The Translator's Tool Box

A Computer Primer for Translators



by Jost Zetzsche

Version 14, January 2020

Copyright © 2020 International Writers' Group, LLC. All rights reserved.

This document, or any part thereof, may not be reproduced or transmitted electronically or by any other means without the prior written permission of International Writers' Group, LLC.

ABBYY FineReader and ABBYY Screenshot Reader are copyrighted by ABBYY Software House. Acrobat, Acrobat Reader, FrameMaker, InDesign, Illustrator, PageMaker, Photoshop and RoboHelp are registered trademarks of Adobe Systems Inc. Acrocheck is copyrighted by acrolinx GmbH. Across and crossMarket are trademarks of Nero AG. AllChars is copyrighted by Jeroen Laarhoven. Any Video Converter is copyrighted by Anvsoft Inc. Apache OpenOffice is a trademark of The Apache Software Foundation. ApSIC Xbench and Comparator are copyrighted by ApSIC S.L. ASAP Utilities is copyrighted by eGate Internet Solutions. AVS4YOU is copyrighted by Online Media Technologies Ltd. Belarc Advisor is a trademark of Belarc, Inc. CafeTran Espresso is copyrighted by Collaborative Translation Networks, LLC. Publisher is a trademark of Alchemy Software Development Ltd. Classic Shell is copyrighted by Ivo Beltchev. ClipMate is a trademark of Thornsoft Development. ColourProof, ColourTagger and QA Solution are copyrighted by Yamagata Europe. CopyFlow is a trademark of North Atlantic Publishing Systems, Inc. Crowdin is copyrighted by Localization Management Platform - Crowdin, LLC. Déjà Vu is a trademark of ATRIL Solutions. Dragon is a trademarks by Nuance Software. dtSearch is a trademark of dtSearch Corp. Easyling is copyrighted by Easyling. ex TranslationFilter is copyrighted by CoDesCo IT Consulting GmbH. ExamDiff Pro is a trademark of Prestosoft. EmEditor is copyrighted by Emura Software inc. Error Spy is copyrighted by D.O.G. GmbH. FileSplit is a trademark of Partridge Software. Flare and Lingo are copyrighted by MadCap Software Inc. Fluency is copyrighted by Western Standard, Inc. Format Factory is copyrighted by SOFTONIC INTERNACIONAL S.A. Google, YouTube and Google Translate are trademarks of Google, LLC. K-Lite is copyrighted by Codec Guide. LibreOffice is copyrighted by The Document Foundation. Logoport and Translation Workspace are trademarks of Lionbridge Technologies, Inc. Fusion is a trademark of Orca Development Corporation. Heartsome and TMX Editor are governed by the GPL v2.0 License. Insert Togglekey is copyrighted by Mike Lin. IntelliWebSearch is copyrighted by Michael Farrell. KeyTweak is copyrighted by Travis Krumsick. Lingofy is copyrighted by Lingofy. LingoHub is copyrighted by lingohub GmbH. Linguee and DeepL are copyrighted by DeepL GmbH. LF Aligner is copyrighted by András Farkas. LINUX is a trademark of Linus Torvalds. LogiTerm, AlignFactory and SynchroTerm are trademarks of Terminotix Inc. Worx is a trademark of Alpha CRC Ltd. Mac and Macintosh are trademarks of Apple Computer, Inc. MateCat is copyrighted by Translated s.r.l. MetaTexis is a trademark of MetaTexis Software and Services. memoQ and memoQ cloud server are copyrighted by memoQ Translation Technologies. Office, Word, PowerPoint, Access, Edge, Excel, Multilingual Toolkit, Outlook, Publisher, Visio, Project, Internet Explorer, Edge and Windows are trademarks of Microsoft Corporation. Mozilla and Firefox are trademarks of The Mozilla Organization. Multi-Edit is a trademark of Multi Edit Software, Inc. Multilizer is a trademark of Multilizer Inc. Norton AntiVirus and Norton Utilities are trademarks of Symantec Corporation. Notepad++ is copyrighted by Don Ho. Nvu and BlueGriffon are copyrighted by Linspire, Inc. Opera is a trademark of Opera Software AS. PDF995 is copyrighted by Software995. PDFCreator is copyrighted by Open Source Technology Group. PerfectIt is copyrighted by Intelligent Editing Ltd. Plunet is copyrighted by Plunet GmbH. PKZIP is a trademark of PKWare, Inc. Poedit is coyrighted by Václav Slavík. PractiCount & Invoice is a trademark of Practiline Software. PrimoPDF is copyrighted by activePDF Inc. Protemos and TQAuditor are trademarks by Protemos. Quadsucker is copyrighted by S-B Software. QuaHill is copyrighted by DEVdivision software. QuarkXPress is a trademark of Quark, Inc. Quicken is copyrighted by Intuit Inc. RC-WinTrans is a trademark of schaudin.com. Replace Studio Pro is copyrighted by Funduc Software Inc. SDL AppStore, SDL Online Translation Editor, SDLX, SDLPhraseFinder, Trados Studio, MultiTerm, MultiTerm Extract, MultiTrans, and Passolo are trademarks of SDL International. SendTo is copyrighted by Trogladite Software Group. SnagIt is a trademark of TechSmith Corporation. Smartcat is copyrighted by Smartcat Platform Inc. SpywareBlaster is copyrighted by Javacool Software LLC. Star Transit, TermStar, Star James and FormatCheckers are trademarks of Star AG. Start10 is copyrighted by Stardock Corporation. Sysfilter is a trademark by Polmann Services. Teleport Pro Tennyson is a trademark of Maxwell Information Systems, Inc. Text United is copyrighted by Text United GmbH. Time Stamp is copyrighted by William Rouck. Toggl is copyrighted by Toggl. T.O.M. Translator's Office Manager is copyrighted by Joachim Voigt. Total Commander is copyrighted by Christian Geisler. Transifex is copyrighted by Transifex. Translation Office 3000, ExactSpent, Projetex and AnyCount are copyrighted by Advanced International Translation. Transmissions is a trademark by Transmissions, LLC. Twins File Merger is copyrighted by Twins Software, Inc. UltraEdit is a trademark of IDM Computer Solutions, Inc. Unicode and the Unicode logo are trademarks of Unicode Inc. Unifier is copyrighted by Melody-Soft. Ventura, WordPerfect Office, WinZip and Paint Shop Pro are trademarks of Corel Corporation. Verifika is a registered trademark of Palex Group Inc. VLC Media Player is registered by VideoLAN. WebBudget and FreeBudget are trademarks of Aquino Developments S.L. Wordbee and Beebox are copyrighted by Wordbee S.A. Wordfast Classic, Wordfast Anywhere and Wordfast Pro are trademarks of Wordfast Ltd. WordFinder is copyrighted by WordFinder Software International AB. WordPress Multilingual Plugin is copyrighted by OnTheGoSystems,Inc. XnView is copyrighted by XnSoft. XTM Cloud is copyrighted by XTM-INTL. XTRF is copyrighted by XTRF Management Systems.

All other product names are trademarks or registered trademarks of their respective companies.

Table of Contents

Introduction 1
- The Purpose of This Book 1
- How to Read This Book 2
- How to Read The Updated Version of This Book 3
- Who the Robot with St. Jerome's Face on the Cover Is 3

Operating Systems 5
- The Benefits of Windows 2000 and Higher 5
- Switching to the Windows 8 or 10 Interface 10
- Working with the Windows 10 Interface 12
- Windows/File Explorer 14
  - *Previewing Files 14*
  - *Folder Paths 14*
  - *Selecting Multiple Files 15*
  - *The Ribbon Menu 16*
  - *Libraries in Windows 16*
- Helpful Shortcuts 18
  - *Sending Files to Other Drives or Programs 18*
  - *To Search with Wildcards 20*
  - *To Copy Files or Folders 23*
  - *WinKey Shortcuts 29*
- Folder and File Structure 31
- Controlling Which Programs Are Automatically Started 31
- Avoiding Expensive Visual Effects 34
- Keeping the Computer Clean 35
  - *The Registry 36*
  - *Disk Cleanup 38*
  - *Finding the Forgotten Space Hogs 41*
- Error Checking and Defragmenting Drives 42
- Starting the Computer in Safe Mode 43
- Restoring Your Computer 44
- Backing Up Files 45
  - *File History in Windows 8 and above 48*
- Backing Up and Restoring the Complete System 51
- Asking for Help 52
- Taking Inventory of Your Computer 54
- Keyboard Languages 54
  - *Installing Additional Keyboards on Windows 7 57*
  - *Installing Additional Keyboards in Windows 8 and above 60*
  - *Mapping Existing Keyboards 62*

Web Browsers 67
Browsing Tips 67
Using URLs to Find Translation Data 69
Using Wikipedia for Language Data 70
File Transfer 71

Reducing Your Risk 75
Common Threats 75
Malware 75
Virus 75
Worm 76
Trojan Horse 76
Ransomware 76
Spyware 77
Backdoor 77
Keylogger 77
Adware 77
Malvertising 78
Rogue Security Software 78
Attacks 78
Phishing 79
Drive-By Download 80
Maliciously Crafted Content 80
Denial-of-Service (DoS) 80
Zombie 81
Botnet 81
Man-in-the-Middle (MITM) 81
Nuisances 81
Pop-up Ads 82
Spam 82
Tracking 82
Tools for Reducing Risk 83
Hardware 83
Router 83
Hard Disk with Built-in Encryption 88
Encrypting Individual Files 88
Privacy Filters 89
Cable Lock 90
Software 90
Firewall Software 90
Anti-Virus Software 91
Anti-Spyware 92
Anti-Phishing 93

Pop-Up Blockers 95
Ad Blockers 97
Anti-Spam 98
A Few Final Words of Advice 100

Text and HTML Editors 103
Text Editors 103
HTML Editors 112
Saving HTML Files from the Internet 114

Utilities 115
Managing Graphics 115
Renaming Files 116
Searching Content 117
Compressing Files 125
Cracking Passwords 128
Converting Measurements 130
Counting Words 133
Time Tracking 137
Managing the Clipboard 140
Taking Screenshots 144
Merging Files 147
Dealing with Help Systems 147
WinHelp 148
HTMLHelp 151
Installing Many Utilities At Once 155

Office Suites 157
Release Versions of Microsoft Office 158
Compatibility 160
Working with Office 161
Working with Word 161
Working with Excel 177
Working with PowerPoint 182
Office Documents With Other Embedded Office Documents 185

Computer-Assisted Translation Tools 189
Translation Environment Tools 192
What Are Translation Environment Tools (TEnTs)? 192
A Little Bit of History 193
Who Should Use Translation Environment Tools? 194
Categories of Translation Environment Tools 195
Tools That Use MS Word as Their Main Translation Interface 196

Wordfast Classic 196
SDL MultiTrans 198
Translation Workspace 199
Tools That Use an Independent Desktop-Based Translation Interface 201
Trados Studio 202
Déjà Vu 206
Star Transit 207
memoQ 209
Across 212
Alchemy Publisher 214
Wordfast Pro 215
Heartsome and Swordfish 216
Fluency 219
MadCap Lingo 221
Text United 222
OmegaT 225
CafeTran Espresso 228
Tools That Use an Online Interface 229
Wordbee 231
XTM Cloud 233
Wordfast Anywhere 235
Memsource 236
Smartcat 240
MateCat 241
Termsoup 245
Lilt 247
Comparing TEnTs 249
Translation Memory Handling 249
. . . . A Word of Caution About Alignment 250
. . . Back to Translation Memories 257
Terminology Handling 262
Work Environment 268
Integration of Machine Translation 271
Quality Assurance Features 279
Quality Assessment 282
Collaboration Features 284
Collaboration Through Exchange Formats 285
Real-Time Collaboration 287
Support of File Formats 289
And How Do I Make a Decision? 290
Standalone Quality Assurance Tools 291
Automatically Tracking Quality Through Editing 296

Source Document Quality Assurance 297
Translation Memory Quality Maintenance 301
Terminology Mining 303
Standalone Terminology Tools 306
Software Localization Tools 313
Binary Software Localization 313
Localization of Text-Based Files 317
Collaborative Localization Platforms 320
Website Localization 321
Management Tools 325

Translating Complex File Formats 331
Desktop Publishing Formats 331
Content-Oriented DTP Programs 333
Preparing FrameMaker Files 334
Preparing Ventura Files 334
Design-Oriented DTP Programs 335
InDesign 336
Alternative Ways of Working with InDesign CS4 and Later Files 337
Processing InDesign Files with MemoQs's Language Terminal 337
PageMaker 338
QuarkXPress 340
Mac vs. Windows for DTP Programs 341
Graphic Formats 342
Pixel-Based Formats 343
Working with Source Files 344
Working with OCR'ed Graphic Files 345
Vector-Based Formats 347
Tagged Formats 349
XML Files with Embedded HTML 353

PDF: Pretty Darn Frustrating? 359
PDF File Types 359
PDF Tools 360
Editing and Commenting on PDF Files 361
Commenting Tools 361
Managing Comments 362
Other Editing Options 363
Enabling Extended Features for Adobe Reader 364
Converting PDF Files to Editable Text 365
Adobe Reader 365
Adobe Acrobat 366

Using Word 2013 and above to convert PDF Files 367
Using Apache OpenOffice/LibreOffice to Work with PDF Files 368
OCR (Optical Character Recognition) Tools for PDF Conversion 368
Post-editing Converted Files in Word/Excel 371
Translating PDFs Inside the PDF (Sort Of) 372
Translating PDF Files Using Translation Environment Tools 373
Using OCR Features for PDF Conversion in Translation Environment Tools 378
In summary . . . 380
Creating PDF Files 380
Creating Translation Resources from PDF Files 381

Multimedia and Audiovisual Translation 383
Some Definitions and Clarifications 385
Codecs and Media Players 387
K-Lite Codec Pack 387
VLC Media Player 387
Media Player Classic 388
AVS4YOU Media Player 388
To run videos with a separate subtitle file ("soft-coding") 388
Other options and settings with VLC and MPC 390
DVD Ripping, Audio/Video Conversion, and Embedding Subtitles ("hard-coding") 395
AVS4YOU Video Converter 396
VLC Media Player 400
Xvid4PSP 5.0 403
Format Factory 406
Any Video Converter 408
Subtitling Tools 410
Subtitle Workshop 410
Aegisub 414
FluencyTranscription 419
Amara 420
SubtitleNEXT 424
Translating Subtitle Files with Translation Environment Tools 424
Sound Editing 426
Audacity 426

Voice Recognition Applications 429

Support 435

Conclusion 437

Index 439

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.

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 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!

Who the Robot with St. Jerome's Face on the Cover Is

Glad—and yet surprised—you asked. As most of you know, of course, this is Jeromobot, my little friend and patron saint. I adopted his image to unite our passion for language and the art of translation—St. Jerome, though apparently unfit for most human relationships, was *very* passionate about languages and felt *very* strongly about his grand translation of the Christian scripture into common Latin—with the new era that we live in. This era requires us to harness Jerome's intensity and excellence to the power of modern technology. May Jeromobot be with you and with us all!

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. Since the release of *Windows 2000* Microsoft has made *Windows* into 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.



Aside from the subsection that immediately follows, I will be discussing only Windows operating systems from Windows 7 onward.

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 no reboots after application crashes. While this is an improvement for all users, *Windows 2000*, *XP, Vista*,

7, 8 and *10*'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, etc), 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 166.

Multilingual display on a webpage thanks to Unicode:

100 000+

العربية • Български • Català • Čeština • Dansk • Eesti • Ελληνικά • Esperanto • Euskara • فارسی • Galego • 한국어 • Հայերեն • हिन्दी • Hrvatski • Bahasa Indonesia • עברית • Latina • Lietuvių • Magyar • Bahasa Melayu • Bahaso Minangkabau • 日本語 • Norsk (bokmål • nynorsk) • Нохчийн • O'zbekcha / Ўзбекча • Português • Қазақша / Qazaqşa / قازاقشا • Română • Simple English • Slovenčina • Slovenščina • Српски / Srpski • Srpskohrvatski / Српскохрватски • Suomi • Türkçe • Українська • Volapük • 中文

10 000+

Afrikaans • Alemannisch • ᨅᨔᨗ • Aragonés • Asturianu • Azərbaycanca • বাংলা • Bân-lâm-gú / Hō-ló-oē • Basa Banyumasan • Башкортса • Беларуская (Акадэмічная • Тарашкевіца) • বিষ্ণুপ্রিয়া মণিপুরী • Boarisch • Bosanski • Brezhoneg • Чăвашла • Cymraeg • Føroyskt • Frysk • Gaeilge • Gàidhlig • ગુજરાતી • Ido • Interlingua • Ирон æвзаг • Íslenska • Basa Jawa • ಕನ್ನಡ • ქართული • Kreyòl Ayisyen • Kurdî / كوردی • Кыргызча • Кырык Мары • Latviešu • Lëtzebuergesch • Lumbaart • Македонски • Malagasy • മലയാളം • मराठी • مصرى • مازرونی • Монгол • ဗြိုဟ်ထောဝ်ထော • Nāhuatlahtōlli • नेपाल भाषा • नेपाली • Nnapulitano • Occitan • ਪੰਜਾਬੀ • Piemontèis • Plattdüütsch • Runa Simi • संस्कृतम् • Саха Тыла • ئۇيغۇرچە / Uyghur • سنڌي • Scots • Shqip • Sicilianu • සිංහල • کوردی • Basa Sunda • Kiswahili • தமிழ் • Татарча / Tatarça • తెలుగు • Тоҷикӣ • ภาษาไทย • ትግርኛ / Basa Ugi • اردو • Vèneto • Walon • יידיש • Yorùbá • 粵語 • Žemaitėška

1 000+

Bahsa Acèh • Адыгэбзэ • Armăneashce • Arpitan • አማርኛ • Avañe'ẽ • Авар • Aymar • Ænglisc • Bahasa Banjar • भोजपुरी • Bikol Central • ཇོང་ཁ • Буряад • Chavacano de Zamboanga • Corsu • Cuengh • Deitsch • ᏣᎳᎩ • Diné Bizaad • Dolnoserbski • Emiliàn–Rumagnòl • Эрзянь • Estremeñu • Fiji Hindi • Furlan • Gaelg • Gagauz • گیلکی • 贛語 • Hak-kâ-fa / 客家話 • Хальмг • Hausa / هَوُسَا • ʻŌlelo Hawaiʻi • Hornjoserbsce • Igbo • Ilokano • Interlingue • Kalaallisut • Kapampangan • Kaszëbsczi • Kernewek / Karnuack • ಕನ್ನಡ • Kinyarwanda • Коми • Dzhudezmo / לאדינו • Лакку • ພາສາລາວ • Лезги • Lëgure • Limburgs • Lingála • Iojban • لوری • Malti • 文言 • Reo Māʻohi • Māori • მარგალური • Ming-dĕng-ngṳ̄ • Mirandés • Мокшень • Dorerin Naoero • Nedersaksisch • Nordfriisk • Nouormand / Normaund • Novial • Олык Марий • ଓଡ଼ିଆ • অসমীয়া • पालि • Pangasinán • Papiamentu • پښتو • Перем Коми • Pfälzisch • Picard • Къарачай–Малкъар • Qaraqalpaqsha • Qırımtatarca • Ripoarisch • Rumantsch • Русиньскый Язык • Sámegiella • Sardu • Seeltersk • Sesotho sa Leboa • ChiShona • සිංහල • Soomaaliga • Sranantongo • Taqbaylit • Tarandíne • Tetun • Tok Pisin • faka Tonga • تۆرکجه • Türkmençe • Тыва дыл • Удмурт • Uyghur / ئۇيغۇرچە • Võro • Vepsän • West-Vlams • Wolof • 吳語 • Zazaki • Zeêuws

100+

Akan • Aǹɔɣya • Bamanankan • Bislama • Chamoru • Chichewa • Ewegbe • Fulfulde • Gĩkũyũ • ᐃᓄᒃᑎᑐᑦ / Inuktitut • Iñupiak • كشميري • Kongo • कोंकणी / Konknni • Latgaļu • Luganda • मैथिली • Молдовеняска • Na Vosa Vaka-Viti • Norfuk / Pitkern • Afaan Oromoo • Полтацкá • ᱥᱟᱱ • Romani • Kirundi • Gagana Sāmoa • Sängö • Sesotho • Setswana • Словѣньскъ / ᐊᓂᔑᓇᐯᒧᐎᓐ ᐊᓂᔑᓇᐯᒪᔭᐍᐏᐣ • SiSwati • ትግርኛ • CWÝ • chiTumbuka • Xitsonga • Tsèhesenèstsestotse • Tshivenḓa • Twi • isiXhosa • isiZulu

There are multi-language versions of these operating systems that allow you to switch the user interface between 95 (*7*) and 111 (*10*) languages. The multilingual versions for *Windows 7, 8* and above (all versions).



You can find a list of all the supported languages for the different versions of Windows at support.microsoft.com/en-us/help/14236/language-packs.

To change the display language in *Windows 7,* select **Start> Control Panel> Regional and Language Options** and make a selection or install an additional language under **Display language**.

To change the display language in *Windows 8* and above, open the *Control Panel*, select **Language** and **Add a Language**.

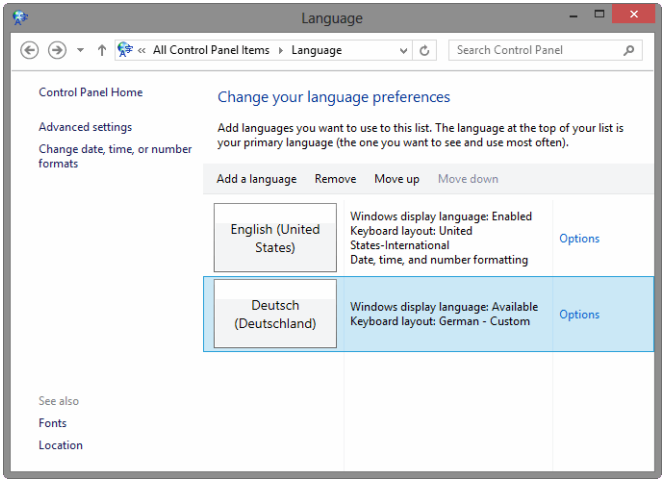


Figure 1: Adding a language in *Windows 8* and above

It's going to take a little while for the configuration to work when you install an additional language for the first time, but when it's done you can easily switch the user interface language by opening the Language dialog (see above), double-clicking on the language you want the user interface to be displayed in, and selecting **Make this the primary language**. You'll be prompted to log on again (no restart necessary!) and everything will be in the language of your choice.

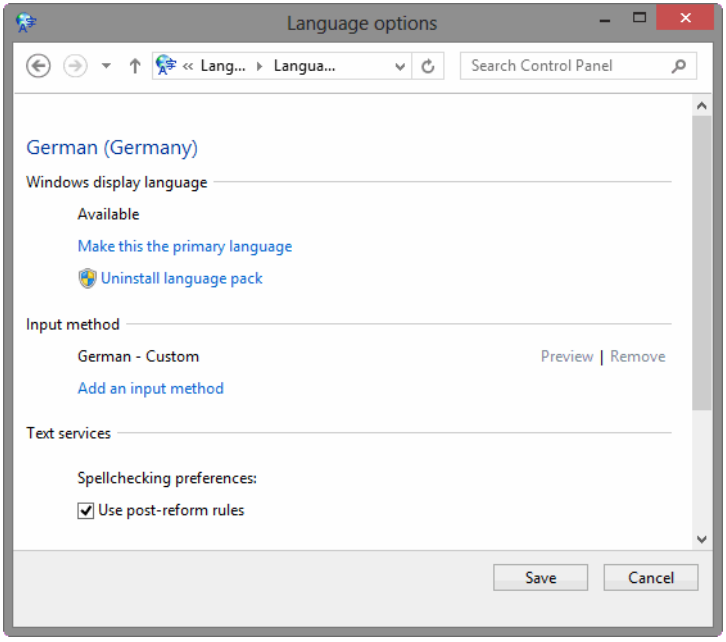


Figure 2: Setting the primary language in *Windows 8*

Windows 10 differentiates between your "display language" and the "preferred language." The "display language" is the language for all *Windows*-specific dialogs, and the "preferred language" is the primary language of your keyboard and your Windows "apps." To open or change the respective setting open "PC Settings" (press ⊞+I for that) and select **Time & Language> Region & Language**.

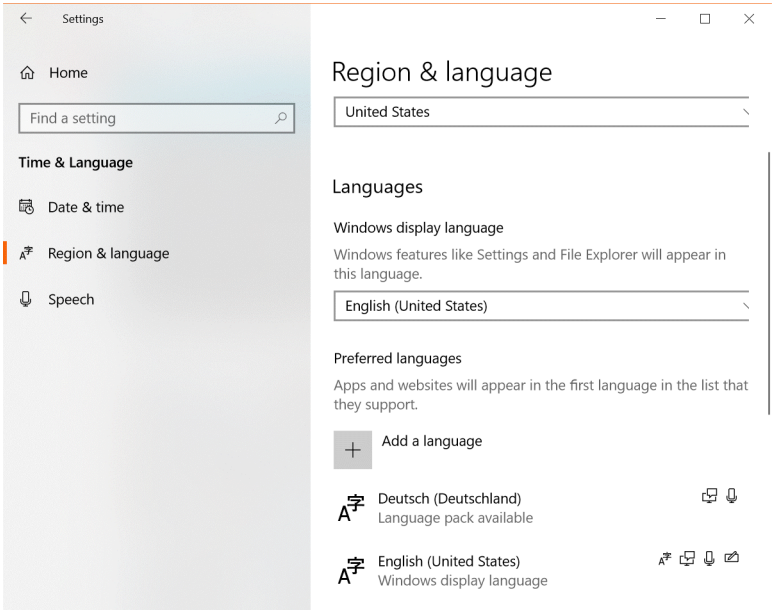


Figure 3: Setting the display language and preferred language in *Windows*



*You can change to the less task-oriented but more precise classic view of the Control Panel under **Start> Control Panel> Classic View**; or in Windows 8, **View by** and select **Small Icons**. (The instructions in this primer are all related to the Classic/Small Icons view of the Control Panel.)*

*In Windows 10, the Control Panel shares responsibilities with the **Settings** menu. Some settings, such as **Windows Update** and the icons for the **Notifications Area**, can be changed only through the **Settings** menu, whereas most other settings can be set in either the Control Panel or the **Settings** menu. The display language happens to be a setting that can be dealt with only through the Control Panel. To open the Control Panel you can right-click the **Windows** icon and select **Control Panel** or you can open the **Start** menu and type* `Control Panel`*.*

Switching to the Windows 8 or 10 Interface

In this book I will use "Windows 8" to refer to both *Windows 8* and *Windows 8.1* unless there are differences between the two.

It might initially be confusing to work in *Windows 8*, especially if you are used to working in earlier versions of *Windows*. If, however, you are like me and are not particularly interested in the "apps world" presented to you in the *Metro* interface, there are two things you can do to make things a lot easier, even to the point where you won't notice much of an external difference between *Windows 88* and earlier versions of *Windows* (aside from the positive developments of the system).

In the *Control Panel* in *Windows 8* (just type "`control panel`" in the metro interface to open it) you'll find the option **Taskbar and Navigation**. This will open the following dialog that will allow you to simply skip the *Metro* interface:

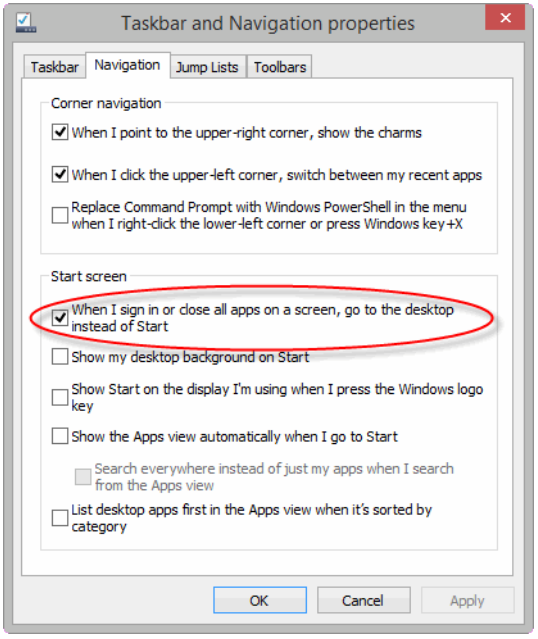


Figure 4: The **Navigation** tab in the **Taskbar and Navigation properties** dialog

You still won't have the familiar **Start** menu in *Windows*, and navigation becomes a little more difficult when you have only a desktop with no apparent way to start or control programs. Here I would recommend installing the third-part tool *Classic Shell* (see classicshell.net). This will give you a number of options to rebuild the **Start** menu in your preferred way, including one like this:

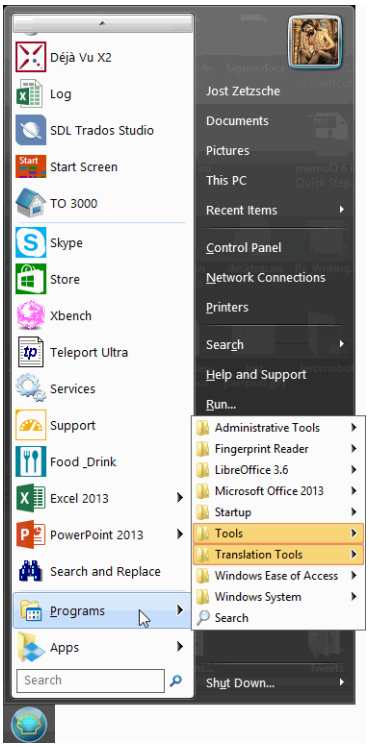


Figure 5: *Classic Shell* **Start** menu in *Windows 8*

There is one other power option available in *Windows 8* that gives you access to all kinds of things (including the ability to turn the computer off). For this you don't have to install anything extra. You just have to press ⊞+X or right-click in the lower right-hand of the desktop and you will see the **Power User menu** with access to all kinds of important places:

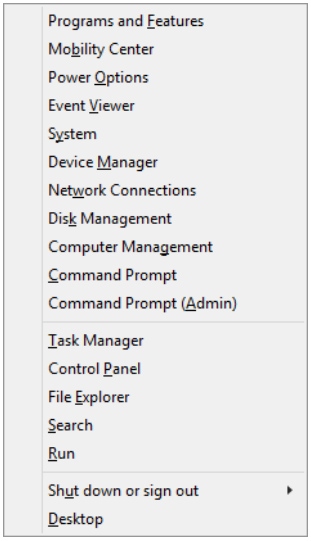


Figure 6: *Windows 8* Power User menu (this menu is identical in *Windows 10*)

Working with the *Windows 10* Interface

Should you upgrade to *Windows 10* as a *Windows 7* or *Windows 8* user? Not if you're very content with your existing situation. Overall, *Windows 10* is probably a smoother-working system than *Windows 8* (not necessarily in comparison to *Windows 7*), but there is always the danger of having to upgrade a program or two because of lacking compatibility with the new operating system, so you might be fine with where you are.

If you decide to upgrade (or purchase a new computer with the new operating system pre-installed), you will mostly like what you see. The **Start** menu is reasonably functional again and there is no artificial separation between a Desktop and "Metro" interface.

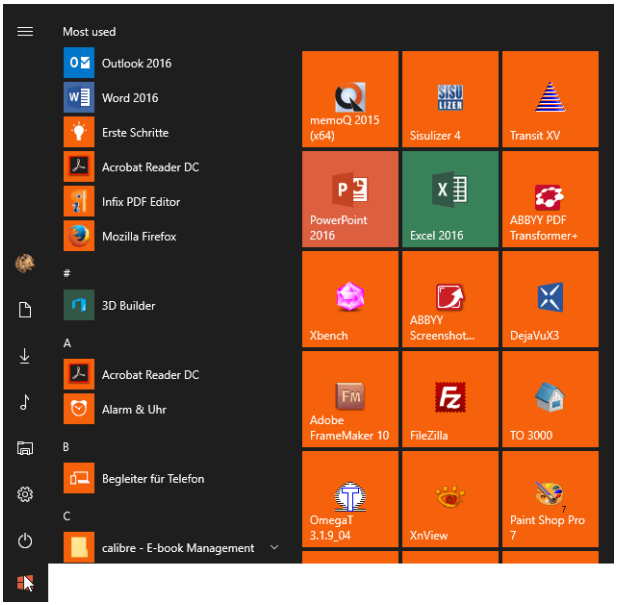


Figure 7: Customized *Windows 10* **Start** menu



If the Windows 10 ***Start*** *menu is not customizable and powerful enough for you, you can also install the third-party product Classic Shell (see page 11) or Start10 (see stardock.com/products/start10/).*

Windows/File Explorer

The most helpful and often-used *Windows* component on my computer is the *Windows Explorer* or *File Explorer* (*Windows 8* and above).



To quickly open the Windows/File Explorer, press the key combination ⊞ (the key with the Windows icon to the left of the spacebar on most keyboards)+E. Other helpful keyboard shortcuts include CTRL+F1 to toggle the ribbon bar, ALT+P to turn the preview pane on or off, CTRL+SHIFT+N to create a new folder, ALT+UP to move one folder up, and CTRL+N to open new instance with same path.

The *Windows*/*File 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.

Previewing Files

The preview feature allows you to automatically preview HTML, *MS Office*, PDF, graphics, and text files within the *Explorer* (you need to enable this by selecting **Organize> Layout> Preview Pane**, by pressing the Preview button on the toolbar of the *Windows 7 Explorer,* or by selecting **Preview pane** on the **View** ribbon in *File Explorer* in *Windows 8* and above).

Folder Paths

An element that might be confusing at first but may become one of your favorite features is the 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.

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

.

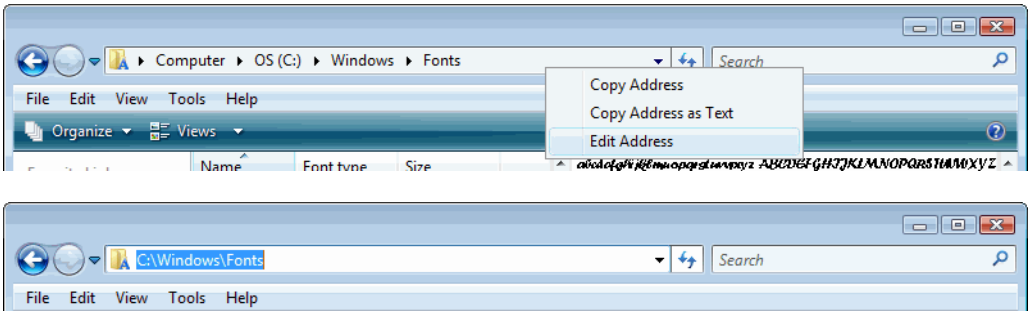


Figure 8: Switching the breadcrumb view to the traditional path view

Selecting Multiple Files

When you previously had to hold the CTRL(+SHIFT) key while selecting multiple files within *Windows Explorer*, *Windows 8* and above has made it easier to select several files or folders in the *File Explorer* at the same time to perform one action simultaneously (such as delete or copy). You can now select little check boxes to the left of the file name that appear as you select the file.

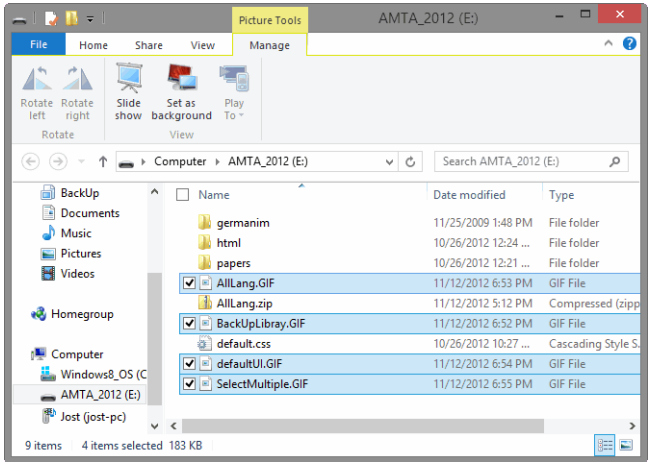


Figure 9: Selecting several files at once in *Windows 8* and above

The Ribbon Menu

Windows 8 and above has also added to the *File Explorer* the ribbon menu that most are familiar with from the last couple versions of *Microsoft Office*. While it might seem unwieldy at first, it adds to your productivity by often automatically displaying interactive ribbons that match your current selection (see the automatically selected **Picture Tools** ribbon because of my selection of graphic files in the image above).

Libraries in Windows

One feature that was introduced in *Windows 7* is the concept of libraries.

Libraries are like virtual folders. Of course this doesn't mean that normal computer folders are physical, but in the traditional computer world, they and only they "contain" the files that are stored in them. Libraries, on the other hand, are virtualizations of that. They can display the contents of other folders from all over your computer, other computers on the network or even a USB flash drive. A library is essentially an organizational principle that monitors other folders and provides a single "location" to work with all their contents.

Out of the box, *Windows 7* and above come with four libraries: *Documents*, *Music*, *Pictures* and *Videos* (as well as possibly *Camera Roll*, *3D Objects*, and *Saved Picture*), each with its obvious content. And again, while the references to those files are stored in the respective libraries, the actual files stay wherever you stored them on your computer.

There are plenty of ways you can use libraries to manage, manipulate, or organize files, but one is particularly helpful: backup.



Note that in Windows 8 and above, the File History backup system uses the preconfigured libraries as its center point of backups (see page 50).

To create a new library, simply right-click on the **Libraries** folder on the left-hand side of *Windows/File Explorer* (the *Navigation* bar) and select **New> Library**. Once you give your library a name and open it, you'll be prompted to add folders from any location you can access from your computer (except read-only media such as DVDs or CDs). Since you don't want your complete

system to be backed up every night, you can pick and choose the necessary folders and then schedule the library for your nightly backup (or you could simply right-click the library before you turn off your lights for the night and select **Send To>** `<your external hard drive>` or whatever you prefer as a backup device).

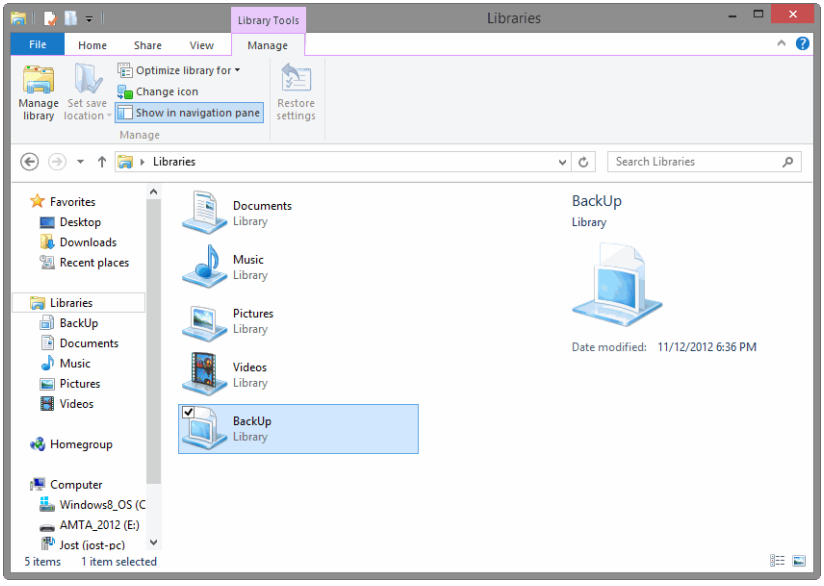


Figure 10: Libraries in *Windows 8* and above

Helpful Shortcuts

Sending 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**.

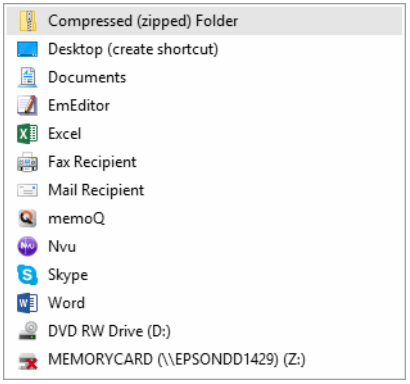


Figure 11: Right-click access to a variety of options

These shortcuts are stored in the `SendTo` folder (under `Users/<user>/AppData/Roaming/Microsoft/Windows`). 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.



*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, ⊞+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.*

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.

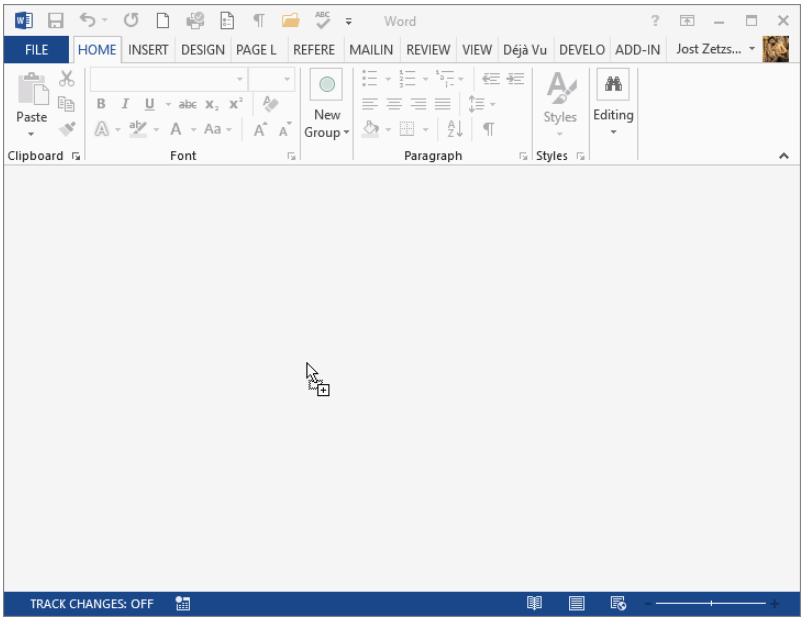


Figure 12: Opening a file by dragging it into an open application

If you are interested in opening more than one application if you want to start work on a specific project (such as a browser, a voice recognition program, and a translation tool), all you need to do is open *Notepad* or another text editor and type something like this:

```
start /d "C:\Program Files\Mozilla Firefox\" firefox.exe
start /d "C:\Program Files\Nuance\NaturallySpeaking10\Program\"
natspeak.exe
start /d "C:\Program Files\ApSIC\Xbench\" XBench.exe
```

Once you entered the text, save the file as a *.bat file. The BAT extension tells *Windows* that this is a batch file that contains a stack of commands that it needs to execute when the file is opened. To open it you'll only need to double-click on it and (in this case) *Firefox*, *Dragon* and *XBench* are

automatically started. Clearly you are probably interested in other programs, so all you need to do is replace the path to the program within the quotation marks and enter the file name of the program afterward. Save it as a *.bat file and you're ready to start the day.

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`.

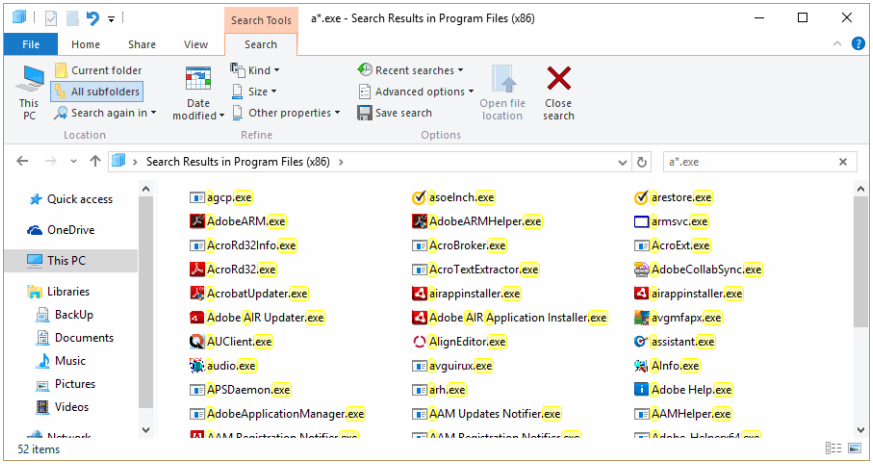


Figure 13: Searching with wildcards in *File Explorer*

The combination of this search feature and the method to open files (see page 18) 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.

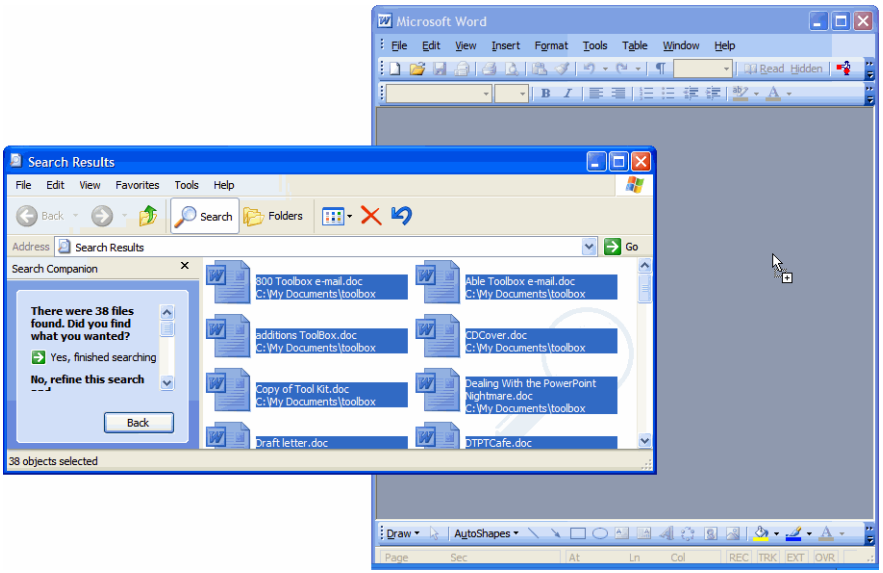


Figure 14: Opening multiple files simultaneously



It's 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* ***Control Panel> Indexing Options****.*

There are also many third-party search tools available. One of the more outstanding ones has to be *Everything* (see 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 any folder or folder group in *Windows Explorer*—just right-click and select **Search Everything**.

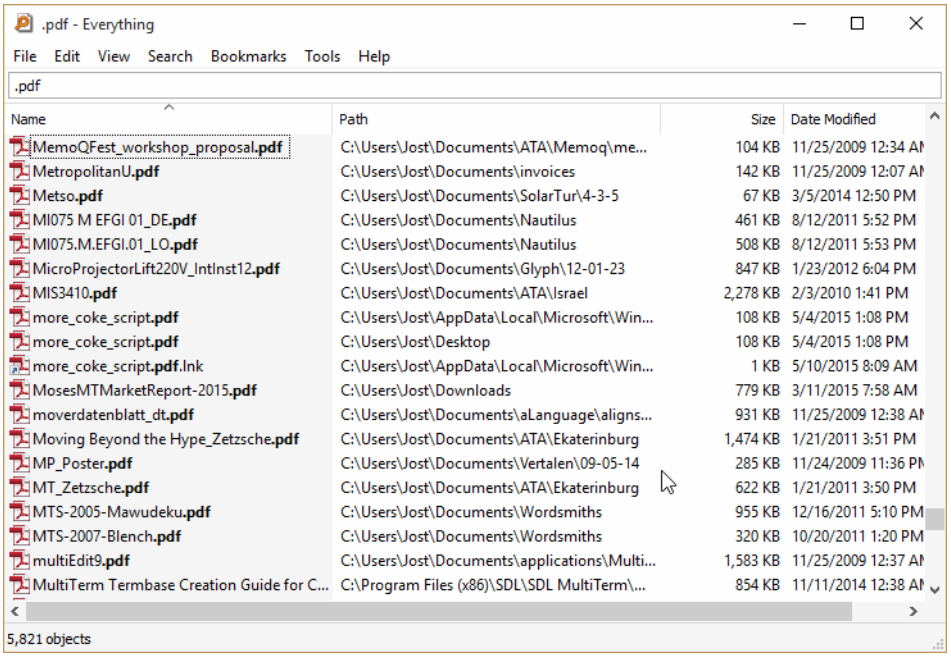


Figure 15: *Everything's* main window with a filter for PDF files

To Copy Files or Folders

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

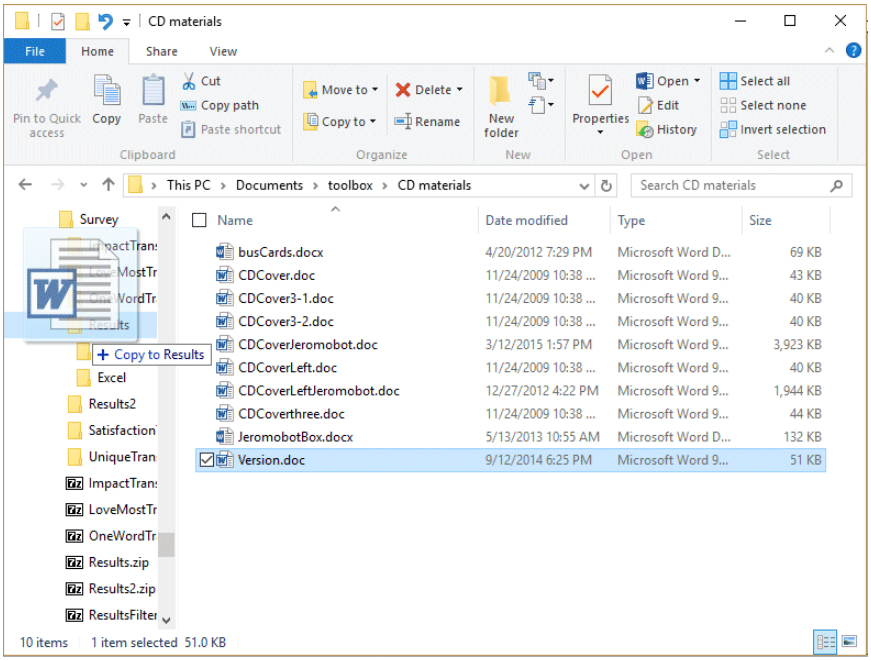


Figure 16: Dragging and copying a file in the *Windows/File Explorer*

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.

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:

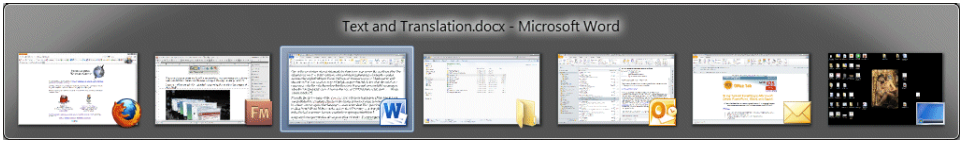


Figure 17: *Windows Task Switcher*

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

Windows 7 (but not *Windows 8* and above) also offers the *3D Flip*, activated by pressing the ⊞+TAB instead of ALT+TAB.

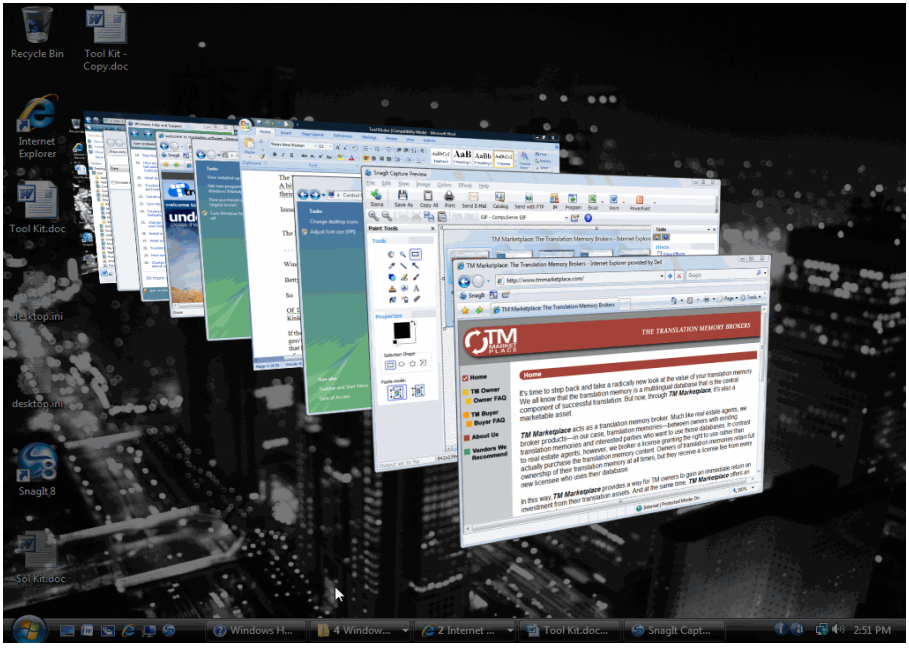


Figure 18: *Windows 7's* 3D Flip

Windows 10 re-instituted the ⊞+TAB keyboard shortcut again, but it leads you to an entirely new concept: the *Task View*—the same place you get to by clicking on the **Task View** button that is displayed by default on the *Windows 10* taskbar.



Figure 19: The Task View button



*Since there really is no good reason for the button to take away any space once you know the shortcut, you can just as well disable it. To do that, right-click on the taskbar and unselect **Show Task View button**.*

The *Task View* is like the *Task Switcher* except that it continues to display after you release the key combination (and will do so until you select one of the applications to view) and it offers access to open a "virtual desktop."

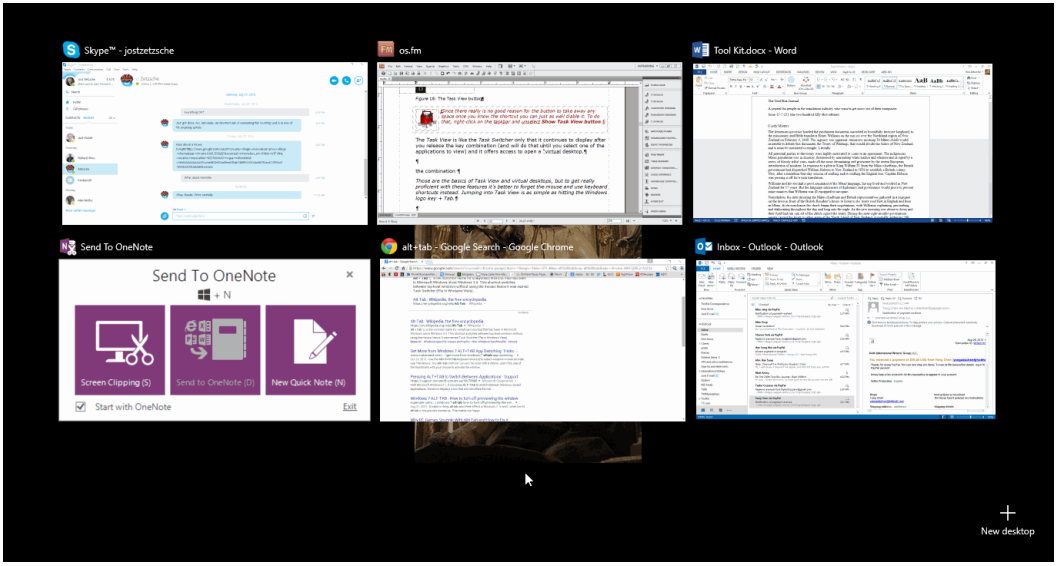


Figure 20: The *Task View* with **New Desktop** link



The Task View also shows up on half of the screen if you "send" the current application to the left or the right side of your computer screen by pressing ⊞+LEFTARROW/RIGHTARROW.This allows you to quickly build up a two-sided view on one screen.

Working with the Virtual Desktop

The virtual desktop is a concept that is very familiar to users of *Apple* and *Linux* computers but was unknown to most *Windows* users before *Windows 10*. Essentially, it's a way to organize your work into different areas without making the current desktop more crowded.

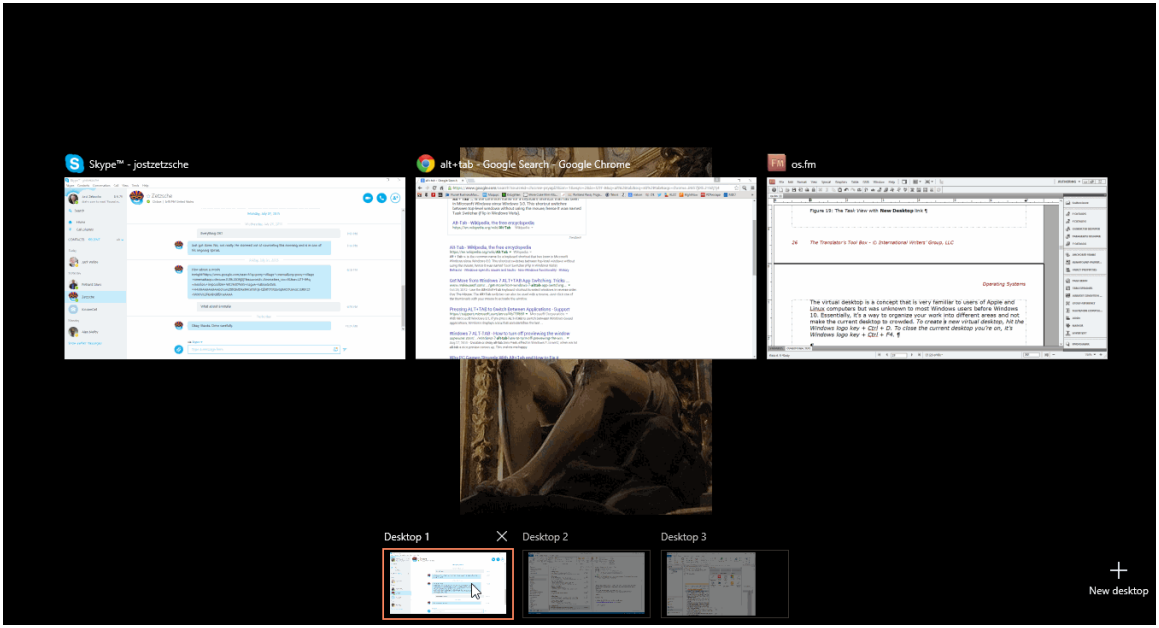


Figure 21: *Task View* with three open desktops

In the example above I have one desktop with some authoring tools, one with a music player (sound is transferred from any desktop, even if it's not the active one) and email (so I don't get constantly interrupted), and a third desktop with an open project in a translation environment tool that I'm presently pausing work on.

If I close any of the desktops without closing their application(s), they automatically are transferred to **Desktop 1**.

To create a new desktop, you can click the link on the *Task View* or press ⊞+CTRL+D. To close a desktop, you can again use the *Task View* or you can press ⊞+CTRL+F4 to close the currently active desktop.

You can jump between the desktops by pressing ⊞+CTRL+LEFTARROW *or* ⊞+CTRL+RIGHTARROW*.*

Working with Jump Lists

The **Jump List** feature allows you to right-click on any icon in the taskbar and select from a number of options, including, and most helpfully, the most recently opened instances of documents with that particular program.

This also works with *Windows/File Explorer*. However, since you typically use *Windows/File Explorer* so frequently throughout the day and there is space only for the last seven visited locations, chances are you won't see the place you need to go to once a day to, say, make a backup of your current project at day's end.

The good thing is there is a special feature for *Windows/File Explorer*: You can manually bookmark some favorites to the top of the **Recent** list by pinning folder locations. Just click on any folder and drag that folder icon to the **Explorer** shortcut on the taskbar. You'll see the message **Pin to Windows/ File Explorer** before you release the mouse button. The folder will now appear under a **Pinned** section of the **Jump List**, and you can remove it by clicking the Unpin from this list icon on the right side of the panel.

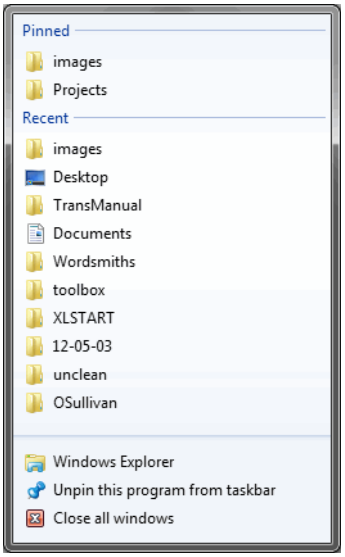


Figure 22: Pinning locations

And because it's so much fun to do this in *Windows/File Explorer*, you can also do it for most browsers or other applications so that you have web pages or documents pinned down in their Jump Lists, allowing you to open those without opening the program first.

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: support.microsoft.com/help/12445/windows-keyboard-shortcuts.

WinKey Shortcuts

One often overlooked set of shortcuts are those associated with the WINKEY (⊞), the key that is typically located on the lower right of the Windows keyboard and displays the Windows logo. Gamers don't like this key because it tends to interfere with their activities, but I really like it because it provides access to a number of features that otherwise require the mouse.

Here is a list of the most important key combinations:

- ⊞: Open the **Start** menu (in *Windows 8* it switches to the previous mode)
- ⊞+E: Display *Windows/File Explorer*
- ⊞+R: Display the **Run** dialog box
- ⊞+M: Display the desktop (additional key in *Windows 8* and above to display the desktop: ⊞+D)
- ⊞+LEFT (RIGHT): Snap your current window to the left (right) (and—in Windows 10—have the Task View displayed on the other half of the screen)
- ⊞+T: Focus on the first and then succeeding taskbar entries (⊞+SHIFT+T cycles backward)

- ⊞+SPACE: Peek at the desktop (in *Windows 8* and above: allows you to switch between the different keyboards you might have installed)
- ⊞+P: External display options
- ⊞+X: Open **Mobility Center** (access to things like brightness, volume, battery, wireless connectivity, external display, etc.) (in *Windows 8* and above: access to the Power User menu—see page 12)
- ⊞+NUMBER KEY: Launch a new instance of the application in the nth slot on the taskbar
- ⊞ + +: Zoom in to 200% (⊞ + - goes back to 100%)
- ⊞+PRINTSCREEN: Save an image of your current screen in the *Screenshots* folder under *Pictures* (in *Windows 8* and *10*)
- ⊞+I: Open the *Settings* app (in *Windows 10*)
- ⊞+G: Open the *Xbox Game bar* to let you record (in video mode) what you do on your screen (in *Windows 10*)
- ⊞+TAB: Open the *Task View* where you can select one of the open applications and create a new desktop (in *Windows 10*)
- ⊞+CTRL+D: Create an additional desktop ("virtual" desktop) without first opening the Task View (in *Windows 10*)
- ⊞+CTRL+F4: Close the current "virtual" desktop—any open applications are automatically transferred to Desktop 1 (in *Windows 10*)
- ⊞+CTRL+LEFTARROW / RIGHTARROW: Switch between open "virtual" desktops (in *Windows 10*)
- ⊞+V: Opens the clipboard history (in *Windows 10*)
- ⊞+. or ; : Opens the emoji panel when typing (in *Windows 10*)
- ⊞+H: Opens the dictation toolbar (in *Windows 10*)
- ⊞+SHIFT+S: Activates screenshot app (in *Windows 10*)

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), 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.

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 117.

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 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** (in Windows 10, open the Settings app by pressing ⊞+I and selecting **System> Notification and Actions**). Here you can set the behavior of each of the icons that are presently displayed or have been displayed in the past.*

There are several 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.*

In Windows 8 and above you can access the Startup folder by pressing ⊞+R and entering `shell:startup`.

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

The *System Configuration* utility will appear.

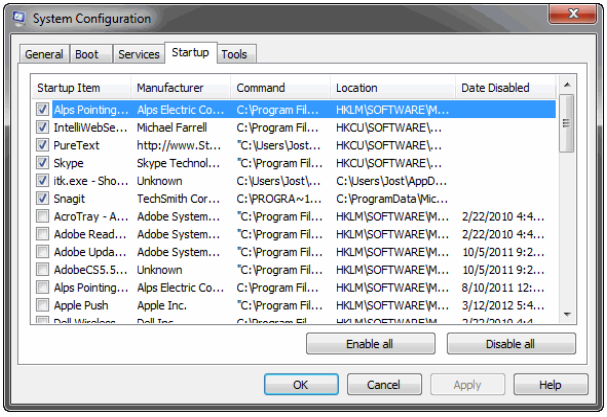


Figure 23: System Configuration Utility

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. You can find one of those lists at pacs-portal.co.uk/startup_search.php.



*In Windows 8 and above you will find a link on the **Startup** tab to the Task Manager from which you can administer the programs that are automatically launched.*

You can also directly open the Task Manager by pressing ⊞+R and entering `taskmgr`.

Many programs are run as so-called services. These are listed on the **Services** tab of the *Task Manager* from where you can stop or start them.

For a better description of each of the services and the ability to decide whether services should be started manually or automatically, or should be disabled, you can open the **Services** dialog under **(Start>) Control Panel> Administrative Tools> Services**, or in *Windows 8* and above you can press ⊞+R and type `services.msc`.

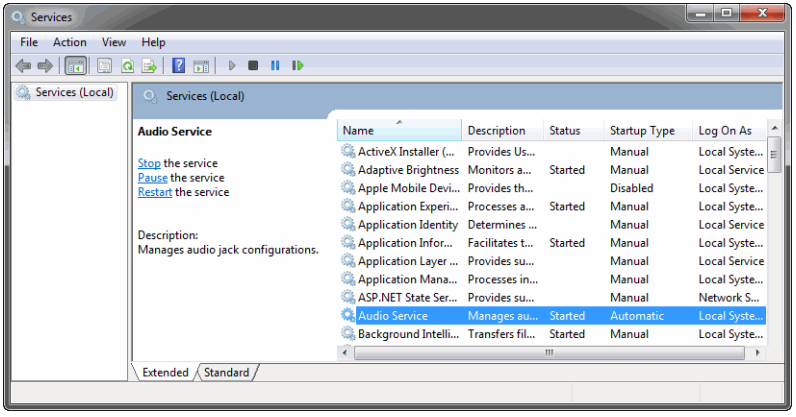


Figure 24: **Services** dialog

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

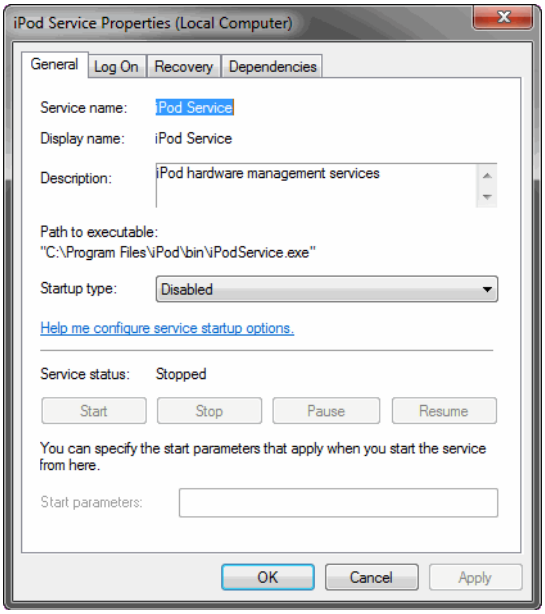


Figure 25: **Services Properties** dialog

Avoiding Expensive Visual Effects

Your computer will need extra resources for the *Aero* interface in *Windows 7*, 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**.

