools that support *InDesign* in its various versions.

ECM-Engineering (see www.ecm-engineering.de) offers an independent filter for all current versions of *InDesign*. This application allows you to export into an RTF format (or, optionally, into an XML format) that is supported by most TEnTs. It gives you the choice to do this in a tagged format or in a WYSIWYG (what-you-see-is-what-you-get) format with intact formatting.

```
<!-- Sysfilter für Indesign 1.0 Build: 13 -->
<!-- Indesign-Dateiname: -->0558005270
<!-- Indesign RTFs.: -->205
<layer>xxxxx</layer>

<!-- Section no: 1-->



<!-- Section end: 1-->

<!-- DateiID: 000000000001_0558005270-4977.rtf-->
```

**This manual provides installation and operation instructions for the following PT-27 torches:**

**P/N 21620 - 25 ft. (7.6 m), PT-27, w/PLUG Pilot Arc Connection**
**P/N 21661 - 25 ft. (7.6 m), PT-27, w/RING Pilot Arc Connection**

```
<!-- DateiID: 000000000001_0558005270-507.rtf-->
```

## *PT-27*

```
<!-- DateiID: 000000000001_0558005270-532.rtf-->
```

## Plasma Arc Cutting Torch

Figure 208: View of ECM-exported *InDesign* file in WYSIWYG format

*The makers of CopyFlow also offer a comparable plug-in for InDesign (see page 301).*

### Alternative Ways of Working with *InDesign CS4* and Later Files

With version *CS4*, *InDesign* introduced the ability to export *InDesign Markup Language* (IDML) files. These are a zip-compressed set of XML files where each XML file represents a "story" (text box). While it's possible to translate these files without any specialized filter (you can export the IDML file out of the original *InDesign* .indd file with **File> Export**, rename the .idml extension

to .zip, unzip the file, locate the XML files that contain the story content—the translatable text—and import or open them with your TEnT), the latest version of most translation environment tools now supports the .idml format directly, and *Trados Studio* in fact supports only the .idml format for *InDesign CS5.*

Yet another approach to working in *InDesign* files is the increasingly popular web-to-print solutions that are made possible through the high-powered *InDesign Server* product (see www.adobe.com/products/indesignserver). The advantage is that the *InDesign* files never have to be exported or converted out of *InDesign* and translators work in a complete WYSIWYG environment— without having to have *InDesign* themselves. .

*You can find a listing of companies with that service under www.adobe.com/products/indesignserver/findsolution. one2edit (see www.1io.com) is one of the solutions that strongy emphasizes the translatability of its InDesign content with both SDL and Across offering plug-ins for that solution.*

## PageMaker

To translate *PageMaker* files (an increasingly rare occurrence because Adobe is trying to push *InDesign* over *PageMaker*) with a computer-assisted translation tool, you can either use *Star Transit* with a separate plug-in that supports *PageMaker* 6-7, or you can use a plug-in that comes with the *Trados* product (only version *2007* and below; *Trados Studio* is not supporting *PageMaker* anymore) called *Story Collector for PageMaker,* which supports *PageMaker* versions 6.5 and 7.

To install the *Trados* plug-in, open the help file under `C:\Program Files\SDL International\Txxxx_xx\FI\PM` for further instruction. Once the plug-in is installed, open the *PageMaker* file in *PageMaker* and you'll find the command *Trados Story Collector* under **Utilities> Plug-ins**.



Figure 209: *Trados Story Collector* for *PageMaker*

Export all the stories into one large *PageMaker*-specific text file, save the original *PageMaker* file (important!), and translate the exported text file with *TagEditor* or any other application that supports the *PageMaker* format. The import process is virtually the same as the export and should go seamlessly.

*For specific illustrations of the differences in exported text files in a text editor vs. TagEditor, see the InDesign-specific illustration on page 296— PageMaker (and QuarkXPress) have a very similar appearance.*

All of the above is true for Western languages and to some degree for Eastern European languages. Any of the more complex languages, however, including the bi-directional languages (Hebrew and Arabic) or the Asian double-byte languages, are flat-out not supported in the Western versions of *PageMaker*.

Though you can purchase language-specific versions for these languages, it would make a LOT more *sense* to convert to *InDesign* and take it from there. Because *InDesign* and *PageMaker* are both Adobe products, the upgrade path is relatively easy (both in terms of purchasing a less expensive competitive upgrade version of *InDesign* when you already own *PageMaker* and in terms of converting the files).

## QuarkXPress

Despite the fact that *Quark* has never been very popular in the translation community (because of a lack of Unicode support until recently and different and more expensive versions for different languages, etc.), it used to be the dominant player in the desktop publishing market, so it is not too surprising that there is decent support for earlier versions of *Quark* among the translation environment tools.

- *Star Transit* offers a separate plug-in that supports the batch processing of the English (and *Passport*) versions 3-8.1 for both the *Windows* and *Mac* platforms.

- *Trados* (version *2007* and below) offers plug-ins for versions 4.1-6 for English (and *Passport*) and version 4.1 for Japanese.

- *SDLX* offers a plug-in for the English (and *Passport*) versions 4-6 for the *Mac*.

All of these plug-ins were preceded by a program called *CopyFlow* (see www.napsys.com) which, just like these programs, allows for the batch export and import of text from *Quark* files.

If you need to translate *Quark* 7 and above files, *Star Tansit* (see above) and *CopyFlow* are  presently the only tools that allow for an export into a TEnT-processable format (both on *Mac* and *Windows*). In fact, many of the TEnT vendors (*SDLX, Trados, Across, Déjà Vu*) now directly recommend the use of *CopyFlow*.

Because of Unicode support in *Quark*'s version 7, it is now possible to write in all languages that are covered by Unicode within *Quark*. However, if you need language-specific capabilities (spell-checking, hyphenation, etc.), you may have to invest in one of the many language-specific versions.

The European language *Passport* edition of *Quark*, which has additional spell-checking and hyphenation capabilities for Western and European languages, is supported by the above-mentioned tools. If you have only the (cheaper) English version, you need to make sure to ask your client to save the file as a "Single Language" file. Otherwise, if the *Passport* edition was used you will not be able to open the file.

*QuarkXPress*'s last Arabic edition was for version 4.1. Fortunately, however, there are *XTensions—QuarkXPress*-specific plug-ins—for the English version of *Quark* that extend its ability to write in Hebrew, Arabic, Farsi, and Jawi. *ArabicXT*, *HebrewXT*, *FarsiXT*, and *JawiXT* are all available at www.arabicsoftware.net for versions 5 and 6 of Quark.



*Unfortunately, only ArabicXT is available for both Windows and Mac platforms, while the others are available only for Mac.*

*ArabicXT has also been released for Quark 7 for Mac.*

It becomes much more hairy with the Asian double-byte languages. While the Japanese version 4.1 is supported by the *Trados* plug-in and several others by *CopyFlow*, it at least means that you have to have several versions of *Quark* for different languages, plug-ins, and platforms.

For a quick overview, here is a chart of which TEnT supports which DTP format:

| | Across | Déjà Vu | Heartsome | MemoQ | MultiTrans | Publisher | SDLX | Swordfish | Trados 2007 | Trados Studio | Star Transit | Wordfast Pro |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **InDesign** | | | | | | | | | | | | |
| v2, CS | | x* | | | | | | | x | x*** | x | |
| CS2-CS5 (.inx) | x | x | x | x | x | x | x | x | x | x**** | x | x |
| CS4-CS5 (.idml) | x | x | x | x | x | x | x | x | x | x | x | x |
| *PageMaker* | | x* | | | | | | | x | x*** | x | |
| *Quark* | | | | | | | | | | | | |
| v4-v6 | | x* | | | | | x | | x | x* | x | |
| v7- | | x* | | | | | x* | | x* | x* | x | |
| *FrameMaker* | x | x | x | x | x | x ** | x | x | -v8 | v8- | x | x |
| *Ventura* | | | | | | | | | x | x*** | | |

\* Through third-party tools
\*\* Without conversion to .mif
\*\*\* Through prior processing with Trados 2007
\*\*\*\* Not CS5

## Mac vs. Windows for DTP Programs

Many files from DTP applications, especially from the graphic-oriented line, originate from a *Macintosh* environment. While generally it's not a problem to open a *Mac* version of *PageMaker*, *InDesign*, or *QuarkXPress* on a *Windows* edition of the program, there are a couple of things that have to be considered.

The most common error is that of missing fonts, which could be either fonts that are truly missing or, just as likely, fonts that have a slightly different naming conversion on a *Macintosh* system than on a *Windows* platform or vice versa. You can choose to remap the fonts on a permanent basis (not a good idea if your client wants to open this on a *Macintosh* again) or on a temporary basis.



Figure 210: Error message indicating missing fonts in *PageMaker*

The other consideration is the differing character set between *Windows* and *Macintosh*, which, if not converted properly, will result in a corruption of special characters. Assuming that you have performed your translation in a text-based format on a *Windows* computer, you have several options to change the character set.

- You can do this in a *Windows* version *Word 2000* or higher (see page 156).

- You can open and save your text file in a *Macintosh* version of *Word 98* or higher, which will automatically convert the *Windows* character set to a *Mac* character set.

# Graphic Formats

For graphic applications, the same common threads seem to apply as for desktop publishing programs: they are expensive, they're not very intuitive to learn, and they present considerable obstacles during translation.

Unfortunately, unlike desktop publishing programs, there is no positive flipside from my perspective as a translator.

Though I know that some of my colleagues have decided differently, I have left it with this and have decided not to invest time and money into graphic programs. I use an old version of *Corel (Jasc) Paint Shop Pro* (see www.corel.com/paintshop), a low-priced power tool that I mostly use to make minor modifications to pixel-based graphics (see page 304). Since Corel bought *Paint Shop*, the tool has morphed into more of a photo-editing tool, so the best current alternative may be *GIMP* (see www.gimp.org), a powerful open-source image editor that may not be particularly user-friendly but gives you everything you would ever desire from a graphics application.