Windows 8 has neither the Aero interface nor the Transparency feature.

In *Windows 10* you can disable the default transparency under **Settings** (⊞+I) **>Personalization> Colors> Make Start, taskbar and action center transparent**. You can also disable some other animated features in *Windows 10* under **Settings** (⊞+I)**> Ease of Access> Other Options> Play animations in Windows**.

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 **Control Panel> Programs and Features**.

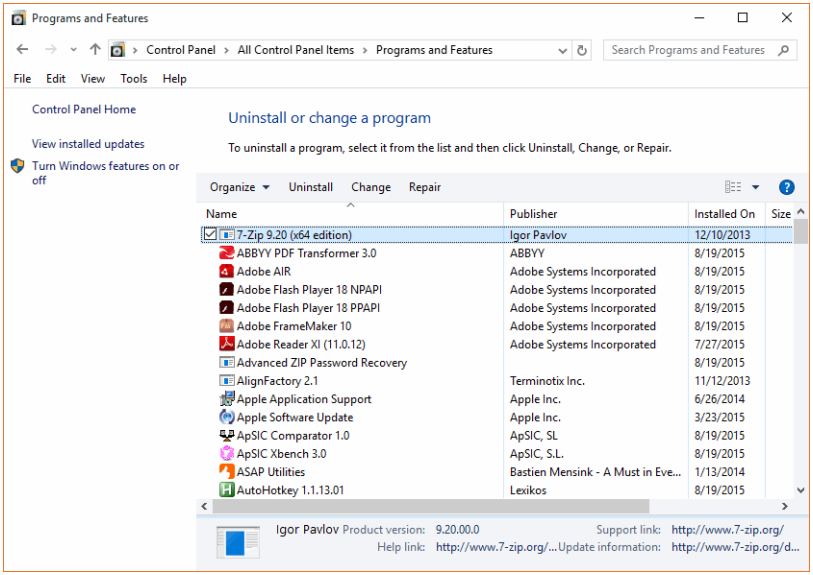


Figure 26: Removing programs

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` or `C:\ProgramData`) and delete them manually in *Windows/File Explorer*.

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 ⊞+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.

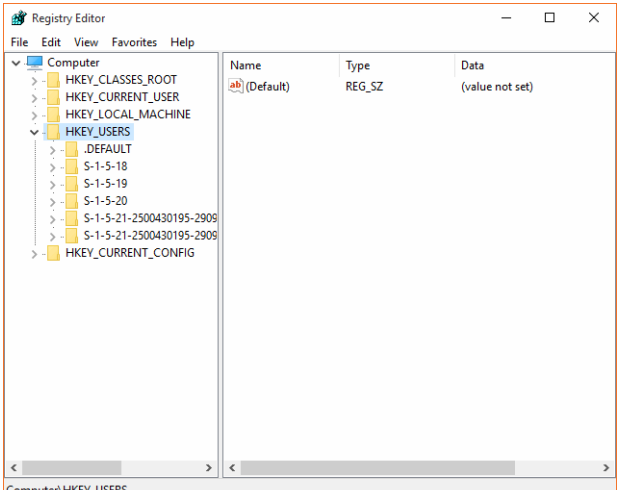


Figure 27: **Registry Editor**

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

A widely used tool for registry maintenance is *CCleaner* (see ccleaner.com). You can see that even this tool recommends backing up obsolete registry (settings) before deleting them:

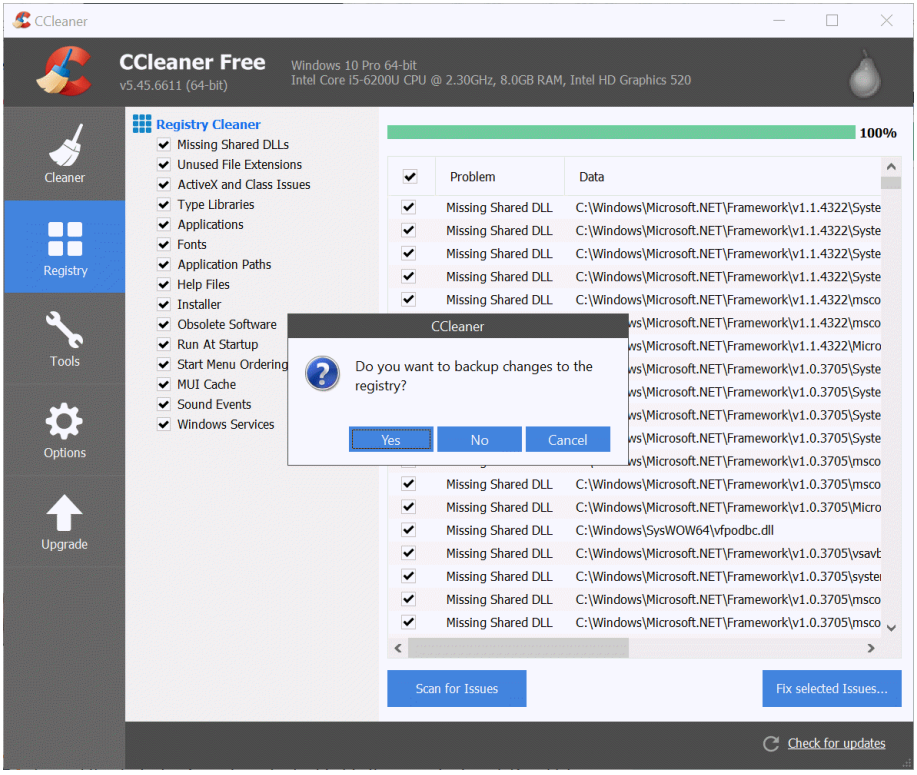


Figure 28: Cleaning the registry with *CCleaner*

Disk Cleanup

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

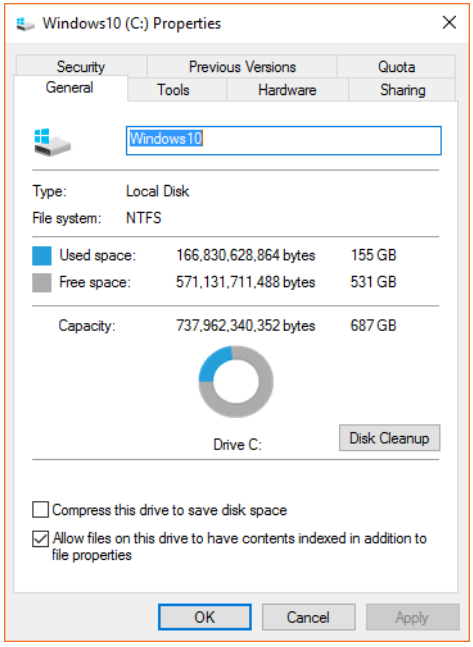


Figure 29: Accessing *Disk Cleanup*

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.

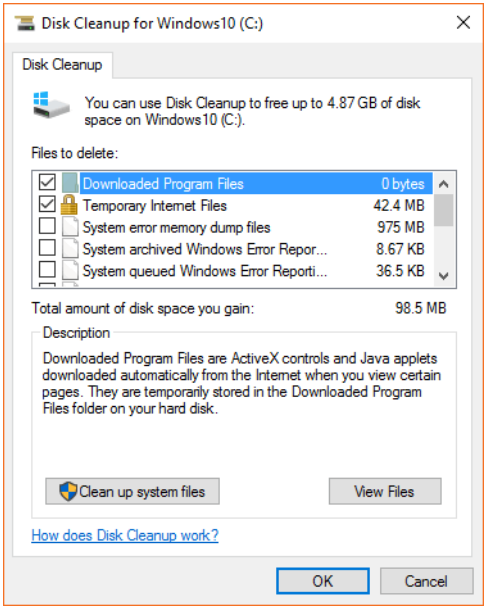


Figure 30: Selecting files for deletion



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.

Starting with the release of *Windows 8.1* you can now also delete copies of old Windows Update files from your hard disk. To do that you will have to select the **Clean up system files** button in the **Disk Cleanup** utility.

Deleting your temporary Internet files will only delete temporary files that have been collected with the *Internet Explorer* and *Microsoft Edge*. 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 and website data)** (*Internet Explorer*)
- **Settings> Clear browsing data> Cookies and saved website data** (*Microsoft Edge*)
- **Tools> Options> Privacy & Security > Cookies and Site Data** (*Firefox*)
- **Settings> Advanced > Privacy and security> Clear browsing data** (*Google Chrome*)
- **Opera> Settings> Privacy & security> Clear browsing data> Delete cookies and other site data** (*Opera)*.

All this can also be done with a tool like the above-mentioned *CCleaner* (see page 37):

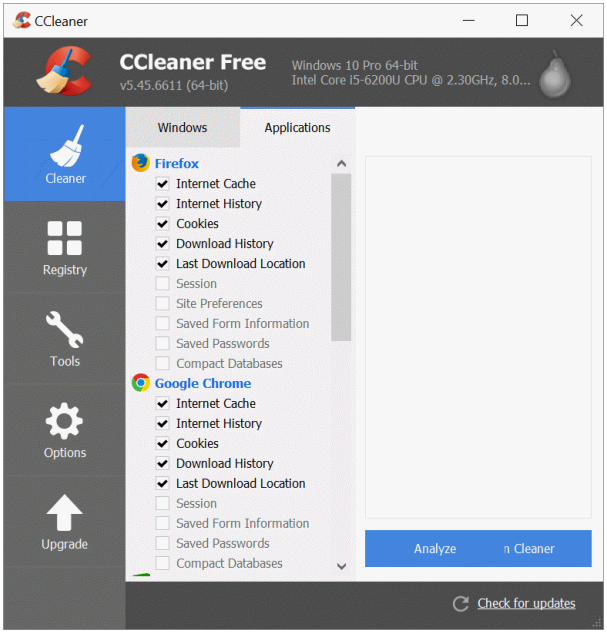


Figure 31: Cleaning browser history with *CCleaner*

Finding the Forgotten Space Hogs

But even after going through all the processes described above you might not have reclaimed the space on your hard drive that you had hoped for. The large media files that you downloaded or created years ago might be sitting somewhere taking up a lot of space that you could very well use in better ways.

There are a lot of programs out there to help you find those large perpetrators, but the one that I find very helpful is ancient by today's standards. *WinDirStat* (see windirstat.info) is an open-source tool—ergo free—and you don't have to take a training class to use it. It has a no-nonsense approach, runs even on the latest operating system, and once it's done analyzing your computer, which takes just a few minutes, it has all kinds of ways to show you where those bad space-invaders can be found, including a very psychedelically colored map.

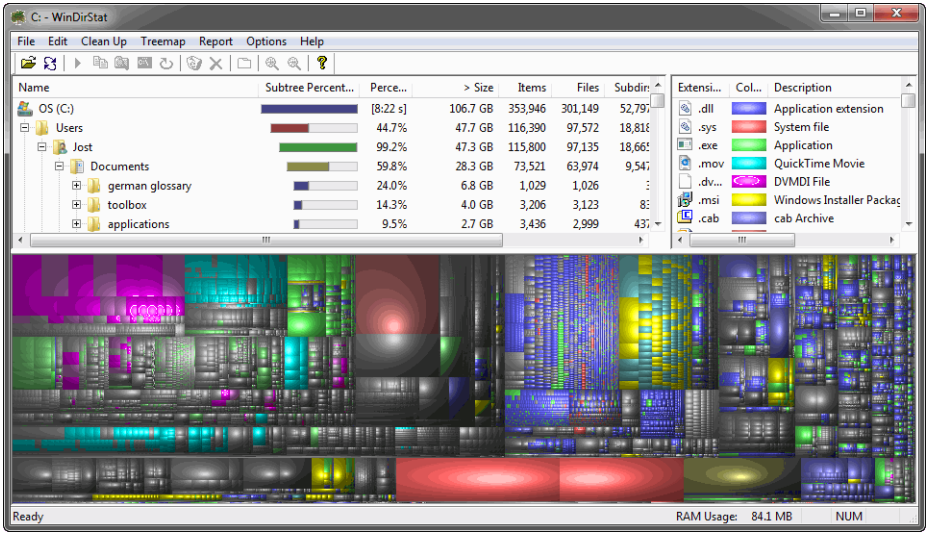


Figure 32: *WinDirStat's* results

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/File Explorer* or **My Computer** and selecting **Properties> Tools> Check** (**now**).

.

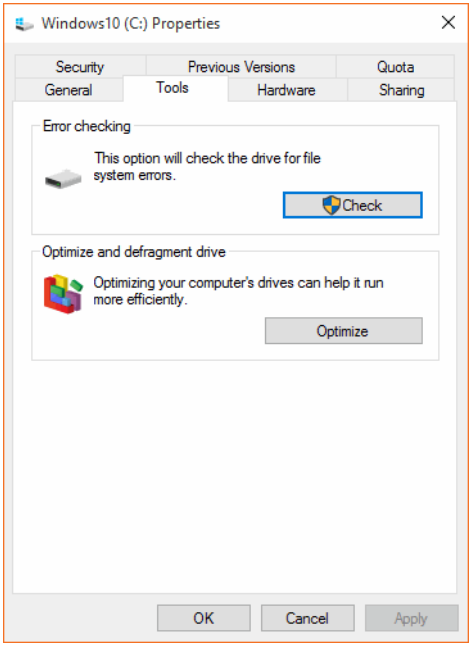


Figure 33: Error checking and defragmenting the hard drive

Once *ScanDisk* has successfully finished checking the drive for errors, you can start the defragmentation in the same dialog. 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).



Note that by default Windows schedules an automatic defragmentation once a week, unless you have a solid state drive. In that case defragmentation is not necessary.

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*.

In *Windows 8* and above, the procedure to boot into *Safe Mode* is slightly different. Open **Settings>** (**Change PC settings>**) **Update and recovery** (**security**)**> Recovery> Advanced startup** and then follow the prompts.

Restoring Your Computer

System Restore is a very helpful feature that 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 at regular intervals), *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** (in *Windows 8 and above*: **Control Panel> System> System Protection**(**> System Restore**)).

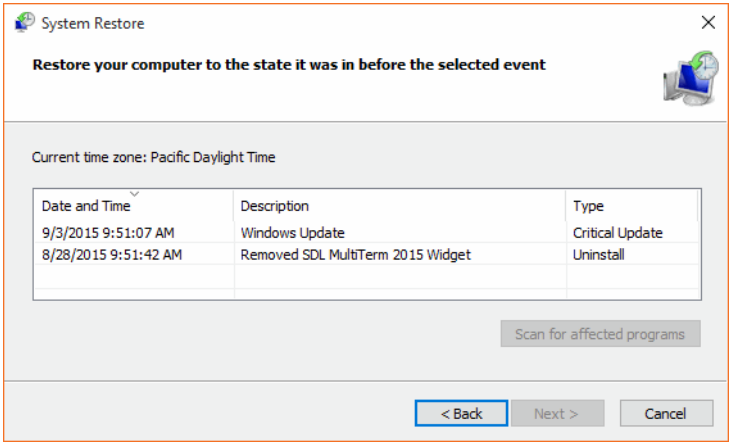


Figure 34: Selecting a restore point

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

In my life as a computer user I, like many of you, have gone through a number of phases when planning how to back up data. For me it started with the 3½-inch floppy disks, which were superseded by Iomega zip disks, followed by CD-Rs and DVD-Rs, then USB thumb drives and now a mixture of external drives and cloud-based systems.

Translators are using a variety of available systems, though they tend to differentiate between highly confidential files (which they back up on local storage options) and other files that they back up on cloud services. I'm sure that this differentiation will begin to vanish once we and our clients become less squirmish about using cloud-based services for translation and storage tasks.

Services like *Dropbox* (see dropbox.com) or *Carbonite* (see carbonite.com) or the *Windows*-internal *OneDrive* allow you to set up an automatic backup schedule so you don't have to worry about manual backups, and you can even access those files from any number of devices.



The only drawback I have encountered is slow upload times for very large files, such as translation memories or email folders.

As far as an automated *Windows*-internal system for offline-backups, one intuitive way to retrieve files is a feature called "shadowing" or—from *Windows 8* forward—"File History." This tool restores files to a previous version by "looking" into the backup location as well as the *System Restore* archives to see whether there is an older version of the file and allowing you to restore it to that version. All you need to do to access the feature is right-click on the file in question and select **Restore previous versions** or select the **History** icon on the **Home** ribbon tab in *File Explorer* (*Windows 8* and above).

Once you select the command, several earlier versions of the file might be displayed (*Windows 7*):

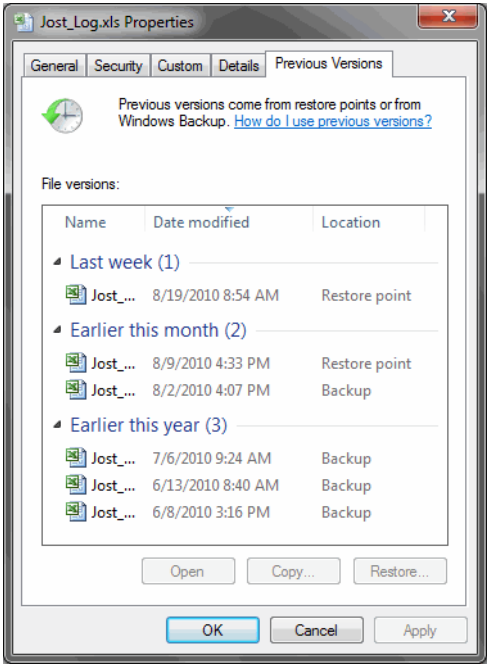
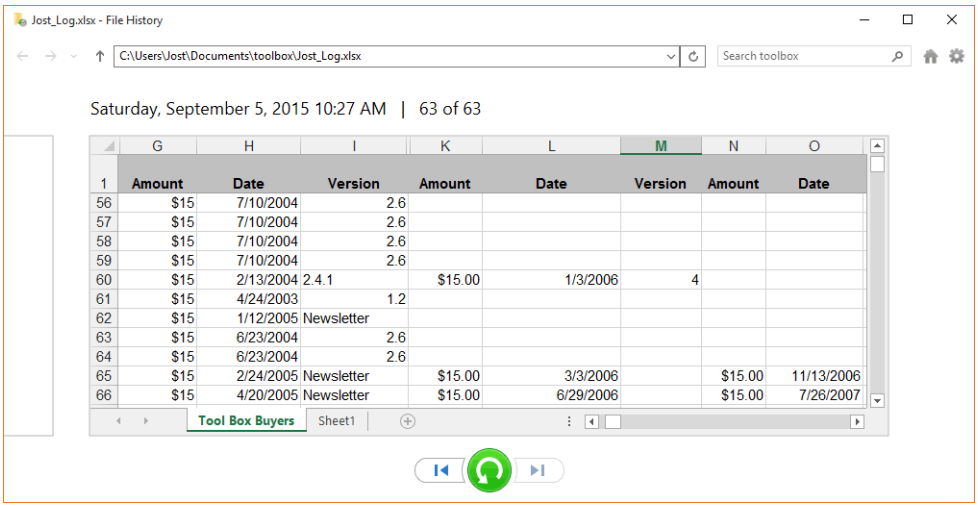


Figure 35: Previous versions of an *Excel* file in *Windows 7*

Or you can even directly view the file's content (works in graphic, text, PDF and office files) and then decide which version to restore (*Windows 8* and above).



	G	H	I	K	L	M	N	O
1	Amount	Date	Version	Amount	Date	Version	Amount	Date
56	$15	7/10/2004	2.6					
57	$15	7/10/2004	2.6					
58	$15	7/10/2004	2.6					
59	$15	7/10/2004	2.6					
60	$15	2/13/2004	2.4.1	$15.00	1/3/2006	4		
61	$15	4/24/2003	1.2					
62	$15	1/12/2005	Newsletter					
63	$15	6/23/2004	2.6					
64	$15	6/23/2004	2.6					
65	$15	2/24/2005	Newsletter	$15.00	3/3/2006		$15.00	11/13/2006
66	$15	4/20/2005	Newsletter	$15.00	6/29/2006		$15.00	7/26/2007

Figure 36: Previous versions of an *Excel* file in *Windows 10*

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



Once you do this, there is no undo: the current version of the file in question is gone and has been replaced. In almost all cases this will be fine. 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.*

The backup system in *Windows 7* was easy to set up and highly customizable, but oddly enough it was used by so few users (Microsoft says about 5%) that it was deprecated for *Windows 8* but, tada, fully rehabilitated in *Windows 10* (you can find it in the **Control Panel** under **Backup and Restore (Windows 7)**).



As you would expect there are a number of third-party products with a virtually unlimited number of options for backup purposes as well. Depending on your translation field, it might be advisable to look into these, and it certainly will not reflect badly if you mention this on your résumé.

File History in *Windows 8* and above

File History in *Windows 8* and above is a tool that allows you to go back an unlimited number of versions for each designated file (see page 45). This sounds good, but there are a few caveats: this feature is not enabled by default and not all files will be backed up.

To enable *File History*, open the **File History** item on the *Control Panel*, select the external backup device (which could be a USB stick or an external hard drive), and select **Turn on**. Under **Advanced settings** you can set up how often you want to run the backup, how much space it's going to occupy, and how long you want to keep the backed-up versions.

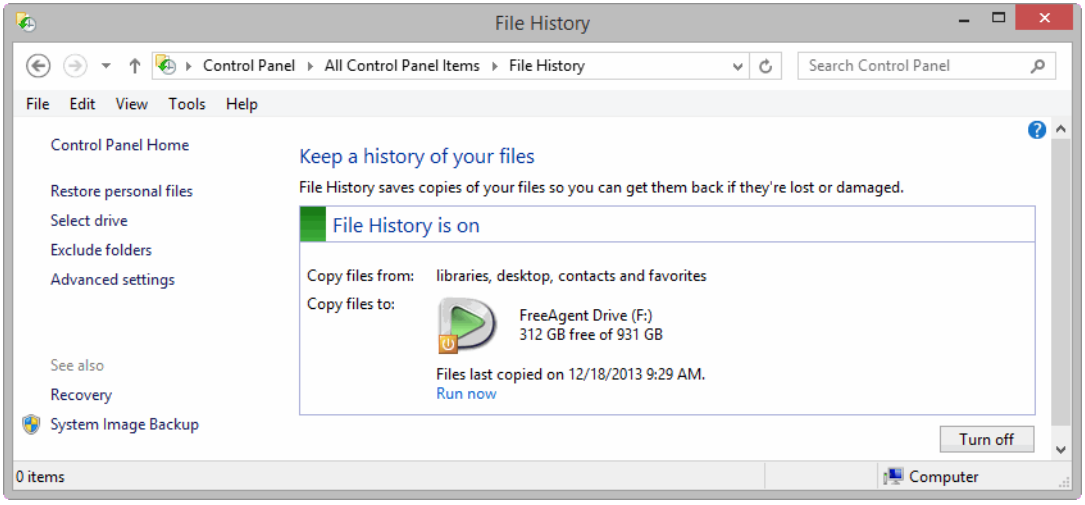
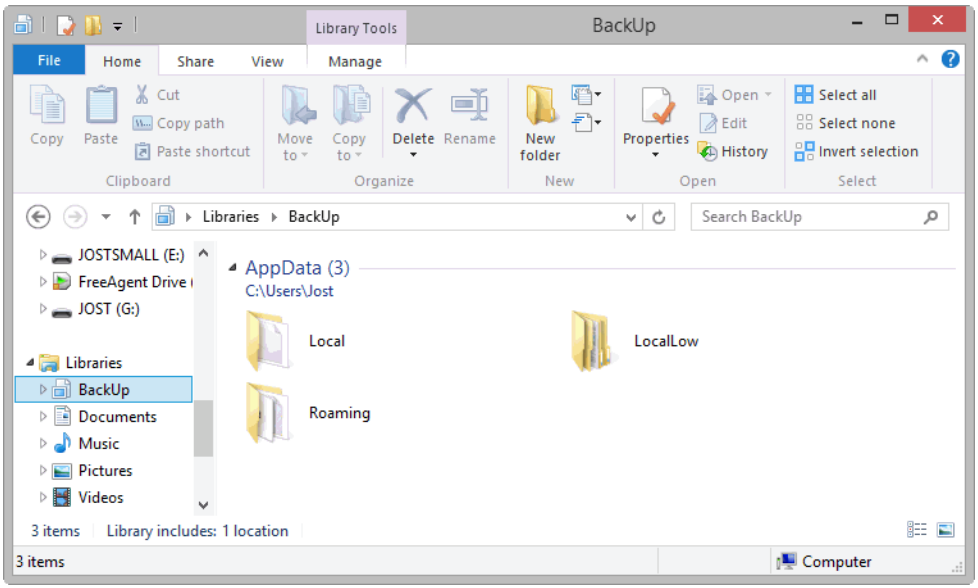


Figure 37: Controlling **File History** in *Windows 8 and above*

The only files that will be backed up are the files in your preconfigured libraries plus content on your desktop, favorites and contacts. This means that any files you want to have backed up that are not contained in the libraries need to be copied into an existing or new library (see page 16 on how to view and create libraries).



I found it very helpful to add all of the data under C:\Users\<user>\AppData into a newly created library (which I called BackUp) so that this is backed up as well. The AppData folder contains a great number of settings files for many programs as well as actual working files for some:





If you are using MS Outlook as your email client, you'll need to be aware that the Outlook files are only backed up through the File History system if Outlook is closed.

Backing Up and Restoring the Complete 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 "imaging" (you may also hear people refer to it as "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.

While this required external third-party programs before *Windows 7*, this feature is now directly integrated into *Windows*. You can access it from a link in the **File History** dialog or **Backup and Restore** dialog ("**Create a system image**"). You are prompted to select a backup location and are then guided through the additional necessary steps.

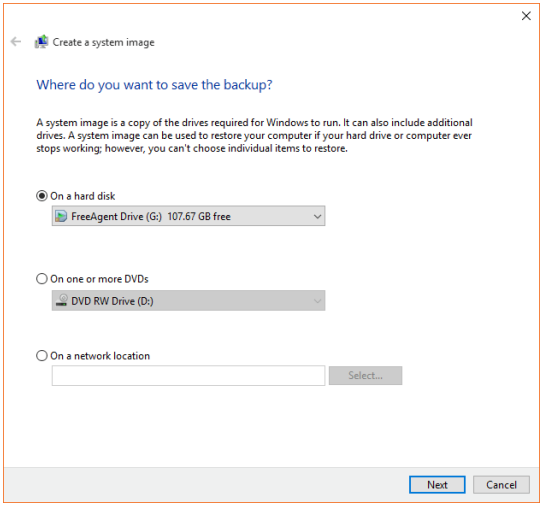


Figure 38: Imaging in *Windows 10*



Imaging is a particularly helpful option if you want to create a copy of a computer in a particular state that you then want to deploy to other computers in your network.

There are two options offered in *Windows 8* and above: to completely reset your PC and delete all data and installed programs in the process *or* to refresh your computer without affecting your files. This second choice keeps your personal data, system settings, and Metro-style applications. Desktop applications will be kept as well if you have previously created a custom image (see below).

To access these possibilities in *Windows 10*, press ⊞+I to open **Settings**, select **Update & Security**, and then select **Recovery**. You will find the following options:

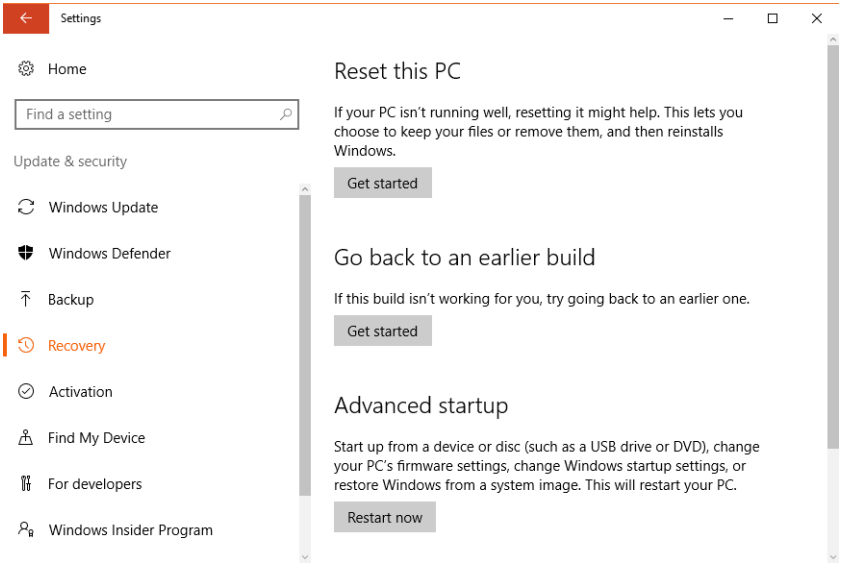


Figure 39: Recovery options in *Windows 10*

Asking for Help

Unless you take your computer into a computer shop when you encounter problems, it's sometimes very hard to explain what went wrong to someone on a phone help line. Of course, there are ways to share your computer with someone else—an easy way to do that is join.me—but another way is to record your problems and send the recording to someone else.

There are a great number of third-party products that do this along with an in-house tool in *Windows 7* and *above*, the *Problem Steps Recorder* (in *Windows 8* and above: *Steps Recorder*). This tool allows you to record everything on your screen (with the exception of text that you enter). Once the recording is done, it is not saved as a movie file but as an MHT HTML archive file and zipped up. Once unzipped, the MHT file can be opened with either *Internet Explorer, Microsoft Edge, Chrome* or *Opera* (or with *Firefox* with a special plug-in). It gives you a screen-by-screen description of what just happened on your computer as well as a narration of the process and operating-specific information.

To start the recorder in *Windows 7* and *10*, click on the *Windows* button, type `psr`, and hit **Enter**. (In *Windows 8*, type `steps` in the **Start** screen and select the *Steps Recorder*.)

Everything else is very self-explanatory.



Figure 40: Windows 7's Problem Steps Recorder

In *Windows 10*, a new self-recording concept was introduced with the Xbox Game bar. To open this, press ⊞+G and start the recording by pressing the **Start** button:

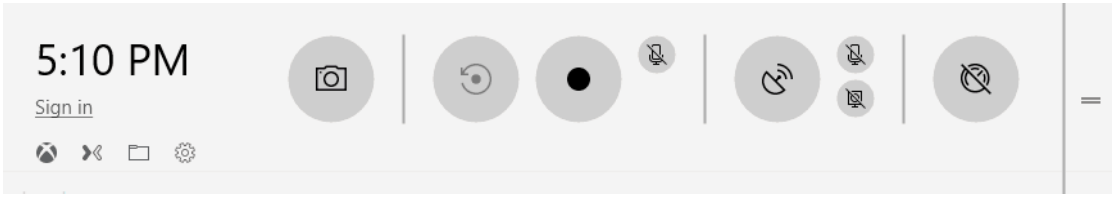


Figure 41: The *Xbox Game bar*

Everything you now do on your screen is recorded in video mode so that you can send a video to your computer support team when "unspeakable" things happen to your computer (or to your client so you can prove it was not your fault when something went awry with a project or a tool).

Taking Inventory of Your Computer

The little freeware download *Belarc Advisor* (see 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.

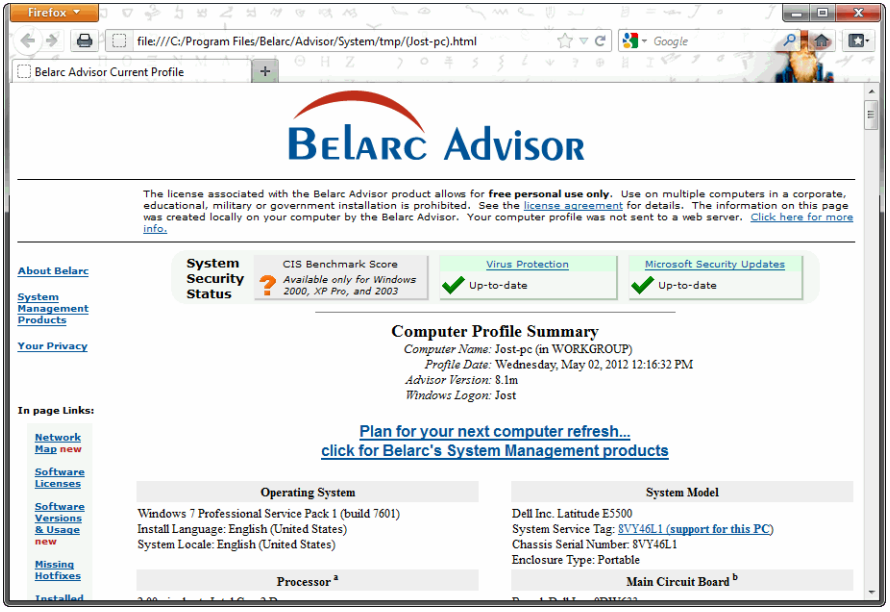


Figure 42: Analysis results in *Belarc Advisor*

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 wordfaqs.ssbarnhill.com/InsertSpecChars.htm or in your *Word* help system. However, there are two problems 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 ASCII 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! Again, 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.

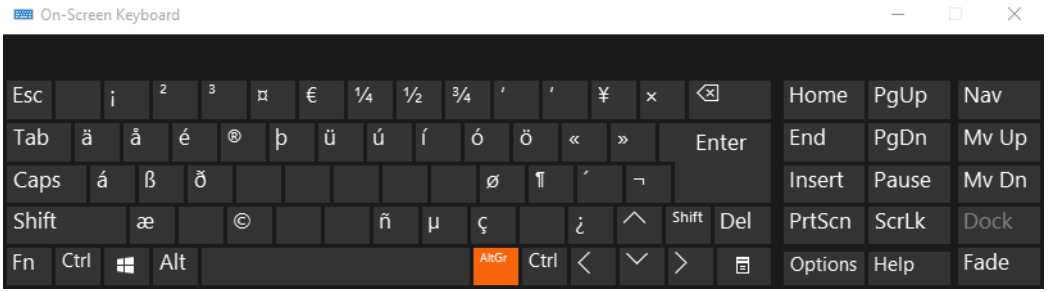


Figure 43: Characters on the US-International keyboard when pressing the right ALT key



*You can find the On-Screen Keyboard of the image above under **Start> Programs> Accessories> Ease of Access> On-Screen Keyboard**; Windows 8 and above: Type `on-screen` in the **Start** screen/menu and select the **On-Screen Keyboard**.*

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.

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."

Installing Additional Keyboards on Windows 7

To install an additional keyboard select **Start> Control Panel> Regional and Language Options> Keyboards and Languages> Change keyboards**.

This or a similar dialog will appear:

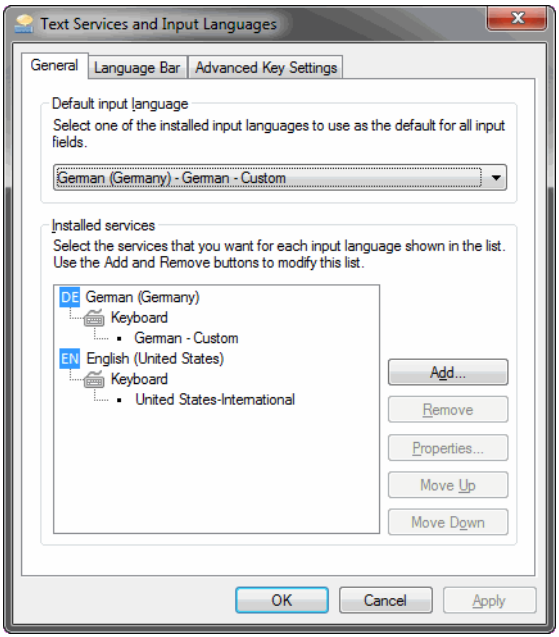


Figure 44: **Input Languages** dialog

Select **Add** and define which additional languages and/or keyboards you would like to have installed on your system.

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

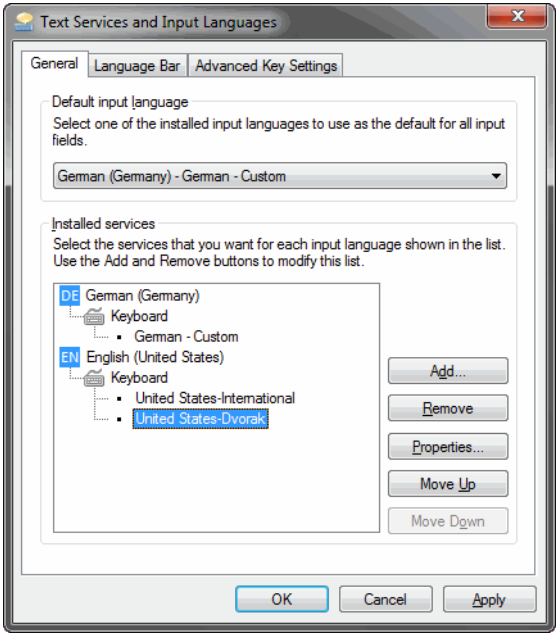


Figure 45: **Input Languages** dialog with a newly added keyboard

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.*

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 47: Examples of East Asian keyboards with access to various features

Installing Additional Keyboards in *Windows 8* and above

In *Windows 8* and above, keyboards are part of the language concept that also gives you access to the multiple language user interface. To select additional keyboards, you'll need to select **Control Panel> Language> Add a language**. This will add the standard keyboard for that language. If you

need to add several keyboards for one language or you prefer something other than the standard one, you can click on the **Options** link to the right of the language and select **Add an input method** in the ensuing dialog.

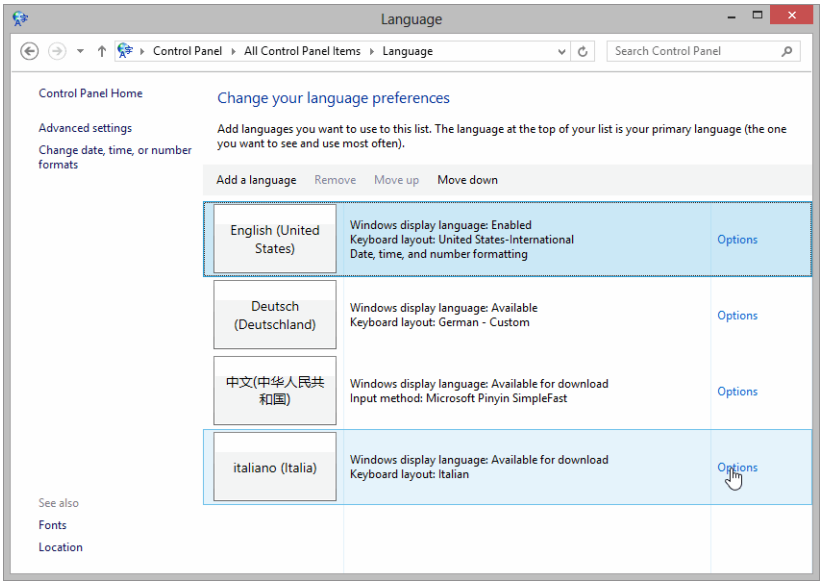


Figure 48: Selecting **Options** for additional language-specific keyboards in *Windows 8* and above

Much like in previous versions you will see a keyboard icon in the system tray once you have more then one keyboard installed. (Left-) clicking on it will show the installed keyboards and allow you to choose a keyboard.

.

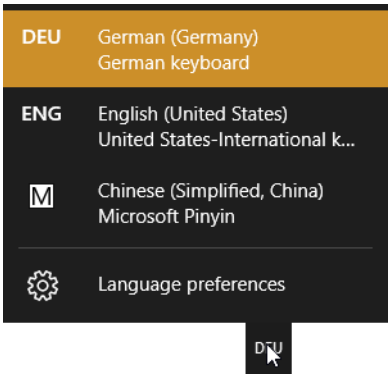


Figure 49: Accessing the installed keyboards



You can also switch between keyboards with key combination ⊞+Space or Alt+Shift.

If you translate into languages that are not covered by the languages offered by *Windows*, you might want to look at *Keyman* (see keyman.com), a free tool (for *Apple*—MacOS or iOS, *Windows*, or *Android* devices) that covers 500+ different keyboards, with many indigenous languages that you won't likely find elsewhere.

Mapping Existing Keyboards

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 microsoft.com/en-us/download/

details.aspx?id=22339). 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 works up to *Windows 8.1* 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.

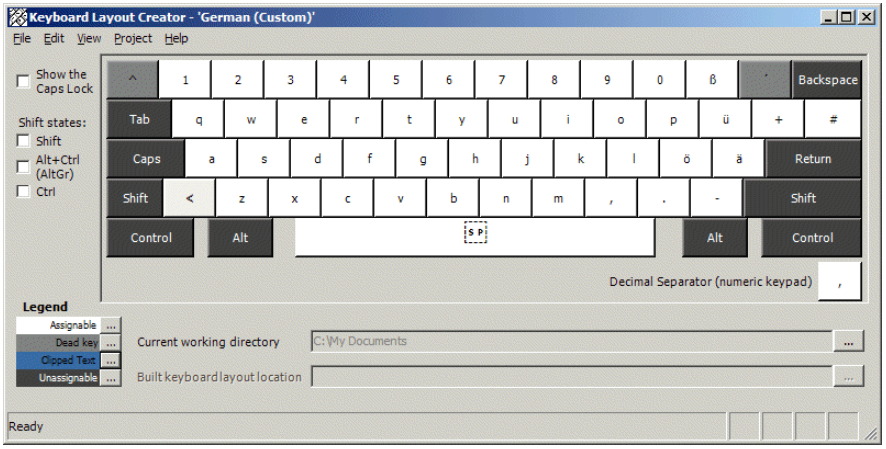


Figure 50: Custom German QWERTY keyboard layout with MSKLC



*MSKLC is unfortunately not directly supported on Windows 10. You can, however, install existing keyboards that were created with MSKLC on an earlier version of Windows. You'll need to copy all the files that are contained in the folder for the customized keyboard layout to your Windows 10 computer, right-click on the setup.exe file, and select **Properties** and then **Compatibility**. Under **Compatibility mode**, be sure to select the last officially supported operating system, click **OK**, and the installation should work just fine.*

A little freeware utility by the name of *AllChars* (see sourceforge.net/projects/allchars/) provides you with a large number of preconfigured keyboard shortcuts (that all follow the pattern CTRL (release!) + one key + second key)

:

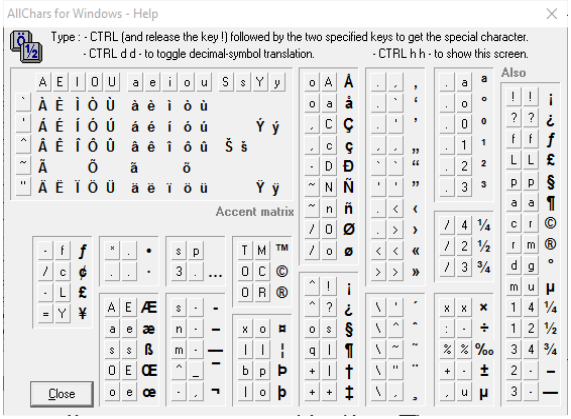


Figure 51: Preconfigured options in *AllChars*

and also allows you to configure your own combinations:

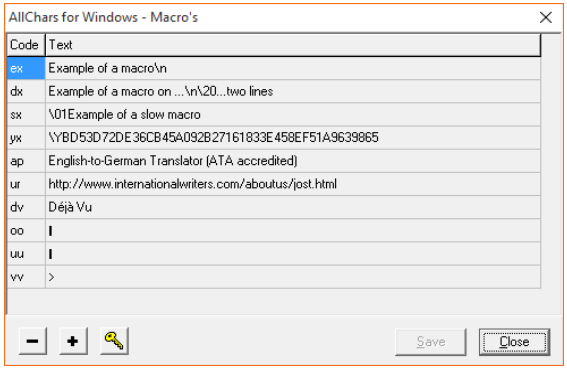


Figure 52: User-defined options for *AllChars*

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



Aside from the character-specific macros, it is also possible to create and use macros to do all kinds of things on your computer. For instance, I use a macro to automatically enter „German quotes.“ The tool that I use for that is called AutoHotkey (see autohotkey.com), the most well-known and well-loved macro generator. Unless you are really into studying the required coding, I would advise you to do it the way I did it: ask a geeky friend to write the macro for you or see whether you can find an existing macro on their website that already does what you need done.

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*). If possible, I use *Firefox* (see mozilla.com) since I prefer to not use one of the tech giants' tools and also have *Opera* and *Google Chrome* on my computer for testing my website and others.



Browsershots (see browsershots.org) is a site that enables you to test any webpage in any thinkable combination of browser and operating system. Simply enter the URL of your page, and you will be shown screenshots of that page in every browser you request and every operating system. It's a little slow for my taste, but then maybe I got a little greedy with the many display configurations. It certainly beats manual testing, and best of all, it's free.

Browsing Tips

When I started translating professionally, the Internet was already a formidable resource that held a lot of translation-related information. But I know the feeling of rummaging through books and other "hardware" 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).
- You can use `~` as an operator for synonyms, so that `~great translator` as a search query returns results with "best translator," "cool translator," or "top translator" (*Google* only).
- `..` can be used as an operator for numeric ranges, so that `Translator's Tool Box $40..60` will find webpages where you can purchase this must-have resource for between $40 and $60 (*Google* only).
- If you want to search for pages that *Google* has in its cache (previous storage), so that `cache:internationalwriters.com "tool kit"` finds pages that have been changed or deleted (*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 example, to see webpages about machine translation from the U.S. or Great Britain, enter `"machine translation" (loc:US OR loc:GB)` (*Bing* only).

- To do research on content of websites similar to the one you have already located, enter `related:<site>`. If you want to find all kinds of translation associations, for instance, you can enter `related:atanet.org` (*Google* 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 URL from the Microsoft help site:

> http://windows.microsoft.com/en-us/windows/create-user-account#create-user-account=windows-8

This URL has one obvious language identifiers—en-US, a mixture of the standards ISO 639-1 and ISO 3166.

To change that page into, say, Japanese, you could just manually replace the URL with the appropriate code:

> http://windows.microsoft.com/ja-jp/windows/create-user-account#create-user-account=windows-8

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

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.

Using Wikipedia for Language Data

The many different language versions of *Wikipedia* are not translations of each other, and that's an important aspect to the multicultural (self-) perception of *Wikipedia*. Articles are typically adjusted (and chosen) by volunteers to fit the target locale. Still, *Wikipedia* is a powerful tool for terminology research—especially when it comes to top-level terms. All translators know that different language entries for each search term are displayed on the left-hand side (or the right if you start out with a right-to-left language such as Arabic or Hebrew), making the correct translation possibly just a click away.

There are also tools that support a more in-depth comparison of different language versions so that you can quickly not only spot the top-level term but some of the terminology that surrounds it. *Manypedia* (see manypedia.com) is a tool that searches *Wikipedia* for a specific term and then looks up the corresponding *Wikipedia* pages in other languages. It will then tell you the percentage of the similarity of the concepts and display the pages you request side-by-side.

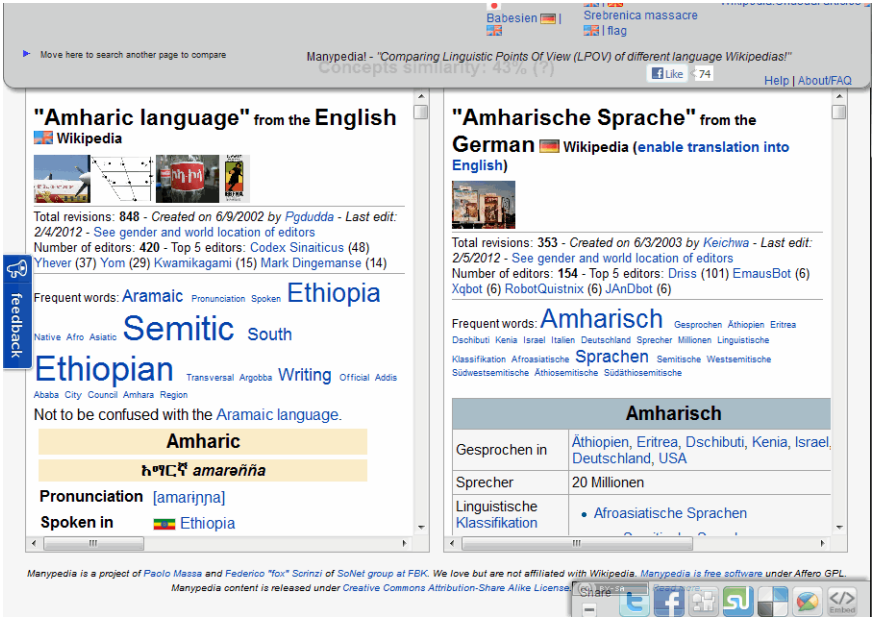


Figure 53: Comparing *Wikipedia* sites in *Manypedia*

Tradupedia (see metadic.com/tradupedia) is heavy on Iberian languages (aside from Catalan, Spanish, Galician, Portuguese and Basque, it also supports German, French, English, Italian and Russian) and it allows you to find equivalent terms and articles in the other language Wikipedias.



Figure 54: *tradupedia's* interface

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 Compressing Files on page 125.

For large files, you should not send attachments by email but via the File Transfer Protocol (FTP), the same protocol used to upload files to websites. While you can use an FTP program ("FTP client") for that (for instance, the open-source *FileZilla*—see filezilla-project.org), it's also easy to connect to FTP servers via *Windows/File Explorer*. You can either enter the FTP address right into the address field of *Explorer,* which will then prompt you for the user name and password (store it from then on so you'll only have to enter it once), or you can use the **Add Network Location** wizard (click on **Add a network location** on the **Computer** ribbon tab in the *Windows/File Explorer*) that will guide you through the process and then add an icon to your **Network locations** (see the graphic below).

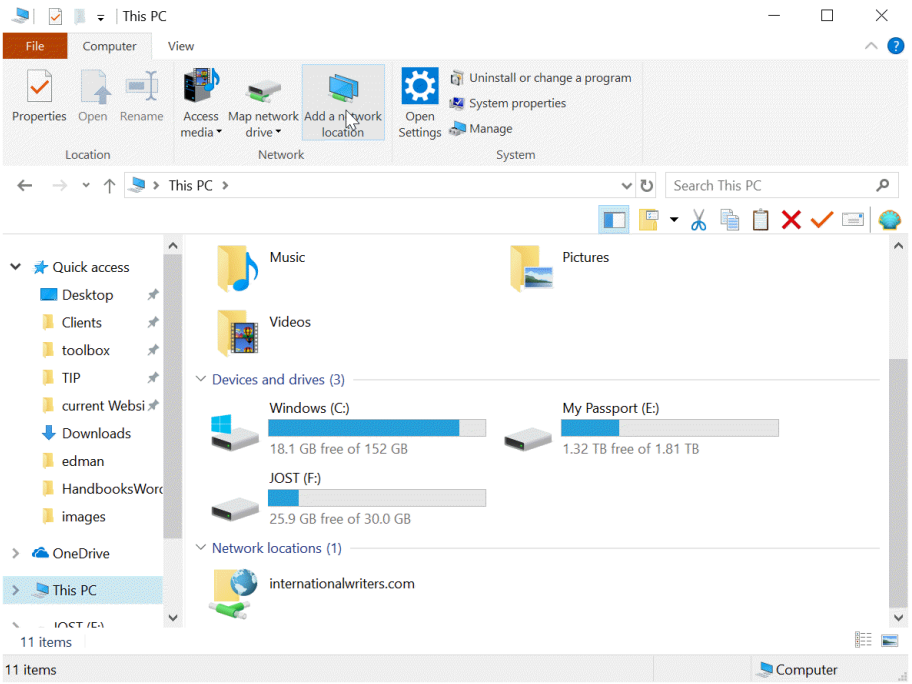


Figure 55: Access to the **Add Network Location** wizard and an added FTP site under **Network locations**

Another, now widely accepted way of sending very large files is through (free) cloud-based services like Hightail (formerly YouSendIt) (see hightail.com). Note that if you directly work with large corporate clients, their corporate network policies might not allow access for their employees.

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, email, 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 81). Often these "zombies" are combined into systems called "botnets" (see page 81).

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 77). This backdoor program either allows unauthorized remote access to your computer by other parties, or it installs a keylogger (see page 77) 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 79) are used to lure you into opening or executing these files and programs.

Ransomware

Ransomware, as the name suggests, demands a ransom from you to get back to be able to control your computer or network. Ransomware typically encrypts all files in a system or network, rendering them inaccessible. Usually a ransom note demands payment in cryptocurrency in exchange for

decrypting your files. If the ransom is not paid, the encrypted files could eventually get destroyed. Most ransomware is delivered through are Trojans. Unfortunately, in some cases, hackers refuse to decrypt files even after the ransom is paid.

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— but you may also encounter some rather obnoxious specimens that generate pop-ups (see page 82) on your computer without any web browser open.

Malvertising

Malicous ad**vertising**, the attacker purchases advertising space allotments from regular advertising networks such as Yahoo and Google and then uses a maliciously crafted ad to exploit one or more vulnerabilities on unsuspecting victim's computers. In addition to keeping your software up-to-date, another step in improving your protection against this type of attack is to use an ad blocker (see page 97).

Malvertising is one of the means (ab)used for drive-by-downloads (see page 80).

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.

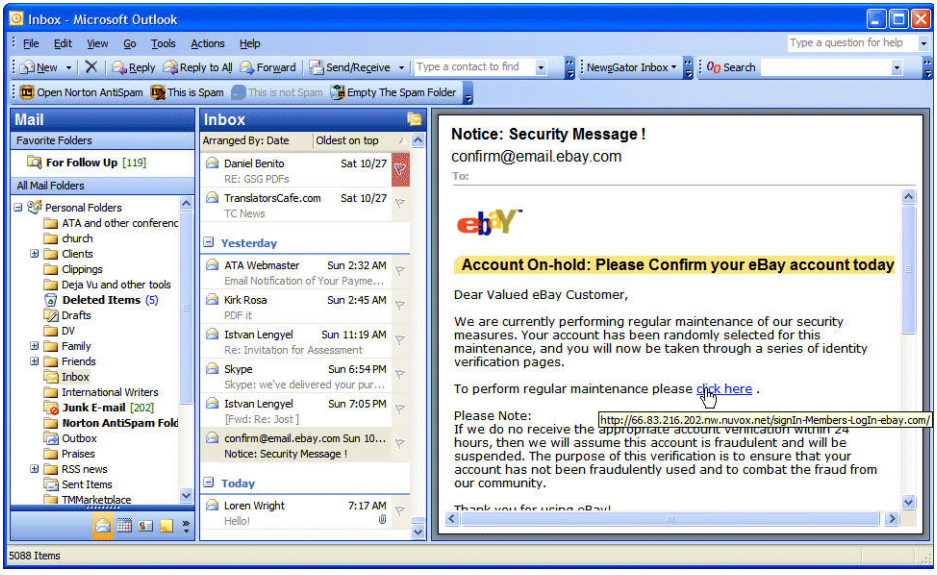


Figure 56: Phishing attempt with fake eBay email

This screenshot nicely illustrates one such phishing technique. At first glance this might look like an email from eBay, but on closer inspection it becomes clear that this is an imposter's phishing email. 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 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 email 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 email client or your web browser.

Maliciously Crafted Content

At times, you might be subjected to a file that contains maliciously crafted content, e.g., a *Word* file containing macros that attempt to perform undesirable or unauthorized activities on your computer.

Similarly, you might encounter PDF files with embedded Flash content that attempts to perform such activities on your computer.

Typically, such activities should be detected and blocked by a good anti-virus product or Internet security suite, but it never hurts to exercise caution when receiving files from an unknown source.

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 81).

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 email spam (see page 82), commit click fraud, conduct a denial-of-service attack (see page 80), 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 email 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 email spam, bulk email, junk email, UCE (unsolicited commercial email) and UBE (unsolicited bulk email) all refer to nearly identical messages sent to a large number of recipients via email.

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 emails. This is usually achieved by placing a reference to a tiny transparent or inconspicuously colored image with a unique name in an HTML-based email. This is unseen by the recipient, but it triggers an access log entry for the image when the email 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 email, 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 *Belkin 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, Belkin Linksys (see linksys.com) and other router manufacturers such as Netgear (see netgear.com) and D-Link (see 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 Belkin 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`.

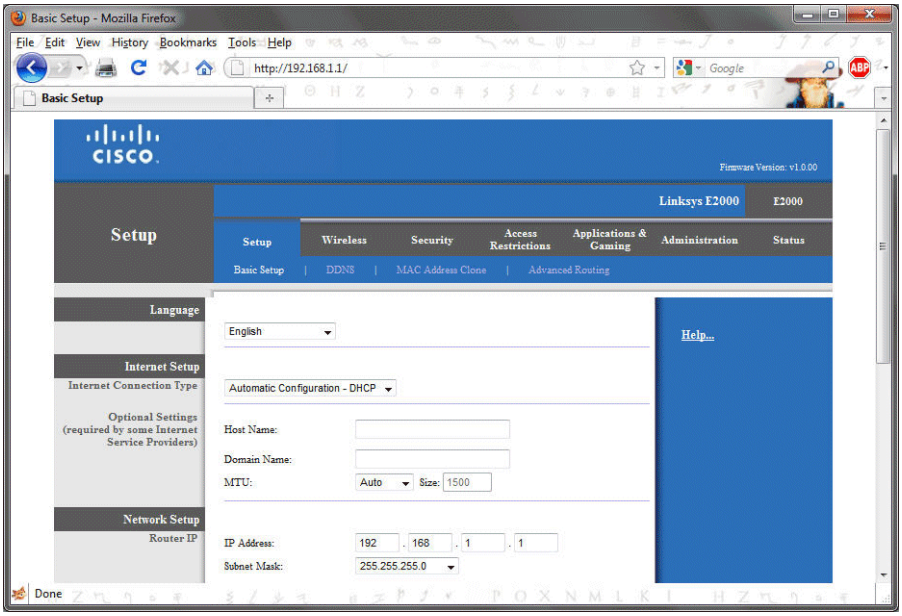


Figure 57: The router's setup interface after successful login

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**.

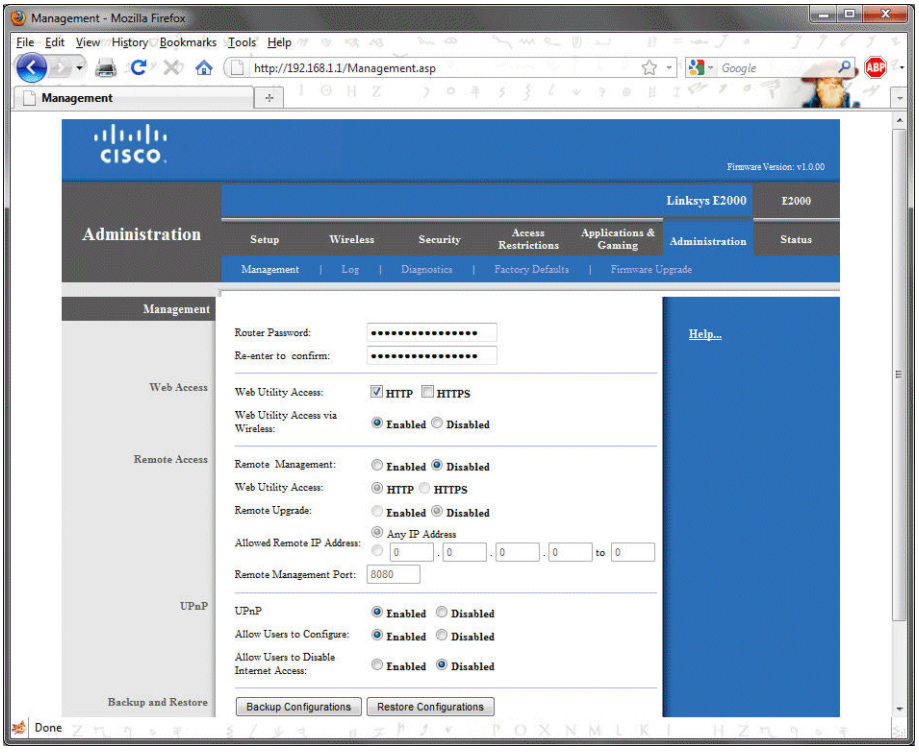


Figure 58: The management interface

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.

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

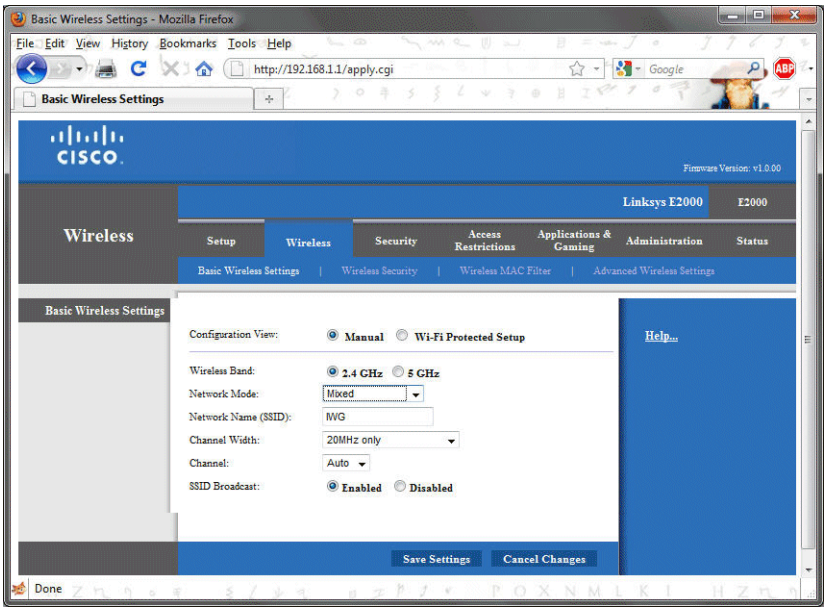


Figure 59: Basic wireless settings

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**.

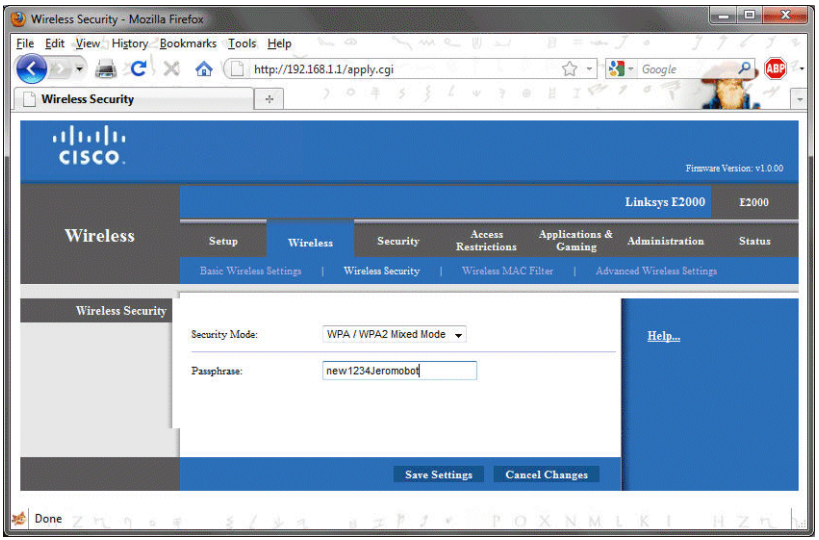


Figure 60: Entering a wireless key

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 seagate.com) are offering hard disks with built-in encryption.

Encrypting Individual Files

Windows (in the *Professional*/*Ultimate* versions) allows you also to encrypt individual files without any kind of fuss.

To encrypt a file or a folder, simply right-click on that particular file or folder in *Windows Explorer*, select **Properties**, click the **Advanced** button, select the **Encrypt contents to secure data** checkbox, and then click **OK**.

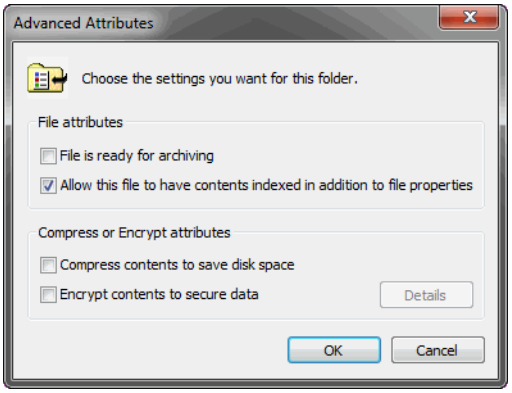


Figure 61: The **Advanced Attributes** dialog

If there is a lot of data in that file or folder it may take some time for the encryption process to complete, but you'll know it's finished when the encrypted folders and files appear in a different color than your other files. And that's really all you will notice—opening, saving or even emailing them will all work as before (they will simply lose their protection once you attach them to an email). Don't believe that they're protected? Try to log on with a different user account and you won't be able to open them (that's also why it's not possible to encrypt commonly shared files such as system files).

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, anti-virus, 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 83) 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.



Microsoft's internal firewall functionality can be instructed by any application to allow traffic through, just as long as the programmer included the respective snippets of code necessary to automatically add a corresponding rule to the firewall's rule set.

In cases where no predefined 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.

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 zonealarm.com),
- *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 f-secure.com) and
- *Symantec's Norton Security* (see symantec.com).

Anti-Virus Software

Anti-virus 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 anti-virus 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, a number of security/anti-virus providers have started to offer versions of their anti-virus products that allow for installation on a number of computers. Among the other contenders for protection against computer viruses are the above-mentioned *Symantec* and *F-Secure* as well as *Bitdefender* (see pandasecurity.com) *and AVG* (see avg.com).

The *Windows*-internal *Windows Defender* provides for a reasonably good baseline protection, it is typically found to be no match for products from Symantec or Bitdefender, which consistently rank at the top. The major differentiating factor in this are so-called zero-day attacks in which an exploit has just become public but is already being utilized heavily.

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 anti-virus software: a memory-resident component plus regularly scheduled scans.

Spybot (see safer-networking.org), Lavasoft *Ad-Aware* (see adaware.com) or SpywareBlaster (see brightfort.com/spywareblaster.html) all offer free versions to personal, non-commercial users and paid versions to commercial users.



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 8 and above* under **Safety> (Turn on) SmartScreen Filter**.

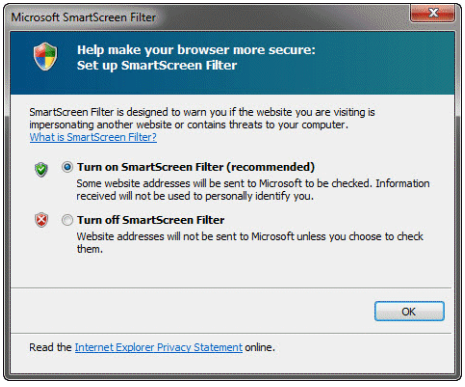


Figure 62: Phishing prevention in *IE*

In *Microsoft Edge* you can find the SmartScreen settings under **Settings> Advanced Settings**.

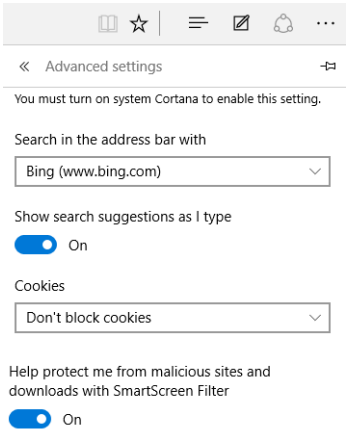


Figure 63: SmartScreen settings in *Microsoft Edge*

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

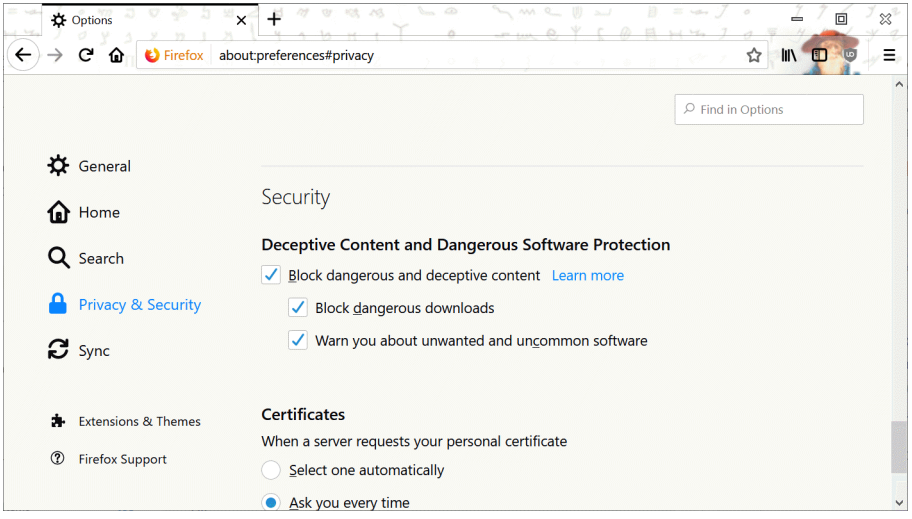


Figure 64: Phishing prevention in *Firefox*

In *Chrome* you can set phishing protection under **Settings> Advanced > Privacy**> **Protect you and your device from dangerous sites**.

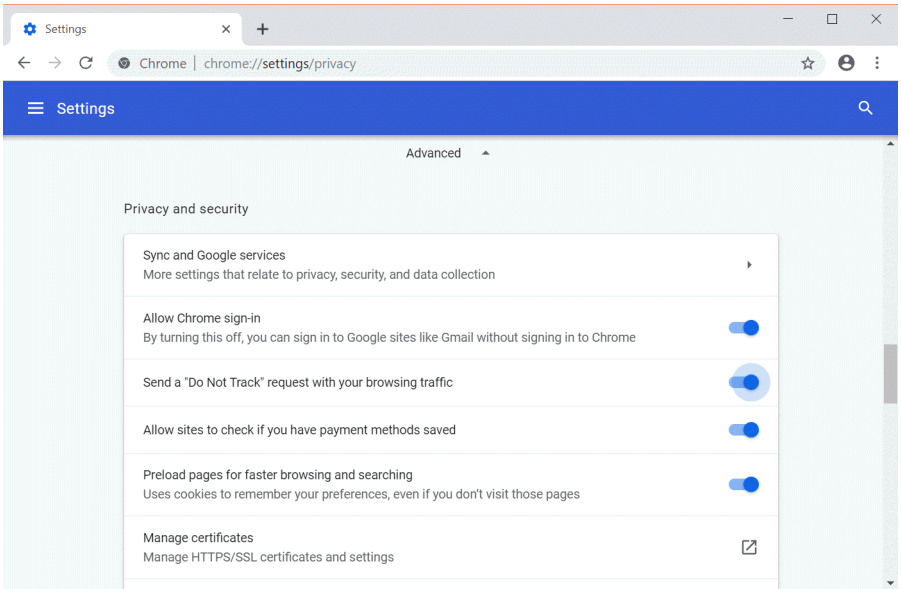


Figure 65: Phishing prevention in *Chrome*

In *Opera*, the protection is enabled by default.

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 browsers, such as *Google Chrome*, even have pop-up protection enabled by default.

In *Internet Explorer*, you can exercise control over the pop-up blocker settings under **Tools> Pop-up Blocker**).

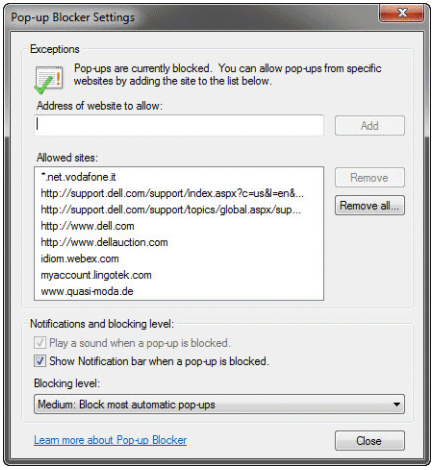


Figure 66: Stopping pesky pop-ups in *Internet Explorer*

In *Microsoft Edge* you can stop pop-ups by selecting **Settings> Advanced Settings.**

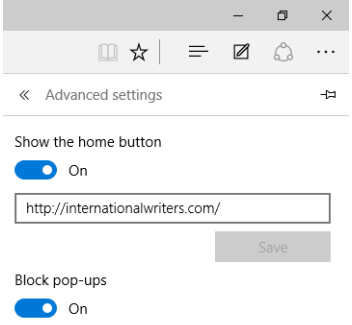


Figure 67: Stopping pop-ups in *Microsoft Edge*

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

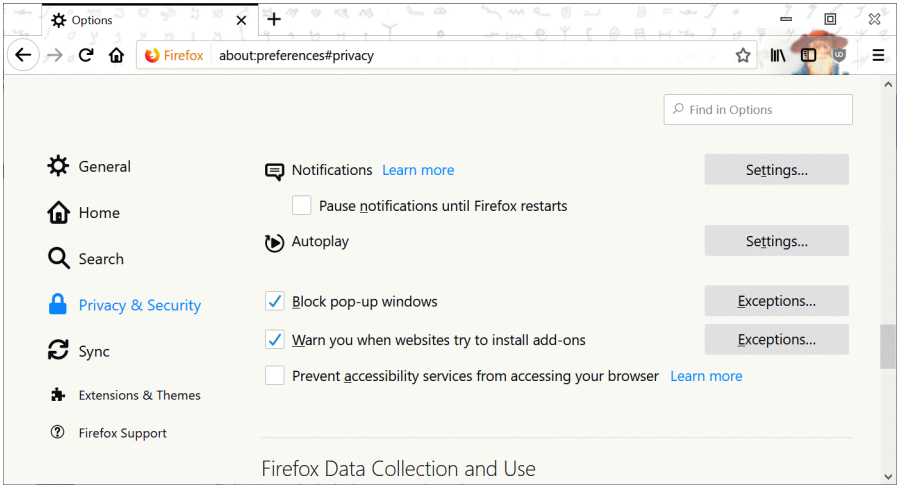


Figure 68: Stopping pop-ups in *Firefox*

In *Opera*, you can gain control over pop-ups under **Opera> Settings> Advanced> Site Settings> Pop-ups and Redirects**.

Ad Blockers

Somewhat related to pop-up blockers are advertisement blockers, which generally rely on preconfigured lists of known advertising sites as well as patterns of path names typically contained in the fully qualified name of the files involved in displaying ads.

One well-known ad blocker is *Adblock Plus* (adblockplus.org), available for all major browsers. Of particular interest with this one is that it offers ad-blocking lists for several other languages, too. When using *Adblock Plus*, remember to remove the check mark in front of **Allow some non-intrusive advertising** in order to avoid falling victim to the afore-mentioned risk of malvertising.

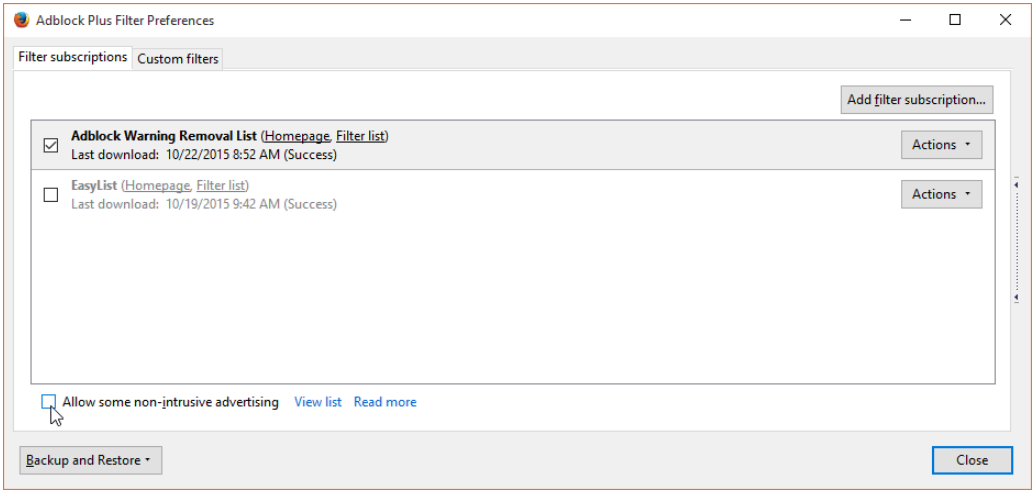


Figure 69: Adblock Plus's preferences



In addition, the Electronic Frontier Foundation (EFF) offers Privacy Badger (eff.org/privacybadger), an easy-to-use tool for blocking ads and not leaving (too many) tracks on the Internet. Privacy Badger is originally based on the Adblock Plus code but uses a slightly different approach.

Yet another option is uBlock Origin, an open-source fork of a blocker that was purchased by Adblock.

Anti-Spam

To curb the huge amounts of unsolicited commercial email 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' email or messages from potential new clients have not accidentally been misclassified as spam.

If you are using applications such as *Mozilla Thunderbird* or *Outlook*, 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> Security**.

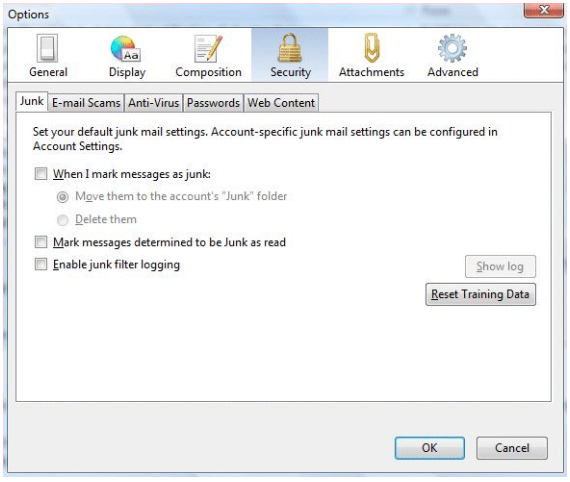


Figure 70: Spam options in *Thunderbird*

To configure "Junk Mail" handling in *Outlook* you can select **Home> Junk> Junk E-mail Options**).

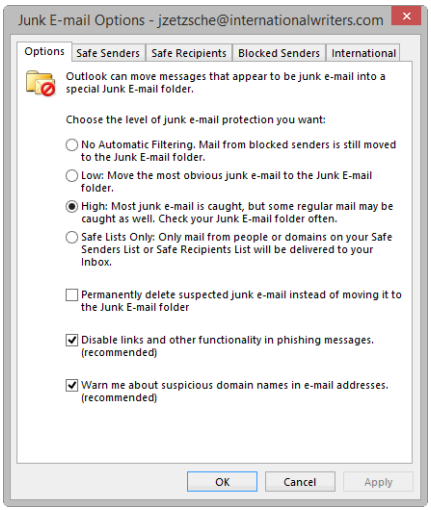


Figure 71: Spam options in *Outlook*

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).

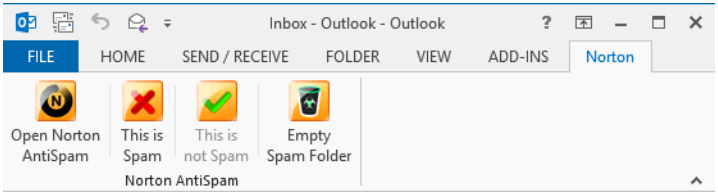


Figure 72: *Norton Anti-Spam* filter in *Outlook*

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, such as *Spybot*.
- Stay up-to-date on firewall, anti-virus 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 and any applications you are running.

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 the text editor *Notepad*, a very limited 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).

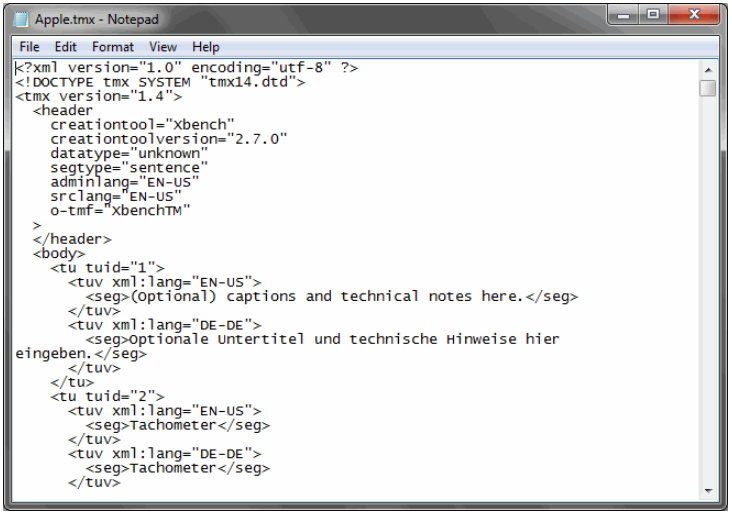


Figure 73: Translation memory exchange (TMX) file in *Notepad*

There are, however, much more powerful (and free or fairly inexpensive) programs out there. Most of these have been developed by developers for developers, so we translators usually see only the surface of what these programs can do. Still, this is usually enough to be duly impressed.

The most commonly known editors for *Windows* are probably *UltraEdit* (see ultraedit.com), *EmEditor* (see emeditor.com), or the free *Notepad++* (see notepad-plus-plus.org). I use *Notepad++* for all Western languages, *EmEdit* for East Asian languages, and (there might be a better choice, but I have not found it) *Windows Notepad* for bi-directional languages, which are not well displayed in the other tools. 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 100 or 200 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 of user interface translations of Microsoft products, access to the Microsoft style guides for many languages, and the ability to download extensive glossaries in the termbase exchange (TBX) format (see page 285). Most 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 309) to convert the TBX files into something more palatable.

While the translation memory files can be accessed on a per-string basis on the Microsoft Language Portal, they can also be downloaded as complete TMs on Microsoft's Visual Studio portal. The least expensive Visual Studio subscription presently (December 2019) costs $1199 for the first year and then $799 annually from then on out. You can find them if you search for "msdn" at microsoftstore.com.

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.

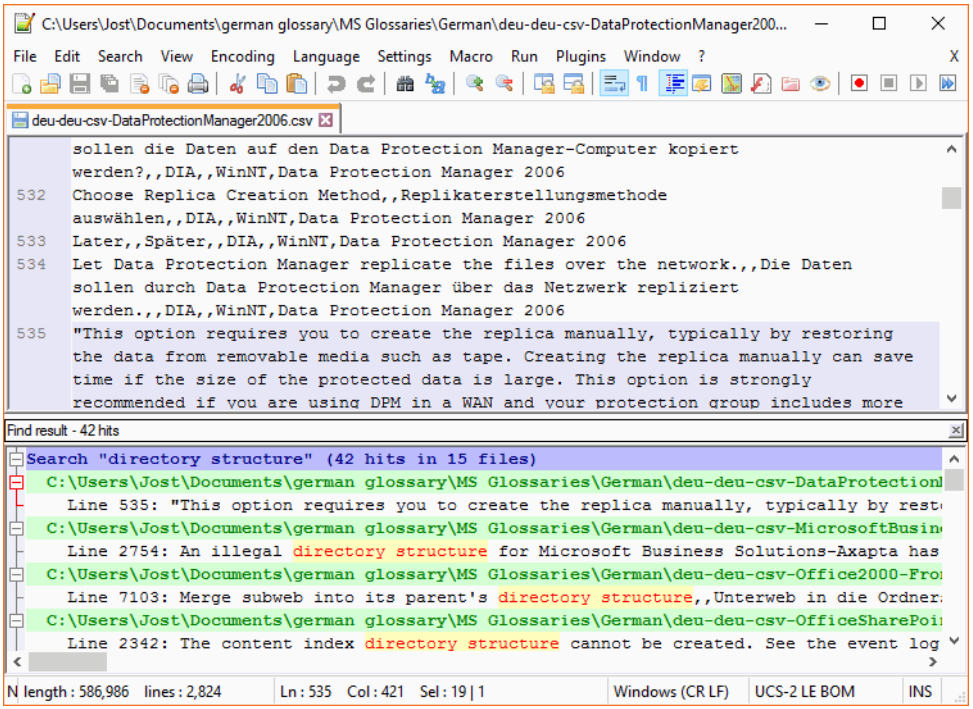


Figure 74: Batch search in *Notepad++*



Any of the searches can be made with wildcards. For wildcards, see page 20.

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.

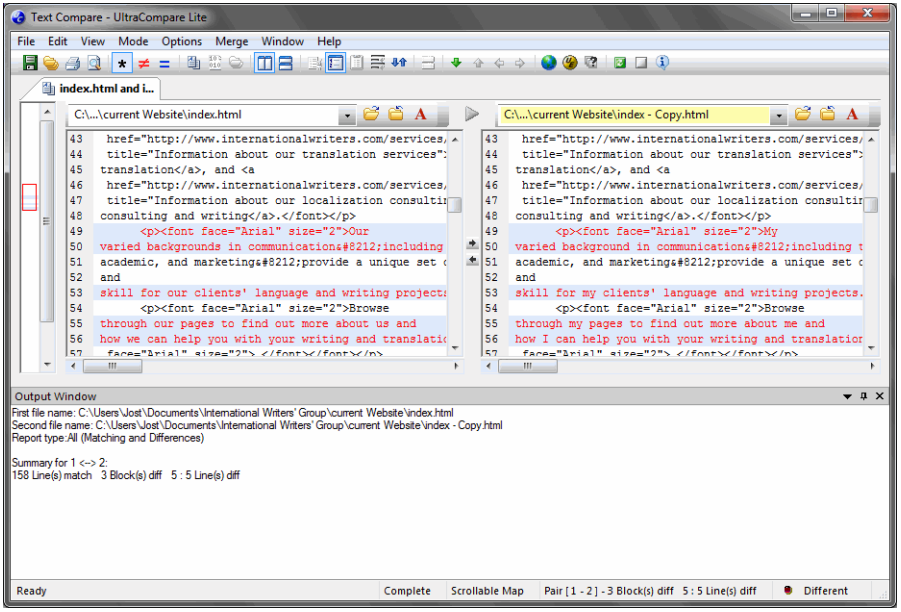


Figure 75: File comparison in *UltraEdit*



If file comparisons are very important to you, you can also use a specialized tool such as ExamDiff Pro (see prestosoft.com/examdiff) or Araxis Merge (see araxis.com/merge). 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.

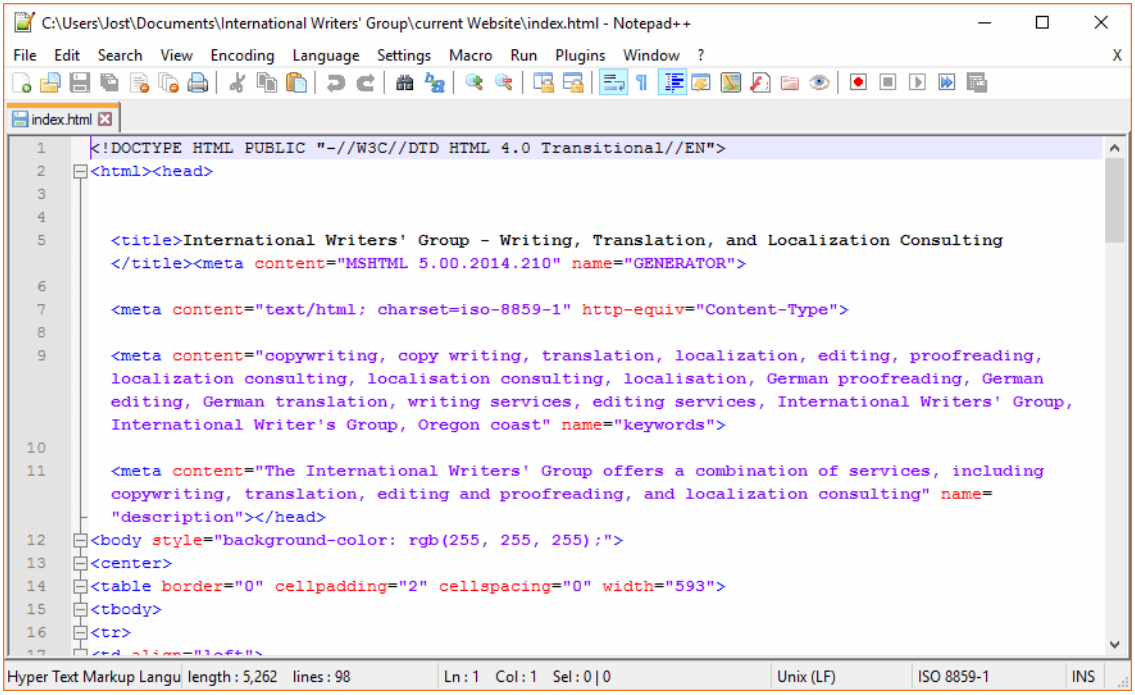


Figure 76: Color-coded HTML file in *Notepad++*

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).

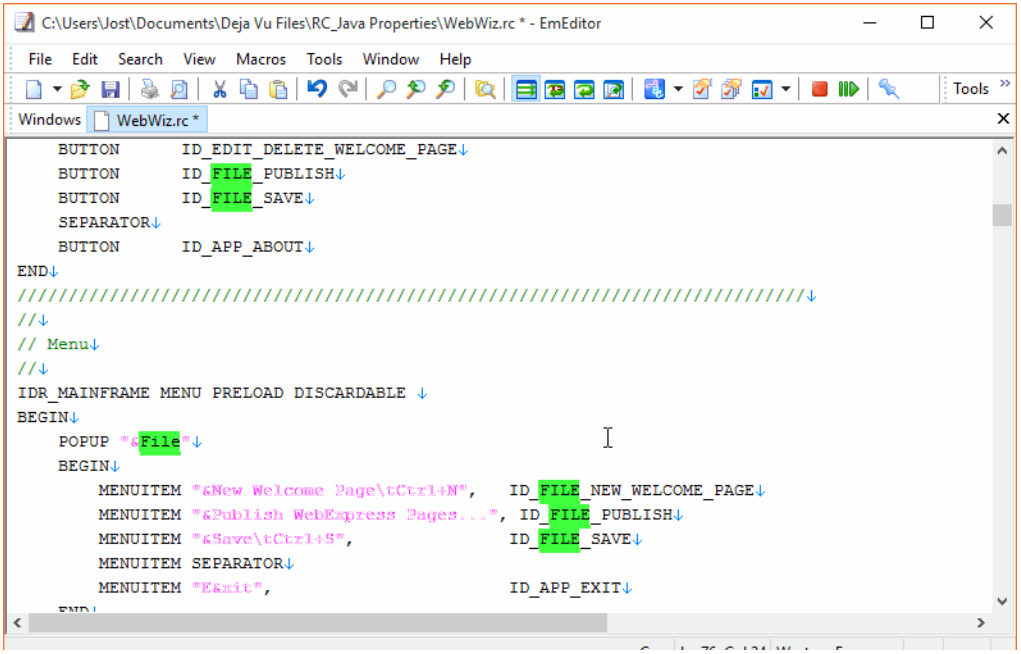


Figure 77: RC file view in *EmEditor*

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*):

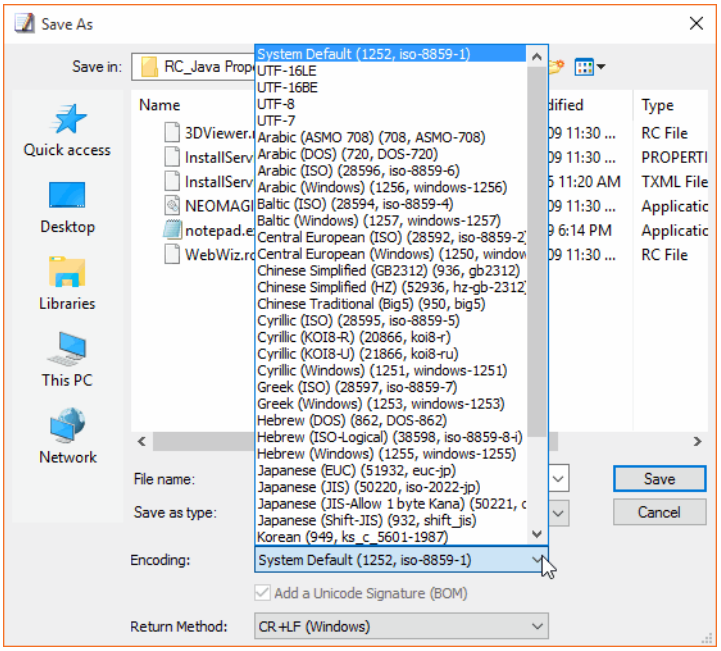


Figure 78: Possible conversions in *EmEditor*



If you need to convert the code page of many files at once, a tool like Unifier (see melody-soft.com/html/unifier.html) allows you to quickly convert any number of text-based files from one code page to another. It does cost a little bit, but it was worth it for a recent project where I had to convert several hundred HTML files in the Japanese code page Shift-JIS into Unicode. It even rewrites the code page tags in the files.

Most of us have received files of indeterminate type—we either don't know the extension (and even sites like 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).

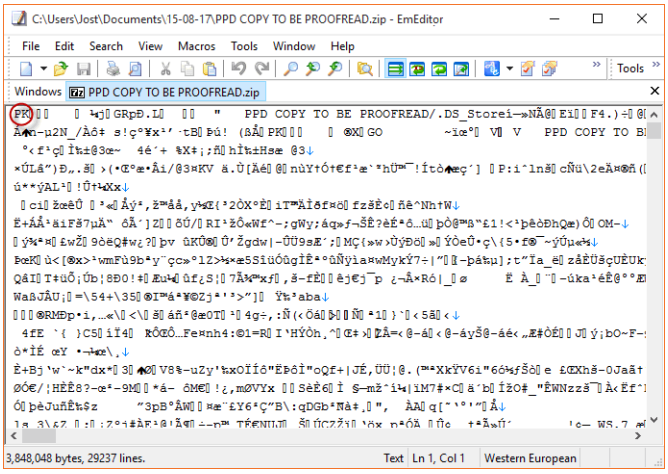


Figure 79: A "cracked" zip file with the magic number "PK" in *EmEditor*

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/File 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).

There are currently two basic flavors of HTML: HTML 4 and HTML5. Typically both kinds are supported by translation environment tools.

Although the afore-mentioned text editors have HTML capabilities (see page 107), 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 *Adobe Dreamweaver* (see adobe.com/products/dreamweaver) to (also high-powered but much more affordable) hands-on products like (the free and powerful) *Nvu* (see nvu.com).



A more current version of Nvu is BlueGriffon (bluegriffon.org), which also supports HTML5. Unfortunately, it's not possible to prevent this editor from making changes to the HTML code of your documents, which makes it a no-go for translation and editing purposes for outside clients (see below.)

If your intention is to translate an HTML file outside of a translation environment tool (see *Website Localization* on page 321), I would strongly recommend using a text editor rather than an HTML editor because that way you do not run the risk of introducing any additional and unwanted coding that your client most likely abhors.