*Another graphic application that has been helpful is the low-cost version of Adobe Photoshop—Adobe Photoshop Elements (see www.adobe.com/products/photoshopelel).*

I have not yet encountered a client who has complained about my lack of a full-featured, high-priced graphics program; in fact, they are usually very glad to supply me with *Excel* spreadsheets, in which I can translate the text of the graphics that can be pasted into the graphics by desktop publishers (probably faster, better, and cheaper than I could do it, anyway).



*If you work in double-byte languages (Chinese, Japanese, or Korean) or in right-to-left languages (Arabic and Hebrew), you will have to make sure that your graphic application supports your language (if you have to enter text into the graphic). GIMP and Paint Shop Pro support these languages on a native operating system or under Windows 2000 and above with the Regional Setting set to the appropriate locale as the default.*

Still, there are some ways to work directly in graphics files for translators:

## Pixel-Based Formats

Most graphic formats (including .jpg, .gif, .bmp, .tiff, and various others) don't contain text. This is true even if it appears to be readable text because the text is nothing more than pixels (little colored dots) on a virtual canvas. While they may form shapes that represent letters, these have nothing to do with the editable letters or words you will deal with in a text editor.

Short of recreating these kinds of graphics from scratch, you will need to get your hands on the "source files." (Yes, I know that clients hate to be asked for that, but typically it helps to mention that otherwise they will have to pay ten times as much.)

Most .jpg-, .gif-, .bmp-, or .tiff-like files were created in a layered file that includes one (or several) layers with real, editable text. Since they were most likely created in *Adobe Photoshop*, they will have a .psd extension and can be opened in, well, *Adobe Photoshop*.



Figure 211: Image file opened in *Photoshop* with active text layer

The nice thing is that Adobe offers a low-priced version of its program (see www.adobe.com/products/photoshopelel) that is more than adequate for translating the text layers that need to be translated. Or you can also use *GIMP* (see www.gimp.org), a powerful open-source image editor that allows you to work with .psd files, though it may not be particularly user-friendly (and it might also mess up some of the text layers—but at least you can access the different layers, delete the text layer, and recreate a new one).

This all may not be good enough, though. Especially if you have a large number of graphics and/or a translation memory database that contains much of the translation embedded within the graphics, you will not want to perform the translation "manually." In this case you can use a number of tools, including those from ECM Engineering (see www.ecm-engineering.de) or

Transmissions (see www.transmissionsllc.com/solutions/photoshop), that allow for the extraction of text from .psd files into RTF or XML formats. These formats can be processed in TEnTs (translation environment tools, such as *Trados*, *Déjà Vu*, *Across*, and *memoQ*) and afterward re-inserted.



Figure 212: Processing *Photoshop* files in *Sysfilter*

More often than not, we don't have access to the source files and have to go through frustrating re-creation or manipulation processes with the graphics. To ease at least some of that (translator) pain, there are a number of management applications for images, such as the open-source tool *Image Localization Manager* (see sourceforge.net/projects/ilmanager/). It's a nifty little tool that allows you to quickly browse folder and subfolders for graphics that are displayed nicely within the tool, determine which contain localizable content, create a list of those image files, and (manually) enter the source text. This can then be exported into a file that you can translate in something

like *Excel* or, if you want to view the graphic as you translate, in *Image Localization Manager*. Of course, when that's done, the translated text and the graphics have to be passed on to the desktop publisher—but as translators, that's not our worry!



Figure 213: Assessing which graphics need translation in *ILM*

## Vector-Based Formats

The above graphic types are pixel-based graphics. Another kind of graphic that is often used, especially in manuals, is vector-based graphics. You can recognize them by their typical extensions, .eps or .ai. They are very different from pixel-based graphics because they are formed by mathematical formulas rather than by simple dots. So, rather than displaying a wheel by arranging a lot of pixels in a circle, a vector-based graphic would calculate it with some kind of pi-based formula.

In contrast to pixel-based graphics, it is possible to translate most vector-based files directly in applications like *Adobe Illustrator* (see www.adobe.com/products/illustrator) or *Corel Draw* (see www.corel.com/coreldraw).



Figure 214: View of an .eps graphic file in *Adobe Illustrator*

If you would like either to batch process the files or to use your translation memory, there are two different options.

The above-mentioned ECM Engineering and Transmissions both offer products for *Illustrator* and *Draw* (only ECM Engineering) files that pre-process the files for use in TEnTs similarly to the way they process .psd files.

The second option is to save the vector-based files into the XML-based SVG format, which is directly supported by *Heartsome*, *Swordfish*, *memoQ*, *Trados*, and some versions of *Star Transit*.

## Taking Screenshots

This does not directly relate to translating graphics, but taking screenshots (pictures of the computer screen or dialog boxes) is often part of our job description as translators—for instance we might have to replace the graphics in the source language in a software manual with those in the target language (provided that the respective software is already translated and functional).

When taking screenshots, I have usually found it sufficient to take them the "traditional way" (ALT+PRINTSCREEN for the active dialog or **PrintScreen** for the complete screen) and then paste that into a regular graphic application.



*One thing to note is the checkered look of the light grey scroll bars in the "classic" Windows look (left image). To fix that, it is usually a good idea to paste some of the darker grey into the affected areas (right image).*

However, there are also programs that specialize in taking screenshots, and while they don't fix everything (such as the "checkered scrollbar issue"), they are good at solving some other pet peeves that have always bugged me about screenshots.

When you help someone to solve a computer problem, the easiest way to describe the problem is often to take a screenshot of the error message or whatever dialog you have problems with, paste it into an email, and send it off. But what most people don't realize is that the images that are pasted into the email are gigantic files in bitmap format. While transmission of large files may no longer be a problem for today's high-speed world, storage of these large files in email is (especially when the email is sent back and forth several times with the image still there, slowly increasing the size of your inbox to obscene proportions). One way around this is to paste the screenshot into a graphic application you use, save it as a .gif or .jpg file, and send that. But that requires a lot of additional steps. . . .

With a screenshot application like *SnagIt* (see www.snagit.com) you can automatically save the image in a file format of your choice (under **Output> Properties> Image File**) and paste it into your email or elsewhere in that format. *SnagIt* also offers the capability to record a video of whatever process isn't going right on your computer. You can then send this off in a commonly used video format if a mere explanation with words and screenshots doesn't help.

There are numerous other possibilities for screenshots in *SnagIt*—so many, in fact, that in my opinion it's a little bit of an overkill in some areas. Because of this, it doesn't hurt to run through the quick tutorial that's offered when you first start the program.



Figure 215: *SnagIt* options

*Another free screenshot tool with not quite as many options is Greenshot (see greenshot.sourceforge.net).*

*If you just need a screenshot tool for simple screenshots, you might also want to have a look at Windows 7's (and Vista's) integrated Snipping Tool:*

# Flash Files

Flash files used to be very similar in their translatability to graphics files: non-editable SWF or FLV files are the compiled versions of an editable FLA file (similar to how JPG/GIF/… files are the non-editable versions of the editable PhotoShop PSD files.) And, also like with graphics files, even the source files were not easily editable in any application, instead you had to have some version of Adobe (earlier: Macromedia) Flash to do this.

*Note that there are a number of tools that "decompile" .swf files into graphical and textual elements. This process is highly unreliable and therefore not recommended.*

*If you google for "Flash localization" you will stumble on Tramigo, a tool that was specifically built for the extraction of text strings for translation purposes. I can't recommend this tool—I have yet to be successful to translate a Flash file with it.*

If there was translatable text within Flash files it was stuck in a layer that was merged with other layers and was therefore not translatable. I'm saying was because nowadays things are (or should be) a lot easier.

Back in 2005, Adobe (back then it was still Macromedia) Flash 8 first introduced the concept that text strings could be stored separately and externally in an XML file. This file can be translated with a normal text editor or (preferrably) a translation environment tool and stored with language-specific extensions so that the correct language version would be retrieved according to the language preference of the user.

Unfortunately this does not mean that everyone is creating Flash files this way and it might be helpful to point your client to this reference (for Adobe Flash Professional CS5) on how to create multilingual text properly: http://tinyurl.com/26by322.

# Tagged Formats

Tagged files are files that are text-based and that typically contain a mixture of "normal" translatable text and "tags," elements that allow for the structuring of the content, page layout, text formatting, insertion of images, etc. Examples of tagged files are the exported text-based formats for the translation of content in some desktop publishing programs (see the example on page 296), but more typically tagged formats include HTML, XML, or SGML files (see the definition on page 150).

*Tags are typically enclosed in <angled brackets>. Internal tags, such as the <b>bold<\b> tag, are embedded in segments, whereas external tags, such as the <p>paragraph<\p> tag, are located outside sentences.*

*It is, however, also possible that tags themselves can contain translatable text. One well-known example is the alt tag for image files in HTML:*

*<img alt="translatable text" src="image.jpg">*

Because tagged text files are "just" text files, they can be translated with a text editor. However, this is typically not a good idea because

- the tags are quite sensitive to corruption, i.e., just deleting or adding a part of a tag may utterly corrupt a file;

- it takes a fairly experienced user to distinguish all instances of translatable text from non-translatable text; and

- though it would be possible to process tagged text files as plain text files in translation environment tools (TEnTs), it would mess up your translation memories with a lot of unwanted coding information; at the same time, you will not really benefit from your translation memory content because there will be very few matches for heavily coded sentences.

Instead, you should be using TEnTs that support tagged text formats, and most of them do. The concept of supporting these formats is to hide and/or protect any untranslatable information and only to display translatables.

This is relatively easy to do with HTML because it is a defined format that does not allow any deviation, but it is more difficult with XML and SGML files. These files are by definition user-definable and require you to "teach" the program how to interpret any give file. Any of these file types refers to a "Document Type Definition" or stylesheet that determines how each element of the file should be treated.

While the .dtd file for HTML is a global declaration that any of the supporting tools refer to, XML gives a somewhat universal access through a supporting technology that describes how to format or transform the data in an XML document, the so-called Extensible Stylesheet Language (XSL). Many translation environment tools offer a predefined XML filter based on a common set of XSL variables that is often sufficient to process XML files.

As SGML files have no such common denominator, you will need to create a specific "filter" or "settings" to process these files.