It is very important to never work in HTML files in *Word* (or another word processing tool) unless you are specifically instructed to do so.

```
<!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>
```

```
<!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">

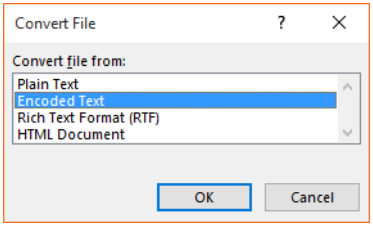
<head>
<meta http-equiv=Content-Type content="text/html; charset=iso-8859-1">
<meta name=ProgId content=Word.Document>
<meta name=Generator content="Microsoft Word 9">
<meta name=Originator content="Microsoft Word 9">
<link rel=File-List href="./indexWord_files/filelist.xml">
<link rel=Edit-Time-Data 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 80: On the left you see an HTML page after being saved in non-obtrusive editor (same as original) and on the right the same HTML page after being saved in *Word*



*To be fair, Word 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.*

*It is also possible to open an HTML file in Word as a text file. To do tha,t open the **Options** dialog in Word (in Word 2007 and above in the **Office/File** menu; in prior versions in the **Tools** menu) and select **Advanced> Confirm file conversion at Open**. Once you do that, you can select the file format as **Encoded Text** during the opening process.*



*In Apache OpenOffice/LibreOffice, you can force-open HTML files by selecting the HTML file in the **Open** dialog and then setting the file type to **Text Encoded**.*

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, Chrome*, *Firefox* or *Opera*, open the web page, press CTRL+S 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 tenmax.com/teleport) or HTTrack (see httrack.com) 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.

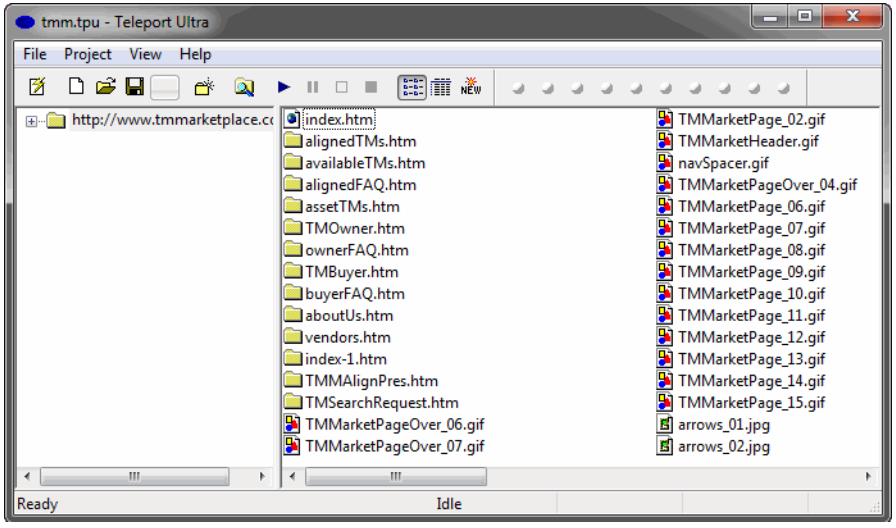


Figure 81: Downloaded website in *Teleport*

Utilities

Before discussing the "real (and often expensive) programs" such as office suites, computer-assisted translation tools, desktop publishing, graphics, or multimedia applications, here is an overview of smaller, inexpensive (or even free) programs—that I'm calling 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.

Managing Graphics

Graphics management has never been one of *Windows'* strong suits. Microsoft must be painfully aware of this since they have been trying to reinvent the wheel with virtually every new release of *Windows*. A good and free third-party option for managing image files is *XnView* (xnview.com), which lets you view, convert, copy, sort and edit any image file.

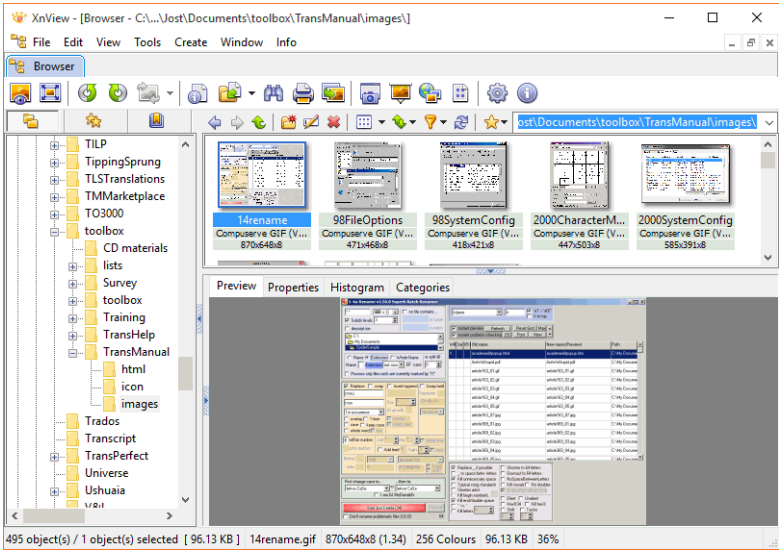


Figure 82: Image management in *XnView*

XnView is particularly useful when dealing with a large number of image files like those 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 Files

A small but useful tool is the freeware *Rname-it* (see 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`).

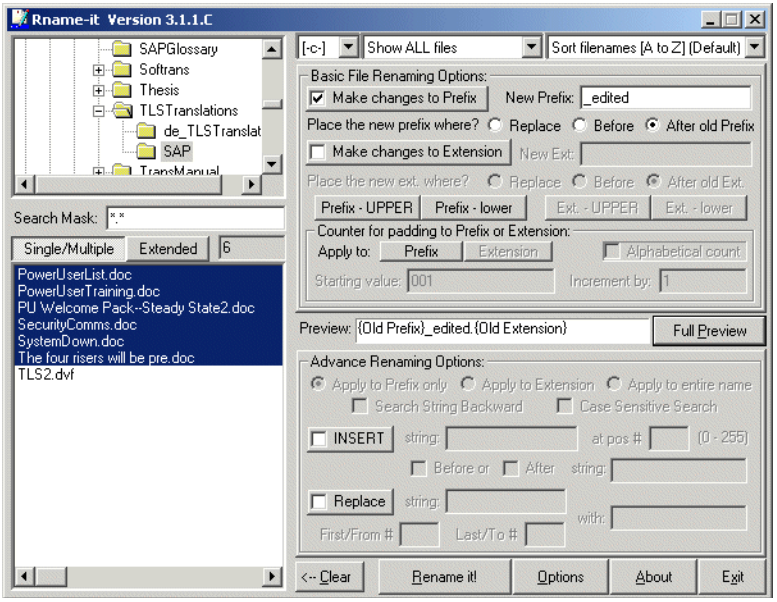


Figure 83: Renaming files in *Rname-it*



*Another useful function of this utility (under **Options**) 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 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.

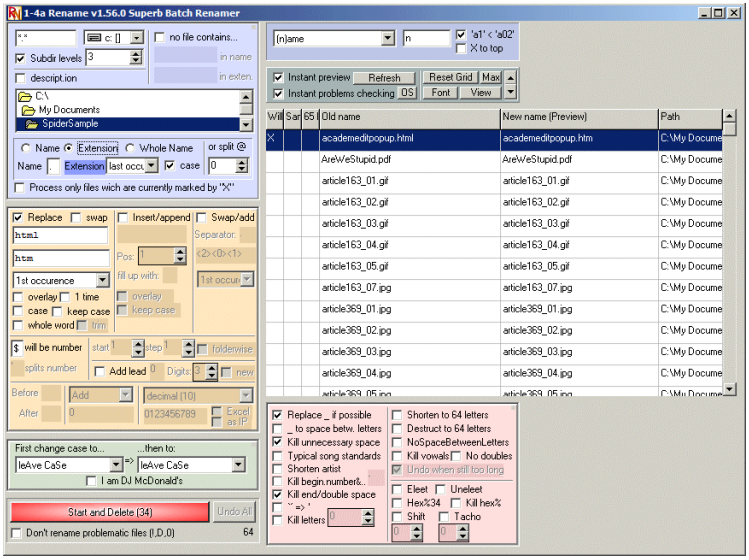


Figure 84: *1-4a Rename's* interface

Searching Content

While the text editors described elsewhere (see *Text Editors* on page 103) have powerful search and replace functions, some utilities specialize in that and offer some additional features that can prove to be very handy.

Replace Studio Pro (see funduc.com/replace_studio_pro.htm), the successor of the well-known *Search & Replace*, can be opened on any individual file or folder from the *Windows/File Explorer* or any other folder view through the right-click menu. It is able to conduct search & replace processes in any text-based file (XML, HTML, TXT, etc.) as well as *Office* "X" files (DOCX, XLSX, PPTX) and *Apache OpenOffice/LibreOffice* formats. You can also search (but not replace) in virtually any other file format.

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—for information on wildcards, see *To Search with Wildcards* on page 20.) and a search string, as well as the string to replace it with.

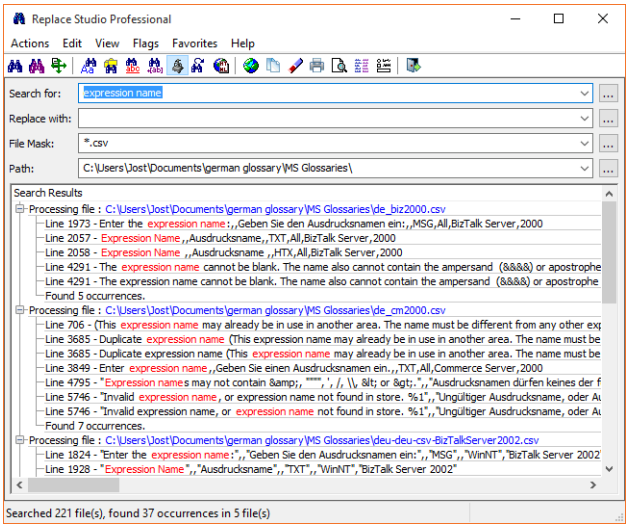


Figure 85: *Replace Studio Pro* after a massive search in CSV files

When R*eplace Studio Pro* 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* (dtsearch.com) or *Archivarius 3000* (likasoft.com/document-search).

Unlike *Replace Studio Pro*, 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*, 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 *Windows* under **(Start>) Control Panel> Indexing Options**.

A search tool for translation-related things outside as well as on your own computer is the wonderful *IntelliWebSearch* (see intelliwebsearch.com). This 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 an unlimited number of customizable search engines, online dictionaries or dictionaries stored on your hard drive.

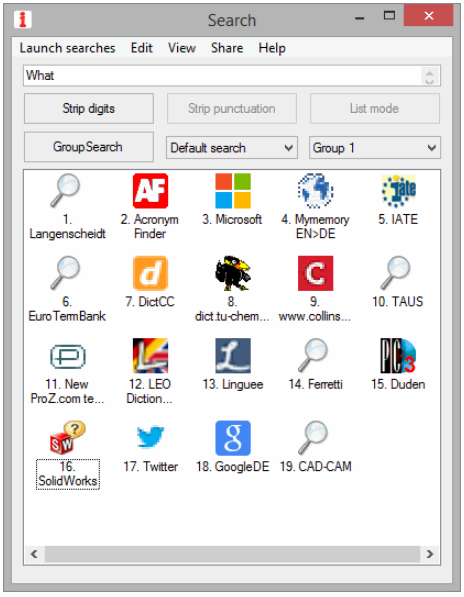


Figure 86: *IntelliWebSearch* with customized search options

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 as well as some good wizards. With a little bit of patience and the help provided under intelliwebsearch/version-5/help, it's quite easy to develop your own searches.



It is also possible to find preconfigured searches at intelliwebsearch.com/version-5/import/#import-from-database or contribute your own searches at intelliwebsearch.com/contributesearch.asp.

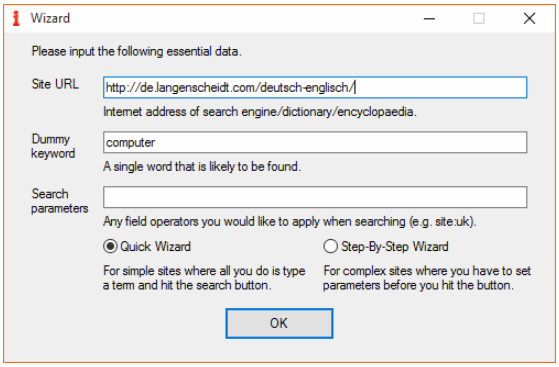


Figure 87: *IntelliWebSearch*'s wizard to define a search

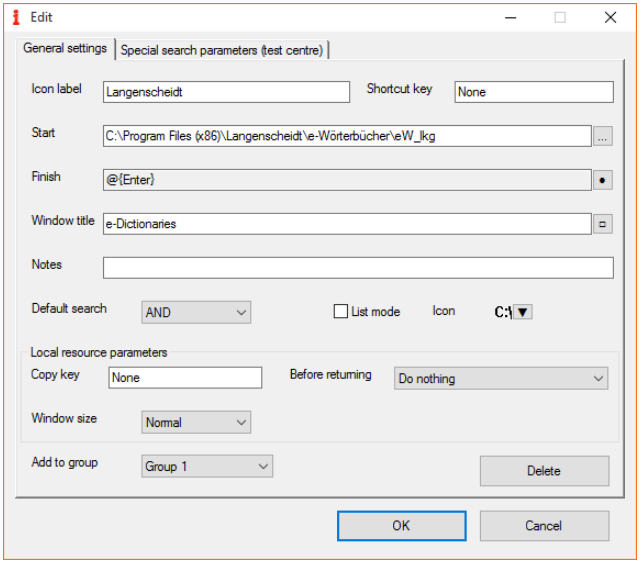


Figure 88: *IntelliWebSearch* with customized search for computer-based dictionary

Another interesting feature in *IntelliWebSearch* is "obfuscation data," which refers to the user's ability to create project-specific lists of terms that you want replaced when sending a query to, say, *Google Translate*, so the data is anonymized and cannot be pointed to a specific client. For instance, if you translate something for Samsung, you can automatically replace "Samsung" with "Apple" so in any segment you send to Google for a pretranslation the two terms are exchanged. And only when you paste it back into your translation is it automatically changed back to the original term.

A very powerful data source (which you can access through *IntelliWebSearch's* interface) is *Linguee* (see linguee.com), a very large corpus of data of and between all European Union languages plus English <> Chinese, Japanese and Russian.

The team behind *Linguee* has found ways to have web crawlers detect translated content online (plus they use all the EU's multilingual data) and match that up with the help of a 50,000+ entry dictionary and other web-based dictionaries. To look up a term or complete phrase, just enter it into the search box; the matches that are displayed are complete segment matches

with the terms in question (both in source and target) highlighted. At first glance the data contains no metadata (origin, subject matter, etc.), but at second glance you will notice the links to the originating sites, giving you all the metadata you could want. You don't have to register to search; as a registered user you can evaluate the translations and correct them, or you can add entries to the dictionary, which in turn are used to fine-tune the matches.

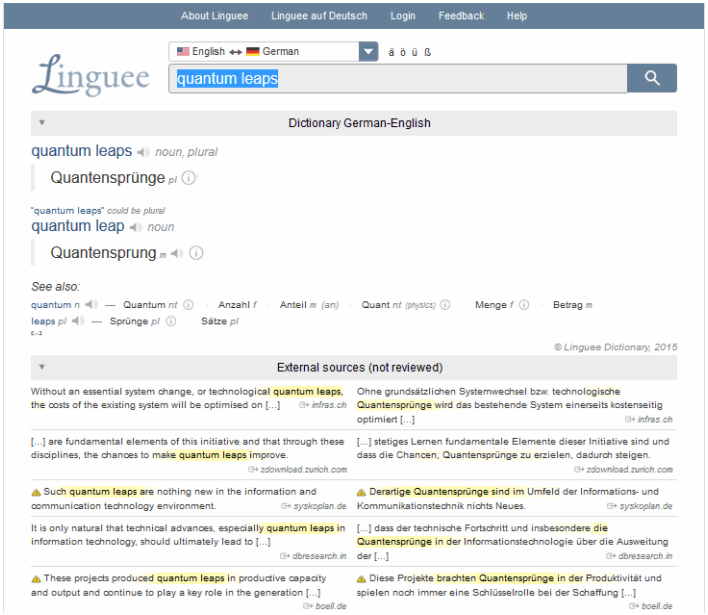


Figure 89: *Linguee's* browser-based search interface

The supported languages all come with in-depth morphological support, so there is an immediate analysis of the grammatical form of the word you are looking for, including a listing of the respective infinitives and their various meanings.

Another resource that can be similarly accessed and provides data of more reliable quality is the *TAUS Data Cloud* (data-app.taus.net). It's a large collection of translation memory data from large translation buyers who mostly come from the software and IT industries. You can search through in a large number of language combinations on a word or phrase level, you can

filter the data by industry and client, and you'll get quick information about the usage percentages of one translation vs. another and so on and so forth. It's really very helpful, and I use it a lot when I work in one of the areas covered by the database.

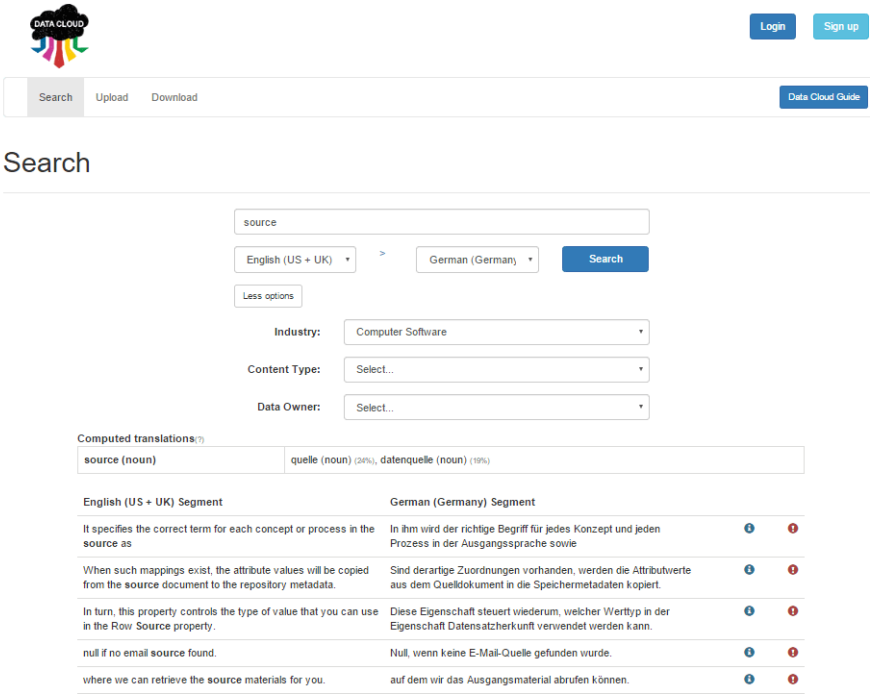


Figure 90: Search results in the *TAUS Data Cloud*

If you are specializing in patents (or really any other technical area), the impressive terminology repository WIPO Pearl (see wipo.int/reference/en/wipopearl) is a must-go-to resource. WIPO ("World Intellectual Property Organization") Pearl contains 180,000 validated scientific and technical patent terms and is available in 10 languages (for both its interface and the data it contains): Arabic, Chinese, English, French, German, Japanese, Korean, Portuguese, Russian, and Spanish. You can do a regular bilingual or multilingual term search that will give you access to a number of filters,

sophisticated definitions of the terms in questions, reliability score, sources, links to Google Images search and the patent database Patentscope, as well as to a "concept map." The concept map is particularly interesting and can be accessed separately as well.

According to WIPO's own definition, the concept map is a "diagram illustrating the relationships existing between concepts in a specific area of scientific or technical knowledge."

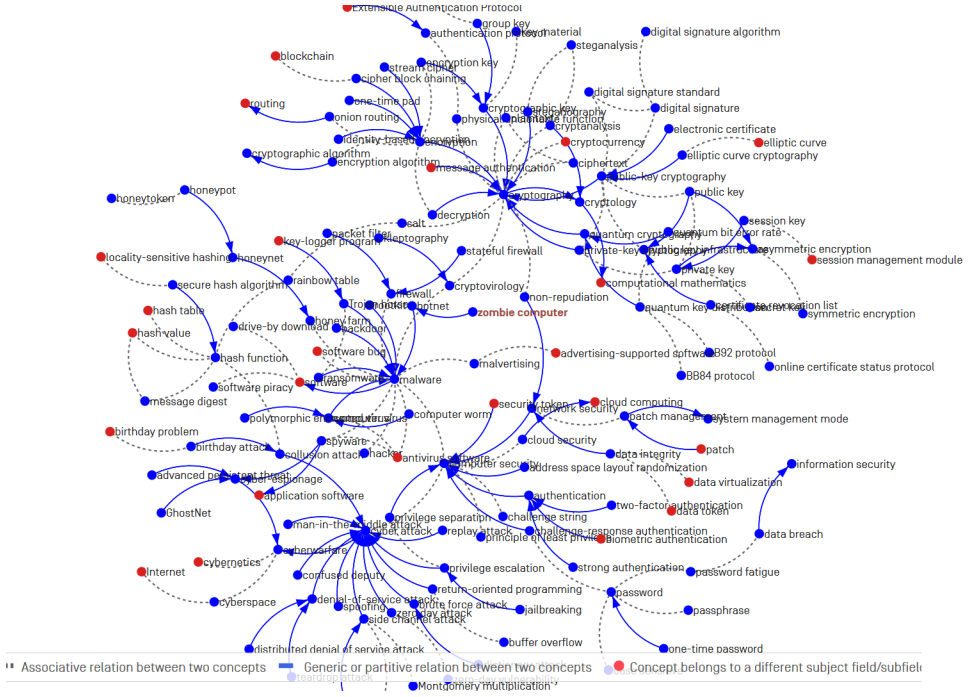


Figure 91: WIPO concept map

Compressing Files

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 pkware.com/pkzip) from the "inventor" of the zip format, and the market leader *WinZip* (see 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 contains its own zip program, but its feature set is very limited.

I'm using the powerful *7-Zip* (see 7-zip.org), which works with a very large 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/File Explorer*, i.e., a right-click on any file, group of files or folder(s) gives you access to the program.



If you exchange a lot of files with other users who use Macs, you might want to look into using Stuffit (see my.smithmicro.com/stuffit-expander-windows.html). Stuffit not only unzips Windows-specific but also Mac-specific compression formats, including SIT and SEA.

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 submenu 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).

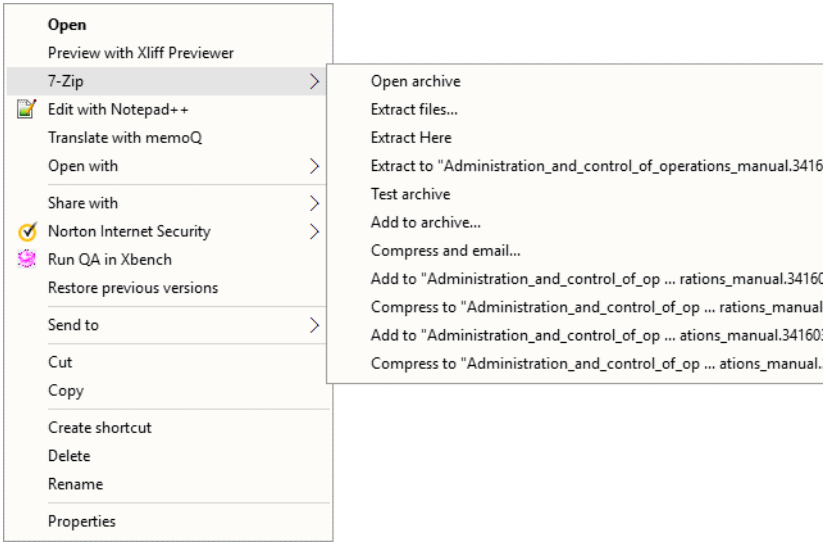


Figure 92: Right-click access to *7-Zip's* options

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.

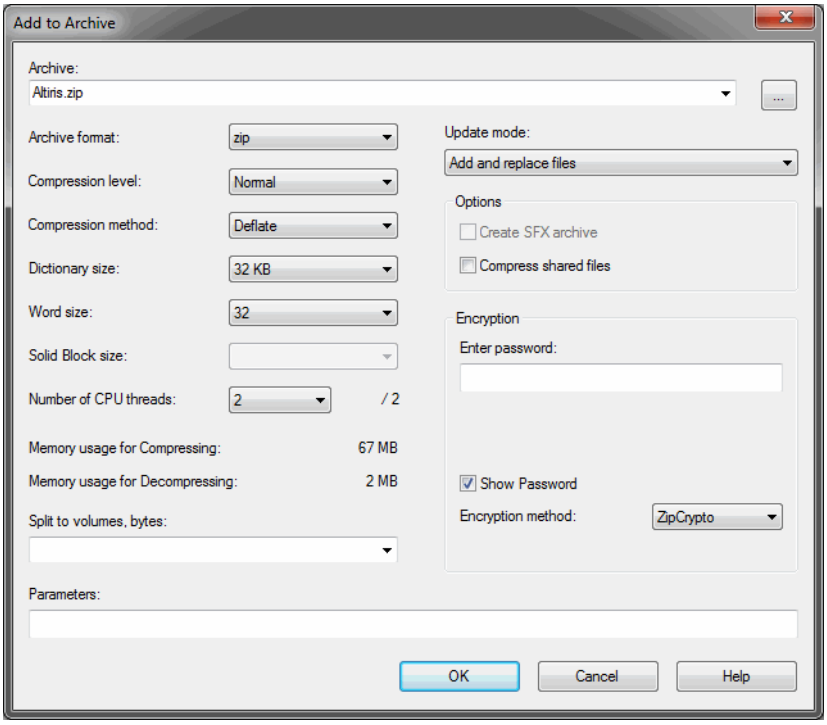


Figure 93: Access to *7-Zip's* password and file splitting features

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.

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, "**Extract to "*\"**," which creates as many folders as there are zip files.

Cracking Passwords

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 elcomsoft.com/tools_for_home_use.html)—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 94: 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.

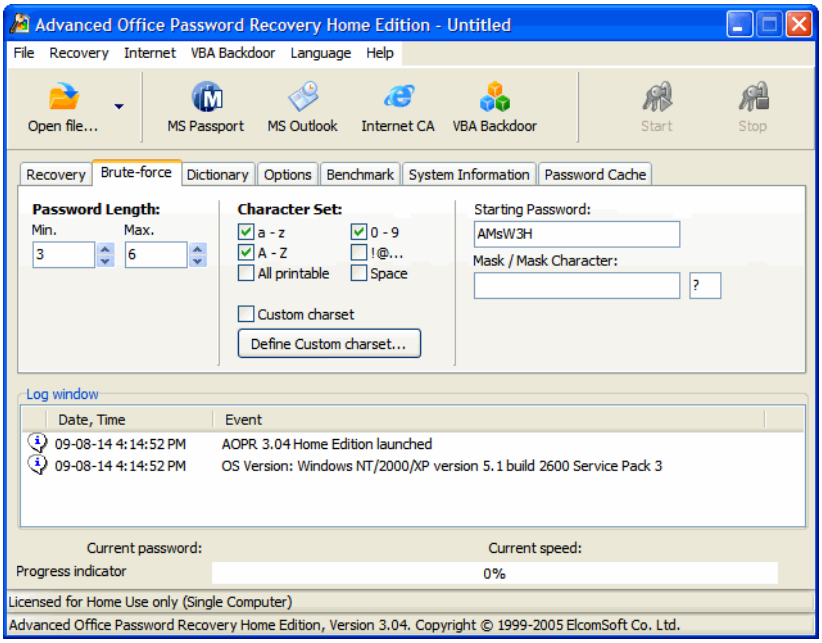


Figure 95: Selecting specific options for a brute force attack for cracking *Office* files

Converting Measurements

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 joshmadison.com/convert-for-windows). *Convert* not only allows you to convert between a multitude of measurements, but even lets you define your own parameters.

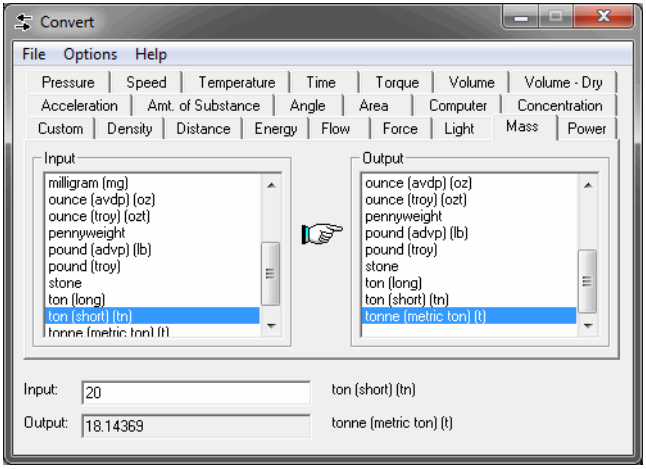


Figure 96: *Convert's* interface—the **Custom** tab gives you access to customizable conversions.

A sometimes overlooked possibility for a wide range of conversions are also presented in the *Windows Calculator* (to start the *Calculator*, press ⊞+R and enter `calc`).

In *Windows 10*, the different conversion types become available when you select the menu icon:

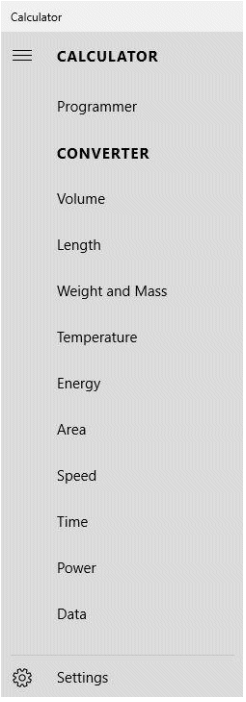


Figure 97: Conversion options in the *Windows 10 Calculator*

In previous versions of *Windows*, select **View> Unit conversion** or press CTRL+U:

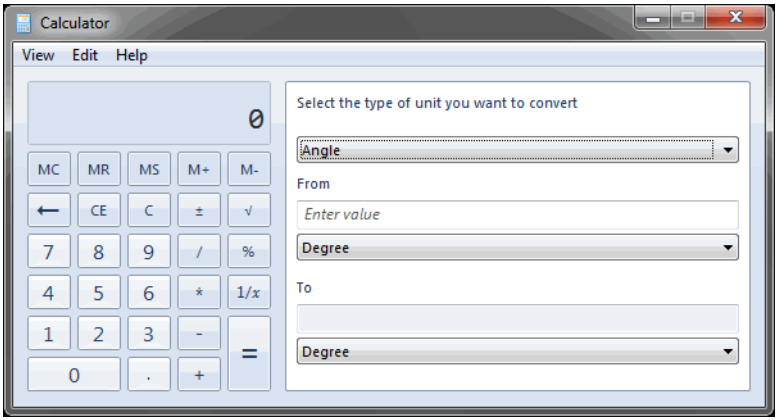


Figure 98: Conversion options in *Windows* prior to version *10*

Counting Words

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, internationalwriters.com/index.html, produced these results with the following tools:

- *Word*: 83 words
- *PractiCount & Invoice*: 154 words
- *Déjà Vu*: 173 words
- *Trados 2007*: 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.
- The text editor *UltraEdit* 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\Firefox` 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.

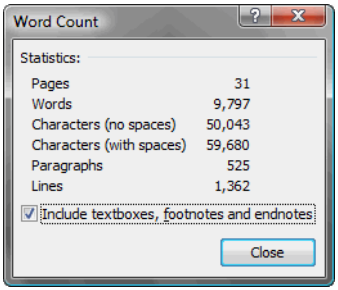


Figure 99: *Word 2007's* word count features



*Word's word count is available at **Review> Proofing> Word Count**. In Word 2007 and above 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. Also, it's not possible to batch count several documents at a time.

I usually choose to do word counts for several documents at a time (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 *AnyCount* (anycount.com) or *PractiCount & Invoice* (see practiline.com), both of which support a very large number of file formats.

Especially *PractiCount* 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.

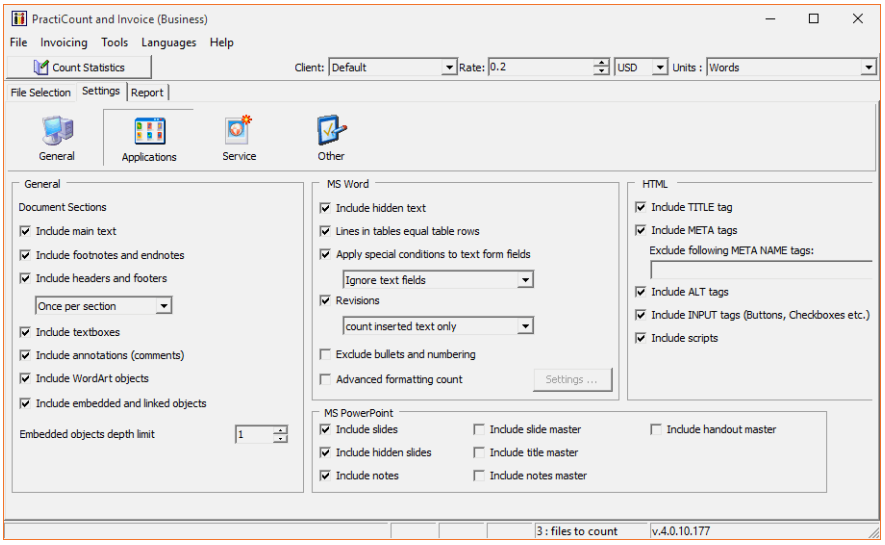


Figure 100: *PractiCount & Invoice's* settings

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>** (**Info>**) **Properties> Statistics** in *Word* and *PowerPoint*.)

Different file types can be mixed for different batch counts:

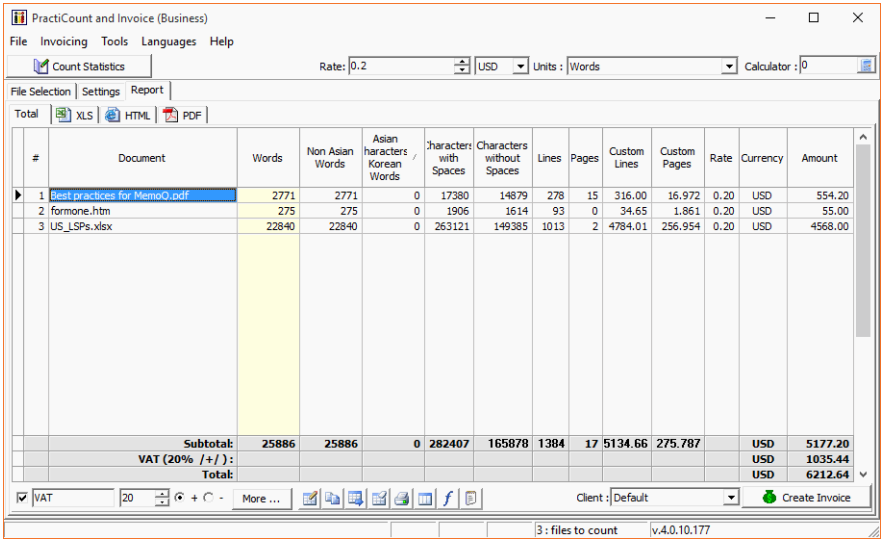


Figure 101: Word count summary view with easy options to save directly in various file formats



*PowerPoint files are counted by PractiCount with PowerPoint's own word count module (within PowerPoint, this is accessible through **Office** button**> Prepare> Properties> Document Properties> Advanced Properties> Statistics** in PowerPoint 2007 and **File> Info** (**Show more properties**) in PowerPoint 2010 and above). Of course, this does not include text on any embedded objects (see page 183).*

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 *Apache OpenOffice/ LibreOffice* under **File> Properties** (or: **Info**)) 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.

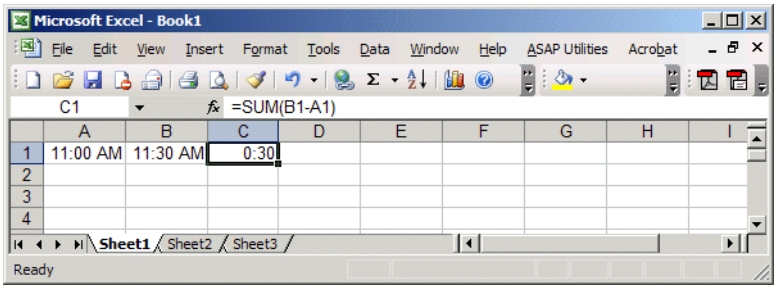


Figure 102: Sample of an *Excel* spreadsheet with formula as described above

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 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.

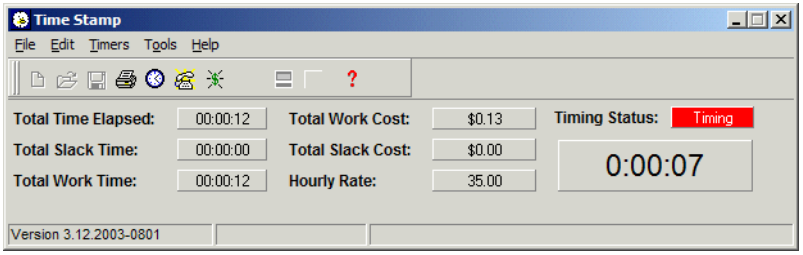


Figure 103: *Time Stamp's* user interface

The Ukrainian software maker AIT released a time tracking tool specifically geared toward language professionals. Similar to its generic counterparts, *ExactSpent* (see 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.

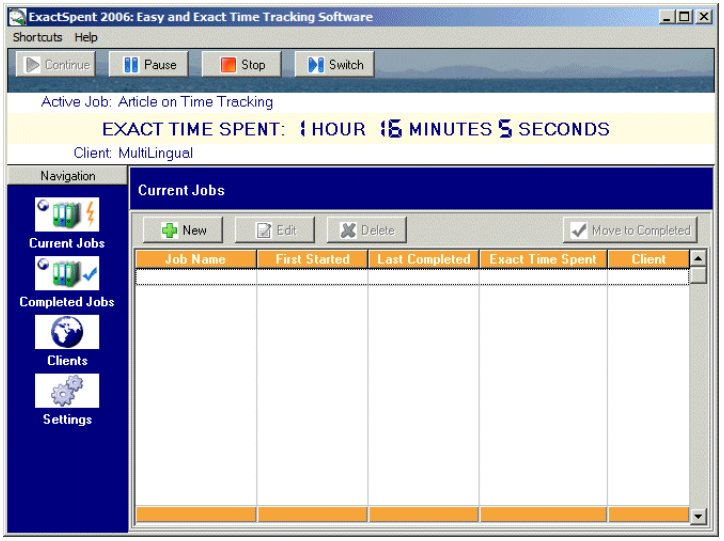


Figure 104: Tracking time with *ExactSpent*

Yet another possibility for tracking time for your various tasks across different machines and devices is with a cloud-based service such as *Toggl* (see toggl.com) or—my current favorite—*Mite* (mite.yo.lk), which many of your fellow translators love to use.

Managing the Clipboard

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

Microsoft Office has included this for some time now under **Edit** (**Home**)**>** (the little pointer next to) **Clipboard**), which allows you to collect up to 24 different clipboard items from anywhere on your computer and paste them individually or all at once into any *Office* document.

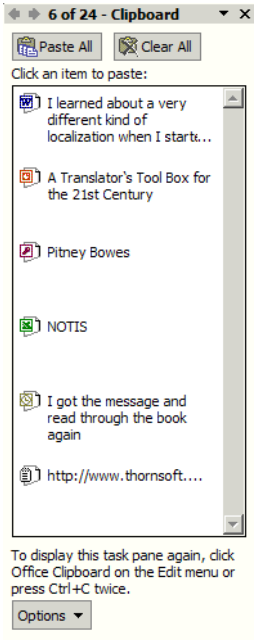


Figure 105: *Office XP Clipboard* with copied content from several *Office* applications

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

In the summer of 2019, Microsoft introduced the *Windows Clipboard* for Windows 10. You'll have to activate it for first-time use with ⊞+V and from then on out access it with that shortcut as well. It will store 25 items at a time for you (up to 4 MB per item); you can store graphics, text, or HTML, it's possible to sync the clipboard across devices; and you can pin items so they don't get deleted (everything else will be deleted when you restart your computer or reach the 25-item limit).

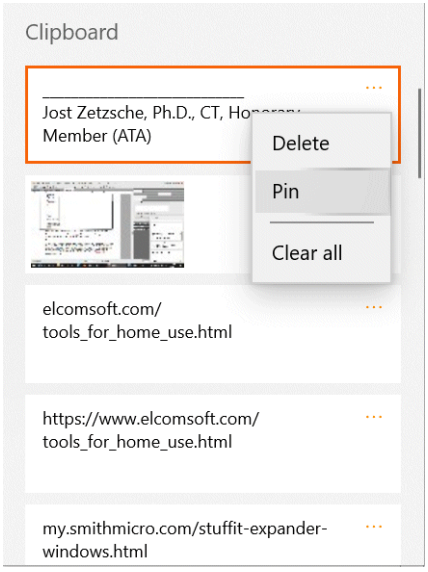


Figure 106: Windows Clipboard

If you want to go (almost) limitless with your clipboard management, including directly printing from your clipboard or storing clipboard entries between different computer sessions (i.e., after I switched the computer on and off) you might want to select *ClipMate* (see 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.

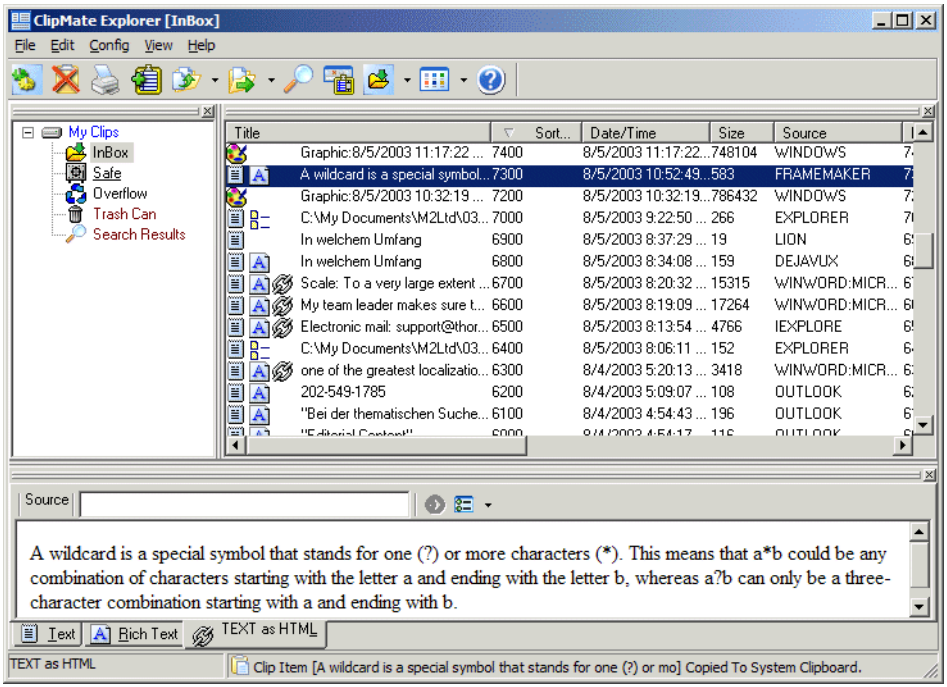


Figure 107: *ClipMate Explorer* with a preview pane (bottom) and a collection view (upper right)

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*.



There are a couple of tricks for Microsoft Word for copying and pasting without actually using the regular clipboard: If you want to copy and paste or cut and paste something within a Word document without placing it on the clipboard (where it would overwrite whatever else you might have there), use SHIFT+F2 to copy (you'll see the message ***Copy to where?*** *in the status bar), place the insertion point in the right location, and press ENTER. (If you want to move (cut) text instead, just use F2 (you'll see* ***Move to where?*** *in the status bar), select the new location, and press ENTER).*

Or you can use the "Spike": Select whatever you want to have moved (cut and pasted), press CTRL+F3, and keep on doing this as often as you want. Once everything is collected, press CTRL+SHIFT+F3.

In both of these cases, the clipboard's content remains the same than what it was before you first copied.

Taking Screenshots

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
- PRINTSCREEN for the complete screen
- ⊞+PRINTSCREEN for the complete screen with an automated save as a graphic files under *Pictures> Screenshots*
- ⊞+SHIFT+S: Activates screenshot app (only in *Windows 10*)

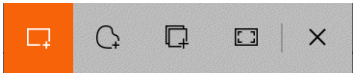


Figure 108: Windows screenshot app

Also available within *Windows* is the *Snipping Tool*:

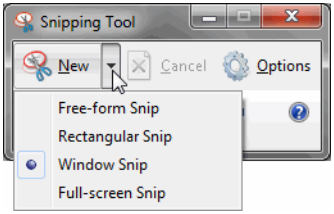


Figure 109: Windows *Snipping Tool*

There are also third-party programs that specialize in taking screenshots, and while they don't fix everything, they are a lot more versatile than what *Windows* offers.

With a screenshot application like *SnagIt* (see 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.

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.

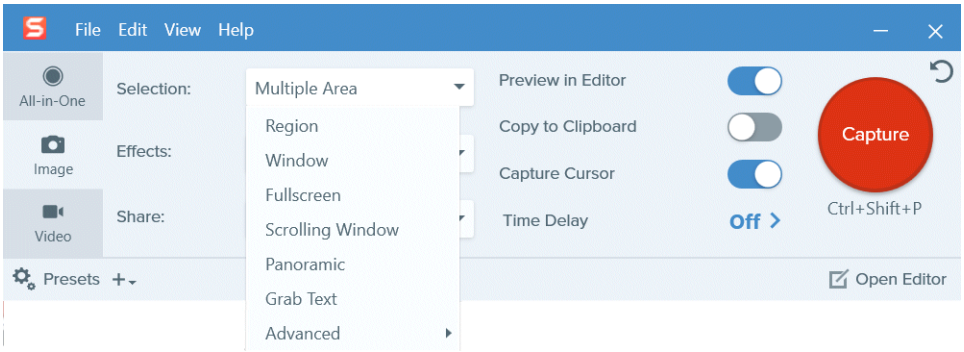


Figure 110: *SnagIt* options



Another free screenshot tool with not quite as many options is Greenshot (see greenshot.sourceforge.net).

A tool that specifically allow you to "harvest" text from screenshots is the *ABBYY Screenshot Reader* (see abbyy.com/screenshot_reader). This can be for instance very helpful in copying chunks of texts from non-copyable PDF or other file formats or from dialog boxes for localization purposes.

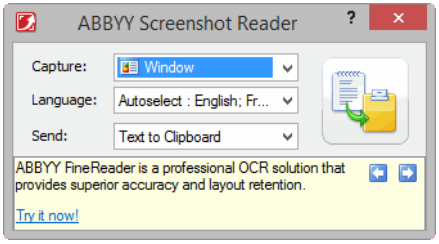


Figure 111: *ABBYY Screenshot Reader*

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 twins-software.com) is no longer supported by its makers, but it's free and it allows you to merge as many MP3, MPEG, text and *Word* files as you would like. 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.

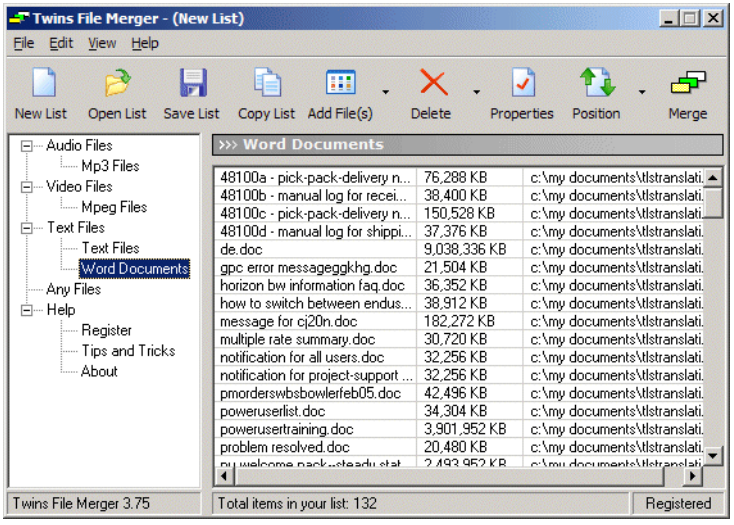


Figure 112: Merging a large number of files in *File Merger*

Dealing with 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, 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).

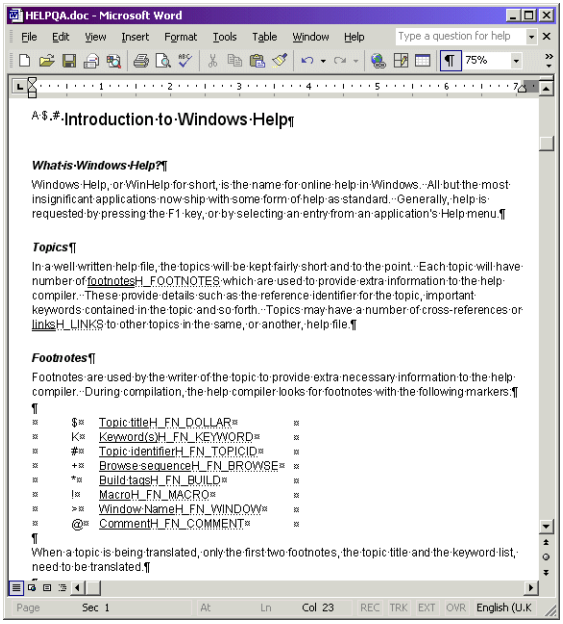
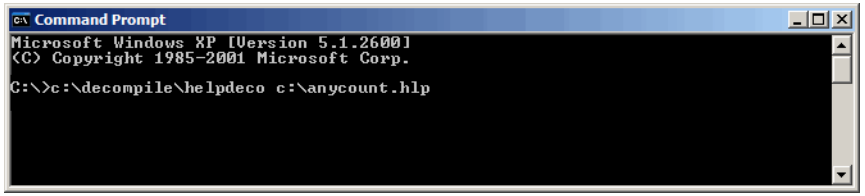


Figure 113: View of an RTF file before its compilation into a WinHelp help system

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 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**.

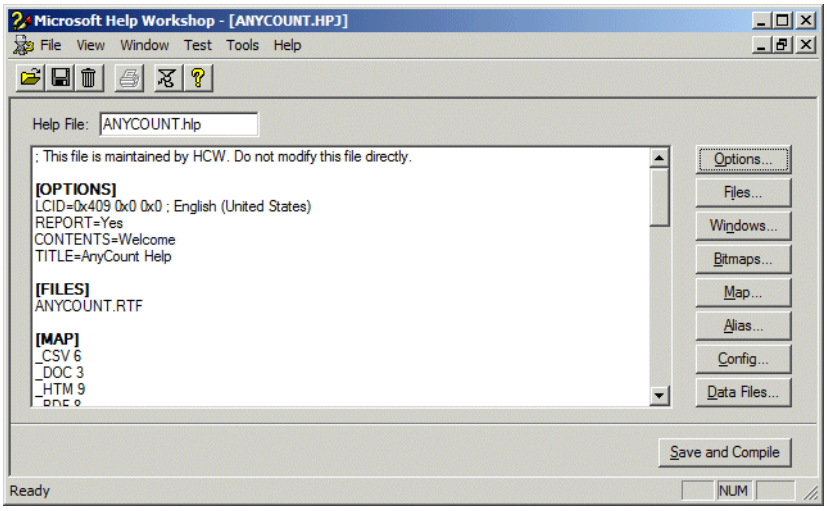


Figure 114: WinHelp compilation in *Microsoft Help Workshop*

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*.

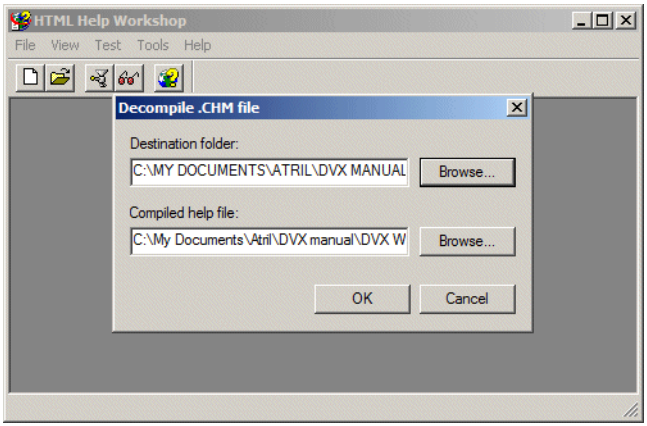


Figure 115: HTMLHelp decompilation in *HTML Help Workshop*



You can download the latest version at docs.microsoft.com/previous-versions/windows/desktop/htmlhelp/microsoft-html-help-downloads.

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 re-compile 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 counterparts 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*, 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 re-compile the individual files with *HTML Help Workshop*.

Once you've fixed any possible 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.

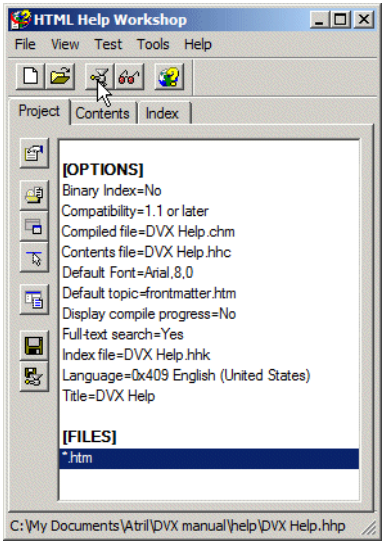


Figure 116: HTMLHelp compilation in *HTML Help Workshop*



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.

Another concept of translating HTML help files is offered by localization tools like *Catalyst or Passolo* (see *Software Localization Tools* on page 313) or with the translation environment tool *Publisher*. Here the decompilation and re-compilation 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 re-compile the CHM files once the translation is done.

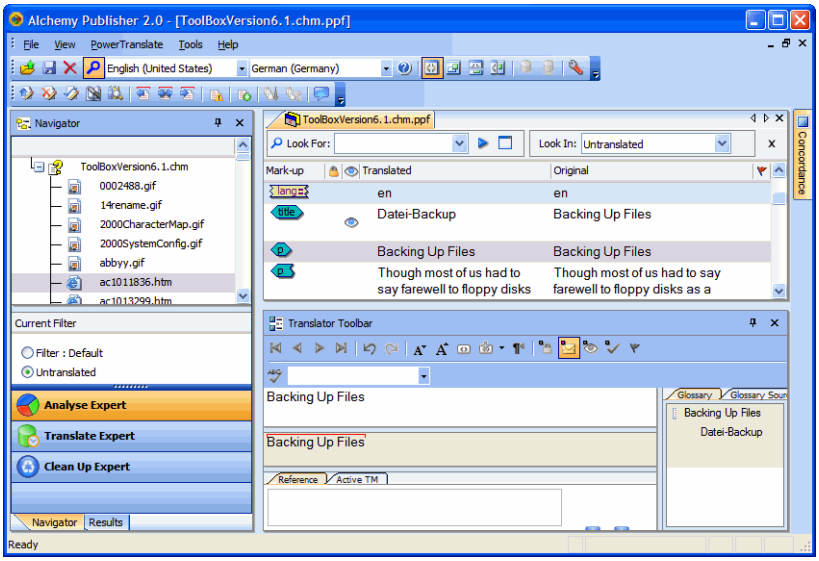


Figure 117: HTMLHelp file translation in *Publisher*



If you are translating a help system that was created in MadCap Flare, MadCap Lingo would certainly be a good choice for a translation environment tool (see page 221).

Installing Many Utilities At Once

If you start from scratch with a new computer and don't want to install all your many programs one by one, there is good news: *Ninite* (see ninite.com) is a super-cool (and free) tool that allows you to select appropriate versions from the available utilities offered on its website bundle into one single download and then install them on your computer in one go. Here is a list of the presently (January 2020) available options:



Figure 118: Available utilities in *Ninite*

Office Suites

It's hard to imagine that a translator could work productively without 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 *LibreOffice* (see libreoffice.org), an active development that stems from the also still available *Apache OpenOffice* (see openoffice.org),
- *Docs*, *Sheets* and *Slides* in *Google Drive* (formerly *Documents*, *Spreadsheets* and *Presentations* in *Google Docs*) (see drive.google.com), and of course
- *Microsoft Office* (see products.office.com).

These products 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 especially *LibreOffice* and *Apache OpenOffice* 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 *LibreOffice*/*Apache OpenOffice* as well as *Microsoft Office 2007* and above 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 112)—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, it's really 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.

Since *Office 2013* grammar and spell-checkers are now freely available for all languages (that are supported in the first place). Why is that even important to mention? Because Microsoft had offered a roller-coaster ride of sorts for how to multi-language authoring which typically involved paying for additional spell- and grammar-checkers. The way it is now is great, and it's made even better by the fact that you are asked automatically whether you want to install a new language once it's detected in your text (if you find the reminders annoying, you can disable them under **File> Options> Language**).

Starting with *Office 2013*, Microsoft also began offering the Office suite as a subscription service called *Office 365*. Since this provides Microsoft with an ongoing revenue stream, it pushes that version very hard over its versions with a perpetual license, often making it hard to purchase anything else.

If you have *Office 365*, your particular version of *Office* corresponds to the most current version of the perpetually licensed *Office* version—or is, in fact slightly more advanced since that version is automatically updated.

Under **File> Options> Language** you'll find the option for a 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.

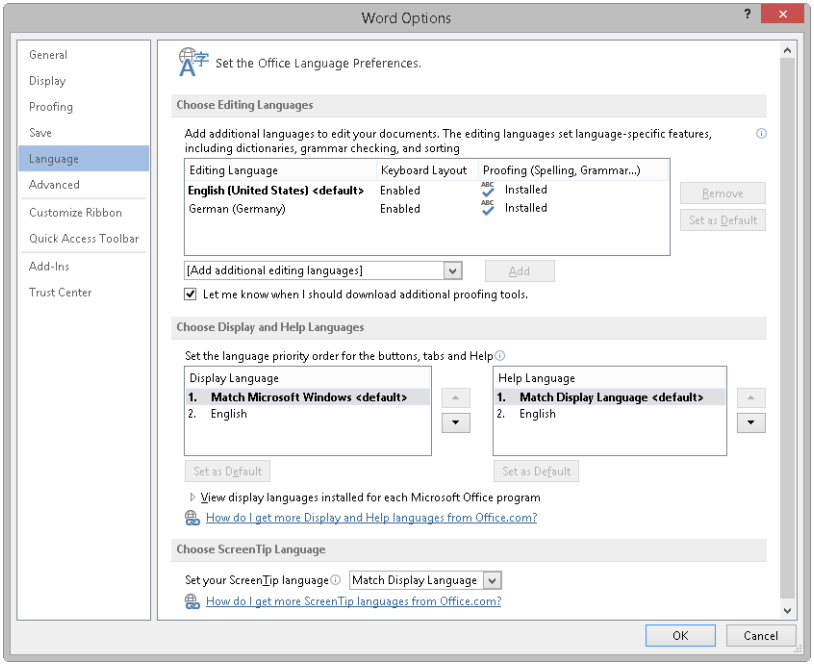


Figure 119: Language options in *Word 2013*

Compatibility

The different applications of *Office 2007* and above 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* and above applications to an earlier format (select the Office button (**File**)**> Save as**).

Working with Office

Aside from *Outlook*, which I use as my email client, as a translator 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. 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*.

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:

The Command Center

The **Options** dialog 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. 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* and above you access it by selecting **File> Options**.

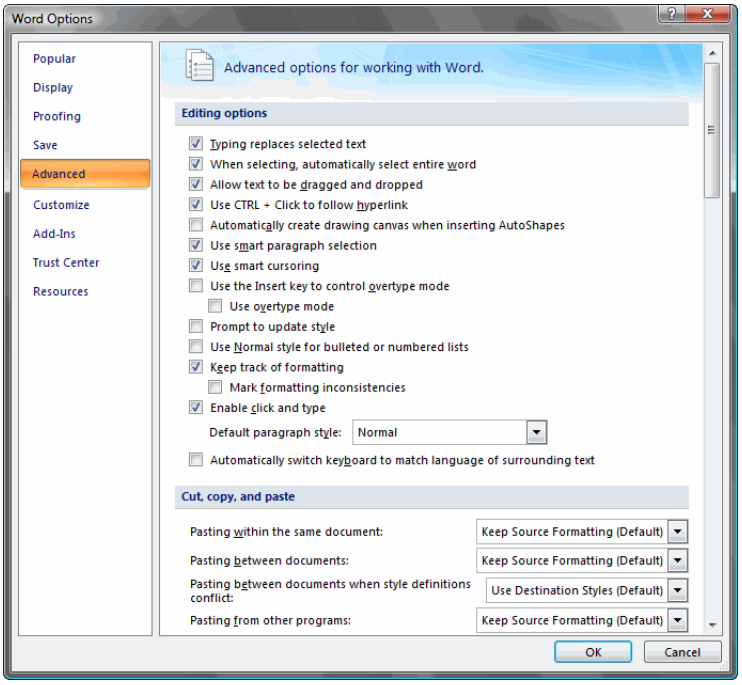


Figure 120: The **Advanced** tab on the **Options** dialog in *Word 2007*

The Paintbrush

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 double-clicking the icon, it 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 CTRL+SPACE.

Editing Environments

Early versions of *Word* (up to *2003*) had a *DOS*/*WordPerfect* emulation mode that allowed you to change the screen to blue and the font to white (under **Tools> Options> General> Blue background, white text**). Many used this for proofreading purposes because it offered a new perspective on the

text and seemed to illuminate typos. For some reason it was dropped from *Word 2007* on, but from *Word 2013* on something similar was silently re-introduced. All you need to do is select **View> Read Mode**, and within the *Read* mode select **View> Page Color> Inverse**.

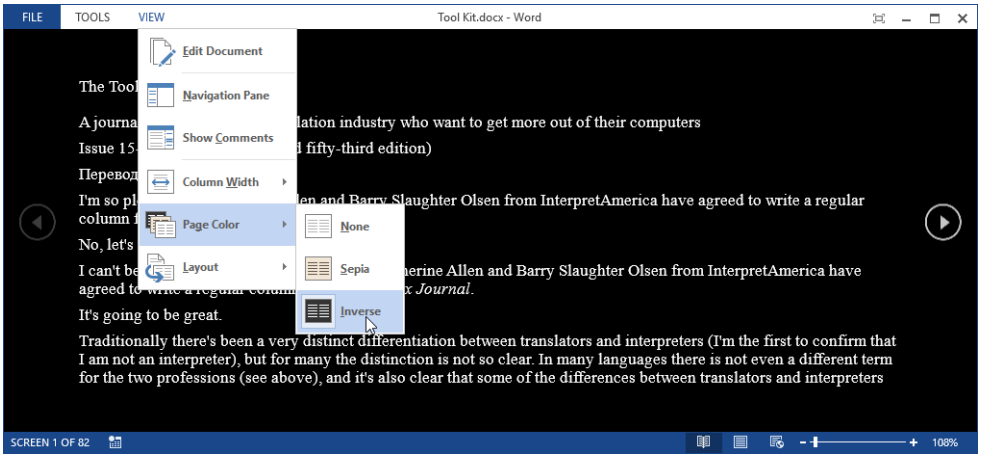


Figure 121: Inverse page color mode in *Word*

In the summer of 2019 these features have been made more prominently available in two options called "Focus" and "Learning Tools" (both accessible under **View> Immersive**).

The **Focus** mode presents you with a full-screen mode (the ribbon bar only appears—in black—when you place your cursor at the very top of your screen), and unlike the (previous and current) **Read** mode, it allows you to write as well as read. On the status bar of *Word* there is even a link to quickly change into that mode.



Figure 122: Focus button on the status bar

Learning Tools comes with its own menu:

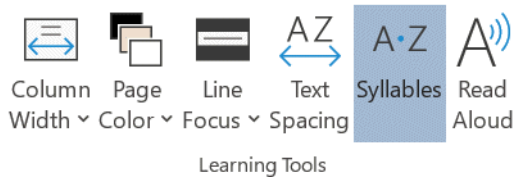


Figure 123: Learning Tools menu

We all have different preferences for how to focus in on a text to catch errors. I like the **Page Color** options (it also changes the font color, but only for the purpose of reading the text without actually changing it for good). Other might prefer to change the **Column Width** (also only temporarily), focus on only a few lines at a time (**Line Focus**), or have the text read aloud (with every word highlighted as it is read).

Naturally not all languages are supported for all options. Text spacing, for instance, is not available for languages with complex or connected scripts, such as Arabic. Syllabification is not available for languages without syllables, such as Chinese, but it is accessible for three dozen European languages.

Word also offers a **View Side by Side** feature (accessible through **View> Windows**) that allows you to simultaneously scroll through two texts at the same time. This can be helpful as you edit or proofread.

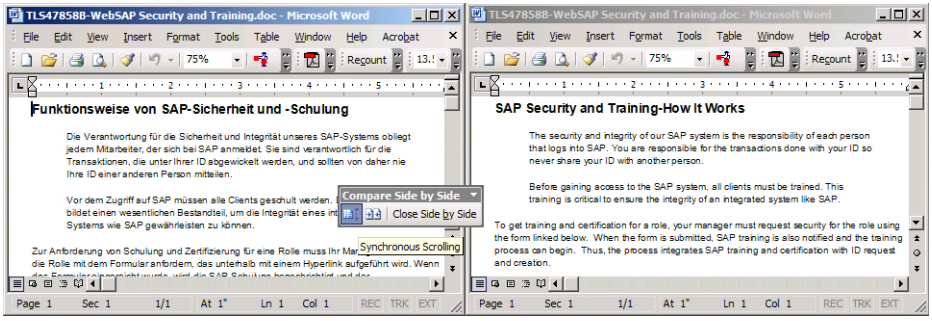


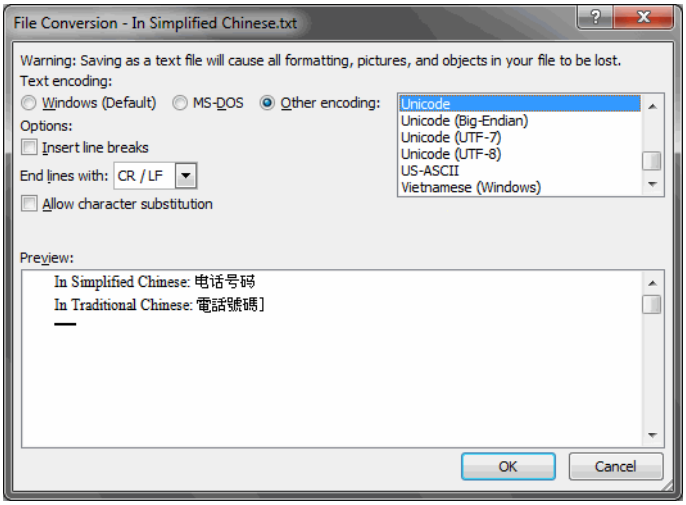
Figure 124: Simultaneous scrolling

Unicode

Office has been Unicode-enabled for a long time. This for instance allows you to save a text in a different code page in *Word*, a feature that comes in very handy in many situations.



*You can access this feature under **File> Save as> Plain Text** and above.*



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.

Language Detection

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 **Review> Set Language**).

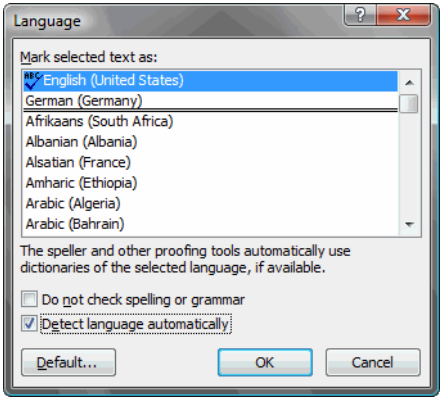


Figure 125: Automatic language detection in *Word*

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.

Track Changes

If you use the **Track Changes** functions (under **Review> Track Changes**) you'll need to 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 the appropriate command under **Review> Tracking**), instead, to make sure that you have deleted all edits in a document, select **Review> Accept> Accept all Changes in Document**.

Of course, 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 2013 and above also has the helpful **Simple Markup** feature (on the **Review** ribbon). Here you can show the location of markups without showing all the markups in the text detail, which, as we all know, all too often makes a document virtually unreadable.

Research and Edit Features

Most *Office* applications have a **Research** command (available under **Review** or with the clever key/mouse combination ALT+CLICK) or **Smart Lookup** (available under **References> Smart Lookup** or as a right-click command) 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.

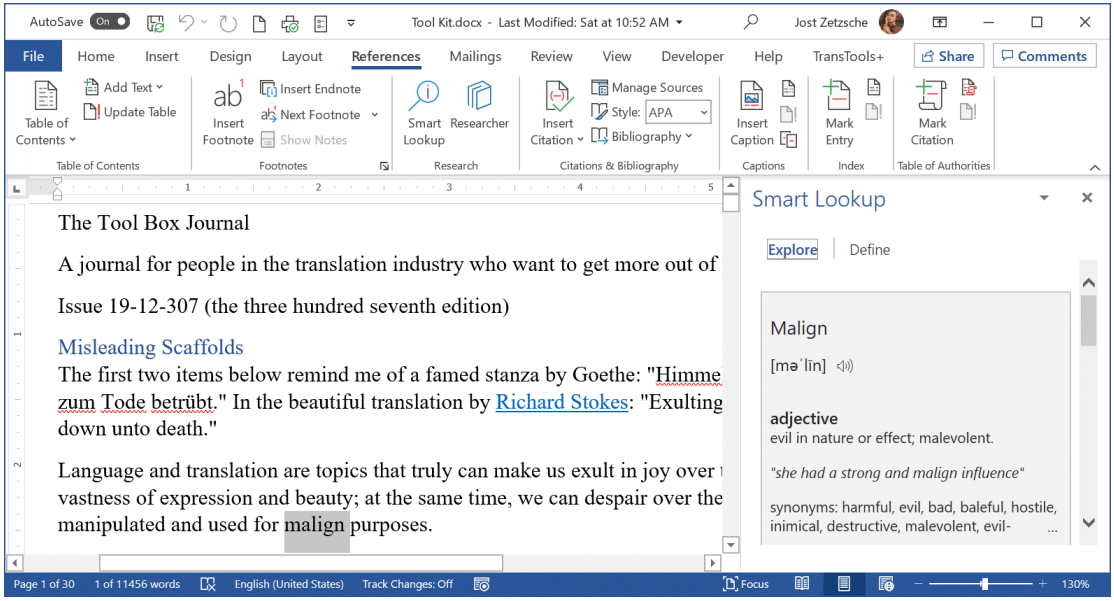


Figure 126: Performing a "Smart Lookup" in Word 365

Office 365 added the "**Rewrite**" feature in late 2019. This feature allows you to right-click on a word or phrase, select **Rewrite**, and see other ways of saying what you meant to say:

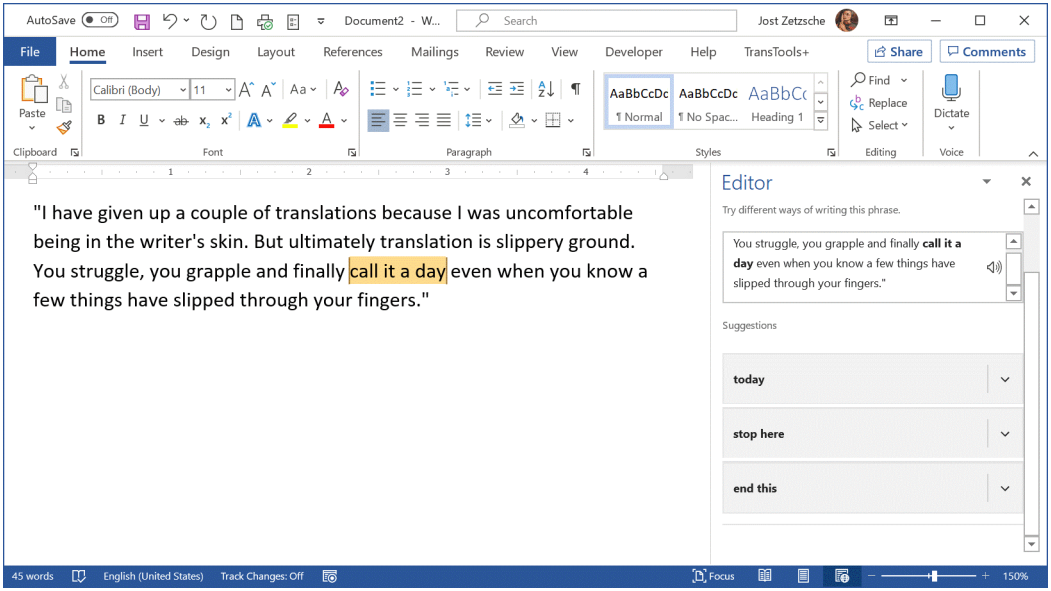


Figure 127: Rewrite feature in Word 365

Of course, this clever quote by Shanta Gokhale in the image shouldn't have to be rewritten but it's still a nice-to-have tool that's at your disposal. (As of January 2020, this tool was available only for English.)



For information on third-party tools for refining and editing documents, see Source Document Quality Assurance on page 297.

Privacy

Office documents store a lot of information about what happened with a document, and sometimes this is not information you might want to share (you might feel that the duration of your work on the document or who also worked on that document should not be anyone else's business). There are

manual ways to remove this information, but these are so manual that it's easy to forget or simply too tedious to do. Here is a quicker way: to remove personal information from a file, open your document and select **Office** button/**File** menu> (*Word*/*PowerPoint*/*Excel*) **Options> Trust Center> Trust Center Settings> Privacy Options> Document Inspector**.



*While you're there (in all versions of Word), you can also find and select the option **Warn before printing, saving, or sending a file that contains tracked changes or comments**—a helpful feature that may avoid some embarrassment caused by sending documents with change-tracked data that was not supposed to be seen by anyone but you, or some frustration when you print a document and forget to turn off tracked changes, thus making it virtually unreadable.*

Search and Replace

Sometimes it's helpful to search and replace something but leave the original text untouched. A scenario where that night be helpful is if you are working on a table where names are listed with the family name first, followed by a comma, followed by the given name:

```
Smith, Roland
Doe, Jane
Kulongowski, Vladimir
```

Now your client wants you to change that for the translated version, and you need to sort this into family name following the given name. To do this, copy the table into a standalone *Word* document, press CTRL+H to open the **Find and Replace** dialog in *Word*, select the **More** button to open up the extended options, and select **Use Wildcards**. Then enter:

```
(<*>), (<*>)
```

(< = beginning of a word, * = 0 or more characters, > = end of a word, followed by a comma and a white space, followed by another beginning of a word, 0 or more characters, end of a word)

to be replaced with

```
\2 \1
```

(second referable field, followed by one white space, followed by the first referable field)

The result will be this:

```
Roland Smith
Jane Doe
Vladimir Kulongowski
```



This feature is also helpful when you want to convert time or date formats.

If you are regularly in the need of very complex and multi-pronged search-and-replace processes and would even like to store them for later reuse, then the *Multiple Find and Replace* tool, which is part of the *TransTools*+ collection of tools (see translatortools.net/products/transtoolsplus), might be a welcome program.

As the name implies, it can run a number of search-and-replace processes simultaneously and, as in the other tools, save that list of processes for later re-use. The *Multiple Find and Replace* tool also provides access to a huge range of regular expressions beyond those offered in *Word*, alongside descriptors that help you choose which one to use:

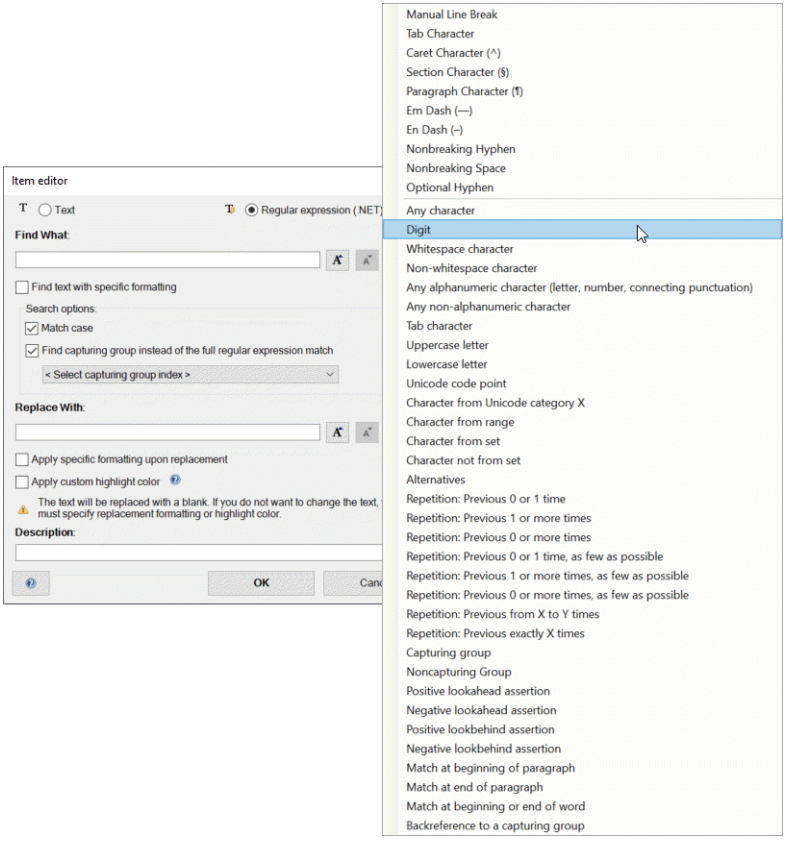


Figure 128: List of preconfigured regular expressions in *TransPlus+'s Multiple Search & Replace* tool

A separate document (see translatortools.net/download/TransToolsPlus_Brochure.pdf) provides a helpful overview of ways to use the various search features, and a preconfigured list of expressions for dealing with common problems after using optical character recognition on a document also serves as a good guide for how to put together a list:

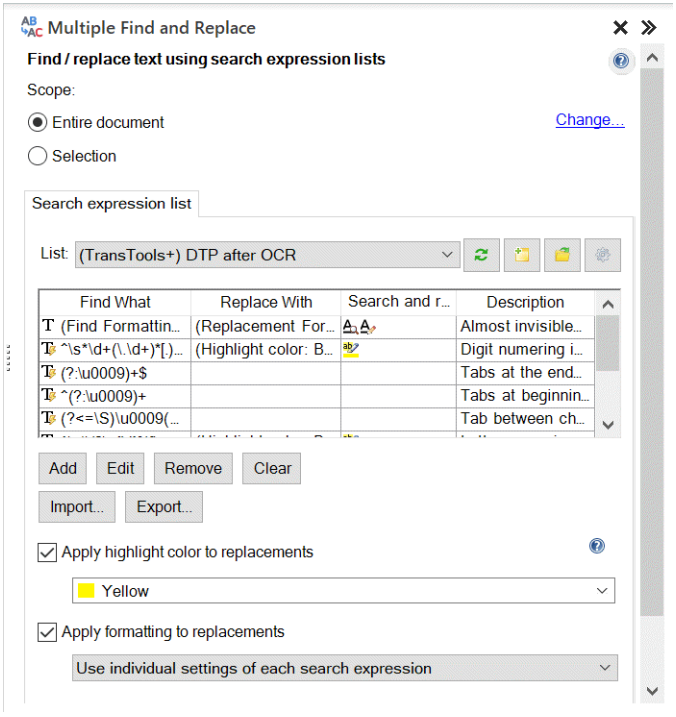


Figure 129: Preconfigured list of search expressions in *TransPlus+'s Multiple Search & Replace* tool

Using Templates

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 **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.

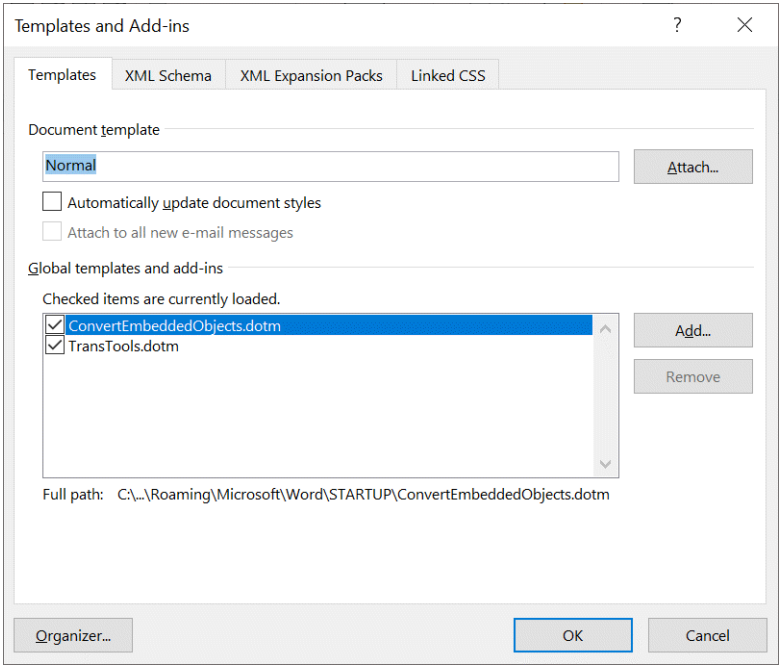


Figure 130: Selecting and deselecting templates in *Word*

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/File 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:\Users\<user>\AppData\Roaming\Microsoft\Word\STARTUP`

and

`C:\Program Files\Microsoft Office\Office<X>\Startup .`



*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 18).*

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

`C:\Users\<user>\AppData\Roaming\Microsoft\Templates.`



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 **Developer> Word Add-Ins**, adding the templates in question, and activating them.

Working with Excel

Browser Integration

Excel provides for an integration into *Internet Explorer*. 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 an *Excel* spreadsheet. This is great for copying glossaries.

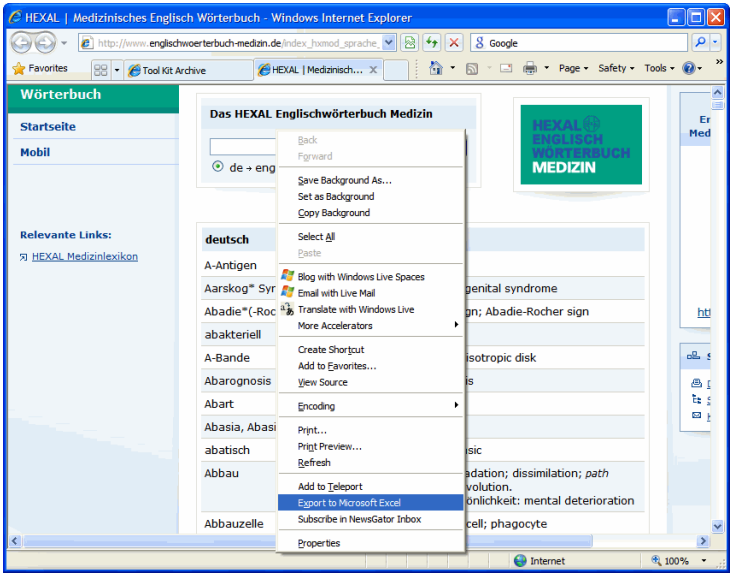


Figure 131: Exporting glossaries from the web to *Excel*

The import of web-based data is also possible right from within Excel: Open *Excel*, select **Data> (Get & Transform Data)> From Web**, enter the URL of the webpage that contains the glossary, select the table in the dialog that is displayed, and click on **Import**.

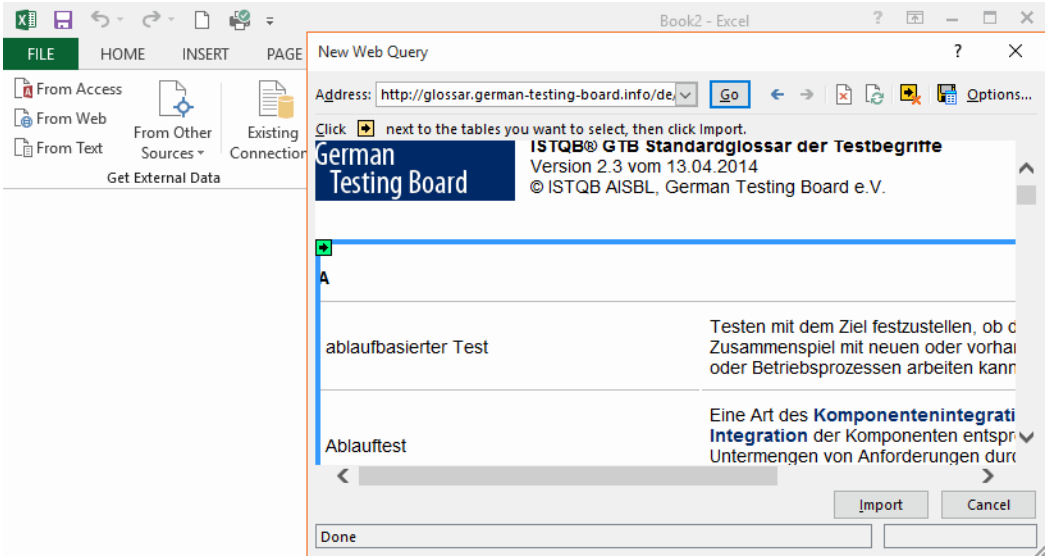


Figure 132: Importing web data into *Excel*

Searching in Excel

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

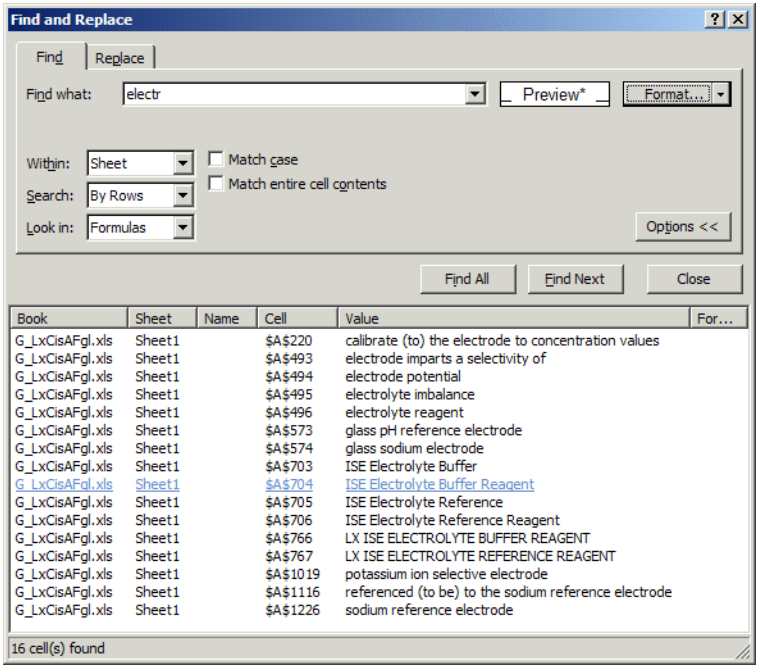


Figure 133: Search results in *Excel 2003*

Different kinds of *Excel*-compatible files

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.



*If you open these files from within Excel, Excel starts a wizard that lets you tell Excel how to segment the text (i.e., how to put the different fields into columns). Well, Excel is actually smarter than you may think, and in most cases it knows how to deal with the file in question. So rather than going through the three- or four-step wizard, you can also force Excel to open the file as it sees fit by selecting **File> Open**, locating the file that needs to be imported, and pressing the SHIFT key while you click **Open**. This way Excel simply uses its best judgment to open the file correctly without the wizard.*

Using the **Fill** feature in *Excel*

If you have a glossary with source and target information, you might want to enter some additional data to that table—such as subject matter, client, or whether the data is approved—before importing it into your terminology database. Rather than going through some convoluted process of entering and multiplying the data in the third, fourth, and fifth columns, you can simply enter the record of interest in the first cell of the respective column, select that cell and all the other cells you want with that data, select **Fill** and **Down** on the **Home** ribbon tab, and there you are. If you would like to do the same with a running number, enter the first number in the first cell and then select **Fill**> **Series**.

A more advanced feature that was introduced in *Excel 2013* is **Flash Fill**. The feature can either be manually activated as one of the options in the **Fill** menu (see above) or with the keyboard shortcut CTRL+E, but it is likely more helpful if it's activated automatically (under **File> Options**: **Advanced> Editing options> Automatically Flash Fill**).

Flash Fill will recognize a pattern once you enter two or more values and then suggest that it automatically fill the remaining column for you. For instance, if you have `<Given Name>` in column A and `<Family Name>` in column B, and you enter `<Family Name, Given Name>` in the first two cells in column C, it will suggest that pattern to you for the rest of the C column.

See the following example of a list of honorary ATA members. The *Excel* preview not only applies the new writing order, it also makes the names upper-case according to the first two entries.

	A	B	C
1	**First Name**	**Last Name**	
2	alison	bertsche	Bertsche, Alison
3	celia	bohannon	Bohannon, Celia
4	lillian	clementi	Clementi, Lillian
5	nicholas	hartmann	Hartmann, Nicholas
6	muriel	jérôme-o'keeffe	Jérôme-O'keeffe, Muriel
7	peter	krawutschke	Krawutschke, Peter
8	freek	lankhof	Lankhof, Freek
9	ann	macfarlane	Macfarlane, Ann
10	rosemary	malia	Malia, Rosemary
11	patricia	newman	Newman, Patricia
12	glenn	nordin	Nordin, Glenn
13	jiri	stejskal	Stejskal, Jiri

Figure 134: Using **Flash Fill** in *Excel*

To accept the preview suggestion, just press ENTER. (As this demonstrates there might be some small errors as in the spelling of "O'keeffe," but those are easy fixes.) This also works with dates, phone numbers, and a host of other things.

Excel Add-Ins

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 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 ribbon bar in *Excel*. If nothing else, you'll enjoy seeing what some of the other 95% of *Excel*'s unused features are . . ..

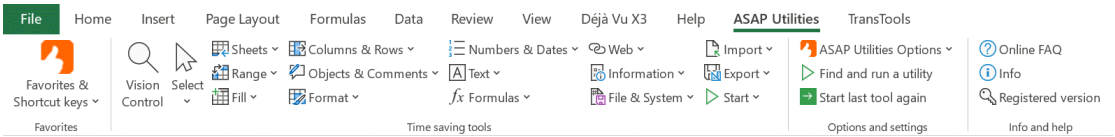


Figure 135: *ASAP Utilities* ribbon bar in *Excel*

A translator-specific collection of add-ins comes with the *TransTools* collection (see translatortools.net/products/transtools#excel). As with *ASAP Utilities*, *TransTools* also installs a ribbon bar that gives you access to a host of translation-specific commands.

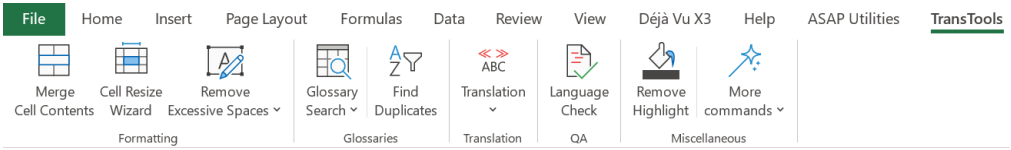


Figure 136: *TransTools* ribbon bar in *Excel*

Most of the commands in the *TransTools* ribbon bar are more or less self-explanatory. One that might be particularly interesting, though, is the *Glossary Search*: it allows you to set up for a system-wide simultaneous search of multiple *Excel*-based glossaries with an independent program that is also installed by *TransTools* and runs in the background.

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.

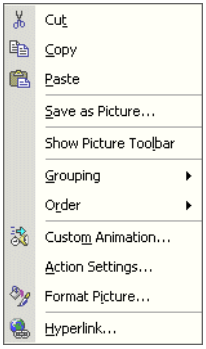


Figure 137: Picture-related commands on *PowerPoint* context menu



For word counts in PowerPoint, see page 137.

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 **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.

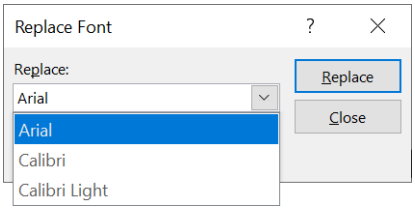


Figure 138: The **Replace Font** dialog in *PowerPoint*

Changing spelling languages in *PowerPoint*

It's easy to change the spell-checking language for individual text boxes in *PowerPoint*, but the program doesn't provide a way to do this for a complete presentation. In previous versions of this book I had a lengthy and very technical description of how to change spell-checking languages in *PowerPoint*. It worked, but describing it as a pain is an understatement.

Fortunately, the *TransTools* collection of utilities (see translatortools.net/products/transtools#powerpoint) provides an easy way out. Even better, it's free.

One of the icons that *TransTools* installs on the **Add-ins** ribbon bar is **Change Language**. (Note that you have to select *PowerPoint* during the installation process to have the *PowerPoint* add-ins installed.)

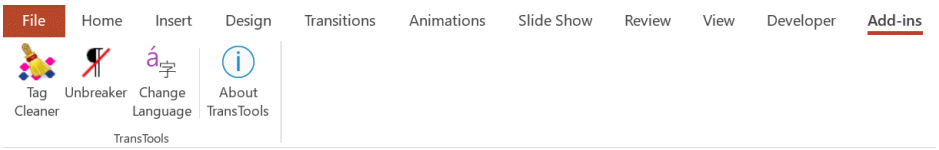


Figure 139: The *PowerPoint TransTools* add-ins

Clicking the **Change Language** icon brings up this dialog, which allows you to change the language setting in your currently open *PowerPoint* presentation.

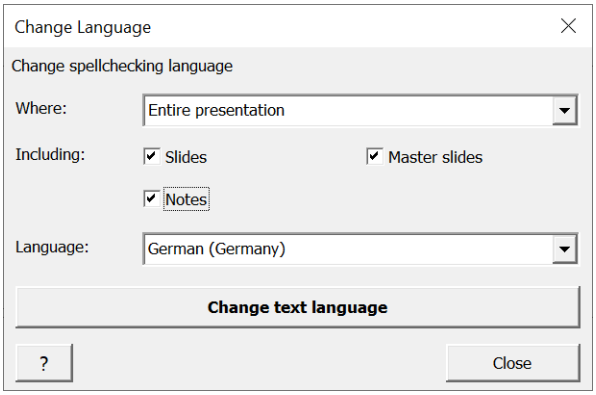


Figure 140: *TransTool's* **Change Language** dialog in *PowerPoint*

Office Documents With Other Embedded Office Documents

There are a number of workarounds for using translation environment tool to process *Office* files with other embedded *Office* files. Often, though you will end up opening the embedded documents, saving them as separate documents, translating them, and re-embedding them.

A few translation environment tools, however, offer a more direct way of dealing with these files.

Star Transit XV and above allows for the direct processing of embedded objects:

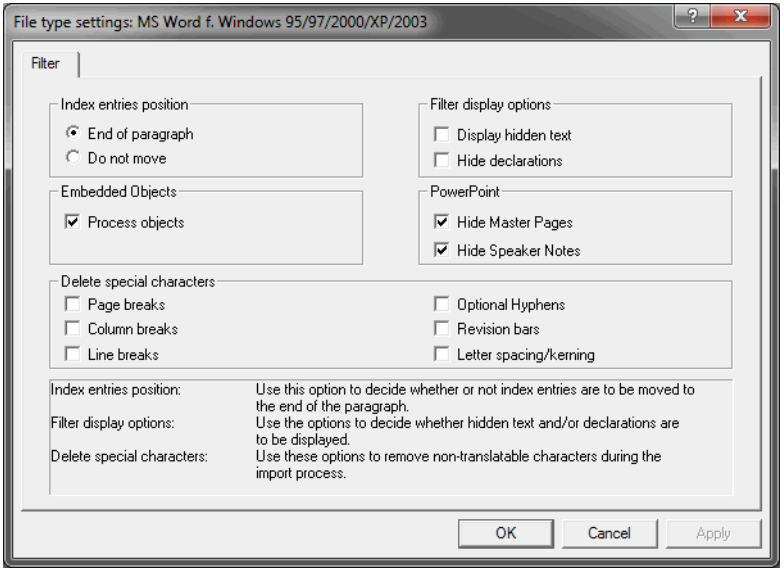


Figure 141: *Star Transit's* option to process embedded objects

And so does *Déjà Vu X2* and above or *Office 2007* files and above:

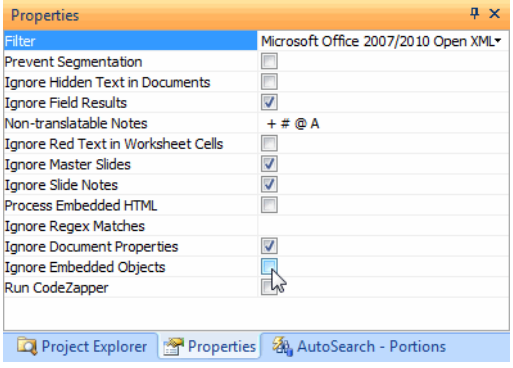


Figure 142: *Déjà Vu's* embedded objects feature

In fact, once you install *Déjà Vu X2* and above on your computer, a **Déjà Vu X2** (**X3**) toolbar will automatically be installed in every *Office* application with the sole purpose of converting earlier versions of embedded *Office* files into *Office* files.

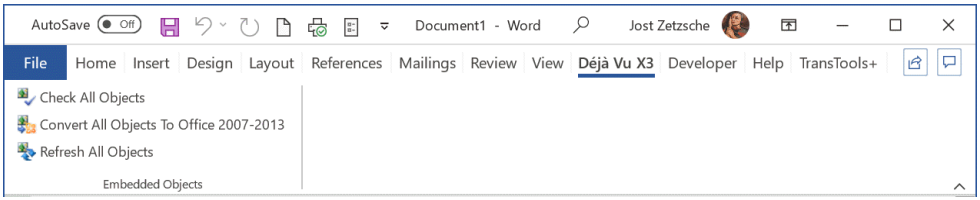


Figure 143: *Déjà Vu X3* toolbar in *MS Word*

memoQ 2014 and above allows for the processing of any embedded supported file type.

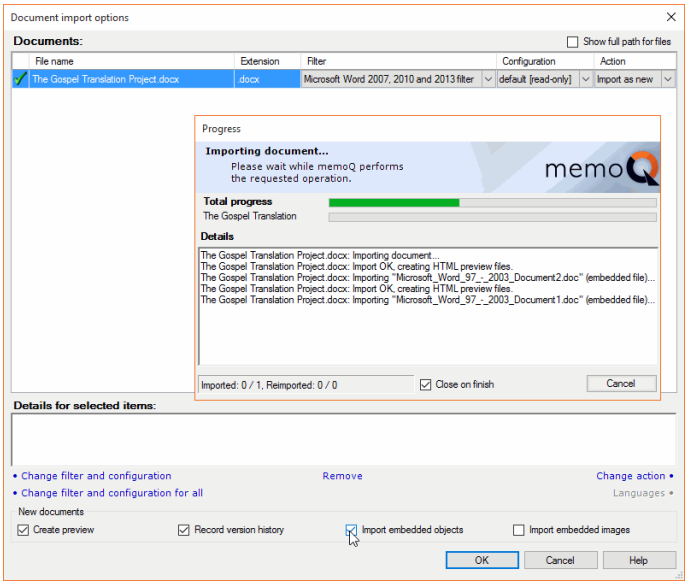


Figure 144: Import of *Word* DOCX document with two embedded *Word* DOC documents in *memoQ 2015*

Trados Studio does not directly support the translation of embedded objects, but you can purchase the third-party app *Extract and Reinject Embedded Objects* (see appstore.sdl.com/app/extract-and-reinject-embedded-objects/434/) that does what its name says.

Computer-Assisted Translation Tools

I am a passionate user of CAT tools and have many of the commercially available (and open-source) 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. That's why 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, Memsource, 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 disproportionately.

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 to 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 wordfinder.com), a cloud-based tool that assembles high-quality dictionary resources and provides lookup from Mac or Windows PCs or mobile devices, and tools like *IntelliWebSearch*, which allows for lookup in online and offline resources (see page 119).
- **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 306). 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 104 and page 309).
- **Project management**—There are a whole slew of tools that provide project management functions for companies, starting with tools like *Plunet*, *XTRF,* and *]project-open[*, and continuing to some of the workflow tools that also provide the features of TEnTs, such as *Star CLM* or the *SDL Trados Business Manager*. 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 325.

- **Word counts**—These are explained in more detail on page 133. 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 313.

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 303.
- **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 desktop publishing or graphic formats. 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 331.
- **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 301.

- **File Optimization**—These tools include utilities like *CodeZapper* or *TransTools* (see page 371) which eliminate most unnecessary codes in *Word* files for easier processing in translation environment tools.
- **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 *. . . A Word of Caution About Alignment* on page 250.
- **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 *Standalone Quality Assurance Tools* on page 291.

And lastly, of course, there are 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 them 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,
- allow translators to connect to a host of other resources, including machine translation, 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.

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).



Curiously, these three tools all were initially developed in the small German town of Böblingen (the home of IBM Deutschland).

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 195) and a number of mergers and acquisitions of translation environment tool vendors as demonstrated by the acquisition of *Trados* by SDL in June of 2005, *Idiom* in 2008, and *MultiTrans* in 2018 (as part of Donnelley Language Solutions), the acquisition of the German *Logoport* by Lionbridge in early 2005, or the acquisition of Wordfast by Translations.com/TransPerfect in 2007.

In 2009, long after IBM had decided to withdraw from the translation environment tool market, another truly big-time player—Google—entered the fray with the release of the *Google Translator Toolkit*. It withdrew it again in 2019.



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

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 262), 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.

Or if you would like to bring machine translation into the range of applications that support your translation work, TEnTs provide for a secure way to do that.

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.



Some of the translation environment tools that have been taken out of this current edition of this book (January 2020) as actively developed tools (though they still maintain websites) are three MS Word-based tools (or LibreOffice/ OpenOffice-based): JiveFusion, Anaphraseus and MetaTexis.

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), terminology database(s), and machine translation engine(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 examples of the more prominent tools to describe the typical features of a TEnT in more detail.



I will not discuss tools that are not accessible by the freelance translator directly but only through a partner in the translation workflow. These include proprietary translation management systems, such as Andrä's ontram (ontram.com), Smartling (smartling.com), Transifex (transifex.com) or Lingotek (lingotek.com); the various LSP-owned translation platforms including Gengo (gengo.com) or One Hour Translation (onehourtranslation.com); and open-source tools like Globalsight Ambassador (globalsight.com). While these tools are becoming increasingly important for our industry, there are a number of distinctions that informed my decision to exclude them. Most of them support exchange standards, but 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 MS 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 sdltrados.com), but since the first release of *Trados Studio* in 2009 it has finally completely abandoned *Word* as a translation interface. Other tools that still support the *Word* interface as their only or their most important interface include

- *Wordfast Classic* (see wordfast.net)—see page 196,
- *SDL MultiTrans* (see sdlmultitrans.com)—see page 198, and
- Lionbridge's *Translation Workspace* (see geoworkz.com)—see page 199.

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 and at tat point 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/TransPerfect.

Wordfast Classic has garnered a significant following among freelance translators, and its reputation 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* (only on *Word 2011* for Mac, though). A drawback is naturally that the amount of supported file formats is limited. While it is theoretically possible (but not encouraged) to tag and process HTML 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 215).

Features that are not immediately apparent in *Wordfast Classic* include the ability to share translation memory data with other translators in real-time (see page 288), a set of relatively sophisticated quality assurance features (see page 279), and an autocomplete feature that completes your entries as you type them (see page 258).

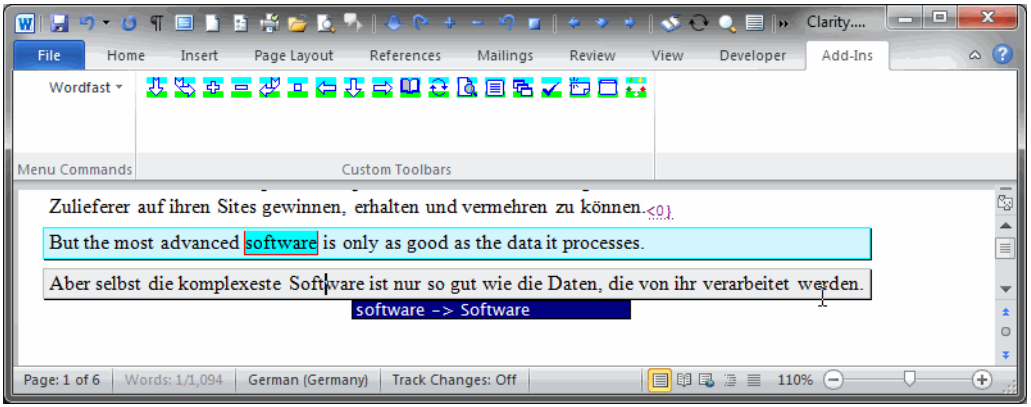


Figure 145: *Wordfast Classic* with match from the glossary

SDL MultiTrans

SDL 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 "TextBase translation memory" tool. Rather than matching on a sentence-by-sentence level, *SDL 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.

SDL 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.

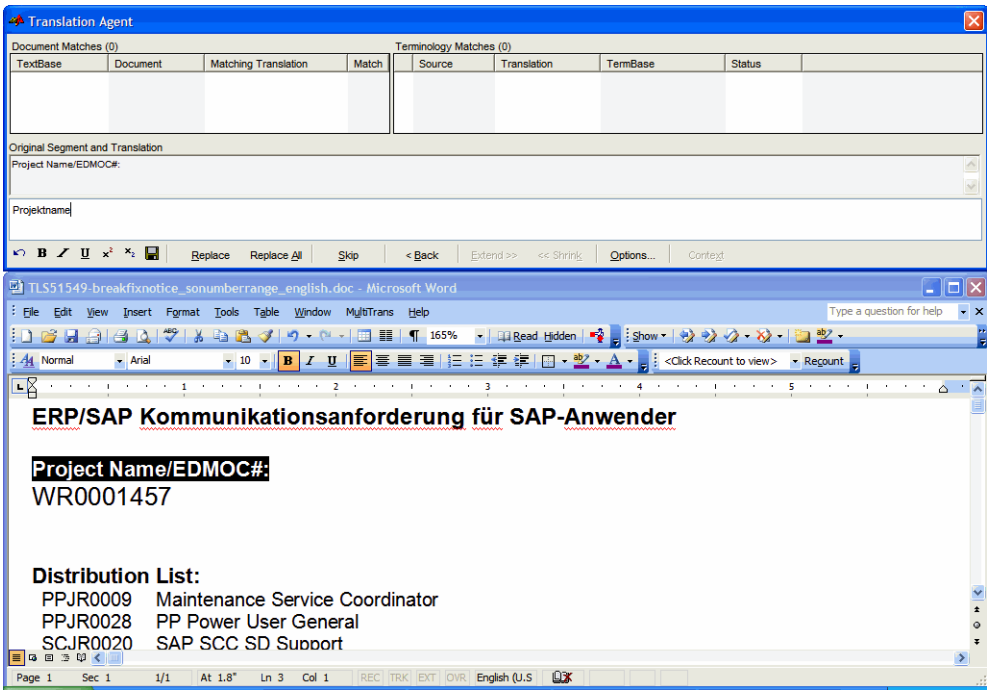


Figure 146: *SDL MultiTrans'* translation view with *MS Word* on the bottom and *Translation Agent* on top



Aside from the Word interface, MultiTrans also offers the translation of files in a PowerPoint and WordPerfect 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. This means that you pay according to your actual usage (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. The offer as *Translation Workspace* as a standalone product has not been very successful.

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* or an independent tool somewhat reminiscent of *Trados TagEditor*, called *XLIFF Editor*.

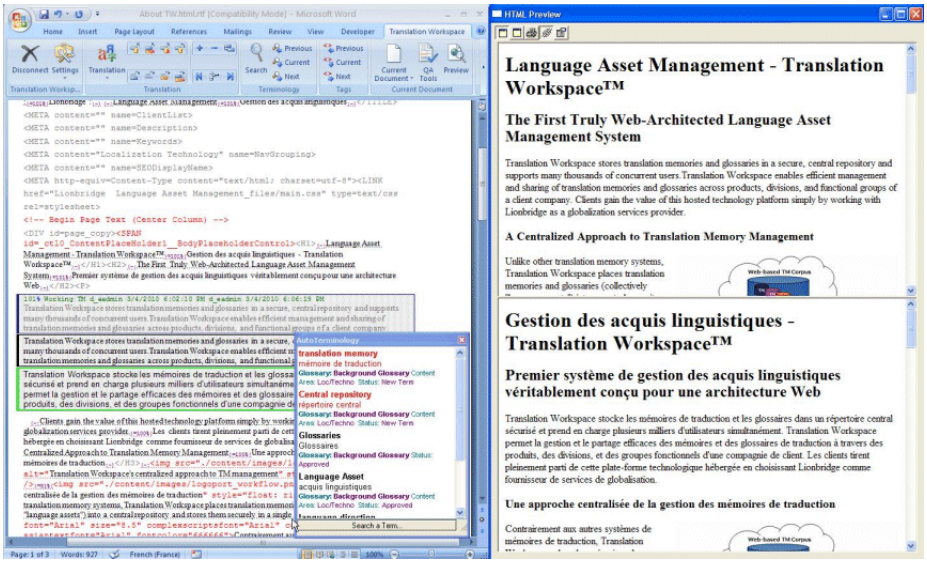


Figure 147: *Translation Workspace's Word* interface with TM and terminology matches and a preview feature

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.

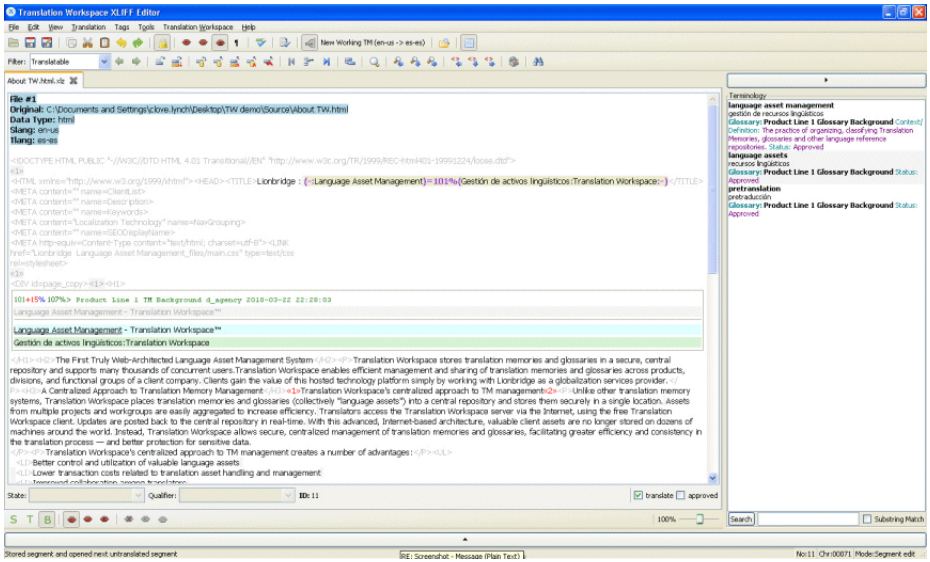


Figure 148: *Translation Workspace's XLIFF Editor* with TM and terminology matches

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.

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

- *SDL Trados Studio* (see sdltrados.com)—see page 202,
- *Déjà Vu* (see atril.com)—see page 206,

- *Star Transit* (see star-group.net/en/products/translation-and-localization.html)—see page 207,
- *memoQ* (see memoq.com)—see page 209,
- *Across* (see across.net)—see page 212,
- *Alchemy Publisher* (see alchemysoftware.com/products/alchemy_publisher.html)—see page 214,
- *Wordfast Pro* (see wordfast.com)—see page 215,
- *Heartsome* (see github.com/heartsome)—see page 216,
- *Swordfish* (see maxprograms.com)—see page 216,
- *Fluency* (see westernstandard.com/Fluency/FluencyNow.aspx)—see page 219,
- *MadCap Lingo* (see madcapsoftware.com/products/lingo)—see page 221,
- *Text United* (see textunited.com)—see page 222,
- *OmegaT* (see omegat.org)—see page 225, and
- *CafeTran Espresso* (see cafetran.com)—see page 228.

Trados Studio

SDL Trados has been the market leader among TEnT vendors for a long time and partly due to the age of the tool and the increased necessity to serve a lot of different markets and users, the tool had morphed into a whole range of connected applications geared toward different file formats (such as the *MS Word* interface for *Word*-compatible files, the *TagEditor* interface for tagged files formats, and the *T-Windows* applications for anything else), different activities (*Workbench* for translation memory purposes, *MultiTerm* for terminology maintenance, *WinAlign* for alignment purposes, *S-Tagger* for *FrameMaker*/*Interleaf* conversion, etc.), and different purposes (translation, project management, workflow design, etc.).

In 2009, when SDL came out with the first version of *Trados Studio*, a completely redesigned version of its tool(s) that combined almost all of the above-mentioned separate applications into one interface, it was a risky move—but one that turned out to be successful.

The *Trados Studio* translation interface is very similar to the now de-facto standard that tools like *Across*, *Déjà Vu* and *memoQ* have always used: a tabular interface with the source text in a left column and the target text in a right column. It would not be true, though, to claim that *Trados Studio* is simply a clone of these tools; there are just too many unique features and innovative features for that. Here are some of them:

- *Trados Studio* was the first tool that came out with an automatic suggestion feature to complete typing for you (comparable to the way you receive suggestions based on previous entries when you enter text into an *Excel* spreadsheet or the address field of a browser). In *Trados*' case, the suggestions are based on a separate "AutoSuggest" database, entries in the terminology database component *MultiTerm* (which remains the only application that is not primarily maintained in the main interface), AutoText entries, upLIFT segments (fragments of translation units within the TM), and machine translation subsegments.
- Also introduced by *Studio* (and now implemented by others) was the *MS Word*-like Track Changes feature that allows you to exactly see what kind of changes an editor or proofreader implemented in any translation.
- With *Trados Studio 2015*, the so-called "Retrofit" was introduced—a feature that allows you to export a translated document into its native file format, make changes, and then reimport it into *Trados Studio* to overwrite the target side in the translation editor.
- In *Trados Studio 2017*, SDL introduced two important features: "UpLIFT," a sophisticated system that increases the likelihood of locating subsegment matches and fixes fuzzy matches; and a self-correcting machine translation system (AdaptiveMT), based on SDL's own Language Cloud MT system (see page 276). Starting in 2019, every user with a valid *Studio* license has free access to SDL's (now: neural) machine translation for up to 500,000 characters/month.

- In *Trados Studio 2019*, SDL mainly focused on better usability. One example is a "smart" **Tell me what you want to do** feature that, like *MS Office* gives you answers to naturally worded questions and takes you to the control in the program you're looking for.

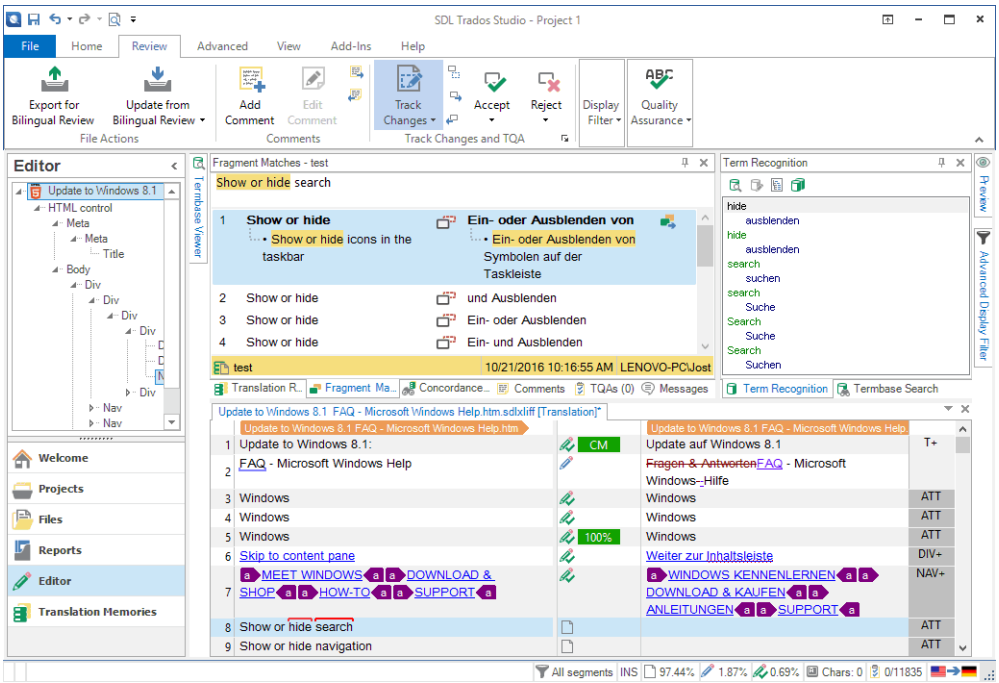


Figure 149: *Trados Studio 2016's* translation interface with activated Track Changes and UpLIFT fragment suggestions

In 2016, SDL also released a completely browser-based translation tool, the *SDL Translation Online Editor* (see sdltrados.com/products/language-cloud/online-editor). While the *Online Editor* is expected to become a fully functional translation editor for professional translators at some point, at this point it is meant for users with an occasional need for translation with limited ways to manipulate translation memories, no terminology management, and very little quality assurance.

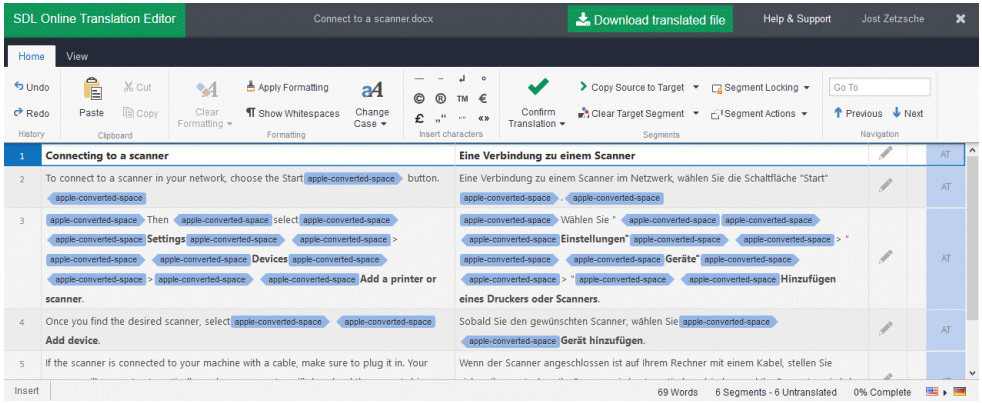


Figure 150: SDL Trados Online Editor

Another "feature" that SDL has introduced and that has so far not been followed by any of its competitors is the online app marketplace *SDL AppStore* (see appstore.sdl.com). Any owner of *Trados Studio* can have access to the API, the application programming interface, for many of the components of *Trados Studio* with which it is possible to develop applications that extend the functionality of the main program. These can then either be used internally or offered on the *AppStore* website for free or for a licensing fee. The introduction of *AppStore* has turned out to be a very helpful move for SDL. Not only have there been many helpful apps developed by third-party developers, but many are in fact placed there by SDL developers who have the option to turn new features into external apps rather than internal features of the main tool, which would make a very complex application even more complex.

In 2019, SDL also bought BaccS, a reporting and invoicing system that is now named *SDL Trados Business Manager*. This tool comes in different editions, geared toward freelance translators ("Lite"—a Studio plugin) to translation agencies and translation buyers ("Desktop" and "Team"). If virtually all of your translation business is happening within *Trados Studio*, this might be an interesting tool. If you are using more than one translation environment tool, it might be a better idea to use an independent tool (see *Management Tools* on page 325).

Déjà Vu

As mentioned above, *Déjà Vu* was the first *Windows*-based translation environment tool, and early on it became the tool of choice for many freelance translators. While it always used the tabular view as its translation interface, it used to consist of separate (but connected) applications for the actual translation, the translation memory and the terminology database. This changed in 2003 when *Déjà Vu X* was released with an interface that offered one interface for all activities. *Déjà Vu X2* and *X3*, released in 2011 and 2015 respectively, were essentially a continuation of the previous version, with the added features of autocomplete ("AutoWrite"), subsegmenting ("DeepMiner"), a deep integration of machine translation and WYSIWYG (what-you-see-is-what-you-get) formatting.

Déjà Vu offers a very large range of supported file formats. While its user group is no longer as passionate and boisterous as it was during the late 1990s when the "flame wars" raged on the Lantra-L list between users of *Trados* and *Déjà Vu* (search the archives at segate.sunet.se/cgi-bin/wa?A0=LANTRA-L), it still is a tool of great value, particularly because of a number of innovative features:

- The assemble feature that *Déjà Vu* pioneered allows for the "piecing together" of translation from the various resources, including terminology database, glossary ("lexicon") and fragments from the translation memory. Provided that the quality of these resources is good, the advantage to the translator can be considerate.
- *Déjà Vu* also pioneered a repair feature for fuzzy matches (*memoQ* and *Trados Studio 2017*+ are the only other tools that offer this), where the terms and phrases within the translation unit that differ from the match in

the translation memory are automatically replaced with the correct term or phrase if that term or phrase exists in one of the resources.

- This fuzzy match repair feature also works with machine translation where only the "offending" part of a segment is translated by a machine translation engine and potentially turns the fuzzy match into a perfect match.

The latest release of *Déjà Vu* (*X3*) added a variety of options that can also be found in other tools, such as WYSIWYG formatting, inline spell-checking and an automatic preview of the translation file.

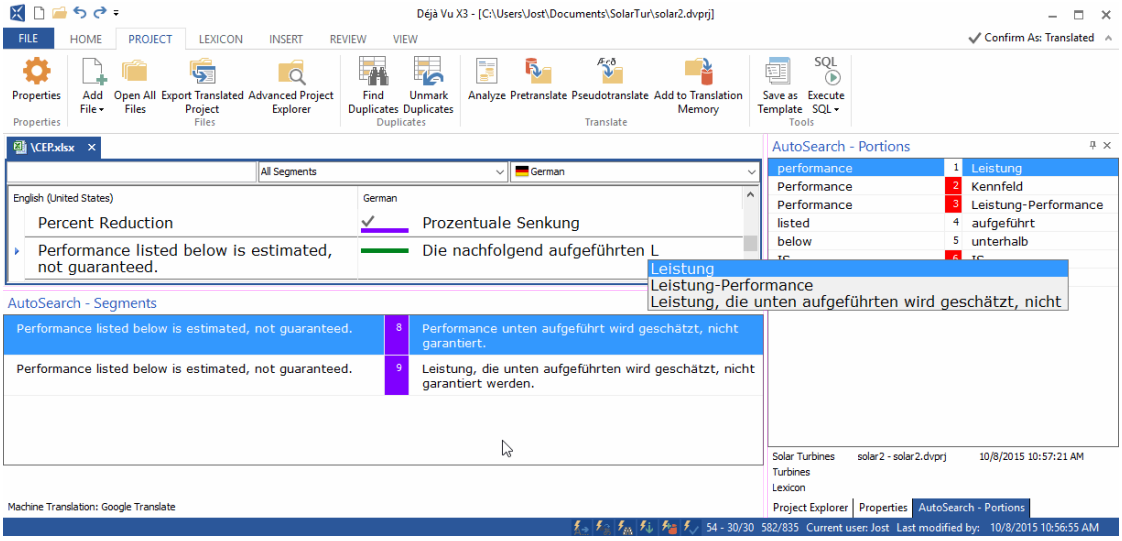


Figure 151: Example of AutoWrite suggestion coming from termbase, TM and various MT sources in *Déjà Vu X3*

Star Transit

Star Transit is a different kind of translation environment tool in many ways. First of all, it does not use the "traditional" kind of external database-based translation memory, but it provides a "virtual" translation memory—referred to as "reference material." It achieves this by associating already-translated

files. The benefit of this is the exact customizability of the translation memory and the inherent availability of context. The drawback lies in the large number of translated file pairs that have to be retained to provide the necessary "reference material."



Starting with Service Pack 7 for Star Transit NXT, a parallel translation memory system with the TM-Container was introduced in the fall of 2013.

Star also does not release many "versions"—*Star Transit 2.7*, the much-loved and very stable version, was introduced in the late nineties, followed by an ill-fated and faulty successor (*Star Transit 3*) that was quickly replaced with *Transit XV* in 2001 and with *Transit NXT* in 2008. These are long stretches without new payable versions for a development company, especially because the development never stopped and was released in the form of Service Packs. To offset this, Star is charging for the support of the following formats: *FrameMaker*, *PageMaker*, *Interleaf*/*Quicksilver*, *AutoCAD*, *QuarkXPress* and *InDesign*. Overall, Star probably has one of the largest number of supported file formats.

What also sets *Star Transit* apart is the morphological support for 15 European languages (incl. English, French, German, Italian, Spanish, Czech, Dutch, Polish, Portuguese, Russian and Swedish), which means that just by entering the infinite form in the (powerful) termbase, other morphological forms are automatically found in the respective languages.

Lastly, the "dual fuzzy" system that Star introduced with *Star Transit NXT* has been very innovative and has so far not been implemented by any of its competitors. The dual fuzzy system is that it not only looks in the source portion of the reference material but also in the target. This means that if

there is no perfect or a fuzzy match for a segment, *Star Transit* compares what you type with all target segments that contain the same translation data you are currently entering—even if it is just two or three words—and displays those sentences to you as options to choose.

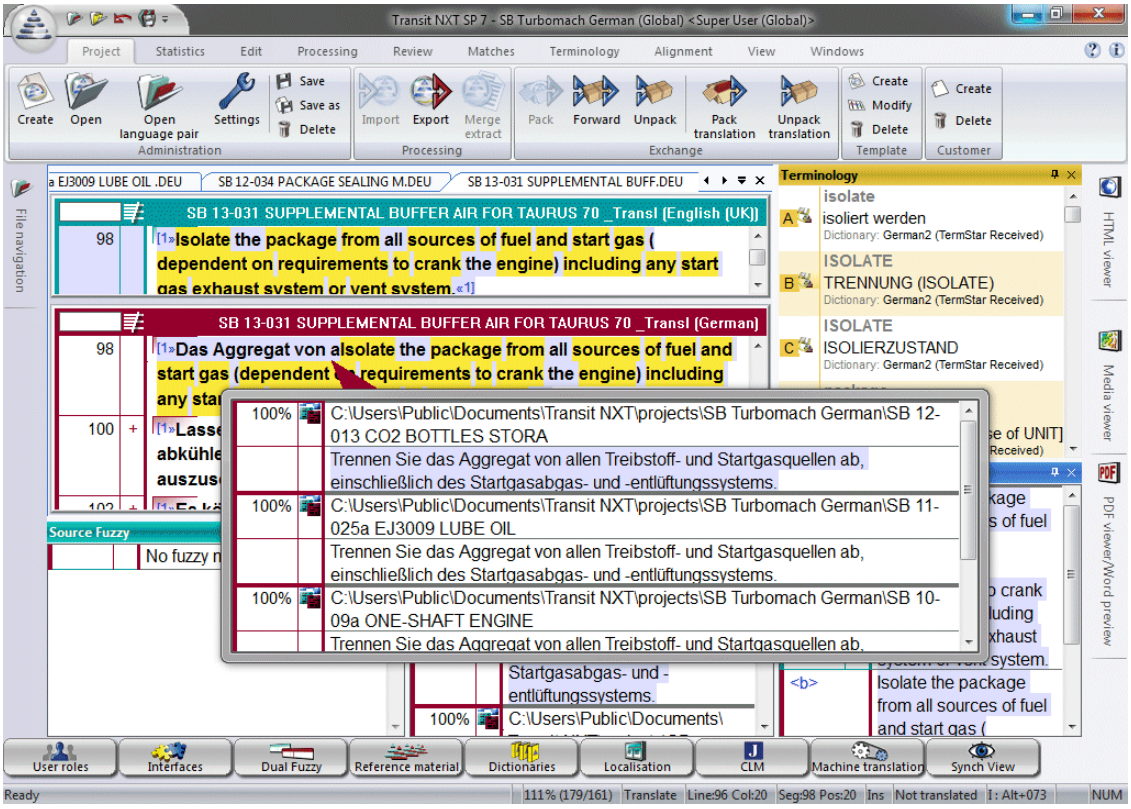


Figure 152: *Star Transit* with target text-based matching

memoQ

The Hungarian *memoQ* is a very process-oriented tool that makes the general workflow user-friendly even for a novice to TEnTs. This is partly achieved by a system of context-sensitive ribbon bars that leads you through the different steps in the processing of each document or project.

In the actual translation interface, the translatable text is—like in most other tools—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 the whole range of formats you can wish for, including project files of most competing tools.

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.
- *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.
- *Language Terminal* (see languageterminal.com) is an online portal that allows *memoQ* users to set up the cloud-based *memoQ cloud*, store projects, share some data, and convert native *InDesign* INDD files into XLIFF files that can be processed with *memoQ* and other tools (for the *InDesign* feature, see page 337). In the latest couple of versions it also allows for the limited (and free) sharing of translation memories, termbases, and projects.
- Tracking of new versions of translation documents as well as document versioning (keeping various versions of one document).

- Import of embedded files, including images that can then be transcribed and translated within *memoQ*.
- An innovative way of detecting abbreviations and the possibility of adding rules and corrections for segmentations on the fly.

Other features, such as a preview, inline spell-checking, WYSWIWYG formatting, track changes, auto-correction of fuzzy matches, AutoSuggest ("Muse"), and a tight integration into various external resources, including machine translation engines, are not unique but well-established within *memoQ*.

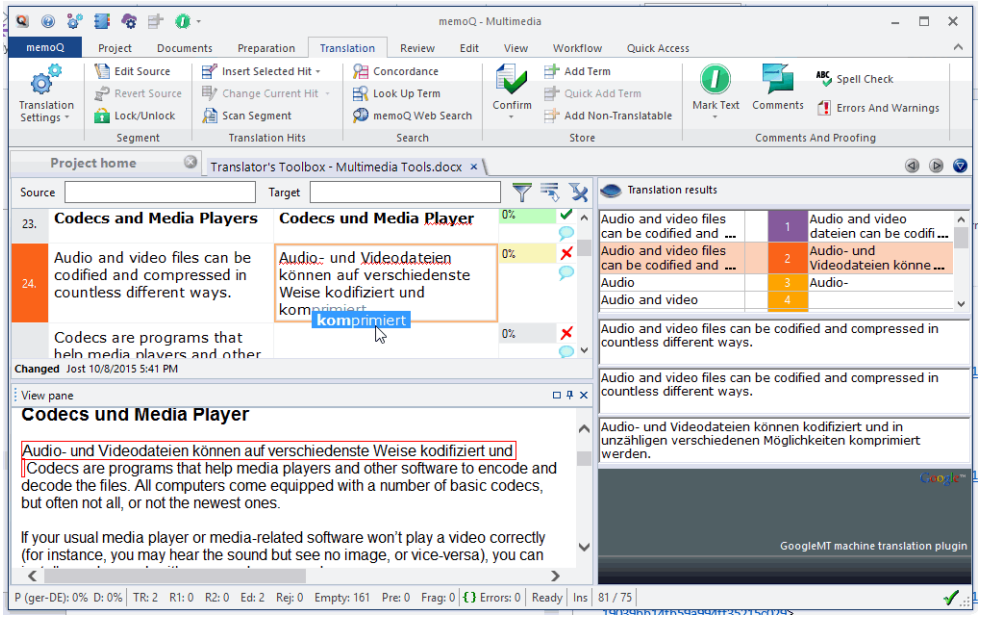


Figure 153: *memoQ* translation view with preview, AutoType and MT features.

Across

The German tool *Across* emphasizes corporate, networked translation but it also equips the translation professional with all the tools (translation memory, terminology database, QA tools, etc.) needed to translate successfully as a standalone translator. And the price tag for freelance translators is particularly interesting: it's free (for freelance translators) if used within a corporate workflow.

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 and for all projects (which means that you don't have to create separate translation memories and termbases for each project).

Also, *Across* is one of the few translation environment tools to offer morphological recognition in its term searches in various European languages, including English, Danish, Dutch, Finnish, German, Norwegian, Swedish, French, Italian, Portuguese and Spanish.

Though the translation interface seems to be in a grid-like interface, the actual translation takes place in a special edit area below the table view.

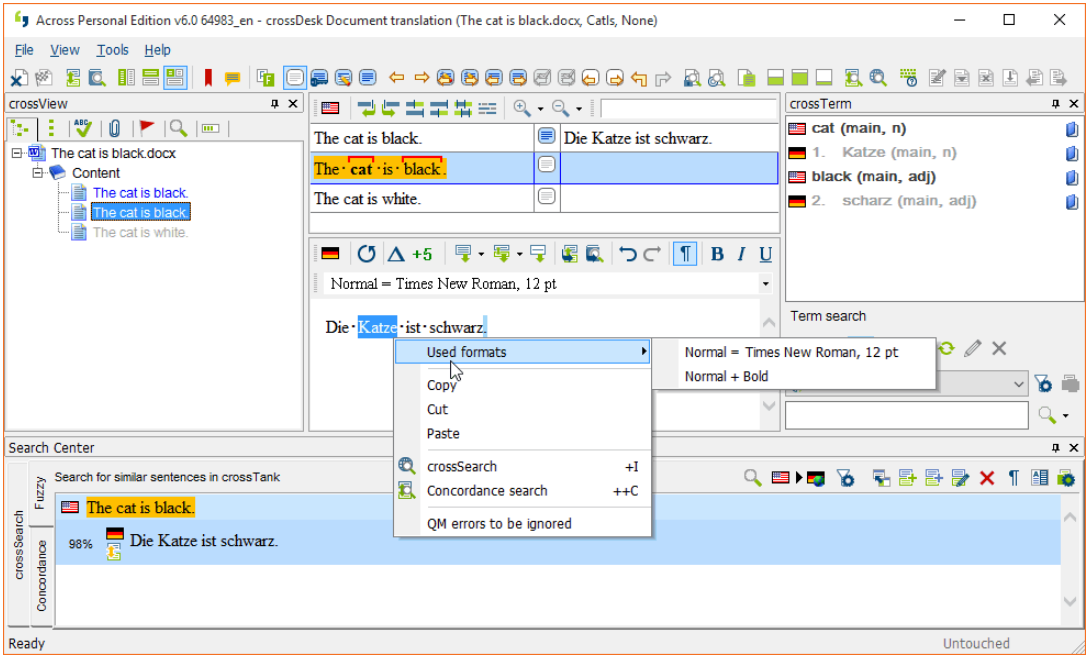


Figure 154: *Across'* translation view

In late 2015, Across introduced a dual system for its freelance product. While the "Basic" version is still freely available, that version neither allows use of the user's own translation memories and termbases, nor is it possible to export documents from it, thus essentially making it only a tool with which to work for an external server-based version of *Across* (typically owned by a language service provider or translation buyer).

To have access to the "Premium" version, which allows access to the above-mentioned features and the ability to use it as a standalone tool, a paid membership to the marketplace crossMarket (see crossmarket.net) is required.

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 or an index maker, for example. The translation memory and terminology databases are simple text-based files. Direct access to *Trados* TMs and termbases is also possible.



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.

In addition to *FrameMaker* and *Word* as well as HTML, XML, HTMLHelp and software-related development files are supported.

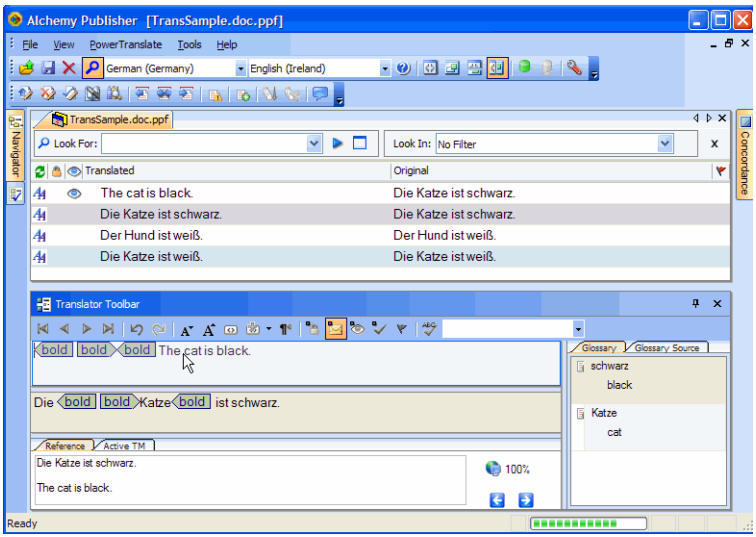


Figure 155: *Alchemy Publisher's* translation view

Wordfast Pro

Wordfast Pro, which originally was supposed to be the successor of *Wordfast 5.5* (now: *Wordfast Classic*, see page 196), has instead become a a parallel version.

The concept of *Wordfast Pro* is very different than the *Classic* version. Rather than using a third-party interface 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—starting with version 4—an XLIFF format that can be processed in virtually all other translation environment tools. The supported translation file formats include *MS Office* formats, HTML, *FrameMaker*, PDF, *Trados* TTX, various software development formats and *InDesign*.

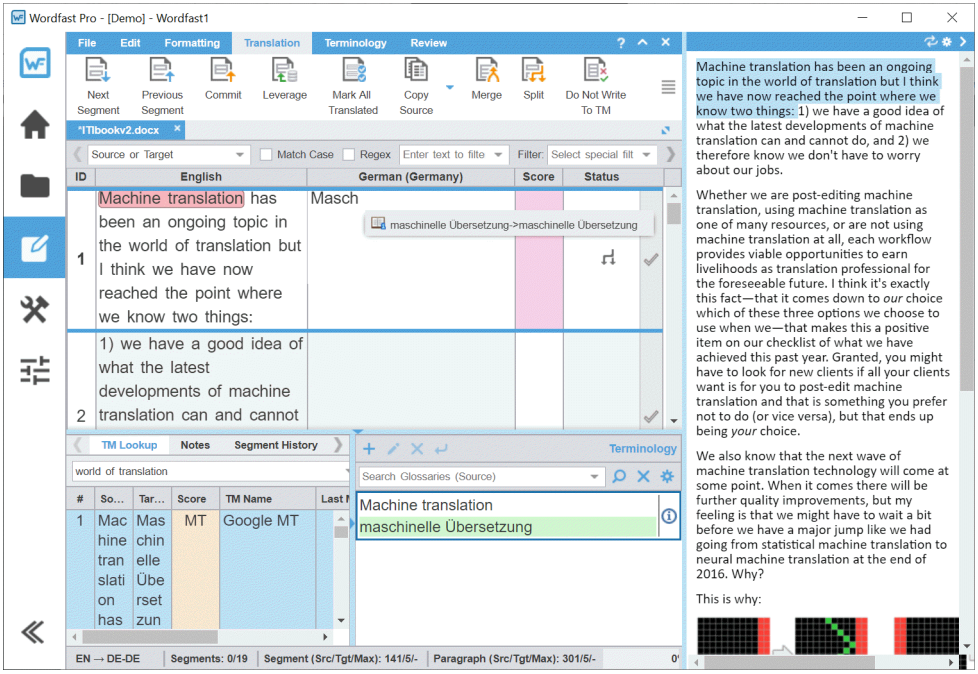


Figure 156: *Wordfast Pro's* translation interface with a preview of the original file on the right-hand side

Starting in version 4, it is possible to export two-column *Word* files. You can translate and/or edit outside the *Wordfast* environment, and a preview feature allows you to preview ongoing translation in its native environment (*Word*, *InDesign*, web browser, etc.).



As a Wordfast Pro user, you can also use the generic "very large TM" (VLTM), which comes in many language combinations (see wordfast.net/wiki/VLTM_in_Wordfast_Pro), and the IATE glossaries (see wordfast.net/wiki/Connecting_to_IATE_glossaries), both of which are also available to users of Wordfast Anywhere or Wordfast Classic.

Heartsome and Swordfish

Heartsome and *Swordfish* are both *Java*-based tools that run on *Mac*, *Linux* and *Windows*. While at this point several other tools use XLIFF (see page 286) as the interim translation format, these tools were the first to go that route. This means that any of the supported file formats (including RTF, *Office 2007 and above*, *FrameMaker*, HTML, *Apache OpenOffice*/*LibreOffice*, *InDesign* and a variety of software development formats) are converted to XLIFF, provide for their translation within that format, and then converted back into their original format.

While both tools share the same origin, the fate of the companies that develop and support them has been different. *Heartsome*'s developers ceased operations in 2014 and donated the tool (and its underlying code) to the translation community (you can download it at github.com/heartsome/), whereas Maxprograms, the Uruguayan company that develops and supports *Swordfish*, is still actively developing its tool.



Heartsome and Swordfish (Maxprograms) have an interesting history. Here it is from Rodolfo Raya's perspective: "Maxprograms has been in the custom software development business since I founded the company in Argentina in 1987. In 2002 Klemens Waldhör, the founder of Heartsome Europe and an ex-Alpnet employee, requested the development of an XML-based publishing system for Samsung, which was completed and delivered in 2003. Once the Samsung publishing project was completed, Maxprograms moved to Uruguay and Geraldine Lim, the owner of Heartsome Holdings Pte Ltd from Singapore and an ex-Alpnet employee, commissioned development services for translation tools.

"Maxprograms designed an XLIFF editor, a TMX editor, and a dictionary editor based on TBX for Heartsome Holdings between 2003 and 2007. In 2007 translation tools development was transferred to I-Len, a translation agency from China owned by Anderson Wang, also a former Alpnet employee. Development continued under the umbrella of Heartsome Technologies, a company registered in Hong Kong that went bankrupt in 2014.

"After ceasing work with the various Heartsome companies, Maxprograms stayed in the translation industry, developing custom software for other clients."

The Heartsome Europe's product was called Araya (still at heartsome.de). Araya was customized for corporate customers with XML- and server-based processes, and the Heartsome tool suite was marketed to the whole range of actors in the translation industry.

Swordfish was first released in 2008. Other tools that were released by Maxprograms include:

- *Stingray Document Aligner* in 2008
- *RemoteTM*, a web-based Translation Memory server, in 2011
- a new version of the *TMXEditor* in 2018
- a number of other tools that can be found at maxprograms.com

In *Heartsome,* all features are processed in one interface:

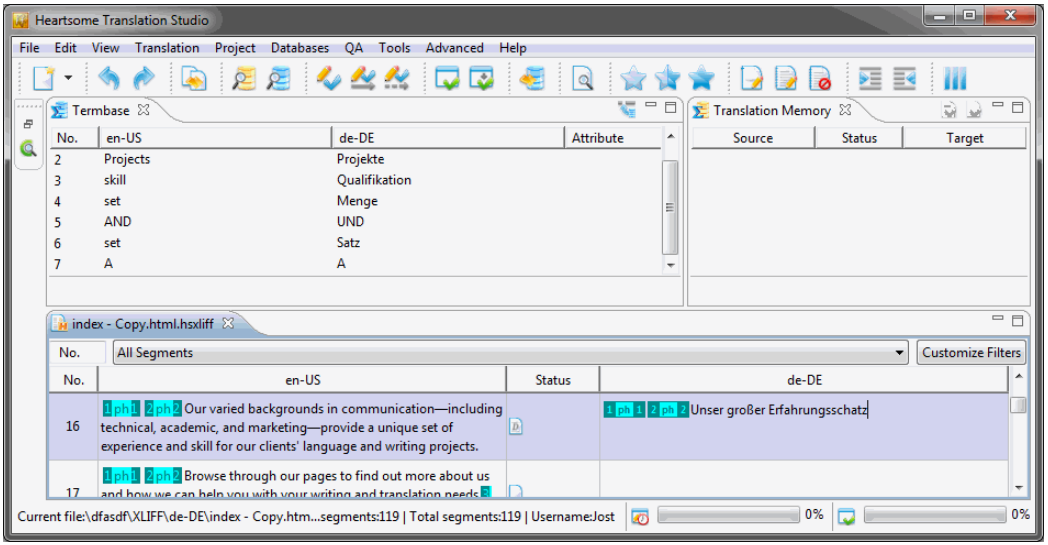


Figure 157: *Heartsome* on a *Windows* computer

Swordfish, on the other hand, comes with a number of interfaces, including the actual translation editor, a TMX and a TBX editor (TMX is the translation memory format, TBX the termbase exchange format), and various conversion utilities (see above):

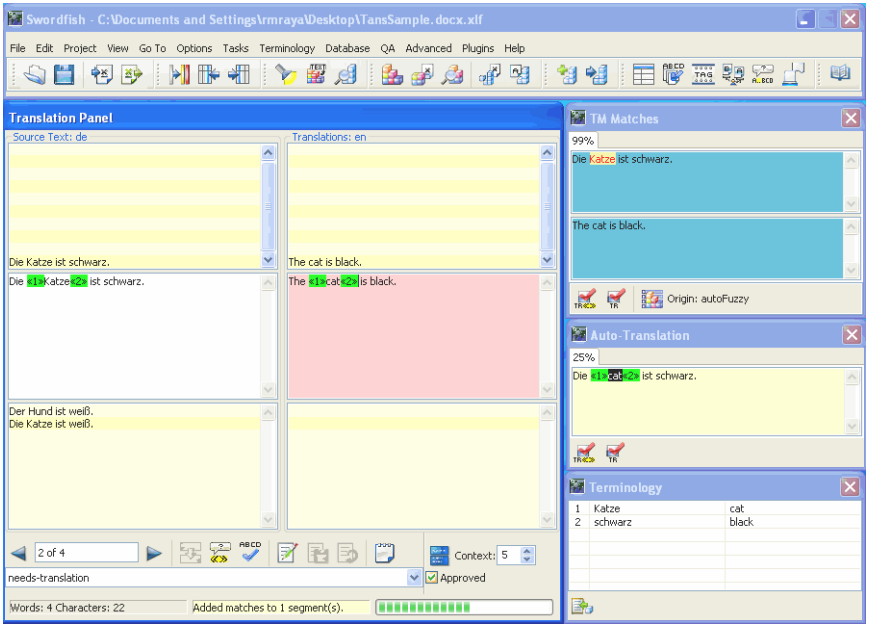


Figure 158: *Swordfish's Translation Editor* on a *Windows* computer



A couple of thoughts about "abandoned" and free tools, which include Heartsome, Felix (see felix-cat.com) or Wordfisher (see wordfisher.com/wordfisher-1.html). While it might seem attractive to use these tools at first, I would only recommend it if you yourself have the ability (and time) to continue their development or someone else does. It doesn't make sense to learn a new technology and build up data assets in it if you know that it won't be supported and developed to match future file formats or processes.

Fluency

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 an editable PDF conversion module (from and to PDF),

optical character recognition (OCR), an interface for manual transcription of items that can't be processed (well) 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 extensive testing (the language-specific resources will be downloaded the first time you use that language or language combination).

Fluency's translation interface is a multi-pane interface with access to a number of linguistic resources, including a pane for translation memory matches, a source and a target, a target preview pane, machine translation views, and a variety of other things. 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 and the target pane that displays segment by segment independently (the current segment is highlighted in the source pane).

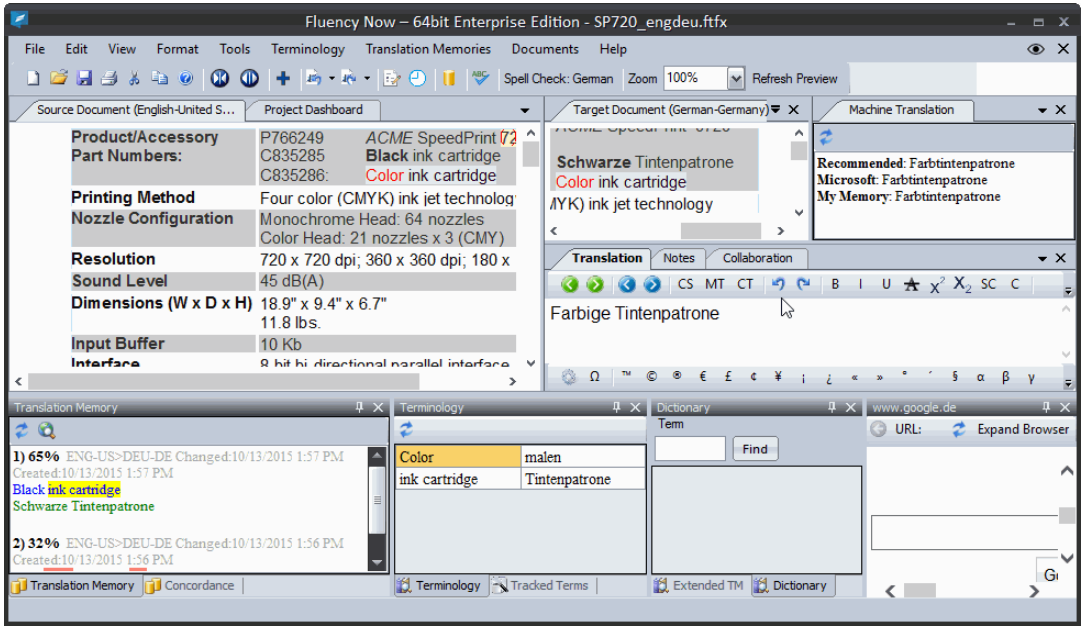


Figure 159: *Fluency's Translation Editor*

The supported file formats include all the expected formats but also some surprising ones, including SRT and ASS subtitling files and *Microsoft Publisher*.



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? Fluency is one of only two tools on the market—the other is Text United—that supports Publisher files.

Fluency also comes in a *Java*-based version for *Macintosh* and *Linux* and is sold with a monthly fee rather than with a perpetual license.

MadCap Lingo

MadCap is the company that split off from MacroMedia (now Adobe) after some of *MadCap's* current owners felt that MacroMedia was treating the help-authoring product *RoboHelp*, which it purchased as part of a larger acquisition, too shabbily. 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), *InDesign*, *FrameMaker*, *Trados* TTX/SDLXLIFF, *Wordfast* TXML, HTML, XML, DITA, JSON, SVG, and RESX files.

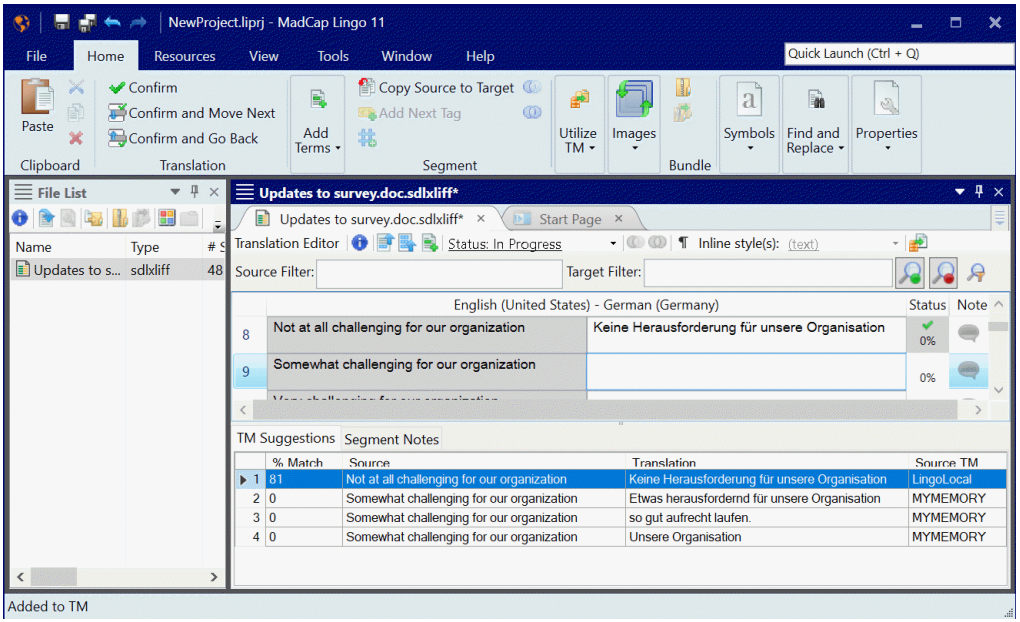


Figure 160: *MadCap Lingo's* translation interface with an SDLXLIFF file

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.

Text United

Text United is (mostly) a hybrid tool. In this case, hybrid means a locally installed Windows-based desktop application that connects to data (including translation files, glossary, and translation memory) sitting in the cloud. By default you'll need an Internet connection to work in the tool, but it's possible to download a local copy of your project and your resources (if you know you'll be offline for a while) and continue to work offline. Once the connection is restored, everything is synced and you can continue to work online.

There are, however, additional alternative interfaces, including a browser-based interface that can be used instead of the desktop app. A key distinction is that you cannot work in the offline mode with this browser-based interface and, if you are a project manager rather than a translator, the browser-based interface does not give you access to many PM-related functions.

A third option for a translation interface is the so-called Overlay Editor. This interface allows you to translate directly in a website (so you can see context and sizing, etc.). This makes sense because *Text United* not only supports a large range of file formats (including *MS Office*—including *Publisher*—*FrameMaker*, *InDesign*, plus the various tagged and software development and subtitle formats), but also uses proxy-based services to translate websites, ecommerce sites, and various other content-managed sites (for proxy-based website translation, see page 327).

Text United is traditionally marketed to companies who buy translation services, but its latest endeavor (2019) encompasses new marketplace and infrastructure plans. Freelance translators can use the tool for free (their only payment would be for the use of machine translation services). The company also provides a marketplace where translators can publish their profile and offer their services, as well as a quasi-self-contained infrastructure that allows not only the use of the technology but provides options for billing, reporting, and project dashboards.