In *Trados 2007 and earlier*, tagged files are processed in the *TagEditor* (thus the name). Upon opening any of these file types in *TagEditor*, the following dialog may appear (if it is not displayed automatically, you can open it through **Tools> Tag Settings**):



Figure 216: *Tag Settings Manager* in *Trados TagEditor*

You can see the predefined HTML and XSL (for XML) settings.

You can change the properties of any of these settings files by selecting **Edit** (only do this if you are unhappy with the results of the existing settings files) or create a new settings file by selecting **Add**.

*Other predefined settings in TagEditor include those for XLIFF, DITA, and .resx files. One of the great advantages of XML is that it is ideally suited as an exchange format—among many others, it is the source format for the translation exchange formats TMX, TBX, and XLIFF (see page 278).*

*DITA (Darwin Information Typing Architecture) is a new XML-based standard for authoring, producing, and delivering technical information, and .resx files are XML-based resource files for .NET applications.*

The creation of a new settings file is largely wizard-based so it is less complicated than it seems. You do, however, need the .dtd file to create the resulting settings .ini file, which in turn allows you to import your tagged text.



Figure 217: Imported XML file in *Trados TagEditor*

As in the previous version, *Trados Studio 2009* contains a number of preconfigured XML filters as well as a generic XML filter.



Figure 218: Preconfigured XML options in *Trados Studio*

If the prepared options are not sufficient for your XML file(s), you will have to create a new filter type based on an XML sample file by selecting **New** under **File Types**.



Figure 219: Creating a new XML file type in *Trados Studio*

As in the previous versions of *Trados*, a wizard will guide you through the different steps of creating the file type.



Figure 220: Imported standard XML file in *Trados Studio*

*Déjà Vu* also contains predefined XML "filters" (available under `C:\Program Files\ATRIL\Deja Vu X\Templates\`), but just as in *Trados* it allows you to either edit that existing filter or create filters for other SGML files. You can access this feature by selecting **File> New> SGML/XML Filter**, and the wizard will lead you through the creation of a very customizable filter file. Unlike in *Trados*, it is possible to forego the import of a DTD file and you can choose to import an SGML or XML file directly to create a filter.



Figure 221: View of an XML filter in *Déjà Vu X*

As you import the XML or SGML file into *Déjà Vu*, you will need to make sure to select the appropriate SGML/XML filter file during the import process under **Properties**.



Figure 222: Imported XML file in *Déjà Vu X*

Most tools, including both *Déjà Vu* and *Trados*, allow the fine-tuning of the filters so that you can exactly determine which parts inside or outside a tag are translatable or to be protected. Typically, it is enough to go through the process of creating a filter or settings file for an XML/SGML project only once because usually all files will adhere to one standard.

## XML Files with Embedded HTML

While most XML files are relatively easy to process, some XML files can present a real headache: those with embedded HTML. In a way they are comparable to, let's say, a *Word* file with embedded *Excel* or *PowerPoint* files. In most of these cases, you will have to find a workaround by saving the files out of the originating file, translating them separately, and later embedding them again. With XML files you may also have to find a workaround, even though the actual process is different.

Consider the following snippet out of an XML file:

```
<Answer ObjectId="(0:0-218428250#216847832, 165)"
Precode="14"><Text LanguageId="7">A &lt;u&gt; new concept&lt;/
u&gt; that makes all the difference
to&lt;br&gt;&amp;nbsp;&amp;nbsp;&amp;nbsp;&amp;nbsp;&amp;nbsp;&am
p;nbsp;&amp;nbsp;&amp;nbsp;&amp;nbsp;&amp;nbsp;&amp;nbs
p;&amp;nbsp;&amp;nbsp;&amp;nbsp;&amp;nbsp;&amp;nbsp;&am
p;nbsp;&amp;nbsp;everyone who is interested in better
performance.</Text></Answer>
```

You can see that the XML tags are enclosed with the typical <less than and greater than> tag markers and they will be easily recognized by your TEnT. The actual translatable text

```
A new concept that makes all the difference
to everyone who is interested in better performance
```

is in the midst of lots and lots of HTML code, for which the less than and greater than tag markers are encoded (&lt; and &gt;) as well as the ampersand sign in the non-breaking spaces (&amp; inside of  ). Importing a file with this segment into most XML-enabled TEnTs results in this:

```
A <u> new concept</u> that makes all
the difference to<br>   
      
      
    
 everyone who is interested in better
performance.
```

Figure 223: XML file with embedded HTML in *memoQ*

The XML codes are protected (in this case hidden), but the encoded HTML codes have been turned into proper HTML codes that are not protected and can thus be easily corrupted. Aside from the danger of corruption these are an incredible nuisance because a) you will have to understand them, b) you will have to translate around them, c) they will make spell-checking a nightmare, and d) they will pollute your translation memory to no end.

If this is what your tool presents you with, there is a (relatively involved) workaround: Open *Word* and select **Tools> Options> General> Confirm conversion at Open** (*Word 2003* and earlier) or **Office button> Word Options> Advanced> General> Confirm file format conversion on open** (*Word 2007 and above*). Open the XML file in *Word* and, when prompted, open as an encoded text file and make sure that you select UTF-8 or whatever the original code page of the file was as the code page. The file will now be displayed as a plain text file. Now, according to the tool that you are using, you can change the style or the formatting of both the XML and HTML tags so that they will be treated as external or internal tags once the file is saved to RTF.

For instance, for *Transit* or *Déjà Vu* you can open the **Find and Replace** dialog in Word (CTRL+H), select **More**, check **Use wildcards** and search for \<*\> and &lt;*&gt; and replace it with hidden text (select **Format> Font> Hidden** for that). After that, uncheck **Use wildcards** and replace &amp;nbsp; with hidden text.

For *Wordfast*, *memoQ*, or *SDLX*, you can do the same, only in this case you can replace the same items with the *Trados*-style tw4winExternal and/or tw4winInternal. (If it is not listed in the list of available styles, copy and paste a couple of characters from another document with that style into your document—it will then be listed—and make sure to delete the characters once you have used the style.)

Once one of the above procedures is done (and I'm pretty sure you can find a style or format that your tool of choice recognizes as non-translatable text), save the file as an RTF file, make sure that you set your TEnT to ignore the text or style of your choice, import the file into or open it with your TEnT,

translate it, export or clean it, open it in Word, and save it as a text file. Make sure that you set the code page correctly and that the extension will be .xml rather than .txt (you can achieve that by placing quotation marks around the complete file name in the **Save as** dialog).



*For those who are on the more technical side, there are also two tools that can you help pre-process these files. One is a standalone tool called PrepTags (see www.your-translations.com/preptags.php) and the other is a free little Word macro called Tortoise Tagger (www.accurussian.net/ tagger.htm). Both of these tools essentially use the above-described ways of tagging the data, but they do it in a way that seems rather overwhelming to the novice who may or, more likely, may not know much about regular expressions. So, be forewarned.*

# Software Development Formats

There is a large variety of software development formats. The most logical distinction between the different categories is the following:

- binary files, or files that cannot be opened and edited with a text editor, and

- flat files, or text-based files that can be opened and edited with a text editor.

The binary files traditionally include formats such as .exe, .dll, or .ocx files. To translate these files, you will either

- need a specific software localization tool that allows the direct translation and necessary strings as well as further language-specific development work and testing (for further information on processing binary file formats, see *Software Localization Tools* on page 276) or

- need to "decompile" these files with a development tool into numerous individual flat files (and later "recompile" the binary file again).

While many software developers view the decompile/recompile process as outdated, others don't, and chances are that you will at some point receive decompiled RC files to translate.

```
///////////////////////////////////////////////////////////////////////
//
// Dialog
//

IDD_WEBWIZ_SIGNONDLG DIALOGEX 0, 0, 174, 103
STYLE DS_MODALFRAME | DS_CENTER | WS_POPUP | WS_CAPTION | WS_SYSMENU
EXSTYLE WS_EX_CONTEXTHELP
CAPTION "WebWizard Admin Signon"
FONT 8, "MS Sans Serif", 0, 0, 0x1
BEGIN
    RTEXT           "Enter &Userid:",IDC_STATIC,15,14,54,8,0,WS_EX_RIGHT
    EDITTEXT        IDC_USERID_EDIT,77,13,63,14,ES_UPPERCASE |
                    ES_AUTOHSCROLL
    RTEXT           "&Password:",IDC_STATIC,15,34,54,8,0,WS_EX_RIGHT
    EDITTEXT        IDC_PASSWORD_EDIT,77,33,63,14,ES_UPPERCASE | ES_PASSWORD |
                    ES_AUTOHSCROLL
    RTEXT           "&New Password:",IDC_STATIC,15,54,50,8,0,WS_EX_RIGHT
    EDITTEXT        IDC_NEWPASSWORD_EDIT,77,52,63,14,ES_UPPERCASE |
                    ES_PASSWORD | ES_AUTOHSCROLL
    DEFPUSHBUTTON   "&Sign On",IDOK,29,78,50,14
    PUSHBUTTON      "&Cancel",IDCANCEL,92,78,50,14
END
```

Figure 224: View of a decompiled RC resource file in a simple text editor

In much the same way that tagged files work, it would be possible to translate RC files in a text editor, but it is not advisable to do that because a) you will most likely overlook text that needs to be translated, b) you may overwrite code where that should not happen, and c) there is just no reason not to use your translation memory for this. In fact, software files are rarely translated on their own. Typically they are translated as a precursor to accompanying documentation—documentation that will be using references to the translated software over and over again—an ideal scenario for the use of translation environment technology!

Many TEnTs support the translation of RC files, including *Déjà Vu*, *SDLX*, *Star Transit*, *Across*, and *Trados*.

Many newer programming languages do not use a compiled format for their resource files. Often this takes the form of XML-based formats such as the .NET .resx format. Many TEnTs (and of course localization tools) support this format, but among TEnTs it's *memoQ* that offers the most advanced support with a feature that not only allows for the translation but also the resizing of dialog boxes.



Figure 225: View of a .resx file in *memoQ*

Java applications typically use the so-called Java Properties files (.properties).

```
# English-language strings for the Content Installer screens

contentTitle = Content Installer

reset = Reset

hiRes = <B>High Resolution Video and Illustrations</B> - \
Provides the highest quality images and full-motion video \
for the selected Titles, but requires the most disk space \
and the most time to download when used by clients.

loRes = <B>Low Resolution Video</B> - \
Provides full motion video for the selected Titles \
at a lower resolution which takes less disk space \
to store and less time to download when used by clients.

poster = <B>Poster Video</B> - \
Provides still images which illustrate the content \
of the selected Titles. This option takes less disk space \
```

Figure 226: View of a Java Properties file in a simple text editor

The structure of these files is relatively simple: typically everything that follows quotation marks needs to be translated, everything that is preceded by a hash mark (#) is to be excluded, and the translatable strings can contain some HTML-based tags.

Java Properties files are supported by *SDLX*, *Trados*, *Heartsome*, *Swordfish*, *Open Language Tools*, *OmegaT*, and *Déjà Vu* as well as most localization tools (see *Software Localization Tools* on page 276).

*Extensions are always a first indication of what the file type could be if you are not sure what format a certain software file is in, but they will often fail you with software files. If you are not sure about the file type, open it in a text editor and study the structure of the file. If translatables are enclosed with quotation marks, try to process the file as an RC file (or, in the case of Déjà Vu or SDLX, you can also test it with one of the other software filters). If the translatables are preceded by an equal sign, try to process them with the Properties filter. As all of these files are text-based, this will not damage the files and very often you will find that you "get lucky," even though the file at hand may not be one or the other.*

Another emerging text-based software standard is GNU gettext .po and .pot files. These are the translatable language resource files used in the free GNU gettext concept for translating software and documentation. GNU gettext is the de facto standard in many open source projects, and it works with a large variety of programming languages. .po files are typically translated or pretranslated files, whereas .pot files are the translatable templates.

```
#: common/catalog.cpp:506
msgid ""
"Free Software Foundation Copyright does not contain any year. It will not be "
"updated."
msgstr ""
"El copyright de la Free Software Foundation no contine ningún año. No se "
"actualizará."

#: common/catalog.cpp:1743
msgid "loading file"
msgstr "cargando archivo"

#: common/catalog.cpp:2011
msgid "saving file"
msgstr "guardando archivo"

#: common/catalog.cpp:3351
msgid "searching matching message"
msgstr "buscando mensaje coincidente"

#: common/catalog.cpp:3646
msgid "preparing messages for diff"
msgstr "preparando mensajes para diff"

#: common/context.cpp:117
msgid "Corresponding source file not found"
msgstr "No se encontró el mensaje fuente correspondiente"
```

Figure 227: View of a .pot file in a simple text editor

Aside from the internal tools that gettext offers (see www.gnu.org/software/gettext), *Déjà Vu, Heartsome, Swordfish*, *OmegaT* and *Open Language Tools* are the only translation environment tools that handle these files.

# Help Systems

Help systems—the documentation resource that is typically part of a software program and can be accessed through the help menu—is a huge topic on its own. I'm not planning to cover this in its entirety, but there are a few questions that I have been confronted with over and over again, and here are some quick answers for those.

First of all, there is a great variety of help systems (*JavaHelp*, *OracleHelp*, *QuickHelp*), but the two most often-used help systems in the Windows world are *HTMLHelp* and the increasingly outdated *WinHelp*.

# WinHelp

The compiled *WinHelp* system typically consists of two files, the .cnt file and the .hlp file. While the .cnt file is a text-based file that contains the table of contents for the help system, the .hlp file is a compiled file that is made up of any number of RTF files.

*These RTF files have to follow strict guidelines as to how they are created so that hyperlinks, index markers, and section breaks function correctly. Most larger translation environment tools (especially those that have been around for a while and seen the heyday of WinHelp) have facilities to accommodate these special features (such as hidden text for hyperlinks or the various kinds of footnotes).*



Figure 228: View of an RTF file before its compilation into a WinHelp help system

In case you receive a .cnt and .hlp file for quoting or even translation purposes, there's an easy way to "decompile" the .hlp file into its RTF components. While there are a number of expensive commercial tools for compiling and decompiling *WinHelps*, under sourceforge.net/projects/helpdeco you can find the *HelpDeco* application which allows you to break apart your help file and analyze and translate the resulting RTF files (and typically any number of image files).

*The downside is that this is not a particularly user-friendly application. To use it, open a DOS window (**Start> Programs> Accessories> Command Prompt**) and point the HelpDeco application to the help file. So, assuming that you have placed the helpdeco.exe at C:\decompile and your help file anycount.hlp is located right at C:\ you would enter this:*



One file that is also created in the process is an .hpj file, the help project file. Though this file is not to be translated, it is important because it contains the information on how to re-compile the project once the translation is done. The free Microsoft program that can be used to do just that is called *Microsoft Help*

*Workshop* and can still be found at www.helpmaster.info/hlp-developmentaids-hcw403.htm (Microsoft itself is not distributing it anymore). All you need to do to re-compile—i.e., to recreate the .hlp file—is to open the .hlp file with *Help Workshop* and click **Save and Compile**.



Figure 229: WinHelp compilation in *Microsoft Help Workshop*

## HTMLHelp

The process for *HTMLHelp* is similar but much simpler. Unlike the *WinHelp* system, *HTMLHelp* consists of only one file, the .chm files. True to its name, most of the translatable content of an *HTMLHelp* system is contained in HTML files. To "get to" the HTML files, you will also need to decompile the help file. Fortunately, both the compilation and decompilation are done with the same freely available and easy-to-use tool: *HTML Help Workshop*.



Figure 230: HTMLHelp decompilation in *HTML Help Workshop*

*You can download the latest version at msdn.microsoft.com/en-us/library/ms669985%28VS.85%29.aspx.*

To decompile an existing help file, just select **File> Decompile**, locate the .chm file, and choose a location to which you would like to export the files. You could receive a great number of different file formats, but the most typical are:

- .hhp: the non-translatable project file (you will need this file to recompile the help),

- .hhc: the translatable (table of) contents file in HTML format,

- .hhk: the translatable index file in HTML format,

- graphic files: these are often translatable and/or have to be replaced with newly created target counter-parts, and

- lots and lots of .html files with lots and lots of translatable content.

Before you start with the translation of your *HTMLHelp* project, here is one thing you should be doing first: Talk to your client about the format in which the authoring of this project took place. Chances are that it was either authored in *FrameMaker* (like this present manual and help system), in some kind of XML form, or even within *Word*. While it is entirely possible and really quite easy to translate the *HTMLHelp* directly, your client may be much better served if you are able to work in the original format. Typically the original authoring environment is set up so that the output can be done in various formats (PDF, printed materials, web based, help systems, etc.), whereas it is much more complicated to do this when you start with a help system.

If your client asks you to translate the help system directly, translate the above-mentioned files, replace the graphics (save them under the same name and the same location), and then recompile the individual files with *HTML Help Workshop*.

Well, usually it doesn't work quite so easily, because chances are some link was corrupted, some graphic is missing, or some file was renamed. So it is advisable to do a quality assurance check which compares the original files and your newly translated files. SDL offers an excellent product for doing just that in *HtmlQA*.



*The equally helpful sibling product for the WinHelp process is called HelpQA.*

Once you've fixed your errors, you can proceed with the compilation in *HTML Help Workshop*. Just select the .hhp file (make sure that it's placed at the root of your project folder), select **File> Compile**, and your help file will be all ready to go.



Figure 231: HTMLHelp compilation in *HTML Help Workshop*


*You can also use HTML Help Workshop to convert existing WinHelp projects. When you convert a WinHelp project to an HTML Help project, the New Project Wizard converts the WinHelp project (.hpj) file to an HTML Help project (.hhp) file, the WinHelp topic (.rtf) files to HTML Help topic (.htm, .html) files, the WinHelp contents (.cnt) files to HTML Help contents (.hhc) files, and the WinHelp index to HTML Help index (.hhk) files.*

A "new" concept of translating HTML help files is offered by localization tools like *Sisulizer*, *Catalyst, or Passolo* (see *Software Localization Tools* on page 276) or with the translation environment tool *Publisher*. Here the decompilation and recompilation is performed right within the tool (with the

support of an embedded version of *HTML Help Workshop*) so that the translator can directly translate the translatable files within the interface of the respective localization tool and the tool will then recompile the .chm files once the translation is done.



Figure 232: HTMLHelp file translation in *Publisher*

## Opening HTMLHelp (.chm) Files

While this is not directly translation-related, I have received a lot of emails from folks who bought this book and could not open the accompanying .chm file help file (which contains the same content as the PDF file but presents it in a more easily accessible format) on either a *Macintosh* computer or even on a *Windows* machine. Here's how to fix that.

For all versions of *Windows*, you might have to open the .chm file on your local computer; this means that you probably shouldn't have it on a network location. In *Windows Vista* or *7*, it is possible that you will be able to view the .chm file table of contents, but you might see the following message on the

right-hand pane: "Navigation to this webpage was canceled." This error is caused by the stringent security requirements of *Windows Vista* and *7* and can be circumvented by right-clicking on the file, selecting **Properties**, and clicking on the **Unblock** button in the lower part of the dialog.



Figure 233: Unblocking option for an HTMLHelp file in *Windows Vista*

For Macintosh computers there are essentially two ways of going about this. You can either download and install the open-source application *Chmox* (see chmox.sourceforge.net), which will allow you to open and use a .chm file, or you can install the *CHM Reader Firefox add-on* (see addons.mozilla.org/en-US/firefox/addon/3235—works through *Firefox* version 3.0), which permits you to view .chm files inside *Firefox*. This add-on is also available as a *Windows* or *Linux* version.



Figure 234: Using the *CHM Reader* within *Firefox*

# Database-Based Data

It's a strange thing with data in databases. So much of today's translatable content is stored in databases for easy and quick user access (this is especially true for web-based content), but translators are often met with a bit of suspicion when it comes to the translation of that data. And there is probably something to that suspicion. Much like software development files that developers often have a quasi-emotional attachment to, database

administrators rightly feel quite protective of their database content, translatable or not. As a result, data from databases is often exported to an exchange format such as .csv (comma-separated value file—a text-based file where data is separated by commas; see page 178) or *Excel* spreadsheets. While these formats typically do not provide many problems to be processed by translation environment tools, there are two significant drawbacks:

- depending on the database, the data may not only not have context but may also be concatenated (one string consisting of many pieces that are not necessarily displayed together), thus making it very difficult for the translator to translate appropriately, and

- it represents a great deal of overhead for the database administrator to not only maintain but also to import and export data in various languages on a regular basis.