The translation interface on the desktop looks very modern with an *Outlook*-like **Home** screen, ribbons instead of menus, and a very lightweight application.

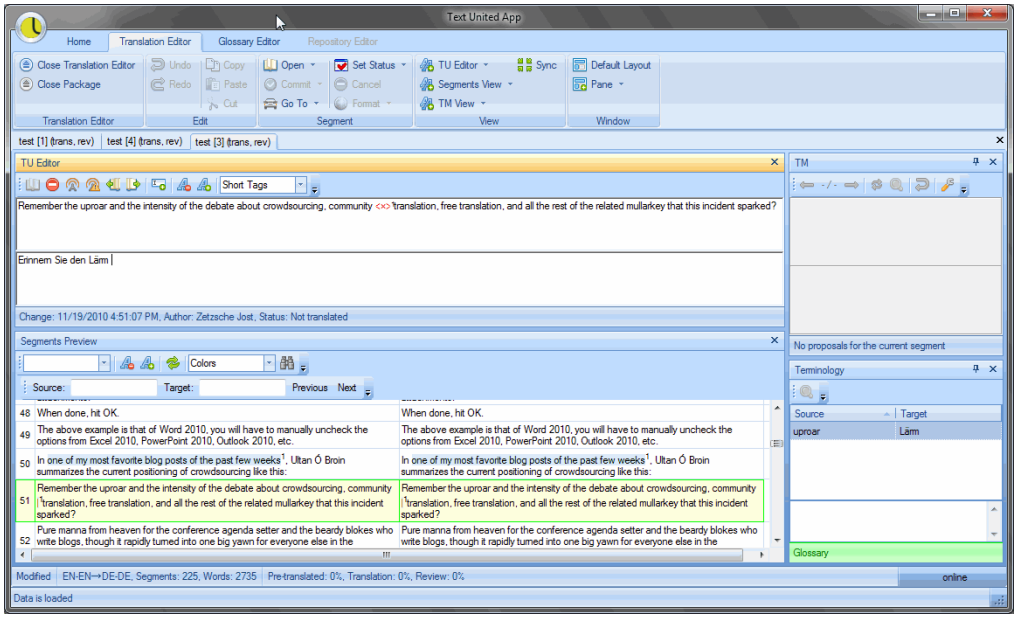


Figure 161: *Text United's* translation interface

Processes are role-based, i.e., a project is assigned to a translator (who can be identical with the project owner). The translator has only as many rights to the project and its resources as the owner assigns to him or her.

When importing a file, a terminology extract of the file's content is performed. This can serve as the base glossary for the translation of the project (for terminology extraction, see also page 303).

OmegaT

There are only a small handful of open-source translation environment tools primarily geared toward the professional freelance translator, and the platform-independent *OmegaT* is without a doubt the king of the hill. *OmegaT* is downloaded more than 10,000 times a month, and even though these downloads include updates of existing users (typically between 8 and 10 updates per year), these are impressive numbers that show that *OmegaT* is not just used by some isolated software geeks.

There is a strange and remarkable dichotomy between the technical and easy-to-use parts in *OmegaT*. When you start the program, the initial screen has information on how to get started with *OmegaT* in five minutes. And they're not kidding. To use the basic features, you just start using the program and it works. When it comes to fine-tuning the *OmegaT* setup, you might find some items available in menus and with an easy-to-use graphical user interface (GUI), but for other features you'll have to manually set up files and alter code. One example: to change keyboard shortcuts, you actually have to create some files that will cause the desired change. If you take your time to think through it, you'll get it done; if not, you'll end up being frustrated.

The interface is super easy and user-friendly: the actual translation is done in a non-tabular, horizontal layout. If you have to deal with inline tags (tags within segments), they are clearly set apart from the translatables. Any panes with access to terminology, translation memory, machine translation, or comments can be arranged like you want and even dragged to a second monitor. And while I wish there were more right-click menus, the actual menus are well organized and give you the necessary access to available features.

The range of other directly supported file formats is very impressive and includes TXT, PROPERTIES, PO, INI, SRT (subtitle), Open Document Formats, (X)HTML, XLIFF, RESX, LaTex, *Wordfast* TXML and *Visio* files. When I say that these files are supported directly, it means that there are other file formats

that are supported indirectly through the open-source *Rainbow* application (see okapiframework.org/wiki/index.php?title=Rainbow). The file types you might want to use with that route include most XML formats, *FrameMaker* MIF, bilingual DOC and *Trados* TTX files.



Rainbow is part of the Okapi suite of tools. You can find a very helpful article on these tools and how they can be used here: atanet.org/chronicle-online/highlights/okapi-tools-how-translators-can-take-advantage-of-them.

OmegaT also includes an interesting project concept: you can have numerous files of various different formats within a project that automatically open one after the other as you translate, and any search-and-replace action can be done simultaneously in all files.

One of the more surprising features is the inclusion of "tokenizers," which provide better morphological recognition in termbase and TM recognition. The supported languages include Arabic, Portuguese, Chinese, Japanese, Korean, Czech, Dutch, French, German, Greek, Persian, Russian, Thai, Danish, English, Finnish, Hungarian, Italian, Norwegian, Romanian, Spanish, Swedish, and Turkish.

Unique to OmegaT is the "Auto-completer," a feature that automatically calls up predictions for auto-completion on the basis of your current project. It is trained on-the-fly, and its data is not stored anywhere but in your computer's memory and is discarded on closing the project.

You can find rich and interesting resources about *OmegaT*—both for novices and for advanced users—at omegat.org.

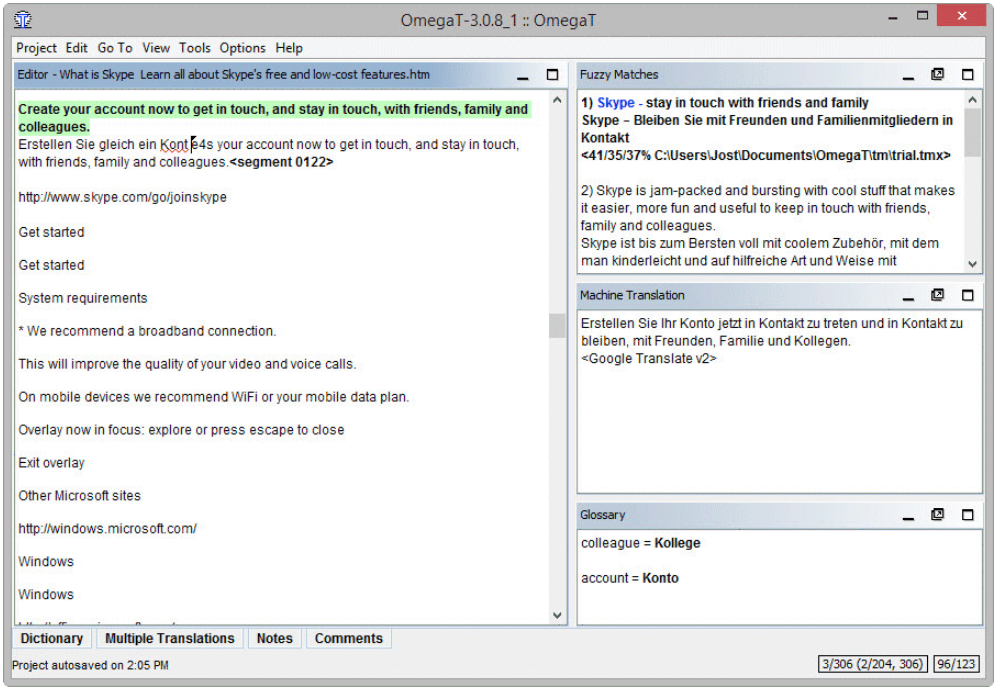

OmegaT-3.0.8_1 :: OmegaT
Project Edit Go To View Tools Options Help
Editor - What is Skype Learn all about Skype's free and low-cost features.htm
Create your account now to get in touch, and stay in touch, with friends, family and colleagues.
Erstellen Sie gleich ein Kont4s your account now to get in touch, and stay in touch, with friends, family and colleagues.<segment 0122>
http://www.skype.com/go/joinskype
Get started
Get started
System requirements
* We recommend a broadband connection.
This will improve the quality of your video and voice calls.
On mobile devices we recommend WiFi or your mobile data plan.
Overlay now in focus: explore or press escape to close
Exit overlay
Other Microsoft sites
http://windows.microsoft.com/
Windows
Windows
Fuzzy Matches
1) Skype - stay in touch with friends and family
Skype – Bleiben Sie mit Freunden und Familienmitgliedern in Kontakt
<41/35/37% C:\Users\Jost\Documents\OmegaT\tm\trial.tmx>
2) Skype is jam-packed and bursting with cool stuff that makes it easier, more fun and useful to keep in touch with friends, family and colleagues.
Skype ist bis zum Bersten voll mit coolem Zubehör, mit dem man kinderleicht und auf hilfreiche Art und Weise mit
Machine Translation
Erstellen Sie Ihr Konto jetzt in Kontakt zu treten und in Kontakt zu bleiben, mit Freunden, Familie und Kollegen.
<Google Translate v2>
Glossary
colleague = Kollege
account = Konto
Dictionary
Multiple Translations
Notes
Comments
Project autosaved on 2:05 PM
3/306 (2/204, 306)
96/123

Figure 162: *OmegaT's* translation editor on a *Windows* computer. Note the squiggly-underlined, interactive spell-checking, morphologically-aware term recognition ("colleague" for "colleagues") and machine translation suggestion.



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.

Other open-source translation environment tools include Virtaal (see virtaal.translatehouse.org) for TMX and PO files and OpenTM2 (see opentm2.org) for HTML, Java PROPERTIES, Apache OpenOffice/LibreOffice, XLIFF and generic XML files.

CafeTran Espresso

CafeTran Espresso is a platform-independent (*Java*-based) tool that is specifically geared toward the individual translator with a large, ever-growing host of features, including:

- automatic assemble using all available resources, including TM, MT and termbase
- several TM systems (the alternative system "Total Recall" is used for quick retrieval from very large TMs)
- translation of the clipboard
- all the commonly supported formats, but also some uncommon ones such as Apple *iWork* files or *AutoCAD* DXF files—and the latter in a rather sophisticated way by exposing the different layers of DXF files
- segmentation correction and rule creation on the fly
- extraction of commonly used terms
- With automatic fragment adjustment for MT matches (under **Edit> Preferences> MT services**), MT suggestions are "corrected" on the basis of matches in your glossary or TM when the term differences are slight (for instance, it changes them to the plural form or other suffix changes)
- and many, many more



In fact, there are so many features in this tool that you will need to do the same thing you would do for tools like memoQ or Trados: plan for an extended gear-up time where you primarily focus on learning the tool rather than adding to your productivity right away. While this might be surprising for a tool that has traditionally been considered a "small" tool, the dynamics of a single developer (Igor Kmitowski) who seems to respond unreservedly to any and every wish of a small but very active user community have led to this ever-expanding set of features.

The range of supported file formats is very large, including *MS Office 2007* and higher files, *InDesign*, *FrameMaker*, a lot of software development file formats, and a great number of bilingual formats coming from other TEnTs, including *Transit*, *Trados*, *Wordfast*, *memoQ* and *Déjà Vu*.

As in many other tools, the translation is done internally via XLIFF, and a fully translated XLIFF file is automatically generated at the end of the project (aside from the actually translated file).

As far as users of operating systems other than *Windows*, approximately 20% use *CafeTran Espresso* on a *Mac* and less than 10% on *Linux*.

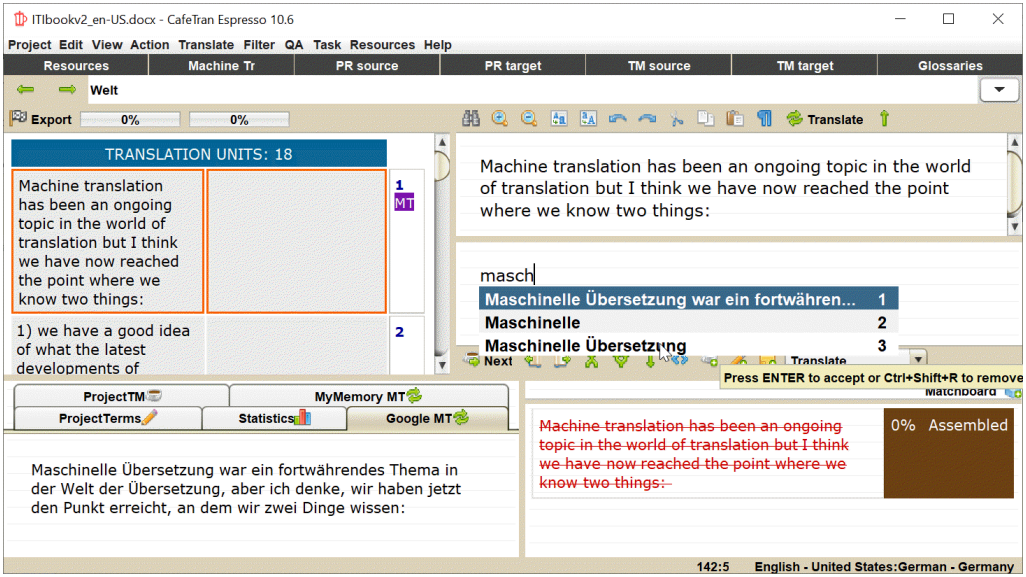


Figure 163: *CafeTran Espresso's* modular and configurable interface

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 lingotek.com) was the first to come out with a new attempt at this kind of interface and approach to a translation environment tool. (After many incarnations, Lingotek is now specializing in the translation of content management systems—CMS—and is really not a directly accessible tool for translators anymore).

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 (the "cloud"); 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.



Many of the developers of the more typical desktop-based tools, such as memoQ, Across and SDL Trados, already are offering web-based tools that are developed in parallel to the desktop environment or are working on solutions like that.

Given at least the first two of these parameters (cloud-based applications and data), it's no wonder that the latest wave of translation management systems, such as the above-mentioned *Smartling* or *Transifex* (see page 196) or the various LSP-owned translation platforms, have all chosen to go this route rather than using desktop tools and translation packages that have to be mailed back and forth.

Following are some of the tools that are currently working (almost) exclusively through a web-based interface and that are directly accessible for translators (i.e., not through a third-party, like a translation agency or a translation client):

- *Wordbee* (see wordbee.com)—see page 231,
- *XTM Cloud* (see xtm.cloud)—see page 233,
- *Wordfast Anywhere* (see freetm.com)—see page 235,
- *Memsource* (see memsource.com)—see page 236,
- *Smartcat* (see smartcat.ai)—see page 240,
- *MateCat* (see matecat.com)—see page 241,
- Termsoup (see termsoup.com)—see page 245, and

- *Lilt* (see lilt.com)—see page 247.



Noticeably missing from this list is Google Translator Toolkit, which was first released in 2009 but then unceremoniously shut down in 2019.

Wordbee

Wordbee is a completely web-based translation environment tool that attempts to combine a translation environment with some relatively intuitive project management capabilities.

Wordbee uses the SaaS—Software as a Service—concept and charges an annual fee which varies to the way you use it. This means that it can be adapted to the needs of any team, from freelance translators to translation agencies, international corporations and institutions.

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.

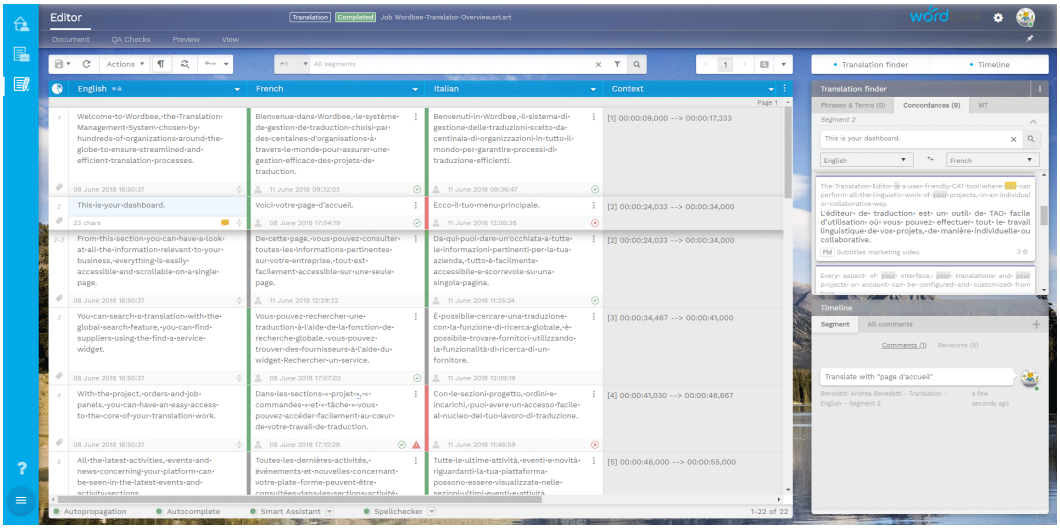


Figure 164: *Wordbee Translator's* translation interface

The list of supported file formats is solid (*MS Office, Visio*, *InDesign*, *InCopy*, *FrameMaker*, *Photoshop*, *Apache OpenOffice/LibreOffice*, RTF, XML, HTML-based formats, as well as various software development formats and formats of other translation environment tools).



Through the CMS (content management system) connector Beebox (see wordbee.com/translation-connectors/) of Wordbee's development team, content from CMSes is also easily accessible within Wordbee as well as any other translation environment tool. For more information, see page 323.

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, 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 also a SaaS-based offering, a Software as a Service. This means that there is no download or installation; instead, 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, *Trados* TTX files, XLIFF and many development formats) 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.

XTM Cloud allows you to start working with essentially no learning curve as a translator. Translation memory matches, (optional) 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.

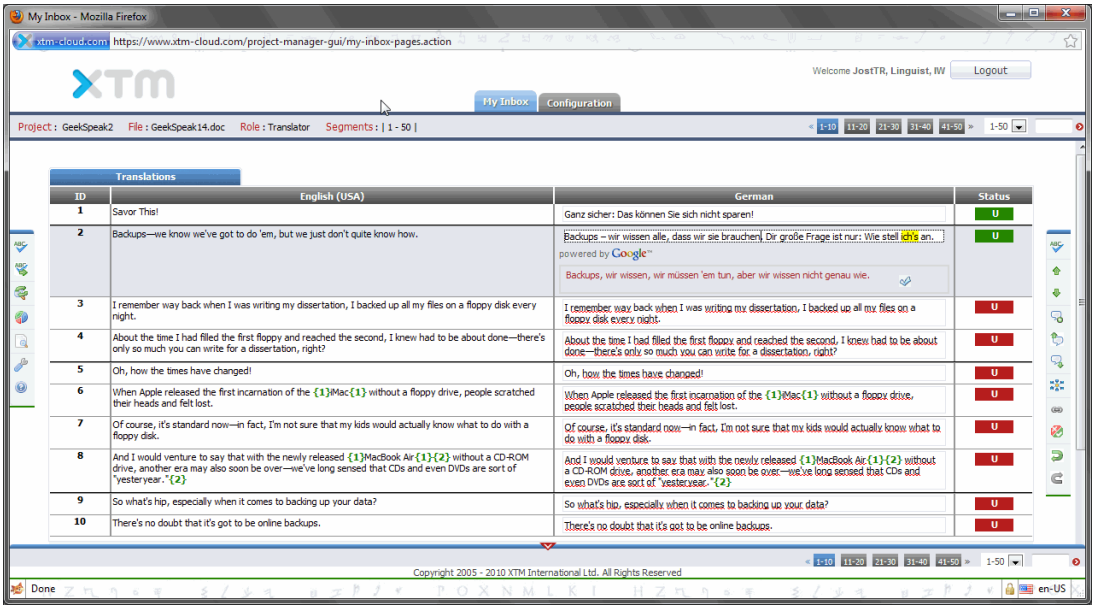


Figure 165: *XTM Cloud's* translation interface

Other remarkable features of *XTM* include "preprocessing," i.e., the ability to correct a badly written text before translation and the use of "pivot languages." Pivot languages come into play when a text is produced in, say, Chinese and has to be translated into 12 languages, and it's just a lot easier (and cheaper) to find the English> Finnish translator than the Chinese> Finnish translator.

The *Visual Editor* is a relatively recent feature. Rather than using the tabular translation interface, *XTM* also allows you to work right within the respective WYSIWYG (what-you-see-is-what-you-get) interface and still have the tools that present you with TM, MT, and term matches presented right in that view. (Presently this is only enabled for *Word*, *InDesign*, HTML and XML files.)

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 215) and *Wordfast Classic* (see page 196). Though browser-based its looks 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*, *Apache OpenOffice/ LibreOffice*, HTML, text, *FrameMaker*, *InDesign*, PDF, XLIFF or the *Wordfast Pro* format TXML) from either your computer or your *Dropbox* or *Google Drive*. 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), the large, public VLTM database (see page 198) and/or MT through various providers. (To set up all these settings, select **Wordfast Anywhere> Setup** and **TMs & Glossaries> Setup**.)

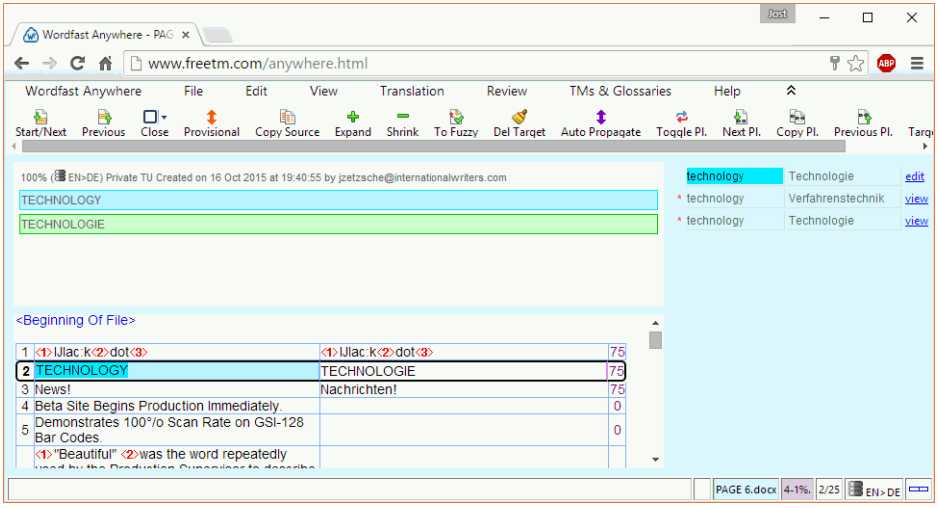


Figure 166: *Wordfast Anywhere's* translation interface

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 exclude *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.



Plus, one reason why really anyone should have an account with Wordfast Anywhere is because of its excellent conversion of image-based PDFs. For more information, see Using OCR Features for PDF Conversion in Translation Environment Tools on page 378.

Memsource

Memsource is a tool with a lightning-fast development. The core concept of cloud-based (online) translation memories and termbases was there from the beginning, but the translation interfaces have gone through a speedy evolution.

It all started off with an *MS Word*-based translation interface, but, unlike other *MS Word*-based tools, this interface was only used to work on translation tables consisting of the extracted text from the supported formats (*MS Office*, *InDesign*, *FrameMaker*, HTML, XML, TTX, XLIFF, SDLXLIFF) and a dynamic view of matches from the translation memory and termbase that were hosted by *Memsource* (using Amazon's AWS services). The setup of the project and preparation of files and the TM and TB all had to be done in a browser-based interface from which you then assigned the translation file to your translator (who could be yourself).

The *MS Word* translation interface was replaced with a very lightweight XLIFF editor ("*Memsource Editor*") in 2011. Just like before, the project and the files need to be prepped in a browser interface, and the online TM and termbase are assigned to the translator who then downloads an MXLFF file (which is an XLIFF file with some specific *Memsource* extensions that it can be translated in other tools as well).

Once you have the correct information to log into the *Memsource Editor* there really is nothing you need to know (well, there are a couple of things, but you can quickly glean those from the menus) and you can start translating and using the resources that are automatically displayed.

As mentioned above, these include TM and termbase data, subsegmenting data from the TM (see page 249) and machine translation data (if that was set up during the project preparation).

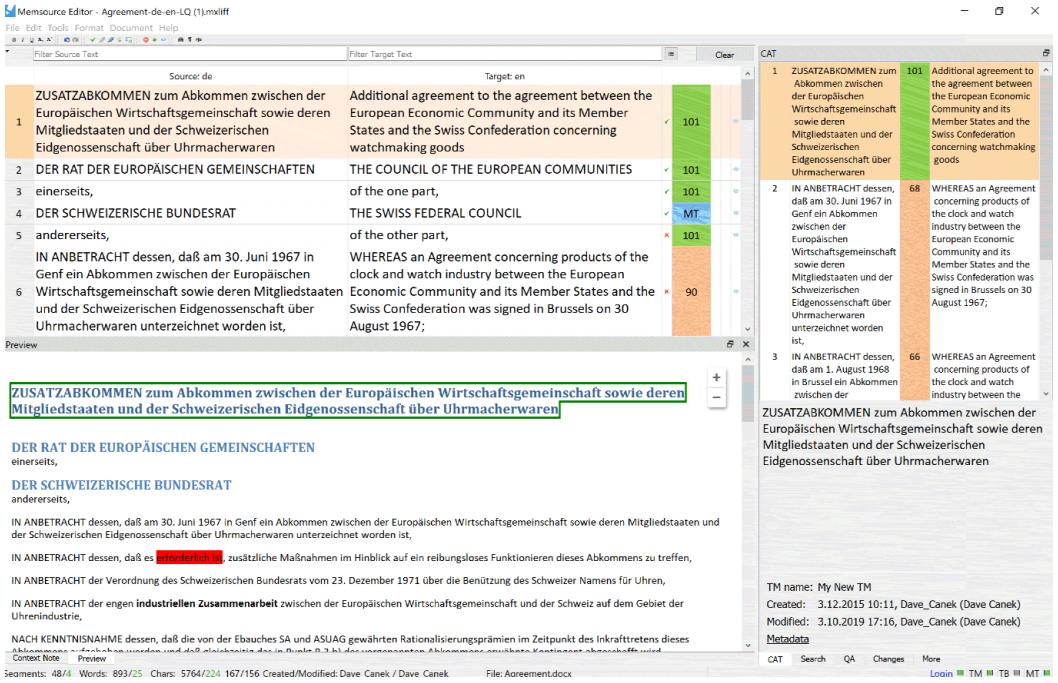


Figure 167: *Memsource Editor's* translation interface with data from TM (101), termbase (TB), machine translation (MT) and subsegments (S)

At the end of 2012, yet another editor was introduced, the *Memsource Web Editor*. The completely browser-based editor is offered in tandem with the desktop-based *Memsource Editor* and is, as shown in the image below, very similar in both appearance and functionality.

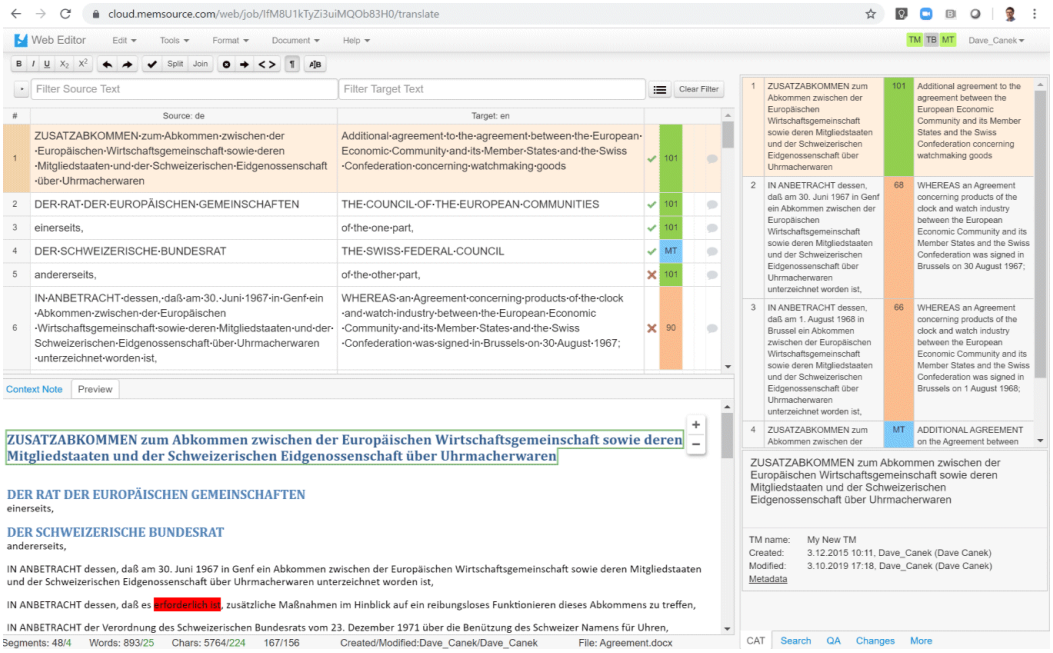


Figure 168: *Memsource Web Editor's* translation interface



One differentiating feature of Memsource Web Editor is that all translation strings of the file in question are loaded simultaneously, which makes search-and-replace and filtering processes easier than in competing browser-based editors.

In 2019, yet another interface was presented: the *Memsource Editor for Mobile*, available for *iOS* and *Android* devices. While many likely feel that a mobile device is not the most productive environment, there are times when it's super practical, and there are also tasks that are likely more practical than

others (editing and proofreading, for instance). Either way, the development team's strategy was to use as much as possible from the functionality and visuals of the web interface while at the same time completely redesigning the interface to make it workable on mobile devices.

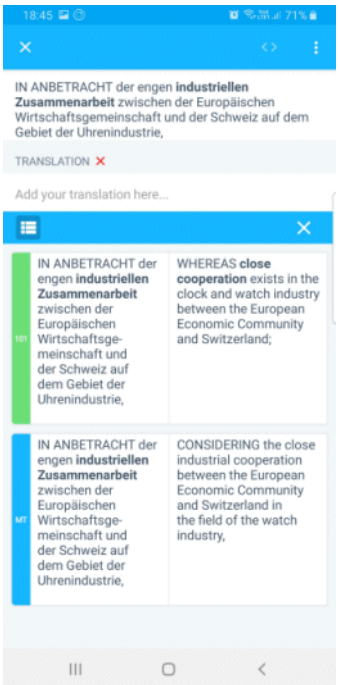


Figure 169: *Memsource's* mobile editor

One other area that Memsource has invested in quite heavily is the use of artificial intelligence (AI) beyond neural machine translation alone. Some features that they already are offering include:

- a language-combination-specific recognition of non-translatables
- MTQE ("machine translation quality estimation"), a per-segment process that uses language-combination-specific data collected across users to estimate the quality of an MT suggestion
- Memsource Translate, an automated way to select the best machine translation engine for each project

Smartcat

Smartcat is a cloud-based platform that attempts to address all aspects of the language service ecosystem. Originally operating as part of ABBYY (typically known to translators as a provider of high-quality optical character recognition technology—see page 368) and released in 2014, Smartcat became independent in 2016, and is now headquartered in Boston.

As a completely cloud-based translation environment tool, everything for *Smartcat* has to be performed in a browser. Once your files are uploaded to a server where they are processed, they are presented to you in the typical tabular (source on the left, target on the right) translation interface.

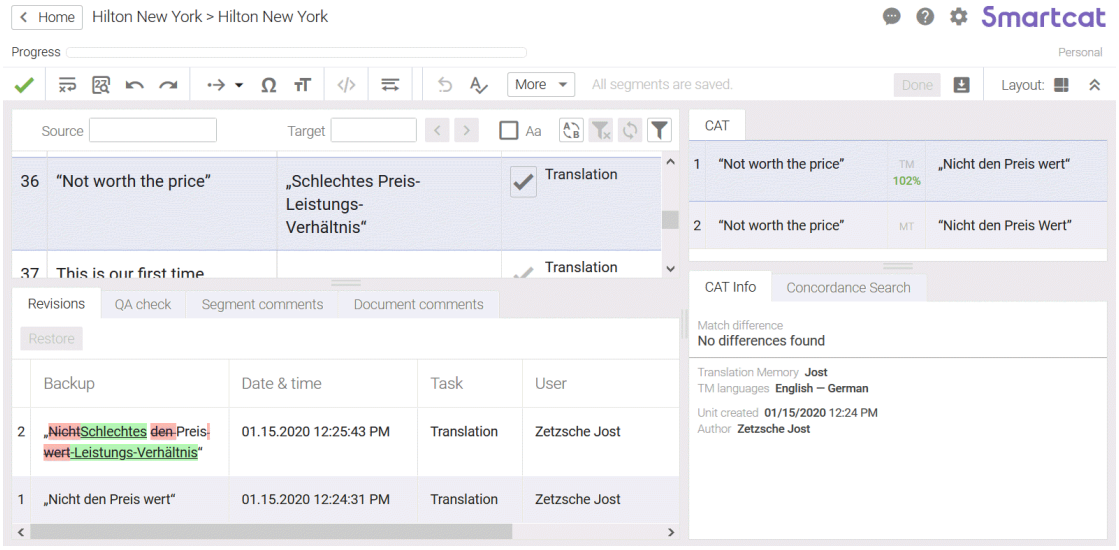


Figure 170: *Smartcat's* translation interface

The supported file formats include a very large range of word processing, software development, subtitling, content management system, and desktop publishing formats (including *Trados* package formats). What makes this list even more interesting is the addition of image formats, including graphic files and image-based PDFs. For these formats, *Smartcat* uses the ABBYY-owned

OCR engine and does a fine job with graphics—if the fonts used on the graphics are regular fonts—and with PDFs (see *Using OCR Features for PDF Conversion in Translation Environment Tools* on page 378 for more information).

If you want to use machine translation services (for the supported engines, see page 275) you don't pay a fee to any of the MT providers directly, instead you pay a fee to Smartcat (as of January 2020, US$20 for 500 pages at 250 words, but that pricing structure is likely to change). To use *Smartcat*, this fee and a comparable fee for OCR services are the only fees you have to pay (if you want to use the MT and OCR services, of course). This is true for free-lancers, translation agencies, and translation buyers alike.

The business model behind this is what Smartcat calls "Connected Translation," a solution that covers all aspects of the translation business, including vendor management (a large marketplace offers access to thousands of freelance translators and LSPs), workflow management and billing services. The latter are fee-based as well if you use *Smartcat* to send payments out.

If you need corporate individual customer service, including customized solutions for non-supported content management systems, there are also monthly or annual paid packages available (see smartcat.ai/pricing).

Another highly interesting feature includes the availability of morphology for 38 languages (see twitter.com/Jeromobot/status/651073023978528768 for a list of the languages) to improve termbase and TM recognition.



Another of Smartcat's highly interesting features is the availability of morphology for 38 languages (see twitter.com/Jeromobot/status/651073023978528768 for a full list of the languages) to improve termbase and TM recognition.

MateCat

MateCat is a tool that was originally developed under an EU grant with the goal of creating an"adaptive" machine translation system that immediately learns from corrections made by the editors to machine-translated segments.

While today there is only a link to that MT system in *MateCat*, it is still owned by the same parent company (Translated) but operated under a different name (ModernMT). Still, *MateCat* is a highly functional translation environment that is offered as a free tool for both translators and LSPs. The catch? For every file and project you translate within *MateCat*, you are offered the services of Translated, the company that now owns and runs *MateCat*, for prices that the majority of professional translators would not start to work for. Naturally you can choose not to accept the offer.

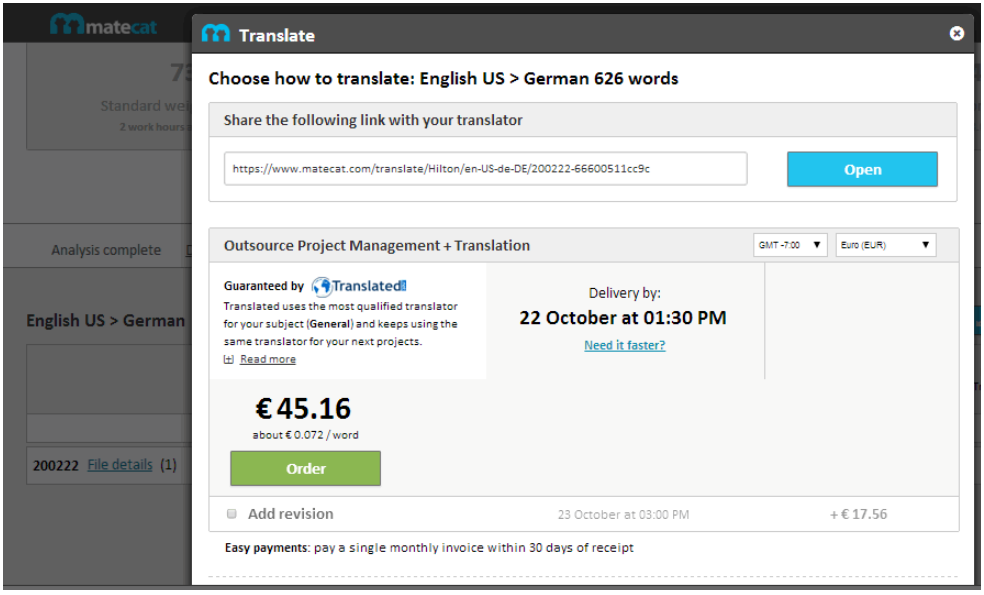


Figure 171: Automatically generated offer sheet by Translated



There is also a truly open-source version of MateCat (see matecat.com/open-source), but it comes without any file filters except XLIFF, so you would have to either use a tool like Rainbow (see page 319) to create translation packages for your project or create all the file filters from scratch.

The range of supported file formats is impressive . . .



Figure 172: Supported file formats in MateCat

. . . and the display of the translation interface in the web browser is easily navigable and transparent.

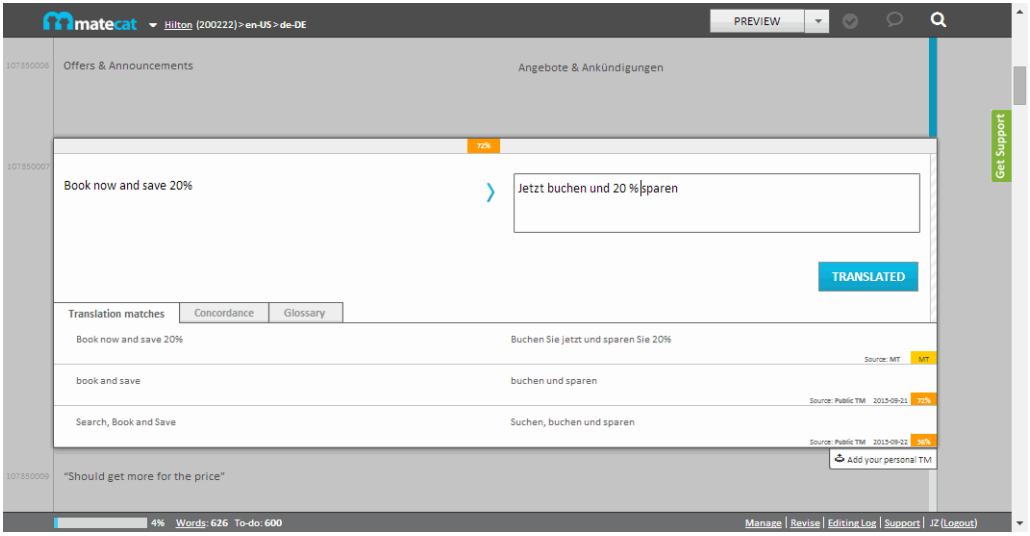


Figure 173: *MateCat's* translation interface

In the default setting, MT suggestions come from either *Google Translate* or Microsoft *Translator*. The TM matches can come from your own TM (which you can decide to keep private or share) and *MyMemory*, gigantic assemblies of bilingual data that Translated has collected (see page 276).

There is a glossary-like feature where you can add terms during translation, but they will be added to the TM rather than a standalone glossary.

MateCat also offers a feature that essentially eliminates the manual placing of tags or inline codes in translation segments in 28 language combinations (January 2020). Instead of showing the tags in the source and target segments, a "Guess Tags" button is displayed; this both enters the source tags and guesses the placement of the target tags. If the target tags are not correctly placed, you can correct them manually. The reason they often will be displayed correctly is because of the underlying word alignment that *MateCat* performs on those language combinations.

Termsoup

Termsoup has been around since 2016, and its market so far has been Asia, with a predominant focus on Taiwan. Taiwan publishes about 10,000 books annually that are translated out of other languages. That significant market is predominantly handled by the publishers themselves who contract with translators and typically do not ask those translators to use any kind of technology.

This is what *Termsoup,* a browser- and cloud-based translation environment tool, was developed for. The idea was for a tool that is not as massive as *Trados* or *memoQ* but more capable than *MS Word* as a translation environment. The name of the tool also reveals that the emphasis is not on translation memory or machine translation but on terminology, including third-party language-specific dictionaries. It uses *MateCat's* filters (see page 241) to convert a very large range of file formats behind the scenes into XLIFF, which is then presented in a bilingual format (either vertical or horizontal) for translation purposes.

What differentiates the display and *Termsoup's* handling of text is that it does not slavishly carry out segmentation on a sentence level (as you can see in the image below).

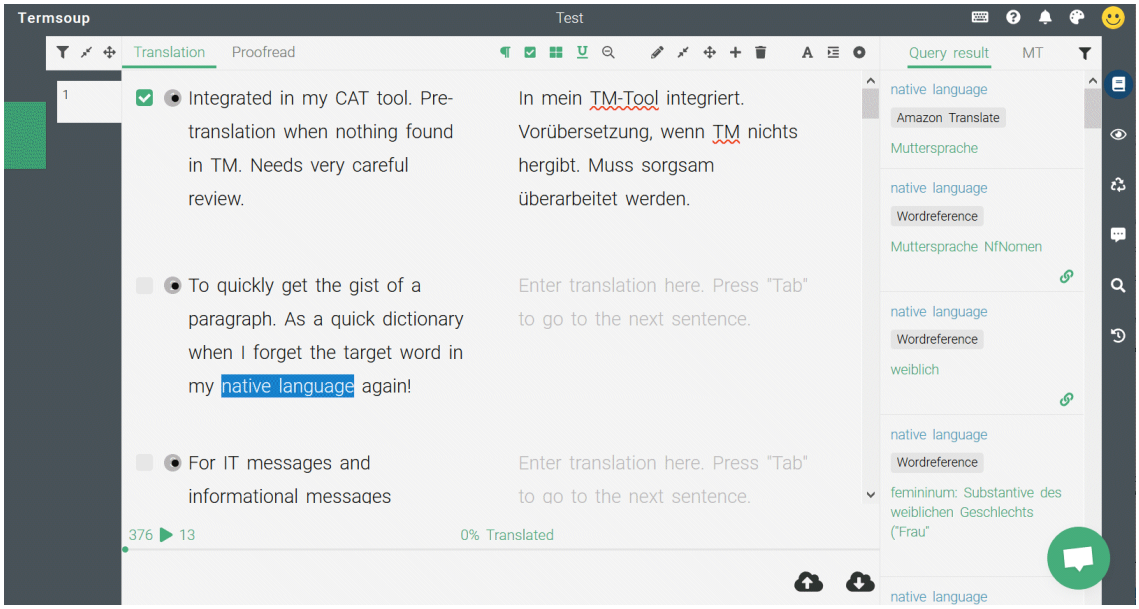


Figure 174: *Termsoup's* interface in any of the supported browsers (*Chrome*, *Firefox*, *Safar*)i

What you can also see in the sparse interface that is shown in the image is the focus on the text and the term search on the right, activated by the highlighting of a term in the source column.

Lilt



I have been going back and forth on whether to include Lilt in this edition as a tool. After all it's available only to translators who work for Lilt's service provider department. I decided to include it because users who have gotten access to Lilt, the tool, because they provided services to Lilt, the service provider, are allowed to continue to use it for their own purposes. Plus, it's just a very interesting tool.

Lilt is a tool that is derived from the PhD work of a Stanford graduate, Spence Green, who researched ways to have a machine translation engine respond interactively to the input of the translator on a per-word level. The machine translation engine (available for English <> Afrikaans, Arabic, Bengali, Bulgarian, Chinese, Croatian, Czech, Danish, Dari, Dutch, English, Farsi, Finnish, French, German, Greek, Hebrew, Hindi, Hungarian, Igbo, Indonesian, Italian, Japanese, Javanese, Korean, Norwegian, Pashto, Persian, Polish, Portuguese, Romanian, Russian, Slovak, Slovenian, Spanish, Swahili, Swedish, Thai, Turkish, Ukrainian, Urdu, Vietnamese as well as German into French and Italian as of January 2020) is a general purpose engine hosted by *Lilt* and built mostly from publicly available data. After entering each word, new queries are sent to the machine translation engine on how to finish the segment.

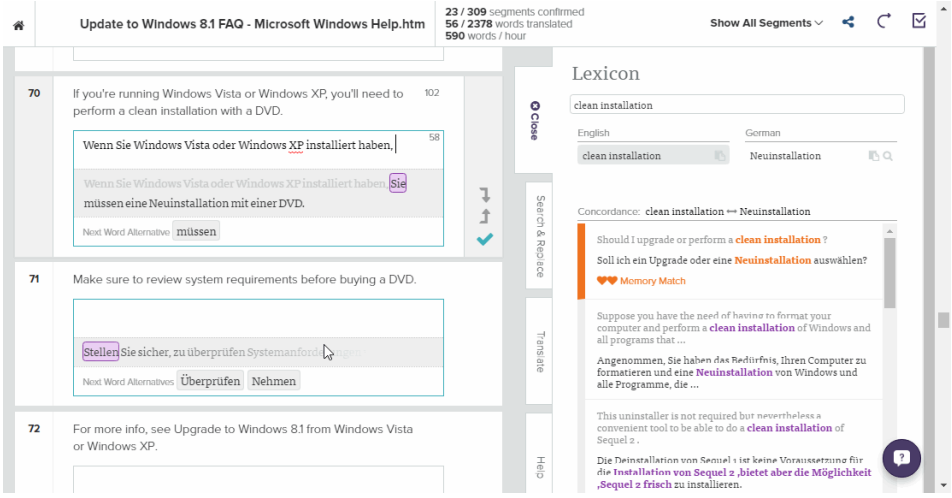


Figure 175: *Lilt's* translation interface

Note how the suggestion in the next screenshot changes (and becomes grammatically more accurate) after entering one additional word:

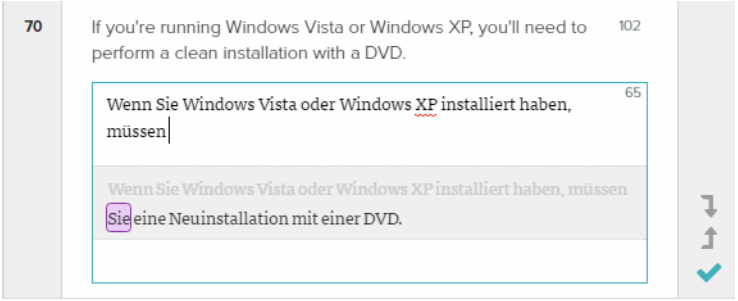


Figure 176: Dynamically corrected machine translation suggestion in *Lilt*

The highlighted term or phrase within the machine translation suggestion can be entered with a keyboard shortcut (ENTER or TAB), after which the next term is highlighted (or the suggested translation changes).

It's also possible to upload termbases and translation memories (and generate them as you translate). If a translation memory match of 85% accuracy or higher is detected, it—rather than the MT suggestion—is provided. A "Lexicon," which consists of a large glossary and a concordance search of public resources combined with the uploaded translation memory and termbase data, becomes available by double-clicking on a source term.

While Lilt uses the terms "translation memory" and "termbase" it does actually not use them in the traditional sense. Any data that Lilt uses is stored as an linguistic resource that the machine translation engine uses for its predictions. Any new entry will be automatically taken into consideration for better machine translation suggestion. This means that the machine translation engine interactively learns in real-time from any translation that is being performed.

Terminology and phrasal matches are being derived from terminology mining processes performed on the same resource (while any data that has been added to the phrase table in the form of glossaries is given preferential status.)



Supported file formats include MS Office 2007 and above, OpenOffice/ LibreOffice, XML, HTML, PHP, XLIFF of various tool-specific formats, SDLPPX, InDesign, FrameMaker, PDF, TTX, TXML, and many text- and XML-based development formats.

And it's the same underlying data that allows *Lilt* to be the only tool that offers morphology for every morphological language (thus excluding Chinese) that it supports. It uses a "neural morphology" engine that creates morphology rules on the basis of the underlying corpus and applies that to terminology recognition.

Comparing TEnTs

There are too many tools out there to make detailed comparisons of every available feature in every available tool. Instead, I will focus on the main features that are present in most tools and show how these are handled in one or two of them. You can use this list of features to evaluate the tools in making a decision for or against a certain tool.

Translation Memory Handling

While the core concept of translation memory functionality is simple—store previous translations on a per-segment basis in a database that will propose perfect or fuzzy matches during new translation—there are surprisingly many variations in how it is implemented and a number of relatively new developments.

Most tools use an external database to store the translation memory. This can happen in various formats, starting from text-based (such as in *Wordfast Classic*) to *Microsoft Access* (as in the case of *Déjà Vu*) or to a large variety of more high-powered databases that are often based on existing technologies (such as the open-source *SQLite* in the case of *Trados Studio* or *Microsoft SQL Server* in the case of *Across*).

Other ways of dealing with translation memory-like features include that of *Star Transit* with its "virtual" translation memory (see page 207) and the corpus-approach of tools like *MultiTrans* (see page 198) or also *memoQ* (here in addition to a more traditional TM—see page 210).

But before continuing with other differentiators in the handling of translation memories . . .

. . . . 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. Virtually all tools 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.

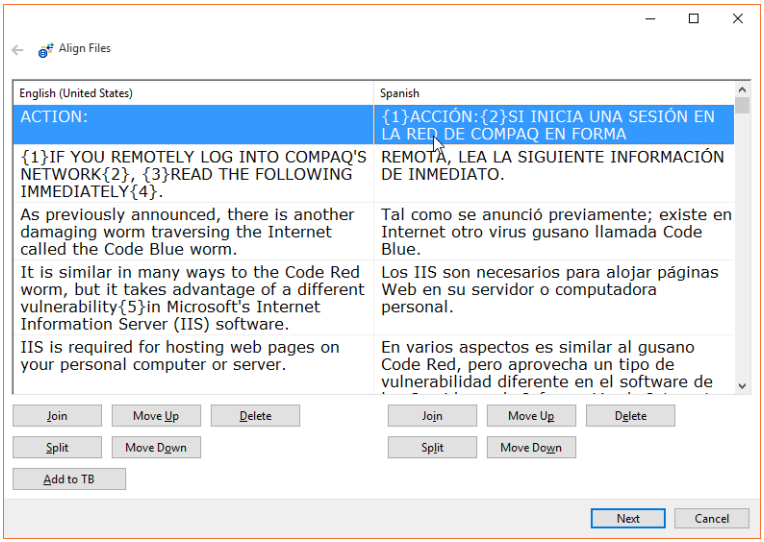


Figure 177: Alignment view in *Déjà Vu* (note that the program split the first sentence incorrectly in the Spanish target)

Trados Studio's initial version *2014* threw out SDL's traditional *WinAlign* alignment tool and offered an integrated alignment tool with an option called "alignment quality value." This is a setting that allows the user to adjust the confidence level of the match. For example, if you had set a high quality level and there was a huge difference in the number of words in the source and target segments, or there were numbers in the source but not in the target, the translation units were rejected.

While this sounded good in principal, it did not produce the results that were hoped for, so in later editions of version *2014* and above, SDL reverted to the old way of reviewing the alignment first before you send the data to the translation memory (unless you choose **Align Multiple Files**—even if it's just a single file pair—and then select **Save alignment result files for later review**).

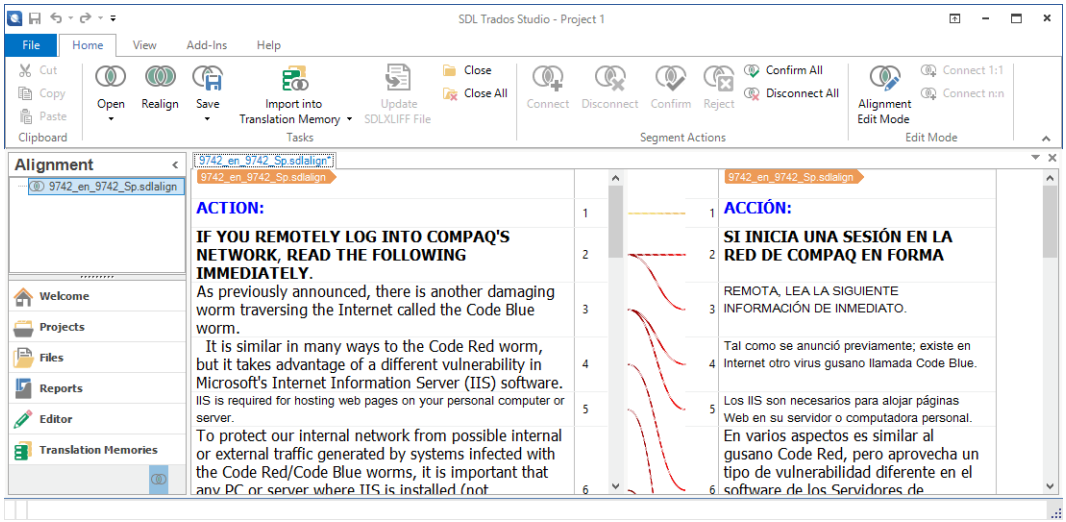


Figure 178: Alignment in *SDL Trados Studio 2016*

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 terminotix.com/index.asp?content=brand&brand=1) 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).

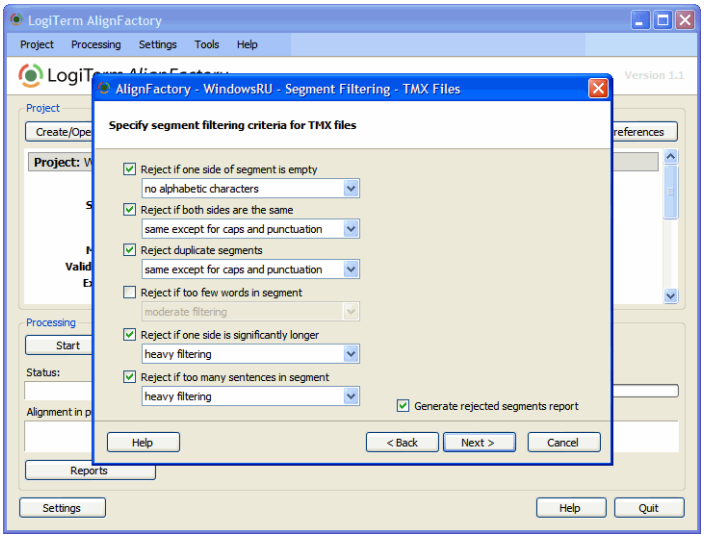


Figure 179: *AlignFactory's* filter settings for alignment

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 (in some editions you can even directly download whole websites and align them). 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.

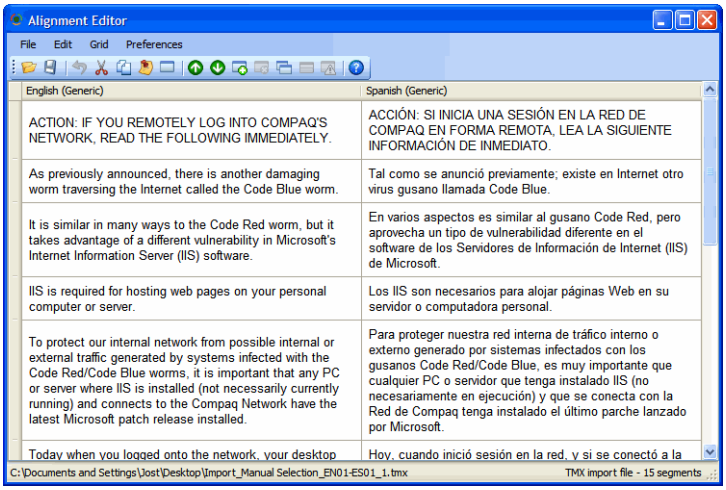


Figure 180: *AlignFactory's* alignment results—note the correct alignment of the first segment



YouAlign (see 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 good idea of what AlignFactory can do.

Another alignment tool that works somewhat differently is *LF Aligner* (see sourceforge.net/projects/aligner). What's different about this tool is a linguistic backbone in the form of dictionaries in a total of 32 languages (all the EU official languages except Irish Gaelic and Maltese, as well as Croatian, Arabic, Hebrew, Chinese, Japanese, Korean, Latin, Norwegian, Russian, Swahili and Turkish), permitting the automatic generation of dictionaries in any combination of these languages. It's on the basis of those dictionaries that the tool decides how to match segment pairs.

The supported file formats include TXT, DOC(X), RTF, HTML, TMX and (in a limited fashion) PDF. Or you can download EU documents or other online documents for alignment purposes directly from within the tool.

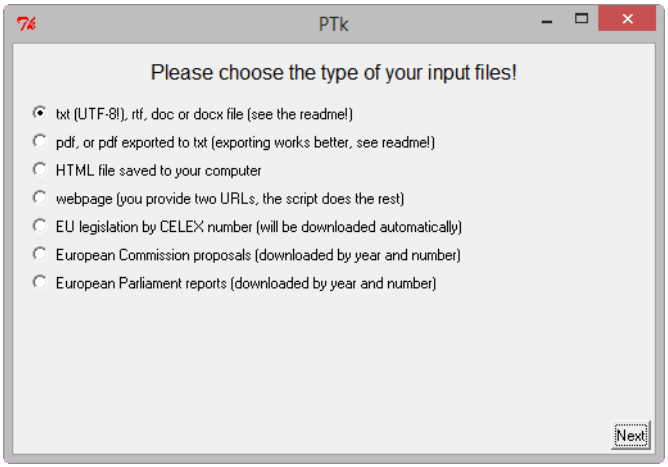


Figure 181: Format options in *LF Aligner*

Yet another tool that has started to use (and improve) the logic of *LF Aligner* is *XTM* (see page 233). Its development team acquired very comprehensive lexicon data in 50 language pairs as the basis of their alignment, resulting not only in better alignment accuracy but also the ability to give a relatively reliable "confidence score" for every proposed segment match. The segment pairs are output in two different *Excel* tables (I know, not my favorite tool for this kind of task either...). One file lists all segment pairs with at least 90% confidence (the confidence is assigned on the basis of the lexicon data, as well as items like function words, numerals, named entities, and of course

punctuation), and another file contains every segment pair and therefore needs further editing. So, if you're in a hurry, you just take the "good" file and import that into a TM; if you really want to squeeze everything out of a text, you take the file that might need some work.

	A	B	C	D	E
1	Src Seg.	en_US	es_NT	Tgt Seg.	Probability
2		Unmatched Not sure Merged Very good			
3	0	ACTION: IF YOU REMOTELY LOG INTO COMPAQ'S NETWORK, READ THE FOLLOWING IMMEDIATELY.	ACCIÓN: SI INICIA UNA SESIÓN EN LA RED DE COMPAQ EN FORMA REMOTA, LEA LA SIGUIENTE INFORMACIÓN DE INMEDIATO.	0,1	1.0000
4	1	As previously announced, there is another damaging worm traversing the Internet called the Code Blue worm.	Tal como se anunció previamente; existe en Internet otro virus gusano llamada Code Blue.	2	1.0000
5	2	It is similar in many ways to the Code Red worm, but it takes advantage of a different vulnerability in Microsoft's Internet Information Server (IIS) software.	Los IIS son necesarios para alojar páginas Web en su servidor o computadora personal	3	0.7586
6	3	IIS is required for hosting web pages on your personal computer or server.	En varios aspectos es similar al gusano Code Red, pero aprovecha un tipo de vulnerabilidad diferente en el softw are de los Servidores de Información de	4	0.7806
7	4	To protect our internal network from possible internal or external traffic generated by systems infected with the Code Red/Code Blue worms, it is important that any PC or server where IIS is installed (not necessarily currently running) and connects to the Compaq Network have the latest Microsoft patch release installed.	Para proteger nuestra red interna de tráfico interno o externo generado por sistemas infectados con los gusanos Code Red/Code Blue, es muy importante que cualquier PC o servidor que tenga instalado IIS (no necesariamente en ejecución) y que se conecta con la Red de Compaq	5	1.0000
8	5	Today when you logged onto the network, your desktop should have received the newest patch if you connect directly to the corporate network.	Hoy, cuando inició sesión en la red, y si se conectó a la red corporativa, debería haber recibido en su escritorio el último parche disponible.	6	1.0000
9	6	If you access Compaq's networks from a remote location using RAS or tunneling, or if for some reason you do not run the corporate login script to log in, you will need to follow the instructions listed below.	Si usted accede a la red de Compaq desde una ubicación remota utilizando RAS o "tunelización" o si por algún otro motivo no ejecuta el script corporativo de inicio de sesión, debe seguir las instrucciones que se detallan a continuación.	7	1.0000
10	7	If you run the corporate login script when you log in, you should not need to read further.	Si para iniciar sesión, ejecuta el script corporativo de inicio de sesión, no necesita	8	1.0000
11	8	Please be advised you are putting Compaq's network at risk if you are running IIS and you do not have the patch installed.	Por favor, tenga en cuenta que si tiene instalado IIS y no instaló el parche, está poniendo en peligro toda la red de Compaq.	9	1.0000
12	9	Please follow the instructions below to ensure you are not putting Compaq's network at risk.	Siga las instrucciones detalladas a continuación para asegurarse de no estar poniendo en peligro la red de Compaq.	10	1.0000
13	10	WHAT DO I NEED TO DO?	¿QUÉ DEBO HACER?	11	1.0000
14	11	Upon RAS/tunnel login, you need to download the MS01-044 patch from one of the following locations and install as instructed:	Una vez que haya iniciado sesión RAS o por túnel, debe descargar el parche MS01-044 desde alguna de las siguientes ubicaciones e instalarlo según las instrucciones:	12	1.0000
15	12	If you do not map to a Corporate S: drive	Si no tiene configurada una unidad S: Corporativa	13	1.0000
16	13	Windows NT 4.0 with IIS installed: http://www.microsoft.com/Downloads/Release.asp?ReleaseID=32061	Windows NT 4.0 con IIS instalado: http://www.microsoft.com/Downloads/Release.asp?ReleaseID=32061	14	1.0000
17	14	Windows 2000 Professional, Server and Advanced Server: http://www.microsoft.com/windows2000/downloads/critical/q301625/default.asp	Windows 2000 Professional, Server y Advanced Server: http://www.microsoft.com/windows2000/downloads/critical/q301625/default.	15	1.0000

Figure 182: *XTM's* alignment results with the confidence rating ("Probability"). Note that the first segment here is correctly aligned.



You can find some more information about the process in this blog post: xtm.cloud/blog/xtm-advanced-text-aligner-3.

If you are "just" interested in massive amounts of bi- or multilingual downloadable data to fill your tools with, there is plenty of that to be had. Be forewarned, though, that this does not necessarily increase the quality of your translation or the performance of your translation environment tool. Here are some sites where you can download such data:

- Various EU translation corpora (ec.europa.eu/jrc/en/language-technologies)
- Hansards corpus of the Canadian parliament (isi.edu/natural-language/download/hansard)

- UN corpus in Arabic, Chinese, English, French, Russian, Spanish (conferences.unite.un.org/UNCorpus)
- OPUS—a vast collection of translated corpora from a large number of sources (opus.nlpl.eu)

. . . Back to Translation Memories

The way translation memories are utilized goes beyond the familiar perfect-fuzzy match scheme of complete translation units. Particularly interesting in this context is the so-called "subsegment matching," predictive typing, and the combination of TM resources with terminology and machine-translated data.

Subsegment Matching

Subsegment matching uses a principle pioneered by *MultiTrans* early on but only now realized by a number of other tools (*Trados Studio*, *memoQ*, *Déjà Vu*, *CafeTran Espresso*, *Memsource* among others) using various methods. The idea is this: The chance that one particular segment (typically a sentence) will ever be repeated is relatively slim, unless you are translating version 1.1 of a manual for which you can use the translation memory that was created translating version 1.0. Still, there is plenty of material in the translation memories that can is useful. Since the very early days, tools offered concordance searches, i.e., manual searches for a particular part of a segment in a translation memory. While these searches were helpful, they were not only tedious but completely based on the translator's intuition of

what might or might not be in the TM. The idea of automated subsegment searches solves that dilemma by automatically displaying all applicable subsegment matches to the translator, and in the process essentially freeing up the 98% or so in the TM that would otherwise be useless.

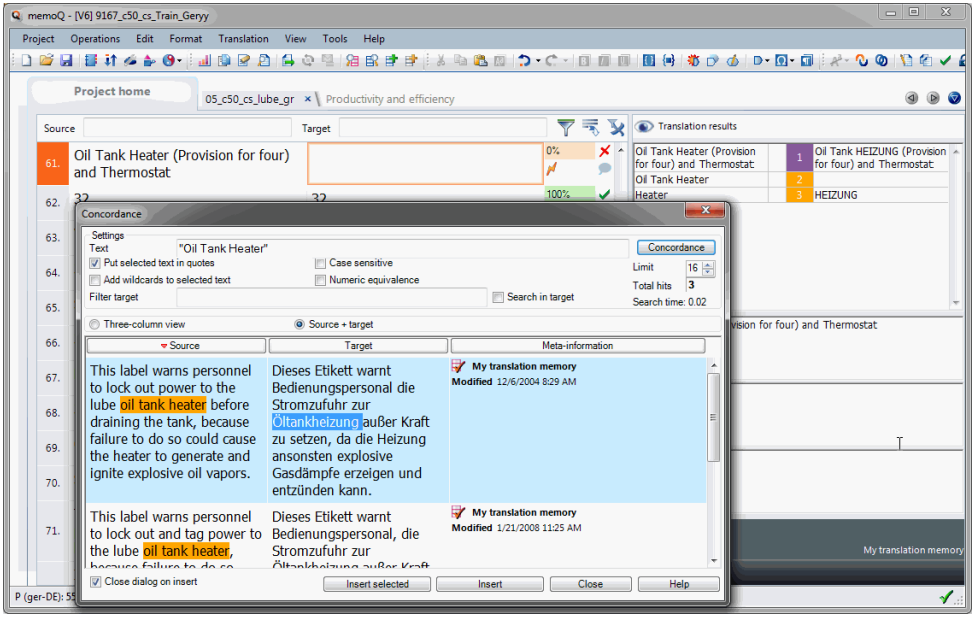


Figure 183: *memoQ's* subsegment matching feature (LSC or "Longest Substring Concordance" in *memoQ* lingo)

Predictive Typing

The second relatively new feature is essentially an extension of the first. We all are familiar with predictive typing in *Excel* or the address field of an Internet browser, typically called *autocomplete*. In those cases, a "database" of previous entries into the *Excel* spreadsheet or the browser is consulted if there are any matches to the first few keystrokes. If so, those matches are displayed and offered to the user for entry with a single keystroke.

The same principle is used in the translation environment tools that offer this feature (among them *Trados Studio*, *memoQ*, *Déjà Vu*, *Across*, *Star Transit, Lilt*, *Memsource* and *Wordfast Classic*), only that they in some way or the other use the content of the translation memory (and/or machine translation) to furnish these suggestions. Depending on your typing habits, this can be a huge time saver, and it increases consistency significantly.

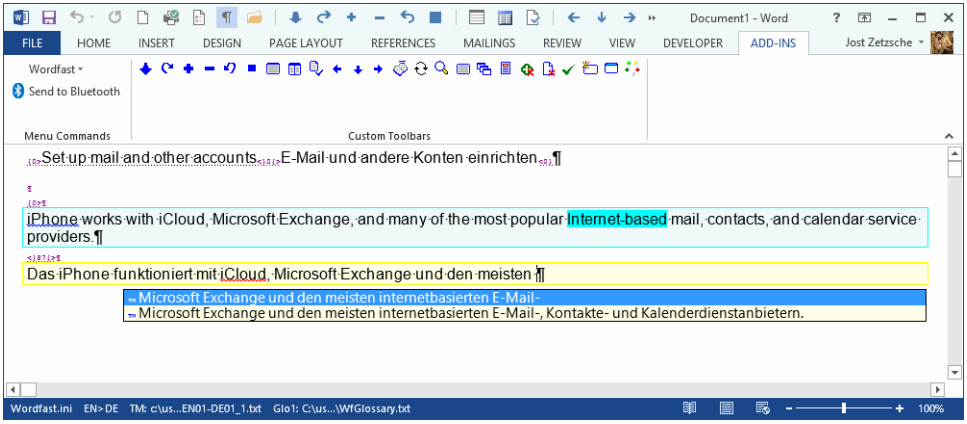


Figure 184: *Wordfast Classic's* autocomplete feature

Combination of Linguistic Resources

An feature that involves the handling of translation memory and other data is *memoQ's*, *CafeTran Espresso's*, *Trados Studio* (*2017*+) or *Déjà Vu's* assemble capability and the possibility of fuzzy match repair with the help of other TM content, termbase entries, or even machine-translated data.

In the following three screenshots, you can see three different levels of repair of fuzzy matches that might give you a good idea of the power and versatility of this feature. While this is obviously a project set up for demo purposes, it still uses "real data." The only content in the translation memory are three segments that have already been translated, and the glossary contains only one term.

Here *Déjà Vu X3* "repairs" a fuzzy match by utilizing data from the glossary ("lexicon"). Because the glossary contains the correct translation of the term that is incorrect in the fuzzy match, it can automatically switch these and turn the fuzzy match into a perfect match:

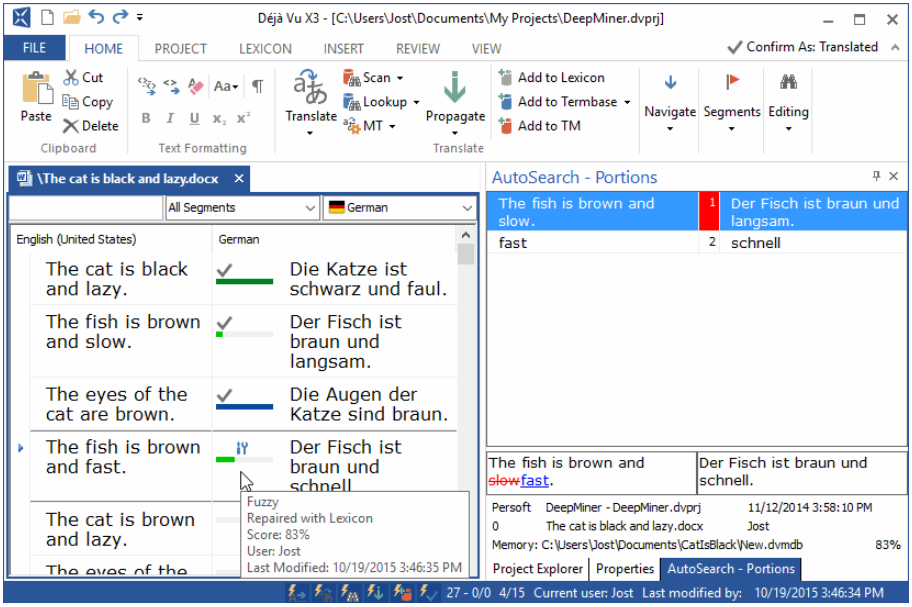


Figure 185: *Déjà Vu X3's* repair feature using glossary data

In the following screenshot, another fuzzy match is repaired even though the glossary was of no help. Here *Déjà Vu X3* subsegmentation ("DeepMining")was able to determine the correct translation because the term in question appeared twice in the segments that are already in the translation memory:

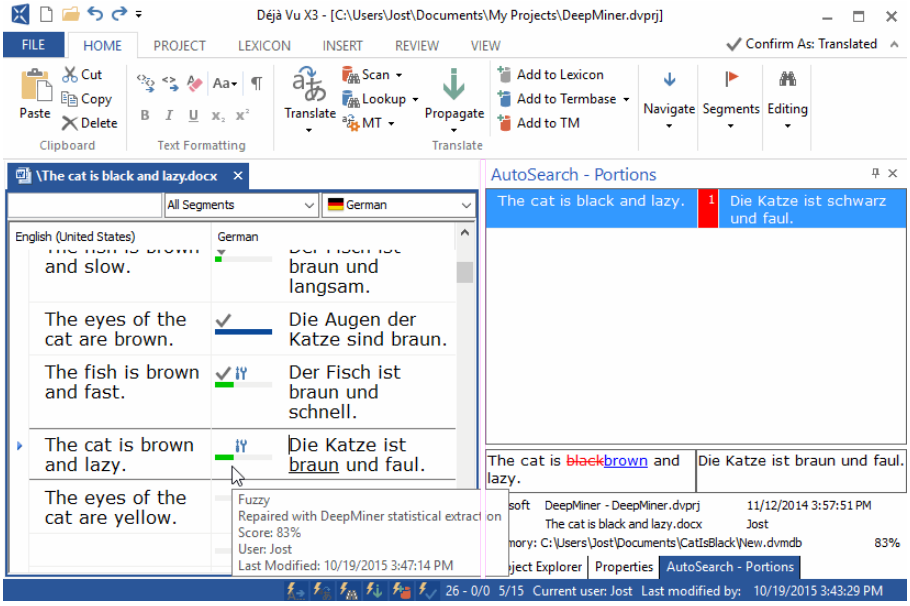


Figure 186: *Déjà Vu X3's* repair feature using subsegment data

In the last example, a term needs to be inserted that is not in any of the local resources at the disposal of *Déjà Vu X3*. In this case, it queries an attached machine translation engine for that single term (and not the whole segment) and fixes the fuzzy match like that:

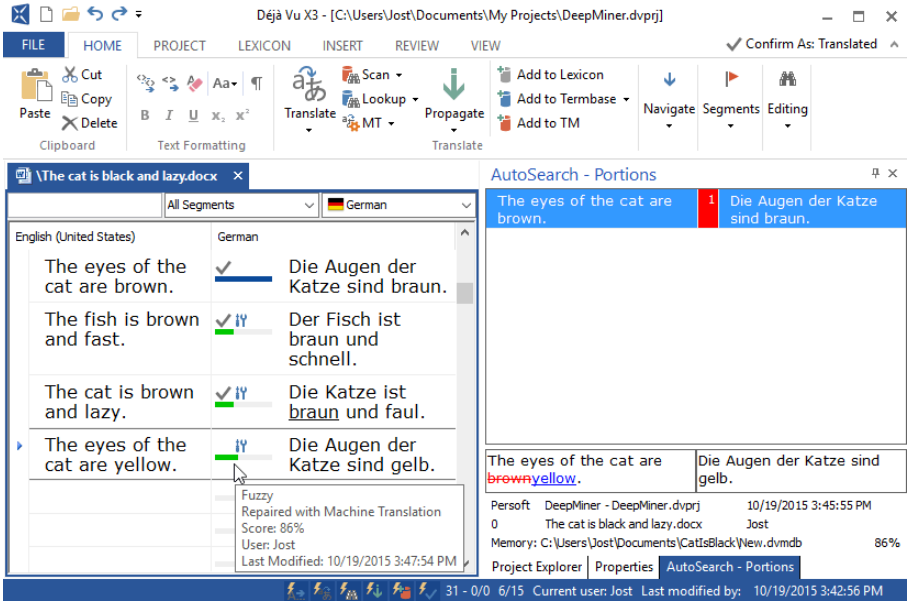


Figure 187: *Déjà Vu X3's* repair feature using machine translation data

Terminology Handling

Only a fraction of the translators who use a translation environment tool today are using the terminology component that all TEnTs offer. That's unfortunate because they are all missing out on one of the most powerful feature of TEnTs.

Of course, there is a reason for this; in fact, there several reasons:

- The often-used name "translation memory" program seems to suggest that the emphasis is on the translation memory.
- There seems to be 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 built up relatively quickly by aligning existing translated file pairs and/or aligning 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!

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 194).

Because of the initial hesitation among translators to show interest in translation technology in the 1980s and 1990s, let alone to cooperate with the technology vendors, some components of the tools, in particular the terminology components, became less translator- and more corporation-friendly.

Trados and *Transit* in particular offer powerful and complex termbase solutions that at least initially failed to win over many translators as users.



SDL in particular has tried very hard to overcome this by trying to make their terminology solutions seem more appealing while still catering to the needs of sophisticated terminology users. The latest attempt to do that is the launch of the SDL Language Cloud Terminology tool.

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.

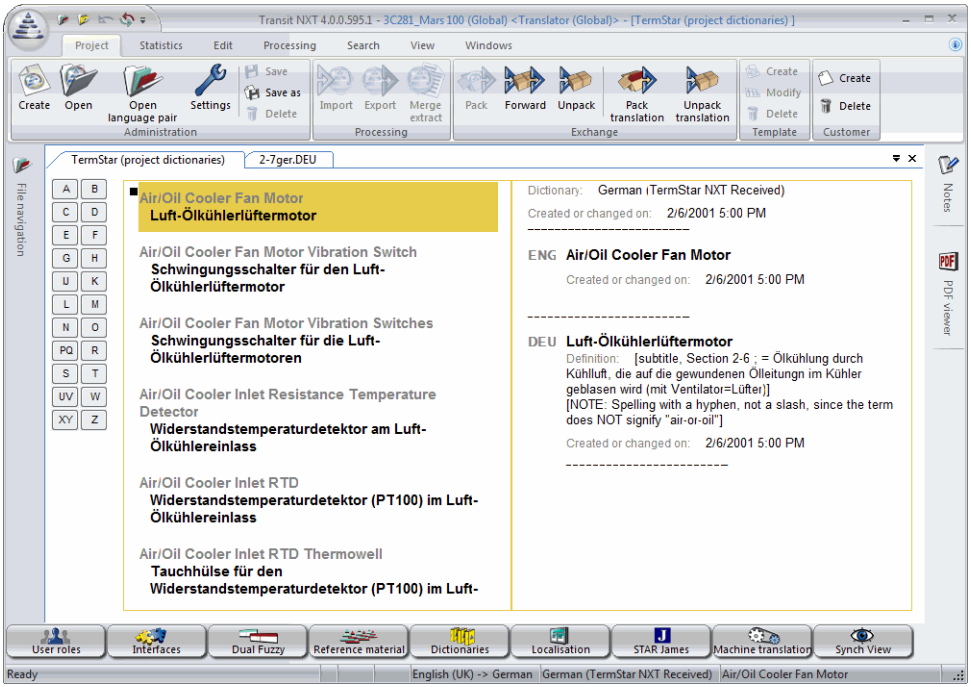


Figure 188: View of open dictionary (terminology database) in *TermStar*

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.

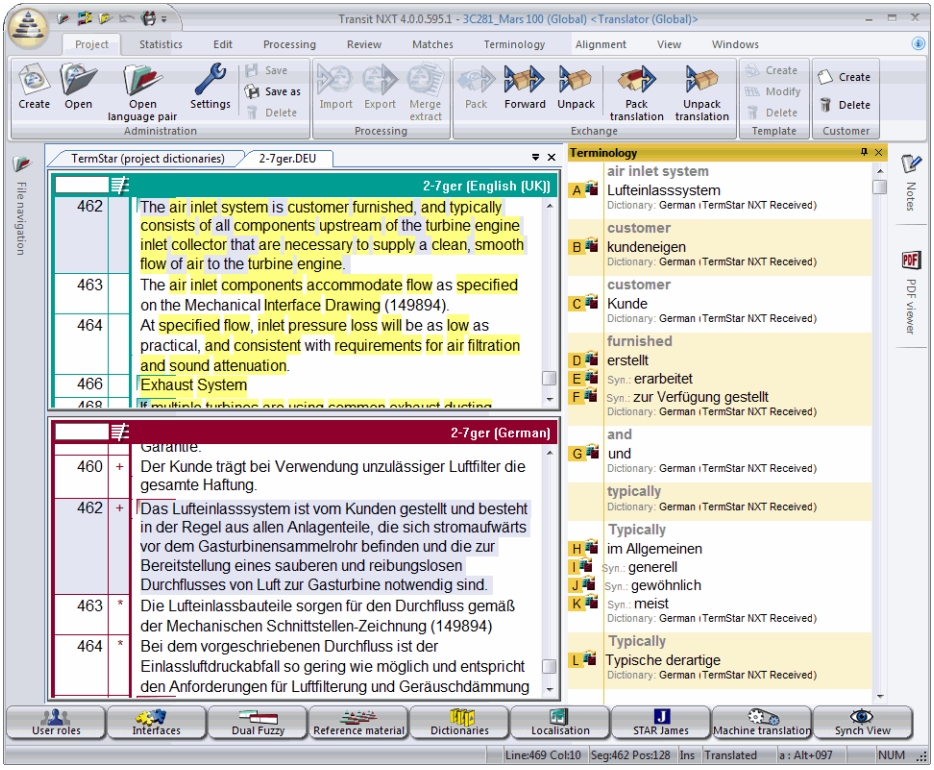


Figure 189: *Transit* translation project with dictionary access (right pane)

Trados's original terminology program *MultiTerm* (5.5 and earlier) was particularly 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 its 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.

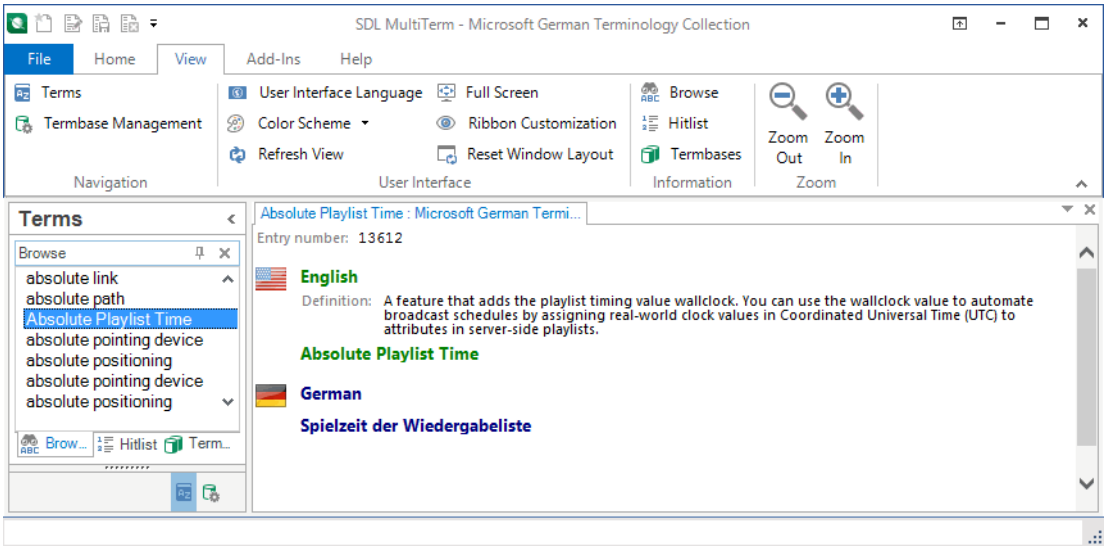


Figure 190: Sample view of *MultiTerm* termbase

In *Trados Studio* 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.

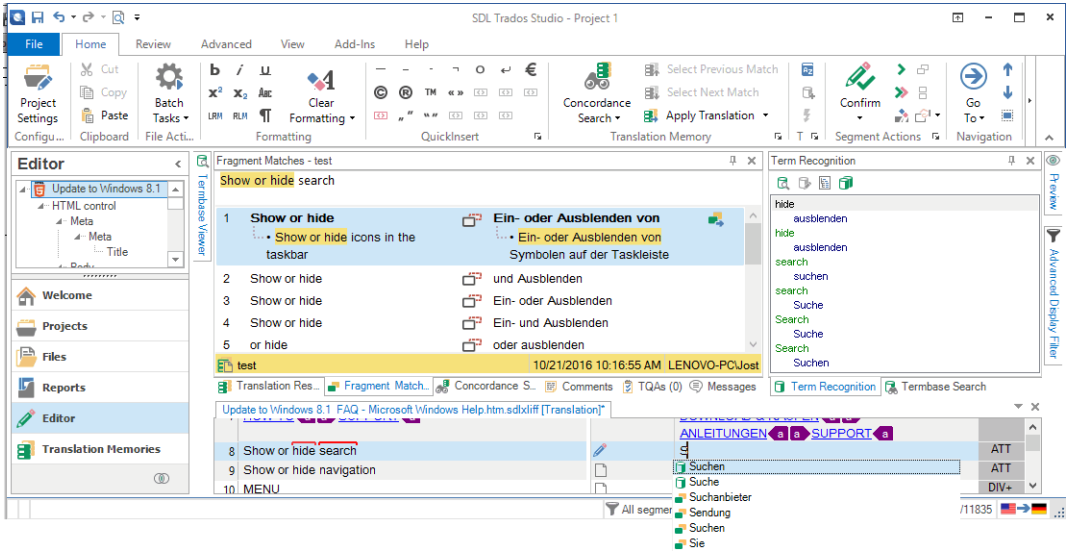


Figure 191: View of *Trados Studio* with an automatically offered terminology match



This autocomplete feature (AutoSuggest in Trados lingo) goes hand in hand with the TM-based autocomplete option (see page 259) and suggestion from machine translation segments.

Of course, there are also standalone terminology tools. You can find more information on those on page 306.

Work Environment

Since the work environment was already used as the main criterion to categorize the tools (see page 195), there may not be too much to add here, but the following might be helpful anyway.

Generally speaking, there are three different kinds of environment:

- the co-dependent environment in a word processing tool (usually *MS Word*)
- the co-dependent environment through a browser interface
- the independent environment through a standalone desktop-based application

Within those general frameworks that were dealt with in the Categories section (see page 195), there are some important differences as to how the translated text is displayed.

One of the arguments by proponents of the *MS Word* interface was always that it's so much easier to translate in an interface where everything is displayed exactly like it's supposed to be, also known as WYSIWYG or "What-You-See-Is-What-You-Get." So all the formatting that will be visible to the reader of the original document and eventually the translated document will be visible to the translator as well.

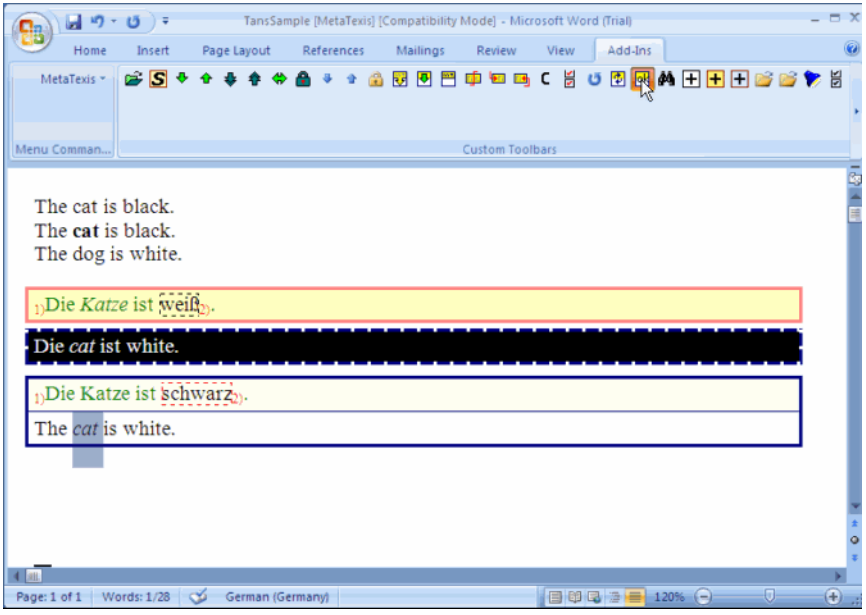


Figure 192: *MetaTexis'* WYSIWYG interface with a *Word* document in *MS Word*

Of course, this only works for documents that were directly compatible with *MS Word*. And while the majority of translators today work in a much larger variety of formats than "just" *Word* documents, some do primarily work in that format, and this might be a good solution for them.

For the makers of tools that display their translation in an independent interface, the question remained: How much visible formatting is necessary or at least helpful when it comes to translation?

Different tool vendors have answered this differently, but what seems to be an emerging trend in the last few years is a semi-WYSIWYG approach. While some of the more common formatting elements are displayed (such as bold, italics or underlining), others are not.

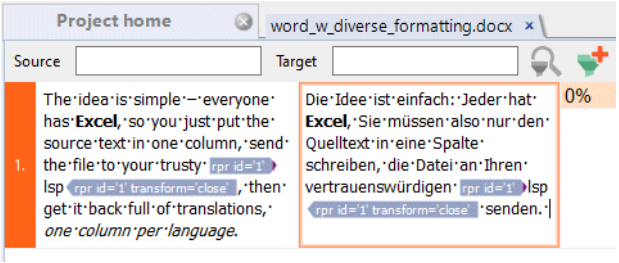


Figure 193: *memoQ's* semi-WYSIWYG interface (note that the bold and italic formatting is preserved but formatting tags are used for the small caps)

The preview option (either in a simulated or the actual originating application of the file) that many TEnTs now offer is another, more comprehensive way of addressing that concern.

Of course, there are other features that support word processing in programs like *MS Word*. These include for instance:

- Inline spell-checking (the real-time spell check with a red squiggly underline)
- AutoCorrect (the automated correction of often misspelled words)
- Track Changes (the visible correction marks)

These features have also already been implemented by many translation environment tools.

Integration of Machine Translation

Generally speaking, machine translation is the translation of a text from one language into another language by a computer without any human involvement. There is plenty of online and offline information about machine translation, but it often falls into one of three categories: it's highly technical and really not meant to address the average reader; it assumes that machine translation has essentially solved the "problem" of translation—often referring to other literary fixes from science fiction movies and books (like "Babelfish" or "Universal Translator"); or it focusses solely on the failures of machine translation.

The following is an attempt to take some of the emotions out of the discussion and yet deal with it in a relatively easy-to-understand manner.

The first attempt at machine translation goes back to the early days of computing in the 1950s, when a so-called rules-based machine translation (RbMT) was developed and first tried on Russian-into-English translation. This form of machine translation consisted of a set of rules about the source and target language and included a dictionary. The transfer between the source and target language in rules-based MT happens either via an "interlingua," a computerized representation of the source text, or directly between the source and target language. Benefits of this kind of machine translation include the relative ease of adding to the dictionary and tweaking the rules. Drawbacks include a typically long ramp-up time to initially build the MT engine (the program that generates the translation) and a usually poor outcome with languages that are not closely related.

Statistical machine translation (SMT) started being used heavily in the early 2000s. The first commercial offering (*LanguageWeaver*, now owned by SDL) was launched in 2002, a widely used open-source engine (*Moses*) emerged in 2005, and providers of publicly available MT offerings such as Google and Microsoft switched from RbMT to SMT in 2007. Statistical machine translation—or, more accurately in the cases mentioned above, "phrase-based statistical machine translation"—is trained on professionally translated bilingual data and monolingual data. It parses the data into "n-grams," which are phrases consisting of an "n" number of words. The same thing happens to the source (original) segment in the translation process. The source n-grams are then matched with target (translated) n-grams, which are then combined

to form whole segments again. Depending on the quality of the training data, the weaknesses of this technique can be the often incorrect choice of the n-grams and their poor combination. This technology also tends not to do particularly well in language combinations with a widely differing syntax. The benefits are that with sufficient amounts and quality of data, the lead time to having some kind of system is relatively short, and the n-grams—the individual fragments—are correct in and of themselves since they originated from professionally translated content and are potentially useful for the re-use by professional translators.

The early 2010's saw the development of a number of hybrid machine translation systems in which both the rules-based and statistical methods were used in parallel.

And starting in 2016 a new wave of development called "neural machine translation" (NMT) began receiving a lot of attention. Much like SMT, NMT programs are fed with large amounts of data. Unlike the former, however, they use neural networks that consist of many nodes (conceptually modeled after the human brain) that relate to each other and can hold single words, phrases, or any other segment. These nodes build relationships with each other based on bilingual training material. Because of these manifold and detailed relationships, it’s possible to look at not just limited n-grams (as in statistical machine translation) but at whole segments or even beyond individual segments. This allows for more educated guesses about the context—and therefore the meaning—of any word in a segment that needs to be translated.

Many companies have been investing heavily in this technology, including Google, Microsoft, Amazon, Facebook, or SDL, but also companies who used to bank on other machine translation technologies.

In some areas, especially with generic texts, neural machine translation has shown to be a dramatic improvement, including in language combinations that previously were not well suited for machine translation. Drawbacks include the need for massive computing power and processing time as well as a "black-box syndrome" where faulty outcomes are difficult to fix.

Machine translation encompasses not only different technologies but also different purposes. Most of the publicly accessible web-based machine translation programs are not trained for specific kinds of texts or types of content. This generic approach means that the data used for the translation results comes from a very large pool of language data and is therefore unreliable for translated texts that will be used for professional purposes. Engines such as *Google Translate*, *Microsoft Translator*, *Yandex*, Amazon Translate, DeepL, *ModernMT* or *Baidu Translate* play a very important role in providing approximate ("gist") translations that make interlingual ad-hoc communication between users of different languages on the Internet possible. While mistakes do happen, the main message is communicated adequately enough to make these welcome tools for those purposes. However, because services typically keep any transmitted data for further training purposes, this open data market makes these tools patently unsuitable for confidential business communication.



As translators, it's important to understand what the different publicly available services do with your data.

At this point (January 2020), at least Google, Microsoft, IBM, ModernMT, and DeepL have committed to not using any of your data if you are accessing the machine translation data through their application programming interface (API) and paying for it. On the flipside, if you are using any of these services through their regular web services, your data will be used. You are using an API if you are utilizing a third-party tool (such as a translation environment tool) to access the machine translation data and have at some point entered a code to get access to the data.

Other machine translation engines—regardless of the underlying technology—are specifically trained for a specific purpose with specific language material. Results achieved with such programs have a much higher reliability when it comes to terminology, and confidentiality is likely not an issue. Still, the resulting translations, no matter how high the investment in training and customization, will not show the same level of quality as a professional translation. "Raw" output from these programs is sometimes used in product knowledgebases, low-level support material, or internal company communications, but the professional translator is needed for most other end-user-facing material.

As professional translators we can interact with machine translation output in a variety of ways. First and foremost, a fundamental decision must be made as to whether it's helpful to use machine translation at all. Just because a company has invested in building a machine translation program does not automatically mean that it's useful for producing all content. For instance, a tool that is good for technical content may be ill-suited for more creative content. If it's determined that machine translation generates enough helpful data to improve translation productivity, that data can be used in a number of ways. In a process called "post-editing," the professional translator edits the machine translation output. This might make sense when the training data is very high quality and the domain of both training data and translatable data is very well defined.

Another way machine translation data can be used is within a translation environment tool alongside terminology databases and translation memories. A very useful aspect in this scenario are the n-grams or fragments that can be accessed automatically through processes like AutoSuggest and can speed up both speed and quality of the translation.

Up through 2011, most translation environment tools came with only a connector to *Google Translate*, but Google's decision to start charging for its machine translation service through its API (if it's integrated into a third-party tool) made most tool vendors look for other solutions alongside *Google Translate*, resulting in a very large number of easily accessible possibilities.

Here are the different offerings of most translation environment tools (as of January 2020; please note that to actually use most of these MT tools you'll need a license key).

TEnT	Integrated MT Engines
Across	*Google Translate, Reverso, Moses, Omniscien (formerly Asia Online), SmartMATE, DeepL, Kantan MT, PROMT, Reverso, SYSTRAN*
CafeTran	*Google Translate, Microsoft Translator, MyMemory, Slate Desktop, DeepL, IBM Watson Language Translator, Yandex*

TEnT	Integrated MT Engines
Déjà Vu	*Google Translate, Microsoft Translator, PROMT, SYSTRAN, Omniscien, MyMemory, DeepL, PangeaMT, Baidu, Ubiqus NMT*
Fluency	*Google Translate, Microsoft Translator, SYSTRAN, MyMemory*
Heartsome	*Google Translate, Microsoft Translator*
Lilt	*MT engine developed by Lilt/University of Stanford*
MadCap Lingo	*Google Translate, Microsoft Translator, DeepL, MyMemory*
MateCat	*Google Translate, Microsoft Translator, SYSTRAN, Tilde, tauyou, Moses, Apertium, Iconic, Deeplingo, ModernMT, Yandex, SmartMATE, Deeplingo*
memoQ	*Google Translate, Microsoft Translator, SYSTRAN, LetsMT!, Omniscien, MyMemory, KantanMT, Crosslang Gateway, DeepL, PangeaMT, MyDutchPal, Globalese, Moravia MT, tauyou, Tmxmall, IPTranslator, Mirai/COTOHA Translator, Intento (see below)*
Memsource	*Google Translate, Microsoft Translator, SYSTRAN, tauyou, Apertium, Omniscien, KantanMT, PangeanMT, Safaba, MoraviaMT, NICT, PROMT, LetsMT!, SDL LanguageCloud*
Multitrans	*Google Translate, Microsoft Translator, PROMT, SYSTRAN, MyMemory, Iconic, SmartMATE, WIPO Translate*
OmegaT	*Google Translate, Microsoft Translator, Apertium, Belazar, Slate Desktop, Yandex, IBM Watson Language Translator, DeepL, Moses, MyMemory*
Smartcat	*Google Translate, Microsoft Translator, Baidu, DeepL, Yandex, Amazon Translate (the above are directly available through the user interface), GTCOM, SYSTRAN, PROMT, Tencent, Intento (available vis inquiries)*

TEnT	Integrated MT Engines
Swordfish	*Google Translate, Microsoft Translator, Yandex*
Trados Studio	*Google Translate, Microsoft Translator, SYSTRAN, SDL Language Cloud (formerly BeGlobal and LanguageWeaver), SDL Neural MT, MyMemory, Sunda Translator, LetsMT!, tauyou, Iconic, KantanMT, Transistent, MyDutchPal,VERTO Lingua Custodia, Convertus, Slate Desktop, Sogou, Youdao, Baidu, CrossLang, OpenNMT, Prompsit, Targoman, Globalese, Lingenio (all of these, with the exception of SDL Language Cloud, have to be downloaded as separate apps from the SDL AppStore – see appstore.sdl.com)*
Wordbee	*Google Translate, Microsoft Translator, Reverso, Lingua Custodia, DeepL*
Wordfast Classic	*Google Translate, Microsoft Translator, WorldLingo, MyMemory*
Wordfast Pro	*Google Translate, Microsoft Translator, Systran, DeepL, WorldLingo*
Wordfast Anywhere	*Google Translate, Microsoft Translator, Yandex, WorldLingo, DeepL, KantanMT, MyMemory, SYSTRAN*
XTM	*Google Translate, Microsoft Translator, Omniscien, KantanMT, CrossLang, Amplexor, SmartMATE*

You can find information on these different MT engines and the language combinations they cover on their respective websites, but two might be worth highlighting because of their different nature:

- *MyMemory* is a mixture of a large collection of translation memories, aligned, web-based data and MT via *Google Translate* and *Microsoft Translator*. The data is hosted and curated by the Italian translation and technology provider Translated (see mymemory.translated.net).
- *Intento* (see inten.to), a tool primarily geared toward larger translation buyers, allows you to get an automated estimate of which machine

translation engine suits your present project best and at what cost. The engines include engines by Alibaba (general and eCommerce engines), Amazon, Baidu, CloudTranslation, DeepL, Google (Basic, Advanced and AutoML), GTCom, IBM, Kakao, Naver, Microsoft (including custom models), ModernMT, Naver, PROMT, SAP, several SDL and Systran systems, Tencent, Tilde, Yandex, and Youdao.

- A tool that is not listed in the table above but can be used within any of the translation environments is *GT4T* (see gt4t.net), a little application that allows you to connect from any *Windows* application (and soon to come: *Mac* application) to the machine translation engines from Google (either statistical or neural), Microsoft, Youdao, Yandex, Baidu, and DeepL. It allows you to search one or many at a time; you can have the machine-translated text replaced; or you can display different translation options in a popup window from which you can choose one. Unlike other tools, *GT4T's* developer acts as a wholesaler between the MT vendors and the translator. The price you pay for a time-based or volume-based license includes the (estimated or actual) cost of the machine translations that have to be bought from the MT providers. Also you can automatically have anything that comes back from the MT providers overridden with your preferred terminology that you can define in *Excel*-based glossaries.

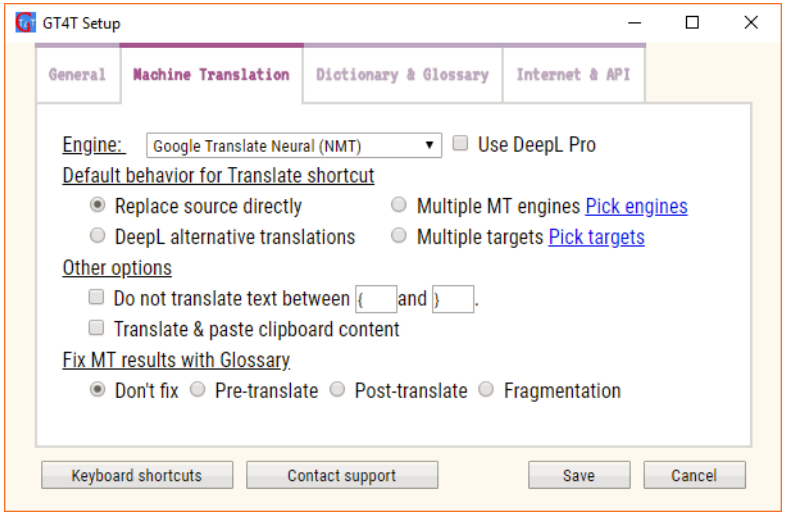


Figure 194: Machine translation options for *GT4T*

So, again, are the many integrated MT engines helpful? Are they used by professional translators? I will leave the answer to the first question up to your preferences (and language combination, and kinds of translation you do, and the many likes and dislikes that you might have about this kind of technology). The answer to the second question is yes, more and more translators are using MT as one of many resources.

Some are using it if no quality TM matches are found (the tools can typically be adjusted so that machine translation is pursued only if no match of, say, 75% or higher can be found). Others are using MT as an extended dictionary for highly specialized terms. And yet others are using it as a source for a variety of suggestions.

Consider this example from *Wordfast Classic* (with machine translations from five different machine translation engines).

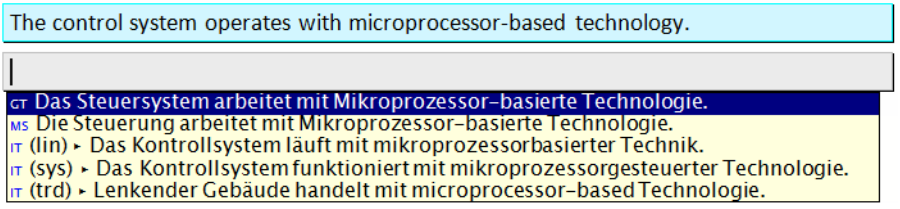


Figure 195: Wordfast Classic with machine translation matches

There is no need to argue about how "good" these matches are, but most of them contain some material that in some kind of combination might be useful in the actual and final translation. You as a translator will have to decide what kind of role this information plays for you. Does it help or hinder? Is it different, for instance, than having a lot of matches from a general TM shown? The answer to that will most certainly depend on your language combination (some language combinations are much more suited to a first machine translation draft than others) as well as your project type or subject matter. It might even be different between different projects.

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. In fact, tool vendors have recognized only rather recently that there is demand for more far-reaching quality assurance features.

Though many think of *Wordfast Classic* as a rather lightweight tool, it offered advanced quality assurance features long before most other tools did. *Wordfast Classic* offers a total of about 15 quality assurance checks, starting with the verification whether the same translated term or phrase was used in the terminology database.

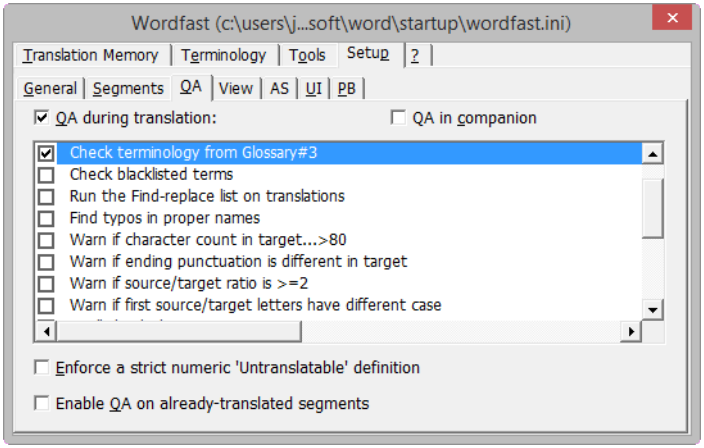


Figure 196: Setting which terminology database is to be used for *Wordfast's* terminology check



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.

The now defunct *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.

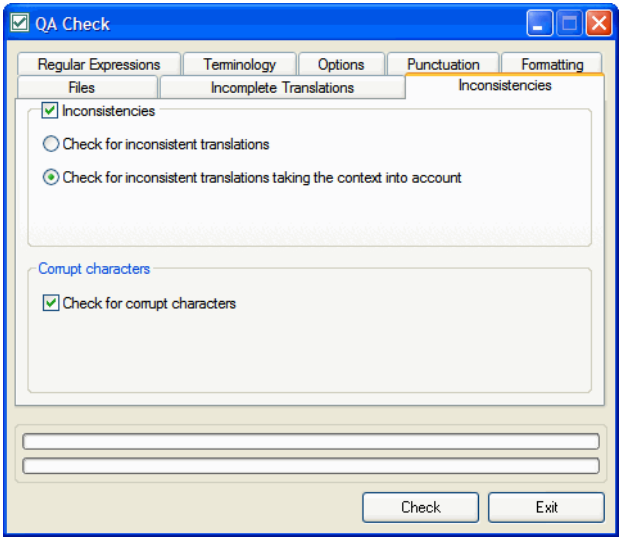


Figure 197: The *QA Check* module in *SDLX*: the **Inconsistencies** tab

However, *Trados* (whose owner also owned *SDLX*) versions *7.1* and above included a larger set of QA features than any of its competitors.

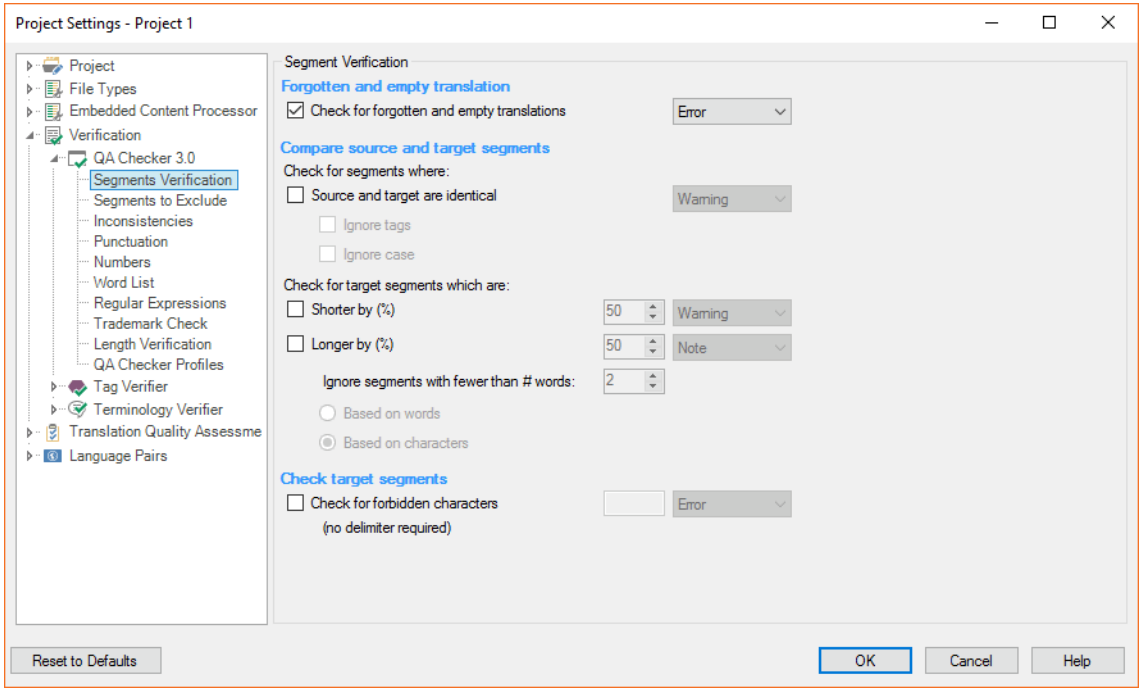


Figure 198: The *QA Checker* module in *Trados Studio*

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*:

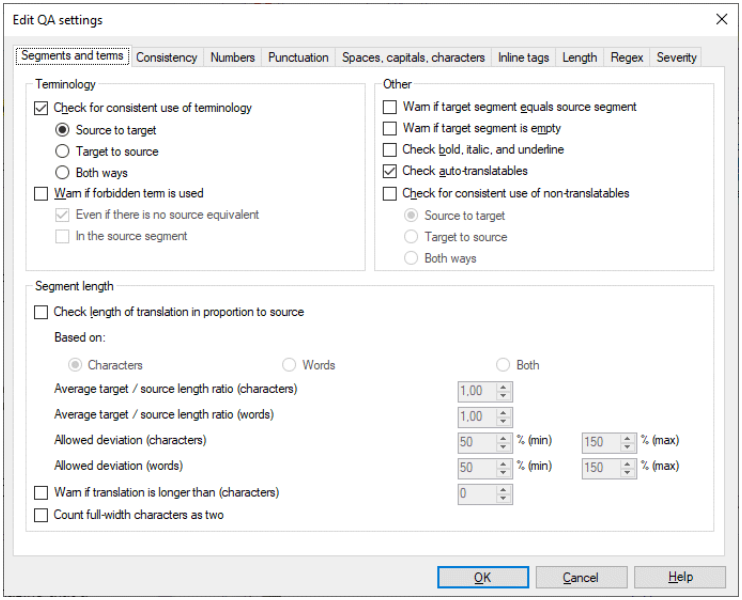


Figure 199: A sampling of *memoQ's* quality checks

For even more comprehensive standalone quality assurance tools, see page 291.

Quality Assessment

In 2011 and 2012, two different organizations started to work on creating quality assurance models that can be used to evaluate translation: TAUS (see taus.net) worked on the DQF or *Dynamic Quality Framework*, and the EU-funded QTLaunchPad (see qt21.eu/launchpad) created the MQM, or *Multidimensional Quality Metrics*. Both of these models are remarkably similar, so similar, in fact, that they have now been merged into one model (though they have retained their separate names). The similarity between the two is not a coincidence. Both organizations looked at quality assessments that were used throughout translation and localization and distilled them into one large scheme.

For both of these efforts, the main goal was to provide a framework to assess the quality of machine translation. The models are helpful in that regard, but they also provide relevant and interesting tools to assess "human" translation.

As translators, we know that quality assessment is an important task, but one that is often contentious and fraught with difficulty. Traditional methods were highly subjective and involved reviewers reading translated texts and marking "errors," but reviewers often disagreed on their assessments. In response, many translation companies invested in developing formalized metrics for assigning errors to different types (e.g., terminology, spelling, mistranslations), counting them, and determining how serious they were. In the 1990s, these efforts led to the creation of widely used specifications such as SAE J2450 (mainly used for the automobile industry) and the LISA QA Model. Unfortunately, these models have not been updated over time and are presented as "one-size-fits-all" models were do not reflect the needs of a rapidly diversifying translation industry. This is where MQM and DQF come into play. Here it’s possible to mold the quality assessment frameworks into text type-specific models in which the reviewer decides which categories are included and what severity is assigned to the various categories.

Various translation environment tools have started to integrate these QA models into their linguistic review processes. *Fluency* and *XTM* for instance support MQM, *memoQ*, *Memsource* and *MateCat* support DQF, and *Trados Studio 2015*+ supports both (you will need to choose the quality assessment as the project template for your project to enable it).

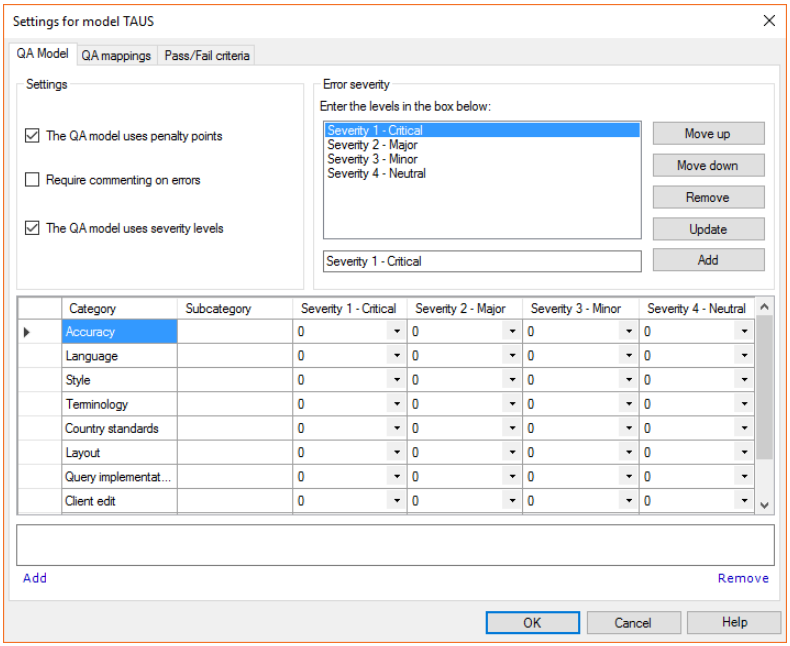


Figure 200: TAUS DQF categories after loading the TAUS template in *memoQ*

Once you specify a quality assessment framework and modify it to your specifications, the reviewer(s) can assess the quality of the translation and deliver a detailed report.

Collaboration Features

When you search for the term "collaboration" in relation to translation environment tools, you will quickly realize that there are a number of different levels of collaboration and different definitions of what this entails.

Collaboration Through Exchange Formats

The first level, which is also the only one that all tools offer in some form, is collaboration through exchange formats. There are a number of existing exchange formats, most importantly TMX for the exchange of translation memories, TBX for the exchange of termbases, and XLIFF for the exchange of translation files. (Another important exchange standard, TIPP—and in a parallel but coordinated fashion Linport—deals with the exchange of translation project packages.)

TMX, TBX and XLIFF are all standards that are based on the same underlying standard, XML, and that's not where the similarities end. All of them play a very important role in the exchange of their respective formats, and all of them have clear limitations to how seamlessly the exchanges take place.

- The two major problems that TMX has are a) the different ways in which the so-called inline tags (tags that contain non-textual information within a segment) are stored in the translation memories of the originating application and b) the different ways different tools segment texts, leading to differences in the way what is considered to be a segment in the translation memory will end up as a match.



SRX is another standard that is concerned with the specific issue of exchangeability of segmentation rules. Unlike the other standards, unfortunately, this is not very widely supported; most notably, Trados Studio does not support it.

- TBX, the standard for exchanging termbase data, has to be able to capture everything that is contained in a termbase. Unlike translation memories, termbases can be *very* complex with literally hundreds of different kinds of fields that describe the terminology data or set it into relation with each other (such as term ABC is a synonym of term XYZ). Naturally the standard to describe that complex data also has to be very complex, which made the adoption of the standard very slow and the actual process of exchanging complex termbases very manual. TBX is not as widely supported as TMX, and many tools that "support" it don't allow for all the different fields to be imported, partly because their own termbase structure does not support many of the different fields.

- XLIFF was construed as a standard to exchange the actual translation file. Since there are so many variables that come with the different file formats and the different ways tools handle translation, there are ways to "extend" XLIFF with custom extensions. This made it into a fantastic format for the internal processing purposes of many tools (*Heartsome*, *Swordfish*, *Trados Studio*, *MultiTrans* and *Memsource*, for instance, all use customized versions of XLIFF as their internal translation file format), and while these formats can be processed by any other tool that also supports XLIFF (virtually all tools support it as a format they can translate and many as a format that they can import into for exchange purposes), there are a good number of properties and attributes that are getting lost in the transfer process. While this typically does not mean that you couldn't translate an XLIFF file from one tool in another, it's highly recommended that you do a trial run to verify that everything works right (many tools offer a feature called pseudo-translate just for this purpose).

  XLIFF 2.0 is the latest incarnation of the previous XLIFF 1.2 version. It was as an official OASIS standard in 2014 and has yet to garner large support among tools. Differences between the two versions of XLIFF are that XLIFF 2.0 is simpler and less expandable by individual tools and therefore easier to exchange between tools.



Several tools, including Trados Studio, memoQ, Memsource and Déjà Vu, also offer a feature allowing the translation file to be exported into a two-column format in a Word file so that translators, editors, proofreaders, or even the client can work on the translation or view the current state of the translation outside a translation environment tool. These files can be re-imported into the tool of origin once the changes are made, and the project within the tool will reflect all changes that have been made within the Word file.

- Lastly there are the package standards. Package standards take care of the complete translation project, including translation files, resources (TM, termbase, and reference material) as well as any kind of meta information related to the project (such as instructions, etc.) and combine them in one zipped up file. The user can open the file in a supporting translation environment tool, which will utilize the individual parts by placing them appropriately on the user's computer and give automatic access to them in the translation process. Once the translation is done, the package file will be sent back to the requestor and will contain all necessary assets.

TIPP (Translation Interoperability Protocol Package) is the first of these standards. Unlike most other standards, TIPP has not been developed by a standards body but by a group of technology developers and practitioners. It has now been passed on to a committee that had independently worked on a similar standard, called Linport (Language Interoperability Portfolio). This committee is now in the process of combining the two standards and submitting it to the International Standards Organization. Once these standards are well-established and widely used, one more cornerstone to tool-independence will be achieved.



The last standard that still needs to be developed is one that allows for an exchange during server-based processes where either the translation data or the resources are placed in an online location and are continuously queried during the translation process. These processes so far are tool-dependent

Real-Time Collaboration

Considering all the problems that are encountered with the exchange formats, the real-time sharing of TM and termbase resources or even the translation file itself would be clearly advantageous.

One possibility is to share ongoing projects on a local network. The (typically) more expensive corporate version of many tools offers this feature, but aside from some corporate or organizational settings with many collaborators (translators, editors, proofreaders, project managers, etc.), this usually does not match the structure of translation workgroups, which in most other cases is virtual. For these kinds of workgroups, either a server-based solution (in which the TMs, termbases and/or translation files are located on a server within the domain of the organizer of the project) or cloud-based solution (in which the TMs, termbases and/or translation files are located on a server outside the domain of the organizer of the project) is required to work together in real-time.

There are only a small handful of translation environment tools that don't support collaboration in these kinds of workgroups.

The majority of the other tools come in a multi-tiered structure: the least expensive—in some cases, free—version is geared toward the freelance translator and supports no workgroups; the higher-priced versions can be used to organize and administer workgroups with real-time collaborations. The collaborators are typically equipped with the freelance editions of the respective tool.



Since the need for collaboration in workgroups is becoming more relevant for all kinds of user groups, companies like memoQ and SDL are now starting to offer more moderately priced workgroup versions for smaller language service providers or groups of freelancers. The memoQ product is called memoQ cloud (see memoq.com/en/memoq-cloud) and the Trados Studio product is GroupShare (see sdltrados.com/products/groupshare/).

There are only a small handful of tools with which you can share resources in real-time with a non-corporate edition. These include *Wordfast Pro/Classic/ Anywhere*, *OmegaT*, *Wordbee* and *Memsource*.

As cloud-based solutions, this feature is very apparent for the last four, but it is less apparent in *Wordfast Classic* and *OmegaT*.

In *Wordfast Classic* you can connect to your *Wordfast Anywhere* TMs and glossaries, which in turn can be shared with others in real-time.

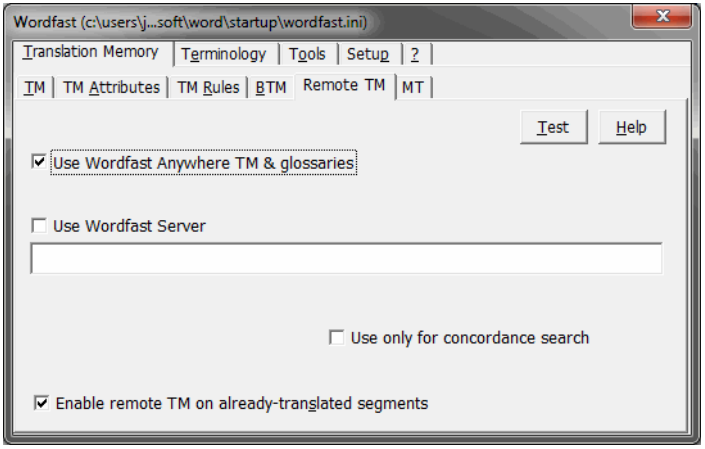


Figure 201: Connecting to *Wordfast Anywhere* resources within *Wordfast Classic*

Real-time sharing of resources is also available in the free and open-source *OmegaT*. You can find a description of the slightly more involved setup procedure at velior.ru/2013/07/11/using-team-projects-in-omegat-as-a-translator/.

Support of File Formats

With the exception of the tools that can work only through the *MS Word* interface and can therefore only process *Word*-compatible files, supported file formats are less and less a differentiator as most of the translation environment tools now support a very similar batch of formats. These include *MS Office* formats (plus often the *Apache OpenOffice*/*LibreOffice* formats), HTML, XML, *InDesign*, *FrameMaker* and a number of software development formats.

At this point, a good number of tools also support PDF files through an internal conversion process (see page 373), and you will have to look closely at the version of your desktop publishing tool to see whether it's supported by your translation environment tool.

File formats that are supported by very few tools include

- *MS Publisher* (*Fluency*, see page 221, and Text United, see page 222)
- Apple *iWorks* files (*CafeTran Espresso*, see page 228)
- Binary software files (*Across*, *see page 212, and Star Transit*, see page 317)
- *AutoCAD* DXF files (*CafeTran Espresso*, see page 228, and *Star Transit*, see page 207)
- XML files with embedded HTML (*memoQ*, *Déjà Vu*, *Trados Studio*, *Memsource,* see page 353)
- Graphic files (*Smartcat*, see page 240)
- CHM help systems without the need to decompile (*Alchemy Publisher*, see page 153)

And How Do I Make a Decision?

So now that you know that a translation environment tool would be a helpful investment to make, how would you choose one of the many that are available?

I would advise you not to actually look so much at the tools themselves but instead see what your particular environment is like.

These should be your first criteria: Who are you clients (or, if you're just starting out, who do you hope your clients will be), what tools are they using, and how do they use them?

- If they use a TEnT and send you preprocessed bilingual files in *Word* or as an XLIFF file, you can work with the majority of tools, no matter whether they match your client's tool or not.
- If they send you the projects in a TEnT-specific package format (a file that contains all the resources and the translation file you need for the completion of the project), it's possible that you can use other tools than the client is using, but you'll need to investigate a little more to know for sure that that kind of exchange works.
- However, if your clients are using a process where the translation memory and terminology data (and possibly the translation file itself as well) is located online, you will have to use the tool that the clients are using (see the note on page 287).

Next, look at what colleagues you are often work with are using. It will serve you well to use the same tool—both for the sake of seamless cooperation as well as some friendly support. And speaking of support, make sure that there is an overall good support system in place (see page 435).

Also, you should inform yourself about training opportunities. For any translation environment tool, especially if you have never used one before, you should consider investing some kind of training.

And lastly, look at the tool itself. Start with looking at the file formats you are translating. Does the tool you are looking at support them all? If it does not support all the formats, would it be OK to not use the tool for some projects? If you are not a Windows user, you will have to make sure that your tool runs on the platform of your choice.

One thing should really not become a major part of the decision-making process: how much the tool costs. Instead, look at the return-on-investment. Any tool that you invest in and can't make good on the purchase price within a few months is a failed investment, no matter the original price. Plus, the initial purchasing cost most likely is the smaller portion of your investment. Training will be the larger.

Standalone Quality Assurance Tools

Tools that offer approximately the same range of features are

- *QA Distiller* from Yamagata Europe (see qa-distiller.com)
- *Error Spy* from D.O.G. (see dog-gmbh.de)
- *Okapi CheckMate* (see okapiframework.org/wiki/index.php?title=CheckMate)
- *Verifika* from the Siberian translation provider Palex (see e-verifika.com)
- AceProof from the Ukrainian translation tool specialist AIT (see aceproof.com)
- *Xbench* (see xbench.net).

All of these tools work on *Trados*, XLIFF and TMX files, as well as some other formats (*Verifika*, for instance, also works on *memoQ* files, *ErrorSpy* on *Transit* files, *AceProof* on Microsoft Helium files and *Xbench* pretty much on every bilingual format—see below). The predefined criteria that are checked

include terminology (not *AceProof*), consistency (*CheckMate* does not check this important item), numbers, completeness, tags, acronyms, typography, and missing translations, and all of these tools also allow the creation of regular expression-based customized quality assurance rules.

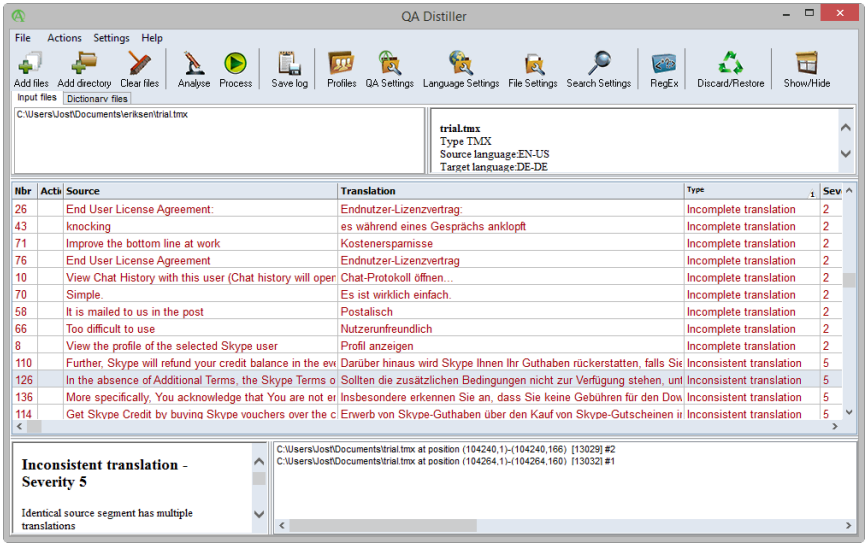


Figure 202: *Distiller* displaying errors

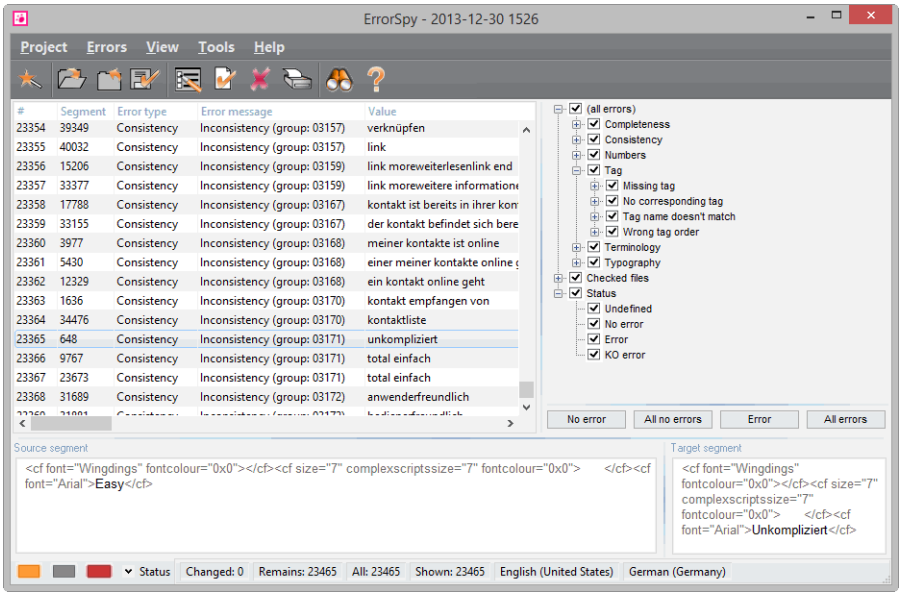


Figure 203: *ErrorSpy* displaying errors

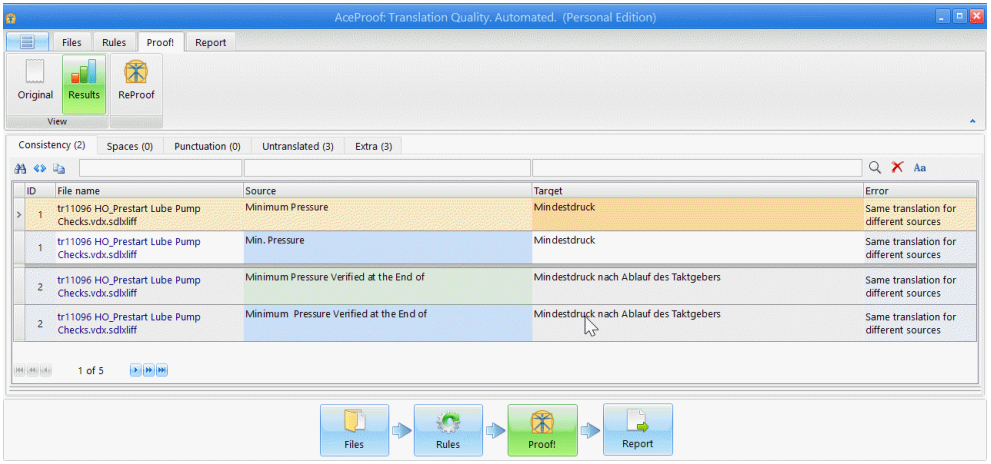


Figure 204: *AceProof* displaying errors

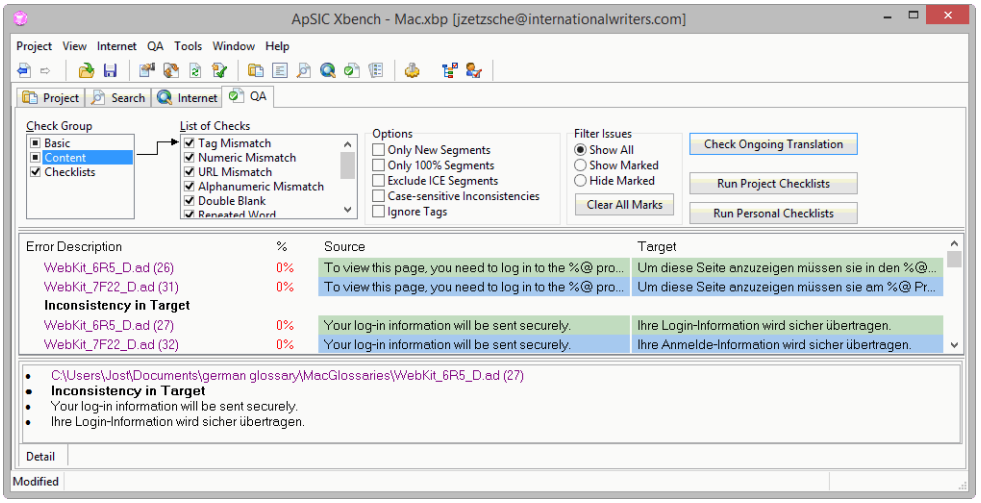


Figure 205: QA features in *Xbench*

As mentioned above, *Xbench* distinguishes itself by the large range of bilingual file formats it supports and has become the preferred tool for many in the field of QA tools.

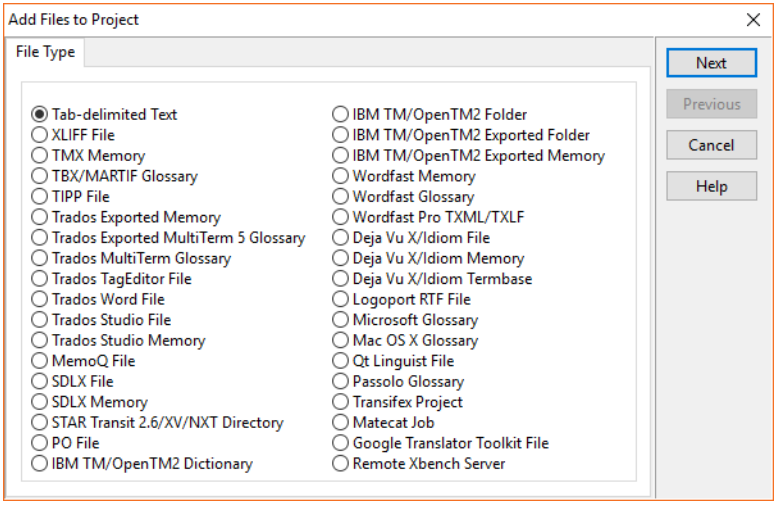


Figure 206: Supported file formats in *Xbench*

Xbench is also directly integrated into a number of translation environment tools through various means. Presently (January 2020) you can find integrations for *Trados Studio*, *memoQ*, *Memsource, Poedit, Passolo* as well as *Transifex*, *Google Translator Toolkit*, *Smartcat*, *XTM* and *MateCat* (the last five through a *Google Chrome* add-in).

A tool that was developed specifically for the application of the MQM quality evaluation framework (see page 282) is the open-source tool *translate5* (see translate5.net). It presently only supports the *Trados Studio*-specific format SDLXLIFF as well as generic CSV files, but since it's open-source it can be developed into supporting a large range of bilingual file formats.

Here is what the interface of *translate5* looks like:

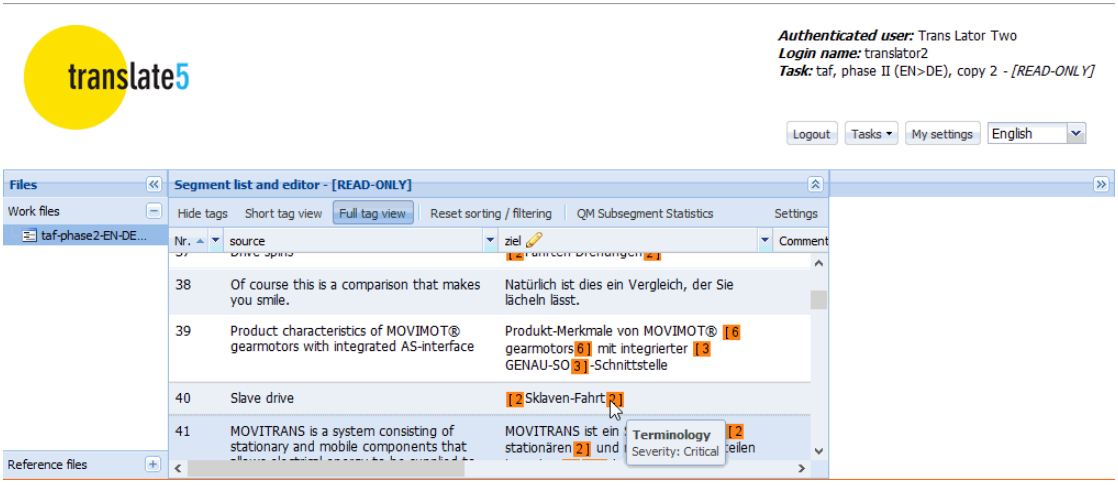


Figure 207: *translate5's* interface with an example of a critical terminology error

Since the interface is highly customizable, it might look very different when you use it. The point is that the reviewer just needs to highlight any part of the segment in question and select the necessary quality category from a hierarchy-based picklist. Once that is done, you'll immediately be able to see tags added to the source or target text around the problematic subsegment. And when you're done with your review, you can filter the text according to the quality metrics and export or view reports and final assessments on them.

This approach is really helpful because it's a completely browser-based, and therefore highly accessible, process (even for the reviewer on the client's side who does not have access to a translation environment tool). And yet, any changes that you or any proofreader or editor makes in *translate5's* interface can be re-imported into *Trados Studio* as tracked changes, and comments that were entered in *translate5* can be transferred to *Trados Studio* as well.

Automatically Tracking Quality Through Editing

TQAuditor (see tqauditor.com) started off as a smallish tool that was designed to compare translated bilingual files, including XLIFF, TMX, Excel, and TTX (much like the useful ApSIC Comparator—see apsic.com/en/products_comparator.html), so project managers could evaluate how much editing was performed to a translated file.

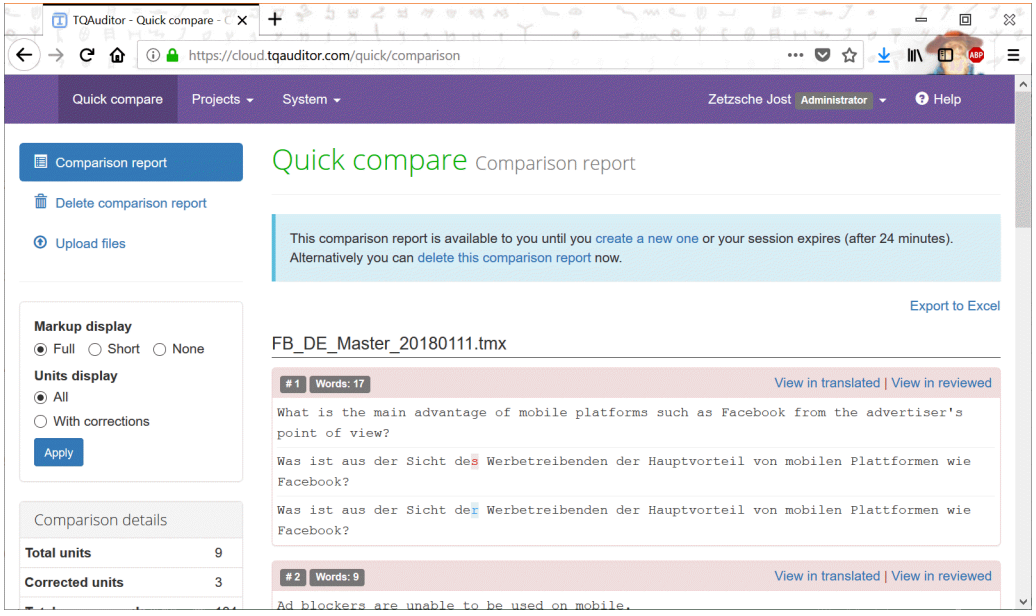


Figure 208: Comparing two versions of one TMX files in *TQAuditor*

But rather than just leaving it at the comparison of individual file pairs, *TQAuditor* aims to build a complete quality tracking procedure within companies that use it. It does that by assigning ratings to translated files that have been edited. Instead of having the editors with their quality assessment

in complete control, however, translators are strongly encouraged to respond to any edit by accepting or contesting it or by asking for a different severity status. If there is no agreement between translator and editor after a certain number of communications, an "arbiter" is automatically called into the process who then makes the final call.