It is therefore not surprising that a number of tools have tried to come up with solutions to translate database content directly within the database environment. Though there are a great number of different database formats, there are also standards that allow communication with almost all database formats. ODBC, Open Database Connectivity, is a native interface that allows access to most database management systems and allows for the use of SQL, Structured Query Language, the universally used language to "talk" to databases. By using this interface and this language, a number of computer-assisted translation tools are now able to translate database content directly, even from as complex an environment as *Oracle* or *MS SQL* databases.

Among localization tools, *Multilizer* (see www.multilizer.com) and *Alchemy Catalyst* (see www.alchemysoftware.ie) support the translation of database content; on the side of translation environment tools, the corporate edition of *Déjà Vu X* does.

Other database formats that have become more common are CMS (content management systems). These systems are database repositories that hold data, typically in XML format, which can be repurposed in a number of ways and can be updated on the fly. While it's no problem to work with XML—virtually all TEnTs support it—it turns out to be not quite as easy for the

translation provider to get into the CMS and access the content. It's possible to export the content into an XML or other text-based format and then re-integrate it, but that's a lot of manual work and also doesn't take care of the often updated content.

The very large translation providers have developed their own interfaces to these systems (or they've outright bought CMS providers, such as SDL's acquisition of *Tridion*), but since so much translatable content is moving into CMSs, the smaller providers need ways to take part in this as well. One way to do this is through middleware tools such as *Clay Tablet* (see www.clay-tablet.com). *Clay Tablet* essentially offers ready-made connectors between commonly used content management systems and systems like *Across*, *SDL TeamWorks*, *Plunet*, *Beetext*, and *]project-open[*, with other connectors in the pipeline. It also offers the possibility to upload CMS content to an FTP server from where the data can be downloaded and processed with any tool, but of course there you have the manual component again.

It's important to note that this tool is not installed on the language provider's end but on the client's end, integrating with the CMS and then connecting with an outside system. But there are plans to offer *Clay Tablet* with an affordable SaaS—Software as a Service—model, essentially forestalling any huge expenditure for anyone.

# PDF: Pretty Darn Frustrating?

This might come as a surprise to some of you that PDF does not stand for "Pretty Darn Frustrating" — even though it often feels that way — but "Portable Document Format." PDF is a multi-platform file format developed by Adobe Systems in the '90s that has become one of the most widely used file formats. The "secret" to PDF's popularity is that it provides a secure and easy method to share documents between users by capturing text, fonts, images, and formatting of documents from basically any application that allows printing. One of the reasons for the existence of PDFs is that they are relatively secure files, a benefit that also makes them very frustrating to work with for us translators.

Frustrating or not, in translation work, we encounter PDF files daily. They can be source text files, documents for proofreading, reference files, and various registration and other forms. We often also need to create PDF files, for example for resumes, invoices, file sharing, and printing/publishing.

## PDF File Types

To be able to utilize PDF files ithe most efficient way possible, it's important to know that, from a practical point of view, there are three different types of PDF files:

- text-based files

- image-based files

- searchable image-based files

In text-based PDF files the text is "real" text; you can copy and paste text from the file (unless it's restricted by the file's security settings) and search for text in the file. Converting these types of files to a fully editable (and translatable, translation-environment-tool-compatible) format, such as to a *Word* file, is less problematic than with image-based files, though it's not necessarily simple as we'll see later.

Image-based files, on the other hand, do not allow copying or searching because what appears to be text is actually part of an image. A scanned or faxed document that has been saved as a PDF file is an example of an image-based file. To convert these types of files to an editable file format one needs to use an OCR (optical character recognition) program, and the result depends on the clarity of the image.

The third type, the searchable image-based file, is kind of a hybrid between the two other types. It's an image file that is searchable, i.e., you can search text even though it's an image. A searchable image-based file can be created from an image-based file using the **Recognize Text Using OCR** function in *Adobe Acrobat* (not available in the *Reader* version). As with any OCR program, the results depend on the clarity of the text in the image. If you have a hard time reading the text, don't think that the program can read it any better. You can also copy and paste text from a searchable image file but again, the resulting text depends on how accurately the OCR program recognized the text.

Why do we need to talk about PDF files and related tools? The better we understand the possibilities and limitations of these files and the related tools, the easier it is to find the best and most efficient ways to handle them. For example, knowing proper tools can save hours of tedious manual editing when converting PDF files to an editable format.

## PDF Tools

*Adobe Reader* is probably already in almost everyone's computer. It allows you to view and search PDF files and also comment on files that have been enabled for commenting (more under *Enabling Extended Features for Adobe Reader* on page 350).

In addition to the free *Reader* version, the *Adobe Acrobat* product family also includes *Adobe Acrobat Standard*, *Adobe Acrobat Pro,* and *Adobe Acrobat Pro Extended* versions.



*Note that here the name "Adobe Acrobat" refers to these three paid versions and "Adobe Reader" to the free Reader version.*

You should make sure to review the additional features that these paid versions offer, such as enhanced editing, commenting, PDF file creation, file conversion, security settings, etc. (see the following table). For many translators, the additional features that *Adobe Acrobat Standard* and *Pro* offer are certainly worth the expense. For a full product comparison, see www.adobe.com/products/acrobat/matrix.html.

| Feature | Adobe Reader | Adobe Acrobat Standard | Adobe Acrobat Pro |
|---|---|---|---|
| Viewing, printing, and searching PDF files | Yes | Yes | Yes |
| Creating PDF files | No | Yes | Yes |
| Scanning and OCR capability | No | Yes | Yes |
| Applying restrictions on printing, copying and altering of PDF documents | No | Yes | Yes |
| Use of commenting tools | No (*) | Yes | Yes |
| Digitally signing PDF documents | No (*) | Yes | Yes |
| Comparing PDF files and highlighting differences | No | No | Yes |
| Enabling commenting tools and to digitally sign PDF documents | No | No | Yes |

(*) Possible if the file has been enabled for this in *Pro* version.

In addition to the *Adobe Acrobat* products, there are many more or less comparable and often less expensive programs that allow you to do many of the same things. For example, PDF Nitro (www.nitropdf.com), Foxit PDF Tools (www.foxitsoftware.com/pdf/), Solid PDF Tools (www.soliddocuments.com), DocuCom PDF Gold (www.pdfwizard.com), Pdf995Suite (www.pdf995.com),

and many others. However, if you consider any of these other tools, make sure that the tool is fully compatible with all current *Acrobat* features. You don't want to spend all that money to find out later that you need *Adobe Acrobat* after all.

I will concentrate on Adobe Acrobat here and will not cover any of these other tools.

# Editing and Commenting PDF Files

## Commenting Tools

One of the most common tasks that translators need to do with PDF files is commenting them as part of a document review (proofreading) cycle. *Adobe Acrobat* offers several tools for this to make commenting easy and clear. Note that these tools are available in the free *Reader* version only if the file has been enabled for that in the *Pro* version by the author (for details about enabling, see *Enabling Extended Features for Adobe Reader* on page 350).

The **Comment & Markup** tools can be accessed in several different ways either through the **Tools** menu (**Tools> Comment & Markup**) or the **Comment & Markup** toolbar (**Tools> Customize Toolbars / View> Toolbars> Comment & Markup**). Most users are familiar with the yellow **Sticky Note** tool and tend to use that for everything. However, in most cases it would be much more efficient and clearer to use some of the other tools. The **Text Edits** tool in particular is very good for indicating text corrections, additions, and deletions. Note that none of these tools actually changes the text in the file—they only indicate what needs to be changed. The actual

changes will then be made to the original file, for example, by a DTP person. Other **Comment & Markup** tools that are often useful in a review process include **Highlight Text**, **Callout**, **Arrow**, **Rectangle**, etc. They help to pinpoint the location where the associated comment is supposed to apply.



Figure 235: Tools for commenting and markup in Adobe Acrobat

## Managing Comments

Sometimes it can be difficult to manage all the comments in a file, particularly if the file is long or includes a lot of comments. *Adobe Acrobat* offers several tools to help to organize and manage comments. Clicking the **Show** button in the **Comment & Markup** toolbar opens a menu that includes several options that allow you to show or hide all comments (**Show/Hide Comments**) or only certain comments based on comment type, reviewer, or comment status.

In addition, you can also open or minimize all pop-ups at once by selecting **Open/Minimize All Pop-Ups**. Individual pop-ups can also be opened simply by double clicking their place holder in the file. In addition, you can access several comment-specific features by right-clicking a comment. One of these

options is **Reply**, which allows you to reply to the comment. That way the reply gets inserted within the comment and will stay with it. If you reply using a separate comment, there's always a risk that the comments might get separated, and other readers don't know which replies and comments are supposed to match.



Figure 236: A comment with a reply

**Show/Hide Comments List** is a very useful command because it opens a separate comments list at the bottom of the window. This list offers several other tools for displaying, sorting, searching, replying to, and managing comments, as well as printing and creating a comments summary. The **Options** button on the upper left corner of the **Comments List** includes options to summarize and export comments.



Figure 237: **Comments List** open in *Adobe Acrobat*

Comments can also be exported (**Document> Comments> Export Comments to Data File**), printed (**File> Print> Comments and Forms> Document and Markups / Document and Stamps**), and summarized (**Comments> Summarize Comments**). Exporting comments and sending

them as a comment file (FDF) can be very helpful with large files. The comment file is usually much smaller than the original file so it's much easier to deal with. However, the recipient needs to have the same PDF file for this to work so that the comments can be transferred to their correct locations.

*It is also possible to check the spelling of comments by selecting **Edit> Check Spelling> In Comments, Fields, & Editable Text** (or **F7**). Spelling settings can be accessed through **Edit> Preferences> Spelling**. Note that the spelling tool does not check the spelling of the actual document content.*

You can access several commenting-related settings through the **Preferences** dialog (**Edit> Preferences> Commenting**), such as the font used in comments and the behavior of pop-ups. If you want to change the appearance of your name in comments, go to the Identity section of the **Preferences** dialog. It's actually a good idea to review many of the other **Preferences** settings as well because they allow you to customize the way *Adobe Reader/Acrobat* behaves in general.

*Adobe Acrobat e-mail-based review or shared review (online on acrobat.com) is an efficient way to deal with multi-party reviews. More info can be found on www.acrobat.com.*

## Other Editing Options

The *Standard* and *Pro* versions of *Adobe Acrobat* also have the **TouchUp Text** tool (**Tools> Advanced Editing> TouchUp Text Tool**) that allows minor editing of the actual text, such as correcting typos and adding or deleting text. This can be a very handy feature to fix small errors or in making small updates, but it often creates text flow problems if used for larger changes. With the **TouchUp Properties** tool you can also change text properties, such as the font. This tool is accessed by first selecting the target text with the **TouchUp Text** or **TouchUp Object** tool, right-clicking the target area, and then selecting **Properties** from the context menu.

Other editing features available in *Adobe Acrobat* include inserting, extracting, deleting and rearranging pages (**Document> Insert/Extract/Replace/ Delete Pages**), and document splitting (**Document> Split Document**). With the **Insert** command you can combine PDF files by inserting pages from one file to another.

In addition to the **Document** menu commands, most of the above-mentioned editing functions can also be accessed using the **Pages** navigation pane on the left. To open the navigation pane, click the **Pages** icon on top of the navigation area. This opens the navigation pane where you can easily

rearrange the pages by dragging and dropping, as needed. Clicking the **Options** symbol on the top of the pane opens a menu that allows you to rearrange, crop, rotate and renumber pages. These can also be accessed through the **Document** menu.



Figure 238: The **Pages** navigation pane

The **Typewriter** tool is handy when you need to type something extra to a document, such as when filling out forms or adding the name of the fax recipient on the page. It basically allows you to type anything anywhere on top of the document.

Figure 239: Acrobat's Typewriter tool

## Enabling Extended Features for Adobe Reader

Some functions are available in the *Reader* version only if they have been enabled for the file in question in the *Pro* version. These functions include the **Comment and Mark-up** tools, the **Typewriter** tool, and the possibility to digitally sign the document. A document is enabled in the *Pro* version by selecting **Advanced> Extend Features in Adobe Reader**. Another way to enable a document is to select **Tools> Typewriter> Enable Typewriter Tool in Adobe Reader**. This latter method will only enable the **Typewriter** tool and the **Comment** and **Mark-up** tools.

*Note that certain functions, such as editing document content and inserting and deleting pages, will be restricted after the file has been enabled for Reader. These functions can be reactivated by saving the file using the File> Save A Copy command. This saves a copy of the document that does not enable the extended features.*

# Searching for Text in PDF Files

*Adobe Acrobat* offers two different ways to search for text in a document. Both methods are also available in the *Reader* version. **Find** looks for text within the current document (**Tools> Find** or Ctrl+F). This opens the **Find** toolbar, if it's not open already, and you can type the search string to the text

box in the toolbar. Clicking the down arrow next to the text box opens a menu with some additional find options. After searching the string for the first time, the **Find Previous** (left) and **Find Next** (right) arrows appear next to the text box.



Figure 240: The *Find* tool

The **Search** function (**Edit> Search** or CTRL+SHIFT+F) is a more powerful way to search text because it can search one or more files at once and displays all hits in a separate **Search** window. It also offers several search options for fine-tuning the search. Many of them can be accessed through the **Advanced Search Options** link at the bottom of the **Search** window.



Figure 241: **Search** window with advanced search options

The *Pro* version offers an additional search feature that's called indexing. With it you can index hundreds of files for very quick searching. Select **Advanced> Document Processing> Full Text Index with Catalog> New Index**, name the index, select directories to be included, click **Build**, and then specify a location for the index file. Then use the resulting PDX index file for searching. It's much faster and less draining on your computer resources than the normal search if you have a large amount of data.



Figure 242: **Search** window after performing a search through an index

You can also convert image-based PDF files to searchable image-based PDF files, as mentioned earlier. This allows searching, because normal image-based PDF files are not searchable. This can only be done using the *Adobe Acrobat Standard* or *Pro* version. With the image file open in *Acrobat*, select **Document> OCR Text Recognition> Recognize Text Using OCR**. The

resulting file looks exactly the same but it will be searchable. This can also be done with scanned images (**Document> Scan to PDF**). As with any OCR program, the results depend on the clarity of the text in the image. If you have a hard time reading the text, don't think that the program can read it any better.

# Printing PDF Files

Print settings (**File> Print**) includes several useful settings to optimize printing. With **Print Range** you can define the printed page range, with **Page Scaling** you can optimize the printout by shrinking, enlarging, or tiling the output, and with the **Comments** and **Forms** selections you can define whether comments and other markups will be printed.



Figure 243: Print dialog

The **Snapshot Tool** (**Tools> Select & Zoom> Snapshot Tool**) is another handy feature when you want to print only a certain area of a document (and maybe print it enlarged). With the **Snapshot Tool** you can define any area in the document for printing. Select the tool, draw the area (rectangle), and select **File> Print**. The selected area is displayed in the **Print Preview** section of the **Print** dialog. All these settings and tools are available in the *Reader* version as well.

## Viewing and Zooming PDF Files

Basic zoom and page display settings are under the **View** menu (**View> Zoom** and **View> Page Display**). Additional zoom tools can be found under **Tools> Select & Zoom**. One of these tools is **Marquee Tool** (**Tools> Select & Zoom> Marquee Zoom**). It's very useful when you want to zoom into a certain area of a document. Just select the tool, draw the target area (rectangle), and the area will be displayed so that it fills up the whole document area. Another handy tool is the **Dynamic Zoom** (**Tools> Select & Zoom> Dynamic Zoom**). It zooms in when you drag your pointer up and zooms out when you drag down. All these settings and tools are available in the *Reader* version as well.



Figure 244: **Select & Zoom** menu

## Electronic Signatures

In general, I try to avoid printing paper copies as much as possible. Using *Adobe Acrobat*, PDF files, and electronic forms or e-faxes often makes this possible and easy. Unfortunately, for some users this works well until it becomes time to sign the document. At that point they feel forced to print out the document and look for a pen. However, it is also often possible to insert an

electronic signature into the document. This can be done in two different ways. The simple and less secure method is to insert a scanned signature (for example a JPG file) into the document by copying and pasting it via the clipboard. It gets inserted as a "stamp," and you can easily resize and relocate it by dragging it to fit to the space available. Note, however, that anyone can copy your signature as an image from the PDF file and reuse it (unless you set the security settings to restrict copying). A much more secure method is to use a digital ID.

Creating a digital ID is a simple process and only needs to be done once. After that, using it is very simple. To create a digital ID, select **Advanced> Security Settings> Digital IDs> Add ID> A new digital ID I want to create now> Next> New PKCS#12 digital ID file> Next**. Fill the information fields as needed and click **Next**. Select a location for the ID file and define a password. Click **Finish** to return to the **Security Settings** dialog and click **Close**. That's it.



Figure 245: Security Settings dialog

You can create several types of digital IDs and even include an image in it. The image could be, for example, a logo or your handwritten signature.

When you want to sign a PDF document, select **Advanced> Sign & Certify> Place Signature**. Drag a rectangle where you want to place the signature. Choose a digital ID from your list of available IDs, type the password, choose an appearance, and click **Sign**.



Figure 246: Electronic signature in PDF

To create a digital ID and to sign a document using a digital ID, you need *Adobe Acrobat Standard* or *Pro*. Signing of documents is possible with the *Reader* version only if the feature has been enabled in the document by the author using Adobe Acrobat Pro (for details about enabling, see *Enabling Extended Features for Adobe Reader* on page 350).

## Converting PDF Files to Editable Text

Converting PDF files is probably the most common problem that translators encounter with PDF files when translating using a translation environment tool. Most translation environment tools can't process PDF files, and even those that can (such as *Trados Studio*, *Wordfast Pro*, *memoQ*, *Alchemy Publisher*, *XTM Cloud* or *Fluency*) don't do it well enough. One problem that is often encountered is that many of the conversion methods insert paragraph marks (hard returns) at the end of every line. This means that these sentences will be incorrectly segmented in a translation environment tool, making translation impossible or the resulting translation units useless.

A good conversion program converts a PDF file to a *Word* file with flowing text but conserves formatting (bold, italics, paragraphs, tables, etc.) without creating text boxes. If the PDF file is an image-based file (such as a scanned or faxed document) the program also needs to be able to convert the image to text accurately. I will briefly review and compare a few options that are available for this task. These include *Adobe Reader*, *Adobe Acrobat Standard/Pro*, *ABBYY PDF Transformer*, and *ABBYY FineReader*.

## Adobe Reader

*Adobe Reader* offers only two possible conversion methods: text can be copied and pasted using the clipboard, or the file can be saved as a text file (**File> Save as Text**). With both methods, each line ends with a hard return (paragraph mark), so they are practical only for a small amount of text.



Figure 247: Text copied using *Adobe Reader*, showing paragraph marks at the end of each line

Some tips for selecting text in *Adobe Reader* and *Adobe Acrobat*: You can select a whole page by clicking it four times. When selecting all (CTRL+A), "all" can either be a whole page or a whole document depending on the **Page Display** setting. If the setting is **Single Page** (**View> Page Display> Single Page**) only the current page will be selected. If any other page display setting is selected, the whole document will be selected. When copying text, sometimes, depending on the file, there might also be an option to copy with or without formatting (right-click menu). You can also use the **Column select** mode to select a rectangle area of text anywhere in a document. It's activated by keeping the ALT button down while dragging a rectangle over the target area.

## Adobe Acrobat

The *Standard* and *Pro* versions of *Adobe Acrobat* offer some additional conversion methods. You can select **File> Export or File> Save As**, which allows saving the file directly in various file formats (such as *Word*, *Excel*, HTML, XML, etc.).

Once you select text within a file, there are also a number of right-click menu options available: **Copy**, **Copy As Table**, **Save As Table**, **Open Table in Spreadsheet**. These table options can be quite handy when trying to convert text into a table format (great for creating glossaries). However, tables can be very tricky to convert with any of the above methods. For example, I have been able to convert a table very well using the Save As Table or Open Table in Spreadsheet options but they usually convert only one page at a time even if I select several pages of the table. The conversion settings can be accessed though **Edit> Preferences> Convert From PDF**. For more information see http://tinyurl.com/r9jn9y (Adobe online help).



Figure 248: Available settings for converting PDF files to DOC format in *Adobe Acrobat*

## OCR (Optical Character Recognition) Tools for PDF Conversion

Some of the most commonly used OCR tools for PDF file conversion are by ABBYY ( www.abbyy.com) and Nuance (www.nuance.com). The product line of both companies includes a full-scale OCR program (*FineReader* by ABBYY and *OmniPage* by Nuance) and a lighter version (*PDF Transformer* by ABBYY and *PDF Converter* by Nuance) that converts only PDF files.

*There is a fundamental flaw in Nuance's Converter product: while it supports a number of different languages, it will always convert into the locale of the operating system that you work under and you can't 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.*

Here I only cover the two above-mentioned ABBYY products but the basic functions and principles are pretty much the same with the Nuance products as well. Both companies offer free trial versions, so it makes sense to try them out first, using files that are typical in your work, before making a final buying decision.

### ABBYY PDF Transformer

*ABBYY PDF Transformer* is a very simple program to use, and you can convert a PDF file with just a few clicks. In addition, it also creates PDF files. The program reads the PDF file and converts it to the desired target file format (*Word*, *Excel*, HTML, text, or searchable PDF). It offers a few settings and tools to customize and improve the conversion process. First, you need to select the correct language(s) and the desired advanced (layout) options. When converting to *Word* format, you have the following three advanced options available:

- **Original layout**: Creates an output document that looks exactly like the original. Text is often placed in text boxes and is difficult to manipulate.

- **Text flow**: Retains the text of the original, but some of its formatting will be lost. The output document will retain paragraphs and fonts, but will not retain columns, exact locations of objects, or spacing, and is easier to manipulate.

- **Keep pictures**: Retains the pictures of the original document.

**Advanced Options** for *Excel* files are **Ignore text outside tables** and **Convert numeric values to numbers**.

If the automatic conversion process fails to recognize text, table, or image areas correctly you can manually change them and rerun the conversion. You can also delete areas that you don't want to include in the target file. The program recognizes 184 different languages.



Figure 249: Converting a PDF file to *Word* format in *ABBYY PDF Transformer*. Green areas indicate text areas and red ones image areas.

## ABBYY FineReader

*FineReader* is a full-scale OCR program and does much more than just convert and create PDF files. It can be used to convert scanned documents and several types of image files to editable format. It has several more options and features compared to *PDF Transformer*, but it's still quite easy to use.



Figure 250: Converting the same PDF file to *Word* format in *ABBYY FineReader*

The conversion process consists of four basic steps:

- Retrieving the image (opening an image file—PDF, TIF, etc.—or scanning a document),

- Text recognition (usually done automatically by *FineReader's* OCR engine),

- Checking and editing of the recognized text (spell-checking and manual editing of the output document in the Text window), and

- Saving the recognized text in the chosen output file format (such as *Word*, *Excel*, *PowerPoint*, HTML, text, etc.).

*FineReader* offers several general and file-specific options for perfecting the conversion process and the output (**Tools> Options**). Many of the tasks can be customized and automated so that a complete conversion can be done with just a couple of clicks (**Tools> Automation Manager**). When converting PDF files to editable file formats, it's important to select the **Document Format Saving Mode** that produces the most suitable output. The availability of the four modes depends on the output file format. When converting to *Word* (DOC, DOCX, RTF) format, all four modes are available:

- **Exact copy**: Formatting corresponds to that of the original but the ability to change the text and format of the output document is very limited. Text is often placed in text boxes.

- **Editable copy**: Formatting may differ slightly from that of the original but document is easy to edit.

- **Formatted text**: Fonts, font sizes, and paragraphs are retained but not the exact spacing or locations of the objects on the page.

- **Plain text**: No text formatting is retained.

In addition, you have options to retain or exclude pictures and footers/headers. You can change all these settings on the fly and see the results immediately in the **Text** window without having to save the output document first.



Figure 251: Available layout options when converting PDF files to *Word* format in *ABBYY FineReader*

The *FineReader* main window consists of four panes, various toolbars, and the menu. The **Page** pane displays the pages of the current document. The **Image** pane displays an image of the current page and allows editing of image areas, changing area types (text, table, or picture) and text properties. In the **Text** pane you can view the recognized text, and check spelling, and format and edit the recognized text. In the **Zoom** pane you can see an enlarged image of the line or image area currently being edited. While the **Image** pane displays the general page view, the **Zoom** pane provides an easy way to view the image in greater detail, adjust the area type and position, or compare uncertain characters with the enlarged image.

### Post-editing Converted Files in Word/Excel

Sometimes additional formatting or "cleaning" is needed after the conversion to delete unnecessary spaces, characters or hidden tags or fix formatting problems.

*For this, the CodeZapper Word macro that eliminates most unnecessary codes can be very helpful. This can also be used to help to clean up converted Word files. You can download it from http:// tech.groups.yahoo.com/group/dejavu-l/files/CodeZapper/—note that you might have to temporarily become a member of that newsgroup for the download.*

### Translating PDF Files Using Translation Environment Tools

It would be great if we could open PDF files directly in translation environment tools without having to convert them first to a *Word* format. This has been on the wish list of many translators for a long time, and recently at least five tools (*Trados Studio*, *Wordfast Pro*, *memoQ*, *Alchemy Publisher*, *XTM Cloud*, and *Fluency*) have introduced PDF file support.

This is good news, but unfortunately in many cases the results of those conversions are less than desirable.

The purpose of a PDF file is usually to be the end product, and they are not really made to be edited. Unfortunately, we are sometimes stuck with a PDF file as the only source file available, and in order to translate it with a translation environment tool, one needs to convert it to an editable file format (such as *Word* or text file) first. There are various tools for that purpose and they work better or worse depending on the tool and the PDF file in question, as explained earlier. Note that the PDF file translation feature in the above-mentioned translation environment tools is not some new miracle that all of a sudden makes PDF files translatable—it's just one of those PDF-to-DOC conversion tools that has been built into them as a filter. For example, *Trados* uses a converter by *Solid Documents*, *Publisher* and *Wordfast* use *BCL*, Fluency uses *TX TextControl*, and *memoQ* employs *Glyph & Cog* to convert PDF files into a *Word* or text file and offer a translation interface for that.

Interestingly, the strategies of the different TEnTs differ.

*memoQ* does not even pretend to save the formatting of the file. Instead, it's "only" converting it into a text file without any formatting (but also without any superfluous hard returns, etc.). While this does not sound attractive in the first place, in many cases you might up saving time (and ugly surprises) even though you will have to spend considerable time formatting the file once you are done with the translation.

*Fluency* uses a PDF converter that tries to retain the formatting of the PDF, but it offers an intermediate step where you can edit the translation file with its possible (and likely) issues before you start the translation process.



Figure 252: *Fluency's* interim editing interface

*Trados*, *Wordfast Pro*, and *Publisher* all try to do everything "behind the scenes." You open a PDF file and the tool tries to keep the layout and presents it to you in its translation interface for you to work in. Once the translation is done, the file is exported into a *Word* or RTF file (which in all likelihood is not the format the PDF originated from). For really simple PDFs, this can work really well. And for others?

See the following painful example:



Figure 253: A PDF file that has been opened directly in *Trados Studio*. An overabundance of tags makes translating the file very cumbersome.

You will be hard-pressed to encounter problems like this in the conversion of PDF files with tools that rely on an OCR process, such as *ABBYY Fine Reader* or *PDF Transformer*. Another problem that is often encountered with *Trados*, erroneous hard returns at the end of lines, is also handled relatively well with the ABBY and Nuance tools. You may occasionally find a rogue hard return in converted files from these solution, but they are few and far between.

*Trados Studio's* PDF filter does allow some adjusting of the conversion process (**Tools> Options> File Types> PDF> Settings**), but the options are very limited compared to those of a proper PDF conversion tool.



Figure 254: PDF filter settings in *Trados Studio*

In summary . . .

• Don't think that you can translate PDF files just like *Word* files even if you can open them in your translation environment tool.

• If you open a PDF file in your translation environment tool, review the converted text in the editor to see if there are problems with tags or hard returns. You can try to fix the problems by adjusting the PDF filter settings of your tool or by first saving the converted source file in *Word* format, fixing the problems in *Word,* and then finally opening the fixed *Word* file for translation. Remember that in most TEnTs you can't edit the source

segments, so the errors need to be fixed before you start translating the file.

- If you need to convert PDF files frequently, consider buying a good conversion program, such as *ABBYY Fine Reader* (or *PDF Transformer*) or *Nuance OmniPage* (or *PDF Converter*).

For more information regarding PDF conversion in Trados Studio, see this blog entry at http://tradoshelp.wordpress.com/2010/05/17/.

## PDF File Password Protection and Permission Control

PDF files can also be password-protected (the password protection can only be applied in *Adobe Acrobat Standard* or *Pro*). There are two types of password protection available. One requires a password to open the file and the other one restricts copying, editing, and printing of the file. These two protection types require separate passwords and can be used independently of one another. You can view a document's security settings by selecting **File> Properties> Security**. The lower half of the tab consists of a **Document Restrictions Summary** showing what's allowed and what's not. You can define security settings (if allowed) by selecting **Password Security** from the

**Security Method** pull-down menu or selecting **Change Settings** if the file already has security settings. This opens the **Password Security Settings** dialog where you can select the desired encryption level, set passwords, and choose which actions will be restricted.



Figure 255: Document security settings in *Adobe Acrobat*. Without the correct password, this document cannot be changed or copied. Only printing is allowed.

Sometimes you might end up getting a password-protected PDF file from a client. If the client is unable to provide you with the password, you might need to use a tool that can unprotect PDF files. There are several tools available, and they can easily be found by searching "unprotect PDF" on the Internet.

# Creating PDF Files

There are many benefits to being able to produce PDF files, including being able to produce professional-looking résumés, invoices, and all kinds of non-HTML web content, and sharing files that are produced in applications that someone else may not have (such as desktop publishing applications).

Nowadays, most people are able to create PDF files with the tools they already own without having to purchase any additional PDF creation tools. This feature is included, for example, in the MS Office 2007 and 2010 suites (**Save As> Adobe PDF**) or the *OpenOffice* suite (**File> Export as PDF**).

*If it's not included in your version of Office 2007, you can download it free from Microsoft Office support site as the "SaveAsPDFandXPS.exe" add-in.*

Of course, it's also possible to convert to PDF files with *Adobe Acrobat Standard* and *Pro* as well as with most other PDF conversion tools. Depending on the tool, the PDF file is created either by saving or printing the original file as a PDF file.

# Creating Translation Memories from PDF Files

In addition to translating PDF files, it's also often desirable to be able to align PDF files to create a translation memory, for example when a previous version of source and target language documents is available only in PDF format.

There are a very limited number of alignment tools that are able to handle PDF files. Based on my experience, the best is *Logiterm AlignFactory* (see *An Excursion: A Word of Caution About Alignment* on page 234).

# Voice Recognition Applications

There are a lot of complaints that speech recognition—the ability to dictate to your computer—is geeky technology. But I think the very opposite is true. How geeky is it to hack on a keyboard to make your computer understand what you are trying to say? Really: think about it. It makes so much sense to be able to speak to your computer, dictate text, and navigate through programs. And the only geeky part about it is that we're not used to it and that it works—kind of.

Though I like speech recognition, I am not a "purist." I use it only when I think I need to speed things up a little, when I have a text that is well suited, or when my fingers just don't work the way I want them to (which unfortunately happens more often than I care to admit). But even when I use it, I don't unplug my keyboard or simply refuse to use it. Some things are just more practical to do on the keyboard, and this is particularly true if you need to switch between languages, which obviously is quite common for translators. . . . (The program that I use—*Dragon NaturallySpeaking*, available at www.nuance.com—supports at least the native language and English in the Dutch, German, Spanish, Italian, French, and Japanese editions; however, unloading one language and re-loading the other takes at least a couple of minutes.)



Figure 256: Selecting one of two possible languages in a German version of *Dragon*

So, which texts are well suited—or better, which texts are not well suited—for speech recognition? The answer to this depends partly on your particular translation subject. In mine it is mostly texts with a lot of proper names and/or loan words. This does not mean that you can't teach the program to recognize the proper names and loan words, but it's one of those judgment things: If you want to use speech recognition (or anything else for that matter) to become more effective, you'd better make sure that you truly are. If you have to spend an hour to train it to recognize a bunch of new terms before translating for an hour and a half on a job that would otherwise have taken you only two hours, that seems like wasted time to me. Plus, while I enjoy translating, I can think of better things to do than training speech recognition. On the other hand, if I can expect that these proper names and loan words will also occur in future projects, I may just as well spend the time to train.

My first rule for success with speech recognition software will probably have the "purists" shaking their heads in agony. After having used the software for some time, I know some of the weak spots of my speech engine (or my pronunciation). Rather than using the "correct" function again and again, I prefer to type those problem terms even while dictating the rest.

My next rule: Take some time to get used to not "thinking with your fingers." Instead, try to preformulate longer segments and then speak them coherently for better results.

This goes right along with the next kind of texts that are not well suited for speech recognition because it's hard to say them naturally: texts with a lot of formatting. Depending on what kind of translation environment tool you're working with and how formatting is handled by the tool, it may be easier to use the keyboard shortcuts for those that you are used to. If there is really a LOT of formatting, it may be easier to just type the whole thing.

Now, technically, there is no formatting function or other fancy maneuver that your speech recognition can't do. That is, if you have the right version. When it comes to *Dragon NaturallySpeaking*, the *Preferred* version comes with all basic formatting in environments like *MS Word* or its own editor, *DragonPad*.

When you use a translation environment tool that makes you work in an interface other than *Word*, you will have a hard time doing everything with voice commands unless you have the *Professional* edition, in which you can easily write macros with virtually unlimited possibilities.



Figure 257: Recorded Macro in the *Professional* version of *Dragon*

The problem is that while the *Preferred* version has a relatively modest price tag, the *Professional* version does not. Once you have the *Professional* version, you can either stay there and pay premium prices for upgrades because you are interested in the slightly better recognition that typically comes with each new version, or you can go the cheap route, downgrading at some point but then losing all your macros. That's the problem that I am stuck in with *Dragon*, so I'm not running the latest version.

*Windows Vista* and *7* also contain an internal voice recognition program.



Figure 258: Starting Vista's *Speech Recognition* in the *Control Panel*

*In Windows 7, this feature is available for the following language versions: Chinese, Japanese, German, French, Spanish, and English.*

This feature has suffered some very public criticism, but I was rather impressed with its accuracy and user-friendliness in a couple of unscientific tests that I ran. I dictated the same paragraphs in both programs and had only a slightly worse recognition in *Vista* than in *Dragon* (96% vs. 98%).

So, unless you are an awesome typist and refuse to change that geeky habit of exclusively using your fingers to enter text, speech recognition is a great alternative way to "type," even before carpal tunnel syndrome hits.

# Support

For most (if not all) of the tools discussed in this book you can make use of support options of varying quality.

Ironically, the best support available is often for some of the tools discussed under *Utilities* on page 117. Many of these tools are created and supported by a small handful of developers who can be passionate about providing excellent support. To access this kind of support, go to the appropriate website and simply send them an email. I have often received detailed answers within a short amount of time, or even an offer to rewrite the program to fit my specific needs.

Obviously this won't work with the larger programs and/or software companies—you will be hard pressed to have Steve Ballmer respond to your question about *Windows* or *Word*. And speaking of Microsoft, though there is a plethora of websites and newsgroups that provide paid or unpaid services, I find my easiest access to information through Microsoft's *Knowledge Base* under support.microsoft.com. The only drawback is that the descriptions in the vast *Knowledge Base* sometimes tend to be slightly technical.



*As a side note, Microsoft has machine translated tens of thousands of knowledgebase articles into several languages. For an example, go to support.microsoft.com/kb/281925/en-us and then click on one of the translation links on the right-hand side. You will see a machine-translated version of the article in the respective language that is preceded by a disclaimer informing the user of possible pitfalls of the translation. The translation is not pretty. But it communicates (most of the time) what otherwise would not have been communicated at all.*

For Adobe products, I usually refer to the excellent user-to-user forums of Adobe's website for their different products (see www.adobe.com/support/forums). You can easily search through the databases of former queries and solutions, and Adobe typically has at least one person per product who oversees the queries.

Translation environment tools have different levels of support for their products. While some products—including *Swordfish, memoQ,* and *Wordfast*—have outstanding support from their development communities, other products are not particularly well-known for this. Either way, the best way to get support for any of these and other translation environment tools is through user groups, almost all of which are located at groups.yahoo.com. Most of these groups are extremely supportive and are very forgiving of even the most basic questions. There are also a number of other translation tool-related *Yahoo! Groups* that are devoted to more general questions rather than any one specific product. I have usually found the specific groups more to-the-point and helpful, though.

# Conclusion

In my work as a translator, I derive at least as much joy from finding creative solutions to translation tasks with my computer as I do from actual translation. I'm not sure that you have to become quite so extreme, but I do hope that some of the tips in this book may have given you new ideas or a new desire to make your working experience with your computer more efficient and less frustrating.

It's the nature of a book about software to become quickly obsolete, so I publish updates on a regular basis. Major upgrades will be announced; minor updates and corrections can be downloaded for free. For more information, please refer to www.internationalwriters.com/toolbox.

If you notice any errors or omissions or have suggestions on what to include or leave out in coming editions of this book or the bi-weekly *Tool Kit* newsletter (see www.internationalwriters.com/toolkit), please send me an email at toolbox@internationalwriters.com.

If you are interested in comparing the different TEnTs, you should investigate TranslatorsTraining.com, a Flash video-based online resource with one-to-one comparisons of the 16 leading TEnTs. I started this website in cooperation with Intrawelt, an Italian translation company that specializes in online training, and it serves as a great resource that gives you a powerful comparison of the main features of the different tools.

# Index

## B

backdoor 75
backing up files
  Windows 7 50
backing up the system 48
backup
  using CD-R 48
  using external drive 48
Beetext Flow
  project management 283
behavior-based detection
  in antivirus software 88
Belarc Advisor
  computer inventory 53
bi-directional languages
  no support in MadCap Lingo 221
  PageMaker support 300
binary files
  software development formats 322
Bing 68
bmp
  graphic extension 305
book
  FrameMaker extension 290
  mission 1
  purpose 1
Botnet 78
breadcrumb trail 11
browsers 67
Browsing Tips 68
browsing tips 68
brute force
  password cracking 132
built-in encryption 85
bulk e-mail 79

## C

cable lock 85
Caps Lock key
  beep when pressed 63
Catalyst
  localization tool 278
  purchased by translations.com 280
  support of databases 336
  translate CHM files 332

categorization
  of CAT tools 188
CD emulators 126
Change Bars
  in FrameMaker 290
changing
  date/time stamp of file 120, 122
Character Map 54
character set
  definition 156
chm
  HTMLHelp extension 330
chm files
  opening in Vista and Windows 7 333
Chrome
  browser 68
Clay Tablet 337
Clear-CAT 191
ClearType 15
click fraud 78
clipboard
  definition 142
  translation with Wordfast Anywhere 232
Clipboard Management Utilities 142
clipboard utility 143
ClipMate
  clipboard utility 143
CMS
  translate 336
cnt
  WinHelp extension 328
code handling
  translation environment programs 243
code page
  changing in Word 156
  definition 156
code pages
  differences between Mac and Windows 304
codes in translation environment programs
  definition 243
CodeZapper
  macro to clean Word files 364
collaboration tools 144
comma-delimited files
  vs. Excel files 177
comments
  deleting in Word 174