Tis is all customizable—starting from what kind of edits and how many of them still make a translation acceptable or good to how many back-and-forths can happen before the arbiter is called in (or whether there will be an arbiter in the first place) to the exact definitions of error categories. All the generated quality data is stored and gives both the translator (and editor) and the translation company an excellent way to track the performance of individual translators as well as company-wide quality development or how the quality in particular industries or subject areas develops.

The reports that the system automatically generates are highly detailed, easy to read, and analyze the data in every conceivable manner. Of course, similar data could be pulled from a cloud-based translation system like *Wordbee*, *XTM*, or *Memsource*, but the reality for many translation agencies is that it's next to impossible to process everything in one kind of translation environment only. By supporting the many translation environment-specific XLIFF formats, this system makes cross-technology quality assessments possible. The same answer applies to you who may be pointing to integrated track changes systems in tools like *Trados Studio* or *memoQ*. If you only and always use tools with integrated features like that, you might be fine without a system like *QATracker* and still have a comprehensive system to internally assess your translation (and translators) by looking at the level of edits they require; otherwise you might benefit from a tool like this.

Source Document Quality Assurance

Quality assurance tools that don’t rely on translation memory technology include *Acrolinx IQ* (formerly *Acrocheck*) (see acrolinx.com) and *FormatCheckers for Word and FrameMaker* from Star (see star-group.net/en/products/data-quality.html).

Acrolinx IQ is a corporate tool to check terminology, style and grammar of documents in German, English and French. That doesn't sound too fancy, but I was really impressed at the program's level of "intelligence." For instance, you can load customized style sheets or modify the preloaded 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.

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 aiming at a better return on translation memory matches and/or better entry of data into the translation memory.

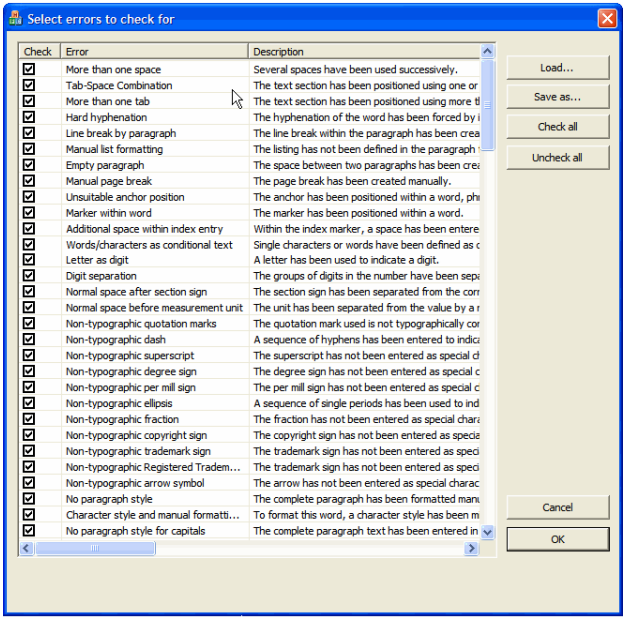


Figure 209: Error checking in Star's *FormatChecker*

A similar but more sophisticated tool might be *PerfectIt* (see intelligentediting.com), an add-on to *MS Word*. The wealth of checks you can do with this tool is staggering. The main categories include hyphenation/dashes, spelling consistency (note "consistency"—it's not a spell-checker), abbreviations, formatting, capitalization, bullets/lists, tables/boxes/figures, and writing style. Under spelling consistency alone, these are the individual checks that are performed: similar words, preferred spelling, spelling variations, accents, numbers in sentences, common typos, phrases to avoid/consider, and contractions. All of these are based on nine different integrated style guides, plus a style guide that you are encouraged to build yourself ("My Style"). At this point, all of the style guides are based on various forms of English, but they are very open to building predefined style guides for other languages (or you can build one yourself).

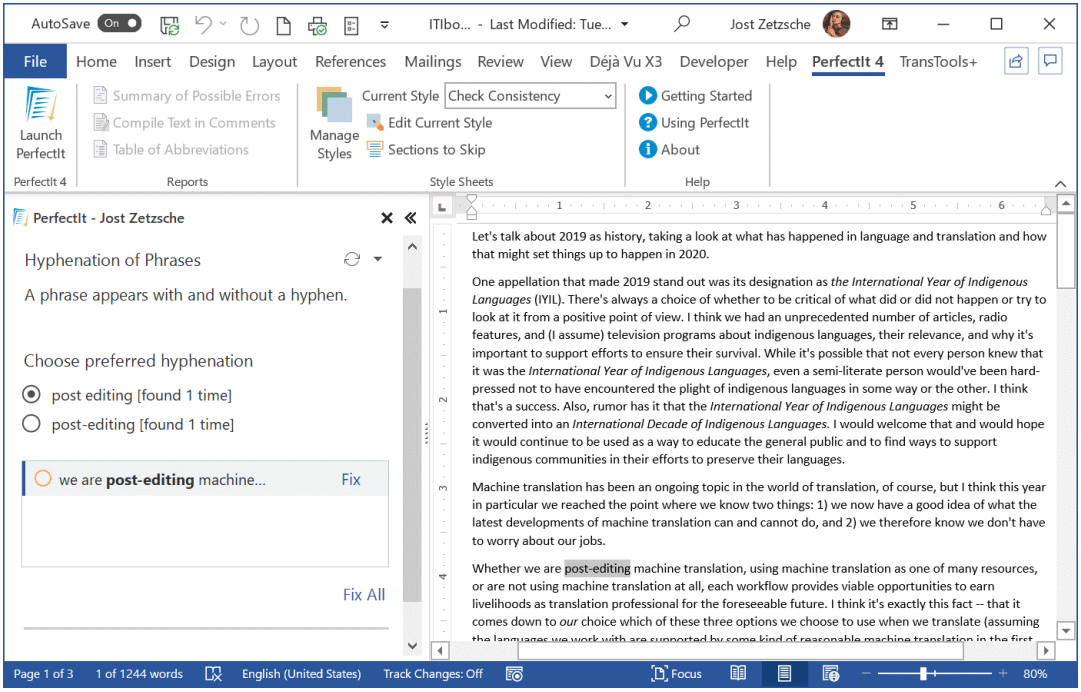


Figure 210: *PerfectIt* within *MS Word*

Another tool, *Lingofy* (see lingofy.com), can be installed not only in *MS Word* but also a number of browsers (*Chrome*, *IE*, *Firefox*) as well as *MS PowerPoint* and *Outlook*.

The underlying style guide of *Lingofy* is the AP Stylebook, and you can add your own entries to that. Unfortunately, there is a limitation of only 3,000 words per proofing pass, and there are very few possibilities to adjust the settings.



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 congree.com/en/product/congree-authoring-server) and Star MindReader (see star-group.net/en/products/authoring-assistance.html).

Though these do not strictly fall into the category of quality assurance, this is a particularly exciting 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 number 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.

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. While some TEnTs offer decent features to do that, many do not. *Olifant* (see okapiframework.org/wiki/index.php?title=Olifant) allows you to open (and save) TMX, *Trados 2007* and *Wordfast* translation memories and filter and manipulate them to your heart's content, all in a clean, multi-column interface.

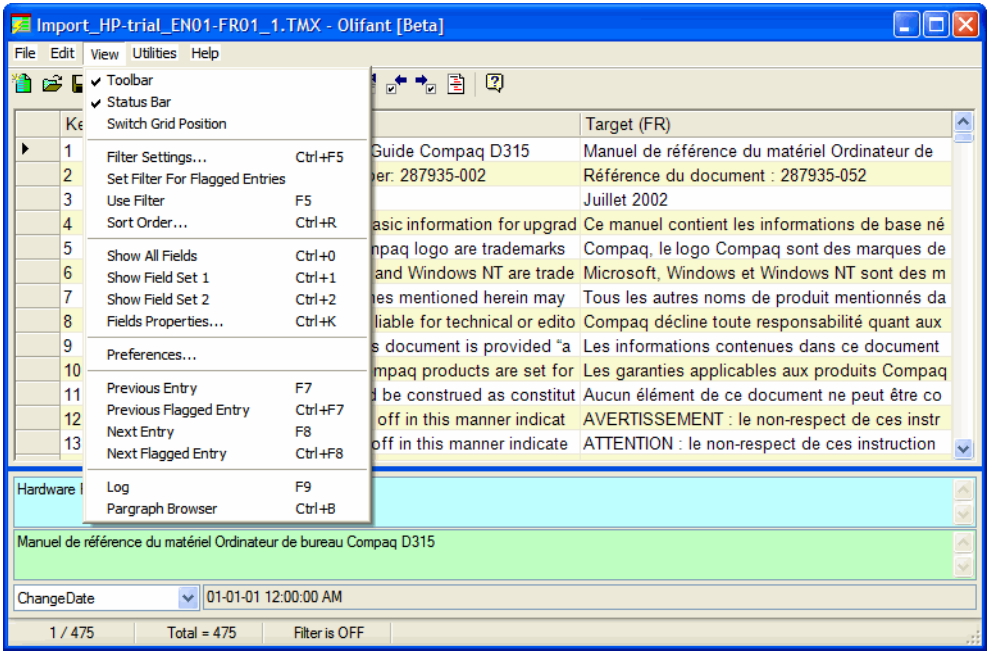


Figure 211: Open TMX file in *Olifant* with the available commands in the **View** menu

Thes—now defunct—TEnT developer Heartsome also developed the *TMX Editor*, an editor for the translation memory exchange format. It's Java-based so it runs on all platforms, and it's surprisingly fast. Just like the translation environment tool (see page 216), it has also been released for free with an open-source license. You can download it at github.com/heartsome/tmxeditor8.

As you can see in the following screenshot, it also offers a quality assurance filter that you apply to your TMX file. Once problematic translation units are found, you can either batch change or delete them or process them individually. Other features include the modification and/or adding of metadata (data about the translation unit), changing of the code page, merging or splitting TMX files, or exporting TMX files into a great number of other formats, including a number of text formats and *Word* or *Excel* formats.

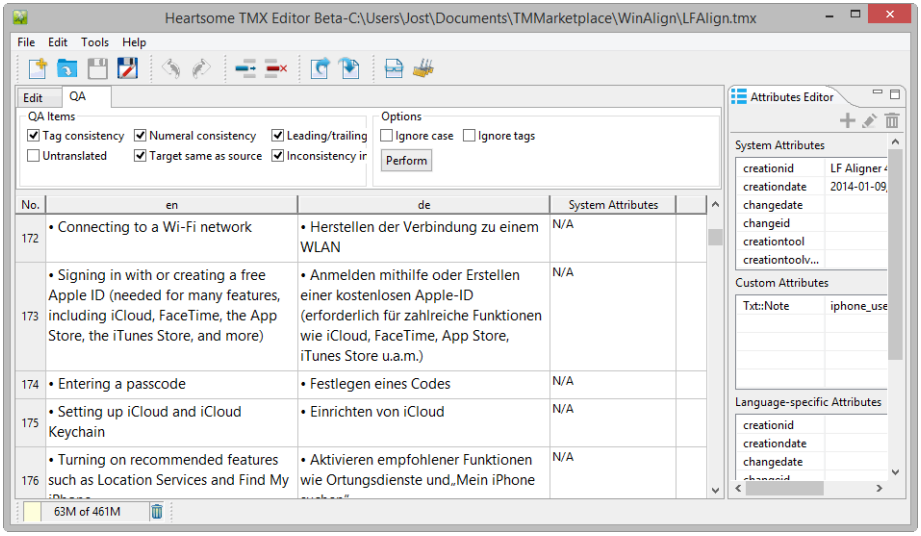


Figure 212: Heartsome's *TMX Editor*



A similar but more updated version is offered by Maxprograms and can be downloaded at maxprograms.com/products/tmxeditor.html.

And lastly, *Apsic Xbench* (see page 309) also allows for the conversion of translation memory formats or other database exchange formats.

Terminology Mining

Terminology mining programs offer the possibility of extracting 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.

A number of TEnTs, including *Across*, *CafeTran*, *Déjà Vu*, *Fluency*, *memoQ*, *Star Transit*, *Text United* and *XTM,* offer monolingual term extraction as a regular part of the ongoing translation process or as a separate feature to build up glossaries or terminology databases.

There are standalone tools for this process as well. SDL's *MultiTerm Extract* (see sdltrados.com/products/multiterm-extract) 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.

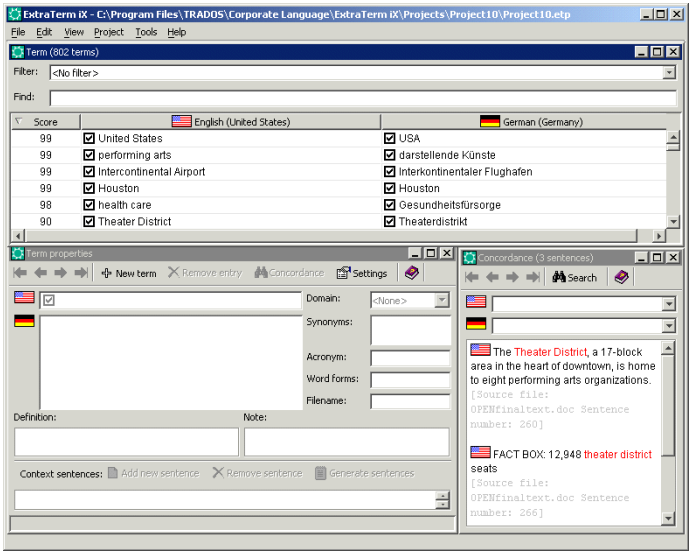


Figure 213: *Trados' MultiTerm Extract* (formerly *ExtraTerm*)

Other tools use a combination of mathematics and linguistic data.

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, which later was acquired by Expert System, effectively halting any development for translation-related purposes.

A terminology extraction that is still available tool is *SynchroTerm* (see terminotix.com/index.asp?content=item&item=7). *SynchroTerm* aligns texts in their various formats to bi-texts, 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.

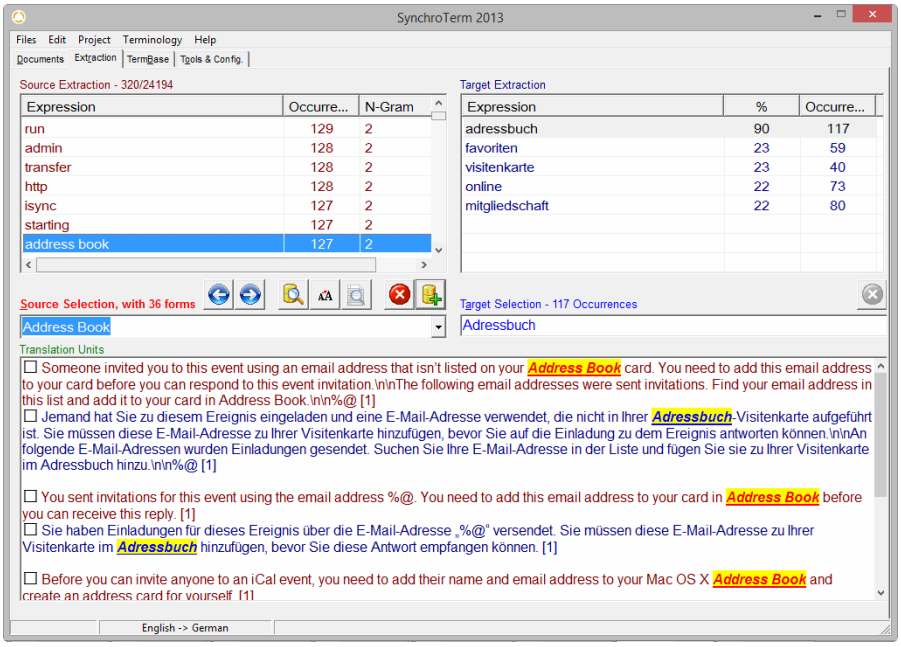


Figure 214: Extracted terminology in *SynchroTerm* from a TMX file

Tilde Terminology (see term.tilde.com) is rooted in an EU-funded project that at various levels supports 24 European languages for monolingual term extraction purposes.

Once you've registered, you can upload one or several files in various formats (PDF, DOC(X), XLS(X), PPTX, RTF, TXT, XLIFF, XML or HTML), have terminology extracted from the file(s), apply content within existing terminology resources to those terms, select from the suggested translations and/or translate the terms, and then export it so you can use it within your terminology database or glossary.

This tool is particularly interesting because of the tools that support the extraction process. These include tools for part-of-speech tagging, lemmatizers, morpho-syntactic patterns, statistical analysis and—for English and Latvian—a tool to normalize terms, which brings terms into their canonical forms (typically nominative singular or infinitive).

When that is complete, the extracted list of terms will be run against a number of (again, optional) resources in the following order: 1. your own personal resources that you might have collected on the site; 2. other users' terminology; 3. the EuroTermBank; 4. the EU's inter-institutional terminology database IATE; 5. the TAUS corpus; and 6. a statistical database that consists of aligned web data. After these databases have been queried for translations, they will be shown as suggestions from which you can choose by just clicking on them and/or you can enter your own translation.

Of course, one of the ideas behind this project is to make it possible to share terminology data. At the outset of each project you can enter a whole lot of optional data, but you will need to make a decision on the language combination, the domain of your text, and whether you want to share the data with other users. The shared data will not include the complete texts that you upload but only the term pairs that you will end up with in your termbases (and only on an individual term pair level rather than complete lists of term pairs). The shared data will also be used for other purposes, including machine translation.



Tilde Terminology is also integrated in a number of translation environment tools, including memoQ and OmegaT.

Standalone Terminology Tools

I have tried to make a strong case for using the integrated terminology components of a translation environment tool (see *Terminology Handling* on page 262). 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* 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* 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—see page 285).

Lingo (see 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 translation environment tool 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.

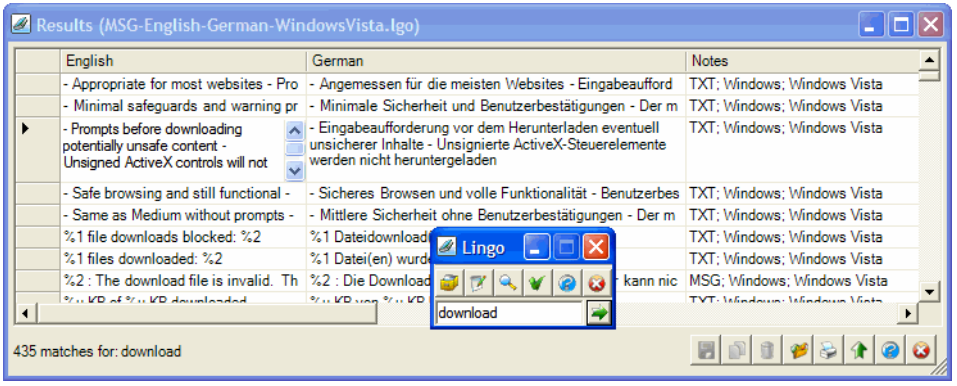


Figure 215: Search results in *Lingo*

A tool with a very similar set of features is *AnyLexic* (see 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.

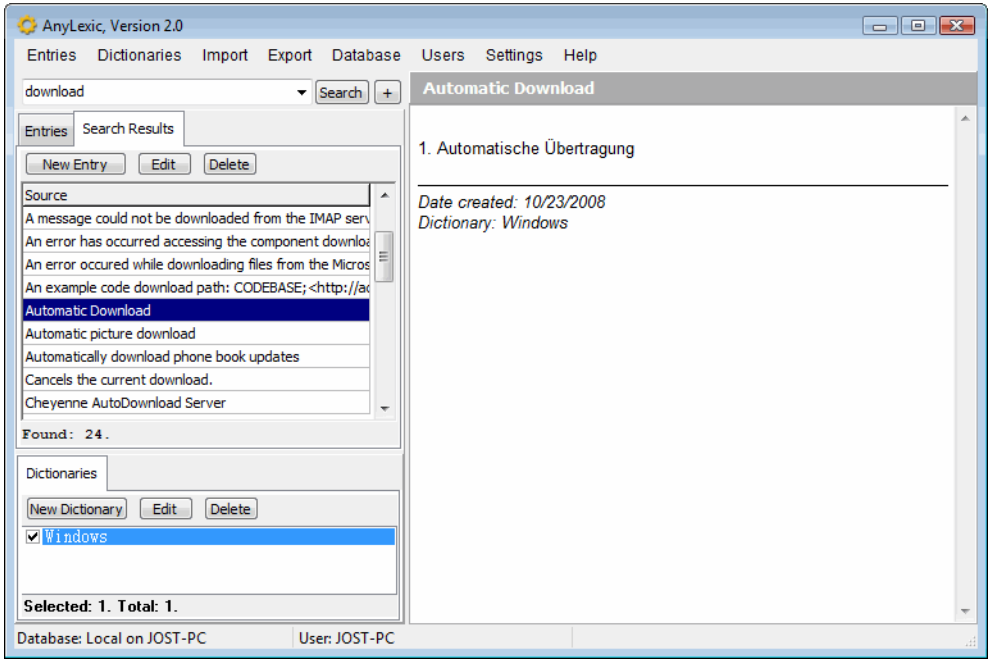


Figure 216: Search results in *AnyLexic*

A slightly different tool is *ApSIC Xbench* (see xbench.net), which was a free tool up through the still available version 2.9 (starting with version 3, you have to pay). *Xbench* performs quality assurance checks on a large number of TEnT formats (see page 309), but it also imports a huge number of bilingual file formats (see page 294).

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.

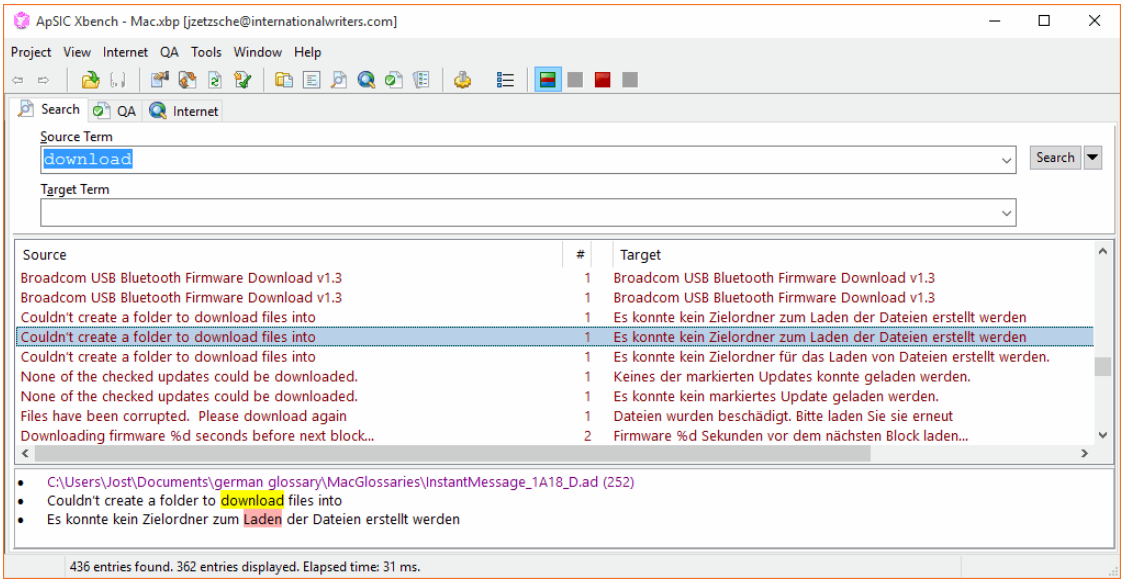


Figure 217: Search results in *ApSIC Xbench*

Lastly there is Terminotix's *LogiTerm* (see terminotix.com/index.asp?content=brand&brand=2), a Canadian tool that does a variety of things (monolingual text searches or terminology database maintenance) but is especially powerful when it comes to aligning texts (it uses the same

alignment module as *AlignFactory*—see page 253) and using these to do terminology searches. The supported file formats for the alignment include *Word*, *WordPerfect*, *Excel*, *PowerPoint*, *InDesign*, HTML, XML and PDF files, and you can store the aligned results as bilingual HTML or XML files.

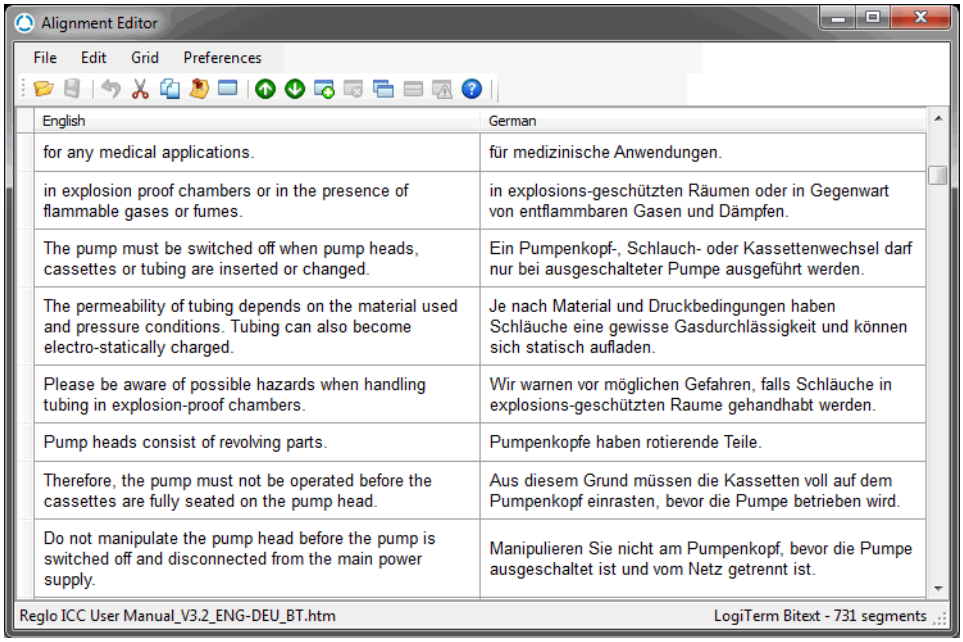


Figure 218: Bilingual HTML text created by *LogiTerm*

Once you have created your bilingual files, you can search in a virtually unlimited number of aligned file pairs for any term or phrase and use a great number of operators in your searches (such as wildcards or for fuzzy matches). You can do this from within the *LogiTerm* interface . . .

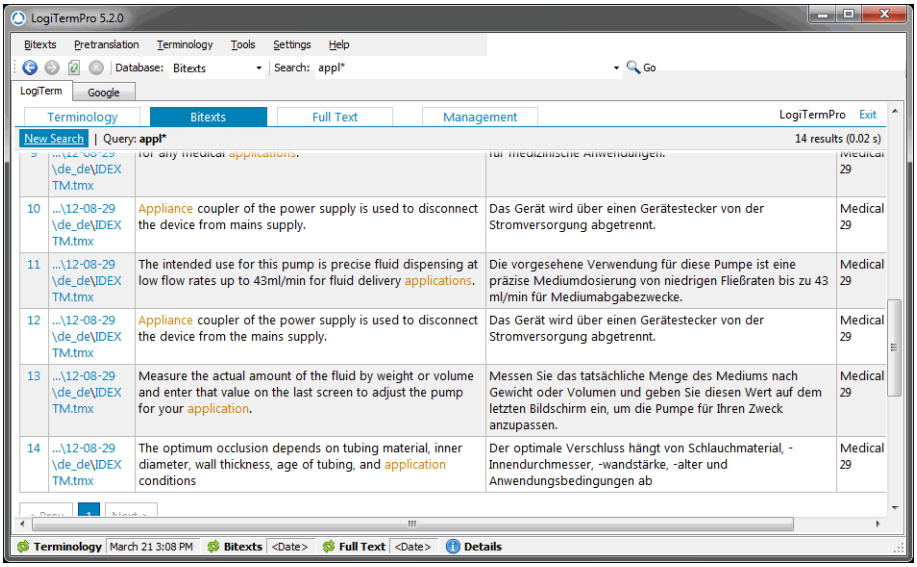


Figure 219: *LogiTerm's* bi-text search feature

. . . or you can do it from within *Word* with a separate *LogiTerm* menu/ribbon.

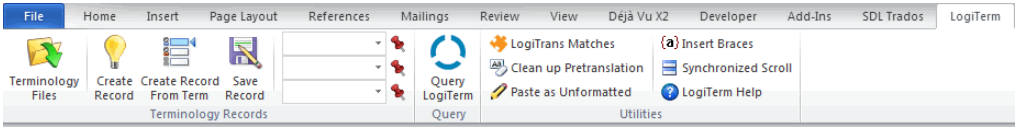


Figure 220: *LogiTerm's MS Word* ribbon bar



You can also use LogiTerm to pretranslate text in a number of MS Office-compatible formats but this is not what it was primarily built for and you'd be more efficient using one of the more traditional translation environment tools for that.

Software Localization Tools

Software translation, or "localization," is such a wide field because there are so many formats and purposes. Originally, software localization referred only to the translation and adaptation of binary files (files that are not text-based and can only be interpreted correctly in specific programs) for foreign language versions of computer programs. However, much of today's localization is happening in text-based files, and the concept and application of localization has gone far beyond large computer programs (anything from small apps to websites). Let's start with the original kind of localization.

Binary Software Localization

In the early years of localization, 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 text change 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 221: 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 also included other development formats, including text-based ones as well as the translation exchange format XLIFF (see page 286).

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 alchemysoftware.com): Supports Win32, .NET, Java, XML, HTML, XLIFF, HTML help, and JSON files, and numerous database types and offers an interface to *Trados* applications.

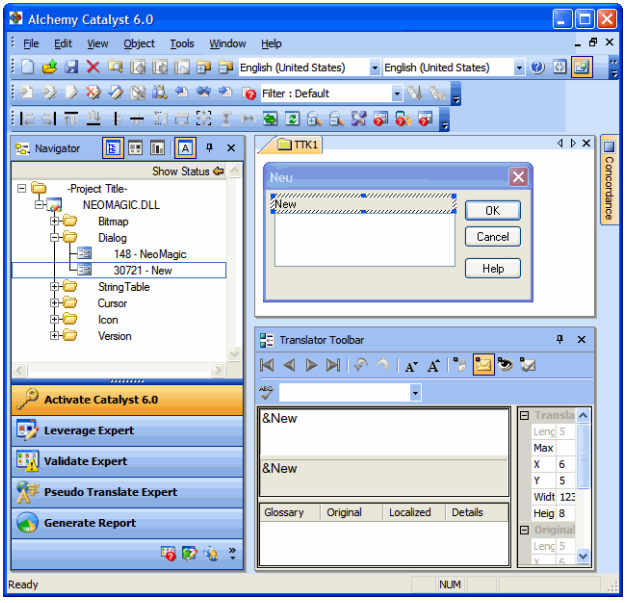


Figure 222: *Catalyst's* translation environment

- *SDL Passolo* (see sdl.com/solution/language/software-localization/passolo): 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, PO, JSON, database and mobile app formats and offers an interface to

Trados and *Star Transit* memories and terminology databases. It also supports all current versions of .NET.

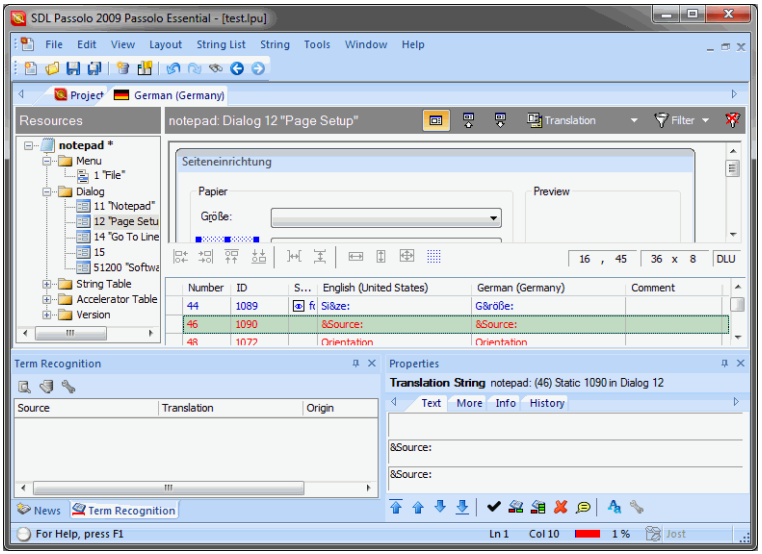


Figure 223: *Passolo's* translation environment



In 2008, Passolo was purchased by SDL and Catalyst by TransPerfect, putting an effective end to the age of independent localization tools.

Users of SDL Trados Studio benefit from the common ownership of both Trados and Passolo by being able to directly work on Passolo files (from Passolo 2018 on).

- *RC-WinTrans* (see 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 multilizer.com): Finnish tool that originally was designed for Delphi. Today it also supports Win32, .NET, Java, XML, mobile app formats and database contents.
- *Sisulizer* (see sisulizer.com) supports a range of formats similar to *Multilizer*.

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 (free) translator edition that excludes some of the more development-oriented functionality, and a developer or localizer edition that contains all the functionality.

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, including the ability to resize dialog boxes etc.

Localization of Text-Based Files

One kind of text-based localization development format that often needs to be translated (especially for computer programs that were developed a number of years ago) are RC files, the result of a decompiled EXE or DLL file (see page 314). It would be possible to translate RC files in a text editor (and that's true for any of the file types in this section), 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 translation environment tools—especially those of the first generation—support the translation of RC files, including *Déjà Vu*, *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. Most translation environment tools (and of course localization tools) support this format.

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 224: 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 a wide range of translation environment tools as well as most localization tools (see *Software Localization Tools* on page 313).



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 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 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 225: View of a PO file in a simple text editor

Aside from *Poedit* (see poedit.net), a free and simple translation environment tool specifically for PO/POT files, *Déjà Vu, Across, Heartsome, Swordfish*, *OmegaT* and *Open Language Tools* are translation environment tools that also directly handle these files. *Trados Studio* can handle them with the help of a free app in the SDL *AppStore* app store (see appstore.sdl.com/language/app/file-type-definition-for-po-extended/868/).



It's also possible to translate PO files by converting them into another translation format such as XLIFF with the open-source software Rainbow (see okapiframework.org/wiki/index.php?title=Rainbow).

When translating apps, you will have to handle different text-based files depending on what platform you are developing for (if you are developing for several platforms, it would be very advisable to use translation memory to not duplicate translation efforts):

- For *iOS*, you'll need to either translate two STRINGS file or—if you're using the Apple's development platform Xcode—one XLIFF file with the text of the STRINGS files.
- For *Android*, you'll need to translate the "strings.xml" file that is located in the `res/values/` directory.
- For *Windows*, the translation will either be done in RESW files (a similar format to RESX) or JSON files.

Most translation environment tools—and all localization tools—support these formats.



You can find a very helpful set of six articles on the different aspects of app localization by the translation company PTI Global at bridgeurl.com/applocalization.

Specifically for the localization of *Windows*-related app and software files, you can also use the free *Multilingual Toolkit* (see developer.microsoft.com/en-us/windows/develop/multilingual-app-toolkit) from Microsoft. Aside from the Microsoft-specific formats, the *Multilingual Toolkit* also supports XLIFF files and therefore has a much broader potential application.

Collaborative Localization Platforms

With the advent of crowdsourcing for some localization projects on the one hand and agile (ongoing) development on the other hand, a new brand of localization tools was developed. These browser-based tools allow a large number of users to all work simultaneously on translation projects, provide the users with the basic features of a translation environment tool and support all or most of the relevant formats.

The most successful among those is *Crowdin* (see crowdin.com). While *Crowdin* was originally a true localization tool that supported only software and app formats, it now supports a wide range of documents formats, such as XLIFF, HTML, text, *Apache OpenOffice*/*LibreOffice*, *MS Office* (X) files, *InDesign*, *FrameMaker*, and SBV and SRT subtitle files.

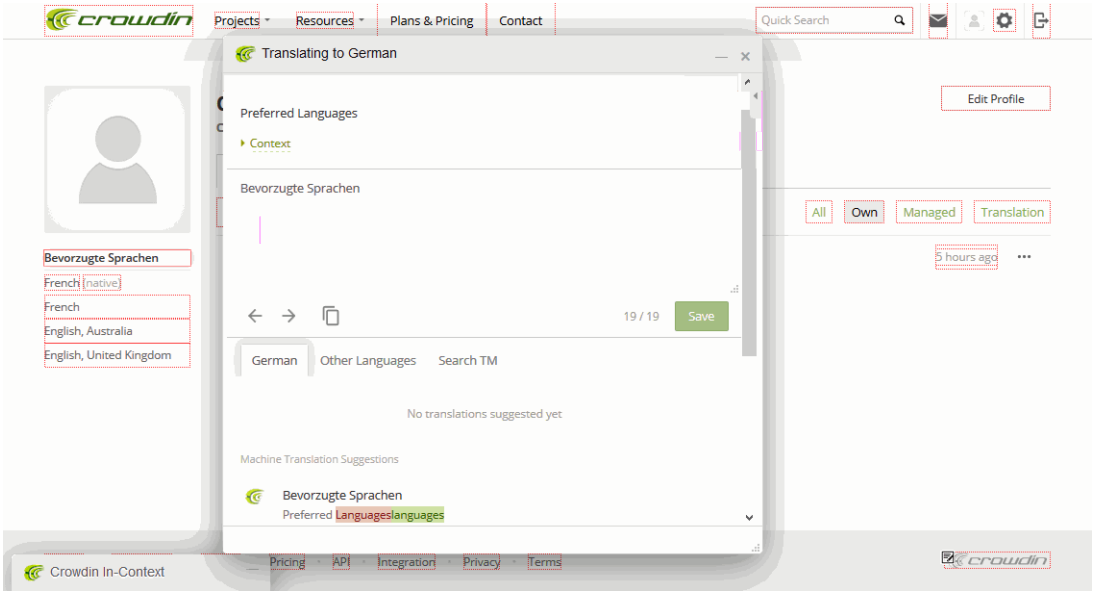


Figure 226: *Crowdin's* interface with WYSIWYG translation

Transifex (see transifex.com) is a similar and equally wide-ranging product, whereas *Lingohub* (see lingohub.com) is primarily geared toward app localization.

Website Localization

Website localization is such a complex topic because there are so many different kinds of websites.

In the early days of the Internet, most websites were static sites. They consisted of HTML pages that contain the very content that's shown in a browser. (Many of today's sites that don't need to be continuously updated are still static.)

In these cases, the fairly straight forward process of translation consists of translating the individual HTML pages in a translation environment tool. Virtually all TEnTs support HTML, either in the version 4 or 5 variety. One important thing to pay attention to when translating HTML pages is the directory structure of the originating site (which likely will have to be rebuilt in the translated language) and the accompanying intra-site linking as well as the code page(s) of the translated pages. Ideally the originating pages were already in Unicode, which means that you don't have to worry about changing the code page. If that's not the case, however, you'll need to check whether your translation environment tool does it for you (many do) and if so, correctly; if not, you will have to do this manually for every page or use a batch converter (see, for instance, page 109).



You can find other helpful tools for dealing with HTML pages under HTML Editors on page 112.

Any kind of website that needs to be updated dynamically according to user choices, or where the owners and creators of the site prefer not to have to enter content into HTML pages, will be a dynamic website where the HTML(-like) pages call up content dynamically from a content management system (CMS).

Some content management systems come with solutions that are easily available and can be used to either directly translate or to generate XLIFF files that can be translated in translation environment tools. For instance, this is the case with the *WordPress Multilingual Plugin* or *WPML* (see wpml.org) for the popular content management and blogging system *WordPress*. In most other cases it is much more difficult to get to the data and have a system in place that automatically monitors the CMS for changes and alerts the translator for newly translatable content.



In the case of WordPress, memoQ also offers out-of-the-box possibilities to translate multilingual WordPress sites..

Wordbee's developers have developed on a solution to access CMS content from many different tools in the form of another middleware tool called *Beebox* (see wordbee.com/translation-connectors). *Beebox* can connect to a growing list of CMS's including *WordPress*, *Drupal*, *Kentico*, *SiteCore*, *Adobe AEM*, *EpiServer*, *Censhare*, *What's On*, *Typo 3*, *SDL LiveContent* and *Sharepoint*. The content can be translated and processed within *Wordbee* as well as any other translation environment tool.

Other translation management systems, including *Lingotek* and *Smartling*, have also invested heavily in connections with CMS's but in these cases the architecture is such that the translation will happen within those tools.



The "memoQ content connector" deals with CMS's by being able to watch for changes that happen within specified folders and importing potential changes automatically into the translation environment and alerting the translator to new tasks.

With this said, it's a pain to deal with translatable content in CMS's and really something that is not easily realizable for the individual translator.

It's no surprise then that in the last few years a completely different process of website localization called "proxy-based website translation" has become increasingly popular. Using this process, website translations are being produced without actually getting into the source of any of the translatable materials. Instead, the user browses in a cloud-based version of the site without actually realizing it. The cloud-based site sends queries to the original, untranslated website every time a user accesses it, which in response serves pages that go through the cloud-based layer where they are translated on the fly and appear in a different language.

In a way, it's not so unlike what is being done to a webpage when it's translated by *Google Translate* or *Microsoft Translator*, only what we are talking about here is not machine translation, and the results are controllable beyond the mere translation. This means that the layout as well as the text being used in the localized website is customizable, and you are free to choose what kind of URL you want to use (provided they're available, of course).

A number of language service providers, including Lionbridge, TransPerfect, Sajan and a number of other companies are offering this as a service (not surprisingly since they are hosting the proxy pages and therefore essentially guarantee the customer's return), and larger technology companies such as *Smartling* and *MotionPoint* are offering it as a product (and offer services as well).

The only tool with this technology available for individual translators at this point offers is *Easyling* (see easyling.com). It offers a solution much like its big competitors, but it really focuses on website translation without intermingling it with services, and it's priced to be affordable for smaller companies and individuals.

Using *Easyling*, you can either translate the website within its own workbench (either in a WYSIWYG view or in a tabular view), use a hybrid model with a preconfigured connection into the translation environment tool *XTM Cloud* (with which the makers of *Easyling* have a partnership) or export everything as an XLIFF file, translate within your preferred translation environment tool and then import it back into *Easyling*.

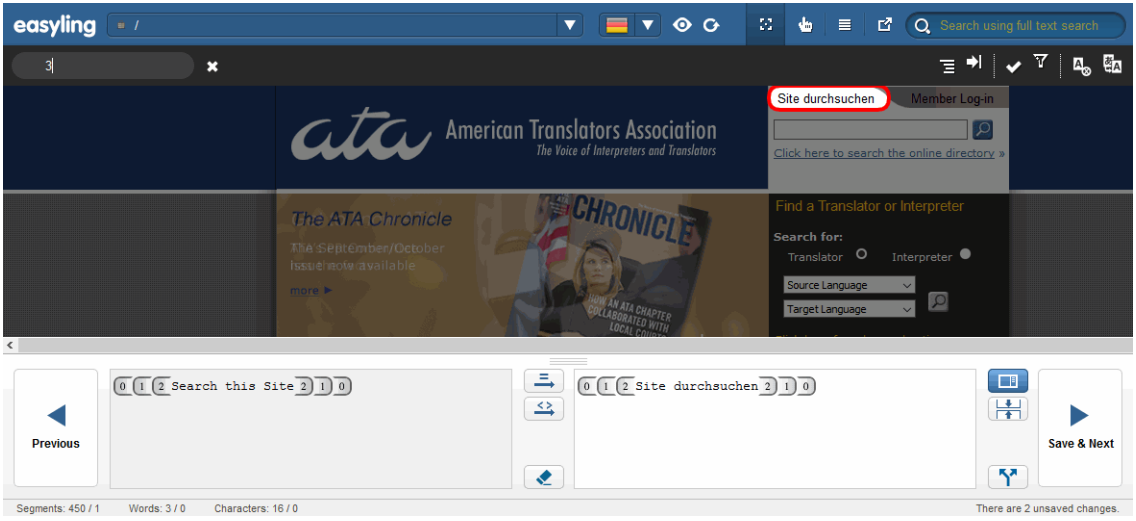


Figure 227: *Easyling's* WYSIWYG workbench

A tool like *Easyling* is an option that might be a good possibility for some customers of yours, especially those who want a very hands-off approach, but others will not like the idea that they won't host their own localized website. Still it's something that might be a good thing to mention as one possibility if the content of a website is otherwise too tedious to get to.

Another product that deserves some attention is *iLangL Cloud* (see ilangl.com). The product provides a relatively seamless extraction of data from content management systems, transforms them into easily translatable files (XML files rather than XLIFF files), and then, once translated, writes the data back into the CMS in a parallel structure to the source language.

In contrast to other than other systems. *iLangL* is not installed within the CMS but sits on a server and you gain access to the different CMS's in whatever way they provide for. The CMS's that are already supported include *Adobe AEM*, *DNN*, *Contentstack*, *Contentful*, *EpiServer*, *SiteCore*, *WordPress*, *Umbraco*, and *Drupal*. For each of these solutions there is a user interface that allows you to decide which content should be translated and in what manner. While the user interfaces are relatively easy to navigate, the setup is system-specific and not for the faint of heart, so you might be well-advised to use iLangL's consulting services to help you with the setup. (The same consulting services are also available if you need a connector to a CMS that might not be listed above.)

Presently, iLangL offers existing connectors to the translation environment tools, *Memsource*, *memoQ*, and *MultiTrans*.

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

- Applications such as *PractiCount & Invoice* for writing invoices and/or reports (see page 135)
- Spreadsheet applications such as *Excel* or applications like *ExactSpent* to track hours and pricing or to use with a good template for invoices (see page 138)

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 projects very effectively 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. Several years 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 ltcinnovates.com/us/products/ltc-worx.html)—*Worx*, now owned by the British Alpha CRC, is the successor of *LTC Organiser*, the first commercial project management program for the language industry.
- *]project-open[* (see project-open.com)
- *Projetex* (see projetex.com)
- *Plunet* (see plunet.com)
- *T.O.M.* (see jovo-soft.de)
- *QuaHill* (see quahill.com)
- Protemos (see protemos.com)
- *XTRF* (see 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*, *Plunet*, *XTRF* and *Protemos* are completely hosted online and are for medium-size to large organizations, and *Projetex* and *T.O.M.* are really more geared toward smaller companies.

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.

And then, of course, there are also tools made for freelance translators. The tool that is probably the leading contender in this group is the little sister of *Projetex*, *Translation Office 3000* (see to3000.com), a no-nonsense database-based solution with little footprint that can significantly minimize your accounting time as a freelancer.

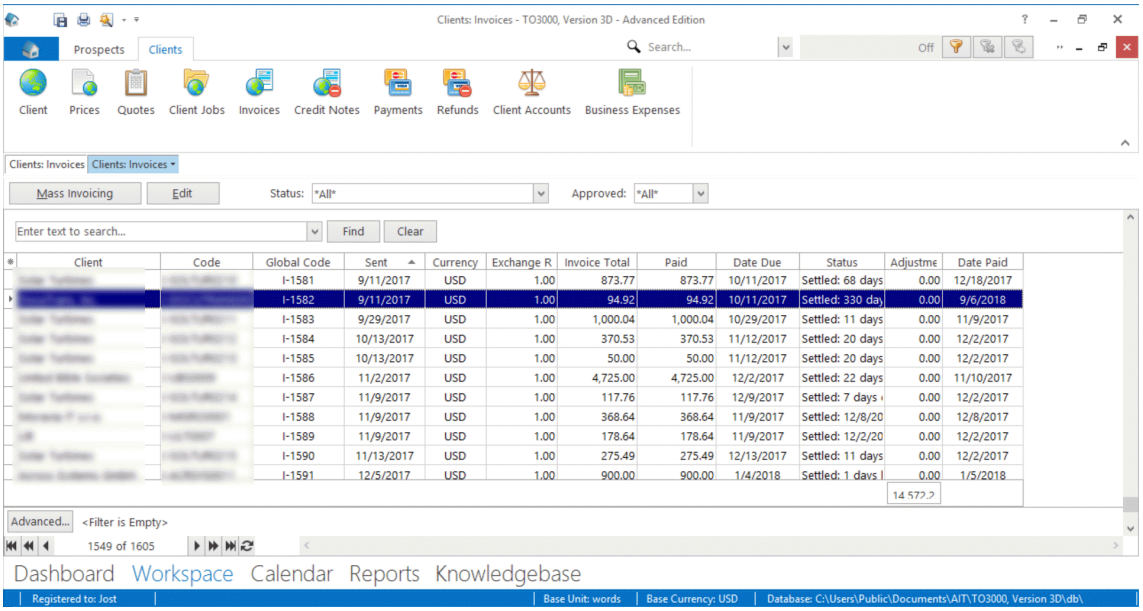


Figure 228: The invoice window in *Translation Office 3000* with easy access to all other modules

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 to see whether the same is true for you.



Other project management tools that are geared toward freelancers include freelancer-specific editions of Quahill and Protemos (see above) and various other tools.

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 unless you exclusively work for clients that prepare everything for you in their preferred translation environment tool, it's difficult to get by with just using word processing application(s).

While I have dealt with office formats in earlier sections (see *Office Suites* on page 157), in this section I have attempted to categorize some of 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.)

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.



In the recent wave of SaaS (Software as a Service) offerings from all kinds of software companies, Adobe, the maker of the most commonly-used desktop publishing programs, is offering various programs to rent the programs on a monthly basis. See, for instance, the Creative Cloud offering (adobe.com/creativecloud/plans.html).

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 quark.com/en/Products/QuarkXPress),
- *Adobe InDesign* (see adobe.com/products/indesign) and (now very infrequently)
- *Adobe PageMaker* (not available anymore).



The tool that has taken PageMaker's place as a lightweight desktop publishing tool is probably Microsoft Publisher which is supported by a couple of translation environment tools (see page 289).

The second group is home to applications such as

- *Adobe FrameMaker* (see adobe.com/products/framemaker) and also with decreasing frequency
- the now discontinued *Corel 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 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 158).

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.

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 free FM2MIF tool (see dtptools.com/product.asp?id=fmfm) 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 most translation environment tools 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.

The only translation environment tool that allows for a direct processing of FM files is *Alchemy Publisher*. 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* editions) is the only TEnT that supports the *Ventura* format—but don't worry, there are very few translation projects in that format and it's safe to say there won't be many in the future since *Ventura* has officially been retired.

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.

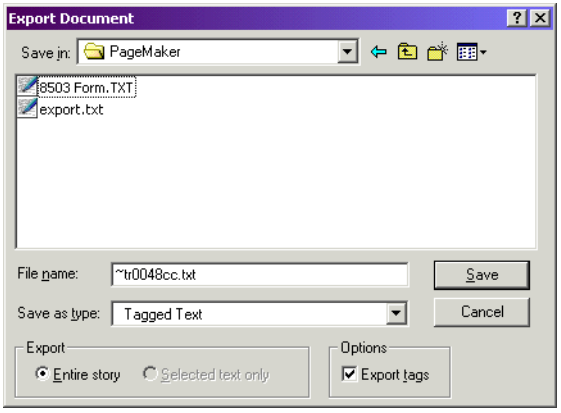


Figure 229: Exporting a story from *PageMaker*

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 napsys.com/products/cfg-for-quarkxpress for *Quark* up to version 9 and *ex TranslationFilter* by CoDesCo at codesco.com/en/ex-translationfilter.html for all versions of *Quark*).

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 *Quark* or *InDesign*, but it is one important way to distinguish yourself from your competition. 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.

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 CC 2020 (15). To translate efficiently in *InDesign* you will need a program that exports all the stories (the above-mentioned text boxes) into one large file that can be processed in a translation environment tool. (Of course, it is possible to translate directly within *InDesign*, but the emphasis was on "efficient.")



Transit XV and NXT also support InDesign CS files through a specially purchased add-in. Third-party vendors, including North Atlantic Publishing (see napsys.com/products/cfg-for-indesign) and Polmann Services (see polmannshop.com/online/en/19-sysfilter-for-indesign), also offers programs to process early (and later) versions of InDesign.

With the release of *InDesign CS2*, the accessibility of *InDesign* files became 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 *CC 2020* your tool officially supports as 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 translation environment tool vendor to update quickly to the latest format of *InDesign*.

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 translation environment tool), the latest version of most translation environment tools now supports the IDML format directly, and many tools, including *Trados Studio*, now support only the IDML format for *InDesign.*

Processing *InDesign* Files with MemoQs's *Language Terminal*

MemoQs's *Language Terminal* (see languageterminal.com) has changed the way translators can work with *InDesign* files. One of the various features of *Language Terminal* is the ability to upload *InDesign* INDD files of any version to a server, which converts these files to a *memoQ*-specific version of XLIFF. This version of XLIFF, with the extension MQXLZ, can be directly processed in *memoQ* or, with a little workaround, in any other tool that supports XLIFF.

For this workaround, you will need to realize that the MQXLZ format is a zipped (compressed) format that contains an XLIFF file (with the extension MQXLF) and a "skeleton" file (which contains all the external data, such as images). To retrieve the XLIFF file, change the extension of the MQXLZ to ZIP, right-click on the file and select **Open with> Windows** (**File**) **Explorer**.



Don't use a compression utility because that might cause problems in the back conversion to InDesign.

Once you see the MQXLIFF file, copy it to an external location and rename it to XLF or XLIFF. Now you can process it in any other tool. Once you're finished with the translation, replace the extension of the XLIFF file with MQXLIFF, open the ZIP file again with *Windows/File Explorer,* and replace the old MQXLIFF file with the newly translated one. Once that is done, close the ZIP file, rename its extension to MQXLZ, and upload it to the *Language Terminal* again to have it converted back to an *InDesign* INDD file. Once the *Terminal* is done with the conversion, you can download a ZIP file that contains the INDD file alongside a PDF with a preview of the translated file.



Make sure that you first run a test with a pseudo-translated file (a file where the characters are replaced with "dummy" characters for testing purposes).

PageMaker

To translate *PageMaker* files (an increasingly rare occurrence because Adobe has long given up active development for *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**.

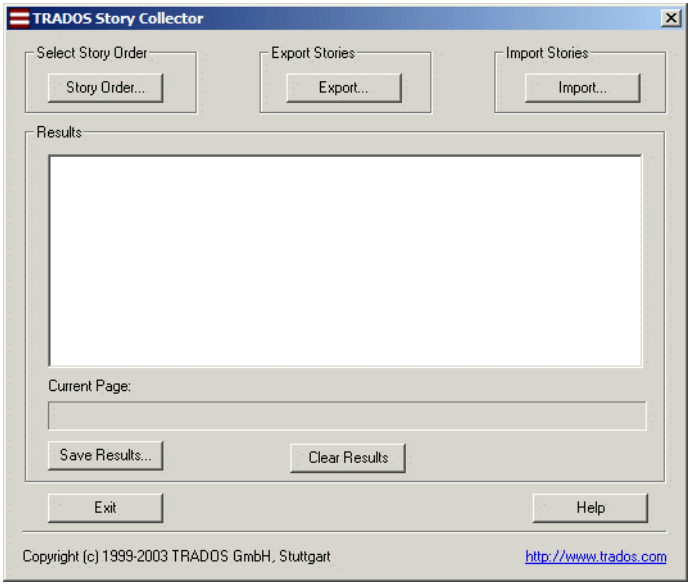


Figure 230: *Trados Story Collector* for *PageMaker*

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.

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.

QuarkXPress

Despite the fact that *Quark* has never been very popular in the translation community (because of a lack of Unicode support until fairly 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 3-9.5 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.

All of these plug-ins were preceded by a program called *CopyFlow* (see napsys.com). Just like those programs, *CopyFlow* allows for the batch export and import of text from *Quark* files (up through version 9). Later versions of *Quark* are only supported by *ex TranslationFilter* by CoDesCo (see codesco.com/en/ex-translationfilter.html).

If you need to translate *Quark* 10 and above files, *CopyFlow* and *ex TranslationFilter* are presently the only tools that allow for an export into a TEnT-processable format (both on *Mac* and *Windows*).

Because of Unicode support in *Quark*'s version 7 and above, 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 Middle Eastern edition was for version 6.5. Fortunately, however, there are *XTensions*—*QuarkXPress*-specific plug-ins—for the English version of *Quark* that extend its ability to write in bi-directional languages. *ArabicXT* is available at layoutltd.com through versions 2015 (11) of *Quark*.

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.

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.

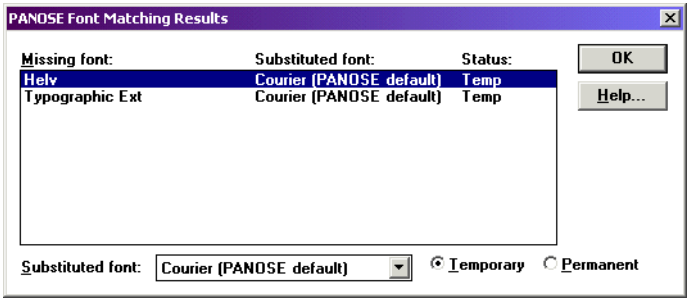


Figure 231: Error message indicating missing fonts in *PageMaker*

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 of *Word 2000* or higher (see page 166).
- 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.



Like its desktop publishing programs, Adobe also offers its graphic application on a month-to-month rental basis, which might be a good option for some projects (see page 331).

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) PaintShop Pro* (see paintshoppro.com), a low-priced power tool that I mostly use to make minor modifications to pixel-based graphics (see page 342). Since Corel bought

PaintShop, the tool has morphed into more of a photo-editing tool, so the best current alternative may be *GIMP* (see 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 adobe.com/products/photoshop-elements).

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 PaintShop 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.

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.

Essentially there are two (smart) ways of dealing with those files. You can either recreate them with optical character recognition (OCR) or you can try to use the editable source files (the preferred method).

Working with Source Files

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*.

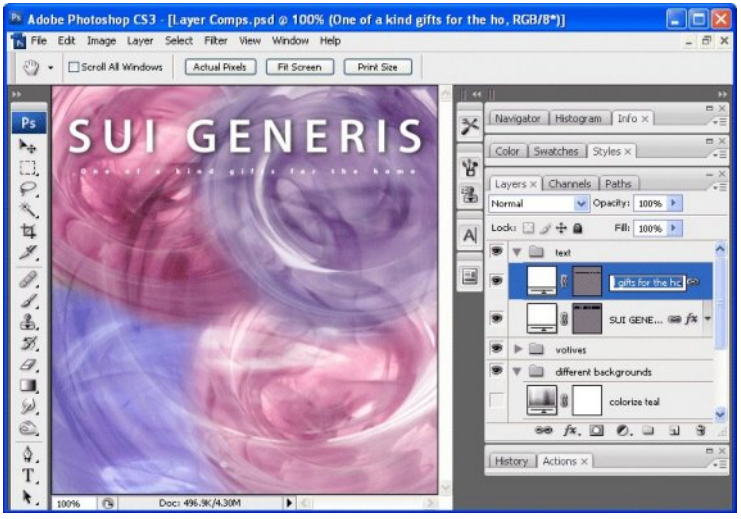


Figure 232: Image file opened in *Photoshop* with active text layer

The nice thing is that Adobe offers a low-priced version of its program (see adobe.com/products/photoshop-elements) that is more than adequate for translating the text layers that need to be translated. Or you can also use *GIMP* (see 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."

At the present time (January 2020) there is only a handful of translation environment tool that directly supports the translation of PSD files, including *memoQ*, *Wordbee*, *Memsource*, and *XTM Cloud*.

If you don't have access to *memoQ* you can use the *Sysfilters* tools provided by Polmann Services (see polmannshop.com) that allow for the extraction of text from PSD files into RTF or XML formats. These formats can be processed in any translation environment tool and afterward re-inserted.

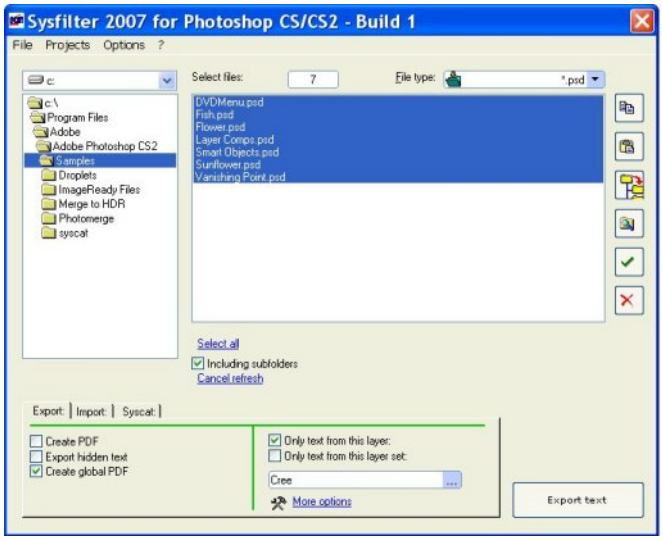


Figure 233: Processing *Photoshop* files in *Sysfilter*

Working with OCR'ed Graphic Files

Since the translation environment tool *Smartcat* (see page 240) has access to the high-quality optical character recognition engines of OCR-provider ABBYY, it's able to process graphic files and deliver decent results.

Look at the example of the following JPG file:



Figure 234: Original JPG file

Once imported into *Smartcat*, the text is displayed in the following manner:

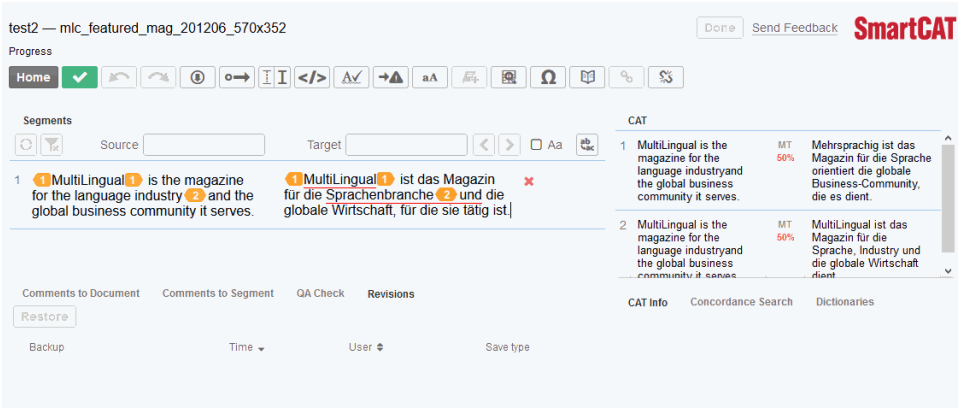


Figure 235: JPG file with OCR'ed and extracted text in *Smartcat*

Once the translation is finished, the default export format is DOCX. *Word* displays the file this way:

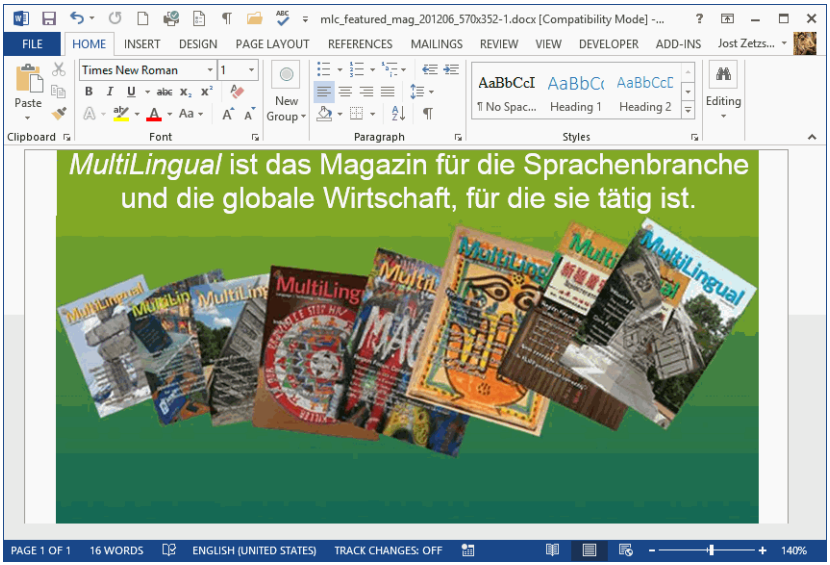


Figure 236: Automatically recreated image in *MS Word*

You can see that there are some problems with graduated color, but overall the image looks surprisingly good.



Note that MateCat and Wordfast Anywhere also have internal access to the same OCR program.

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 adobe.com/products/illustrator) or *Corel Draw* (see www.coreldraw.com/product/coreldraw).

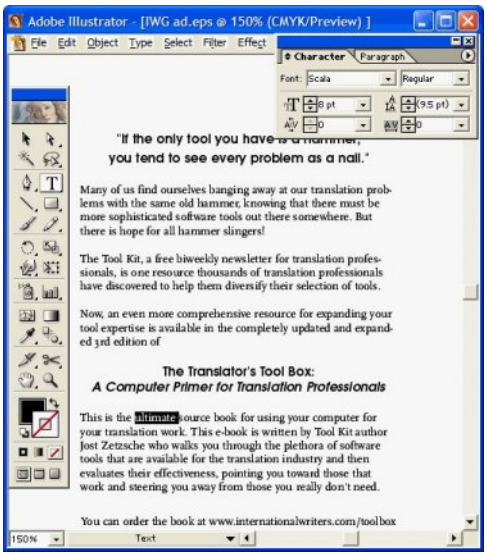


Figure 237: View of an EPS graphic file in *Adobe Illustrator*

If you would like either to batch process the files or to use your translation memory, there are two different options.

The above-mentioned Polmann Services offers products for *Illustrator* and *Draw* files that preprocess 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*.

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, but more typically tagged formats include HTML, XML or SGML files (see the definition on page 158).



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, 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 given 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.

Trados Studio contains a number of preconfigured XML filters as well as a generic XML filter.

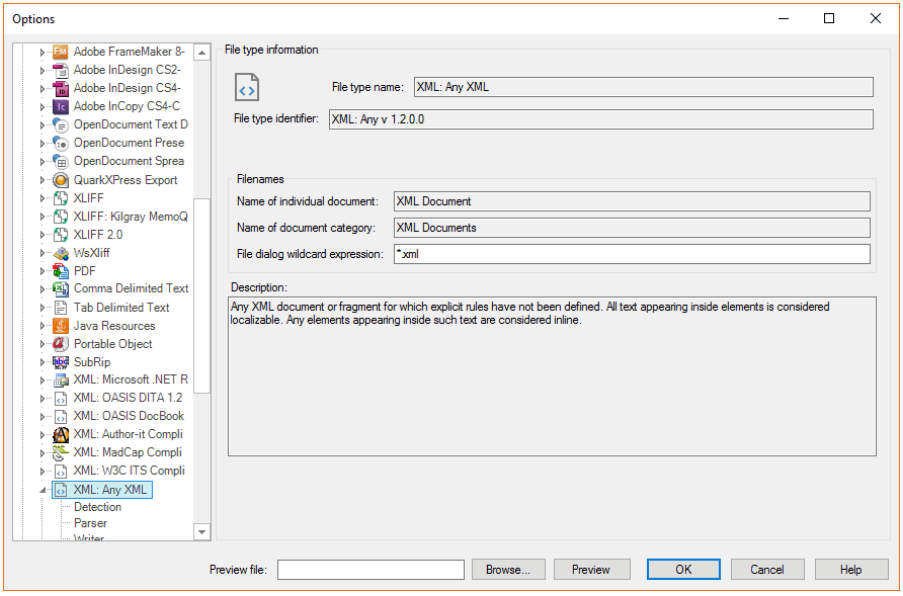


Figure 238: Preconfigured XML options in *Trados Studio 2017*

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**.

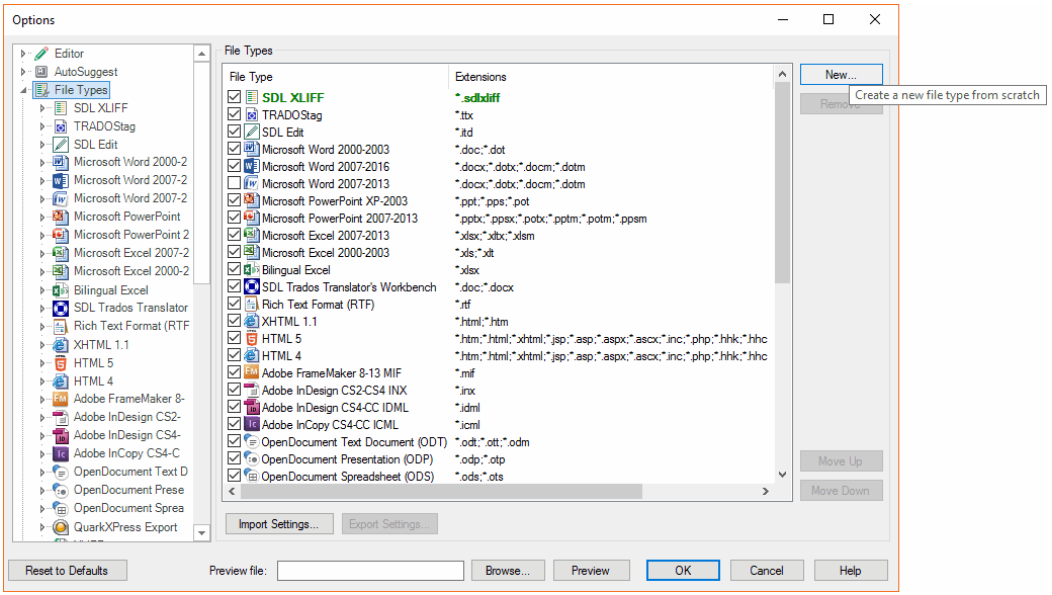


Figure 239: Creating a new XML file type in *Trados Studio 2017*

As in the previous versions of *Trados*, a wizard will guide you through the different steps of creating the file type.

Déjà Vu also contains predefined XML "filters" (available under `C:\Program Files` (*Windows 7* and above: `ProgramData`)`\ATRIL\Déjà Vu X(2)\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. 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.

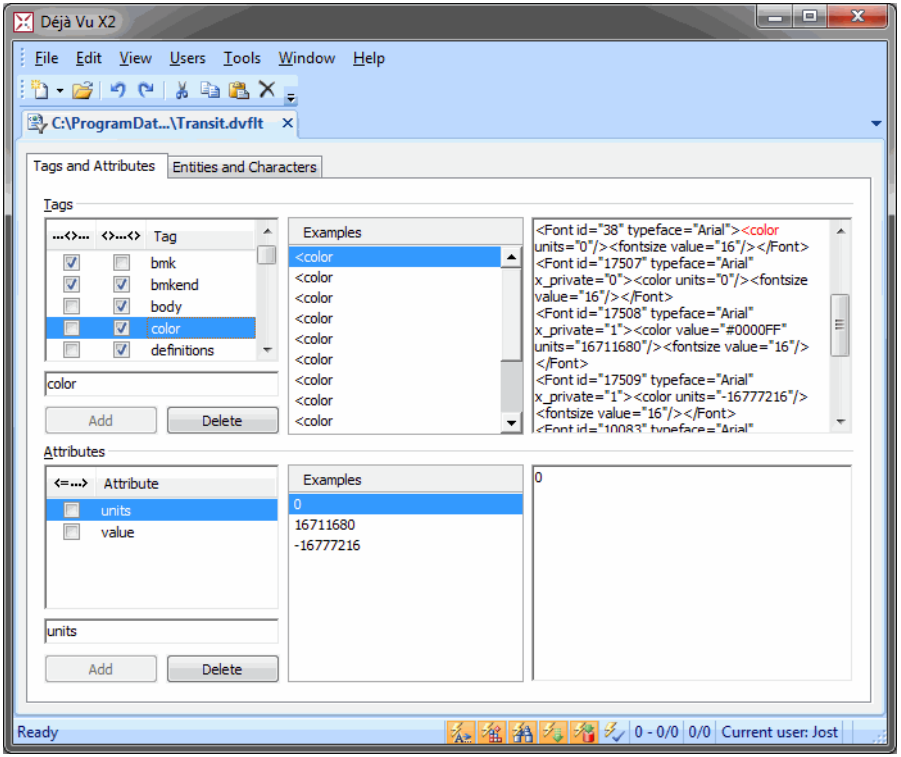


Figure 240: View of an XML filter in *Déjà Vu X2*

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**.

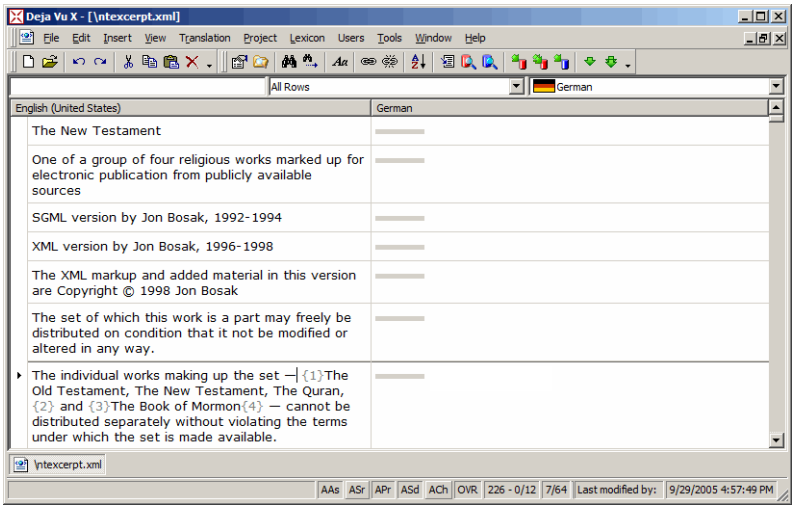


Figure 241: Imported XML file in *Déjà Vu X*

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 have traditionally presented a real headache until very recently: those with embedded HTML.

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;nbsp;&amp;nbs
p;&amp;nbsp;&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 (< and >) as well as the ampersand sign in the non-breaking spaces (& 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 242: XML file with embedded HTML in early versions of *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.

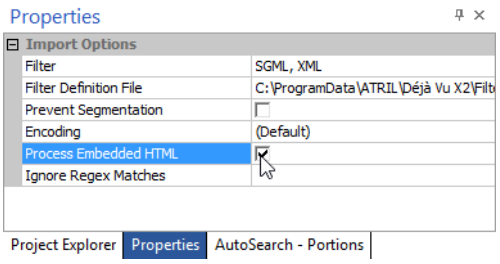
To avoid this scenario, long and tedious workarounds were needed that involved the conversion of the XML files into *Word* files and the semi-manual preprocessing of the XML and HTML tags.



There were even a couple of tools on the market that were specifically designed to aid with that process. One was a standalone tool called PrepTags (see your-translations.com/preptags.php) and the other is a free little Word macro called Tortoise Tagger (accurussian.net/tagger.htm).

Three of the leading translation environment tools have finally put an end to the misery by offering better solutions.

The most straightforward routine comes with *Déjà Vu X2/3*. Here you simply check **Process Embedded HTML** when configuring the import of the file:



memoQ has chosen a slightly different path that has applications for other scenarios as well. Here you can select to use cascading filters for the import of the file so that several routines are applied in the filtering process:

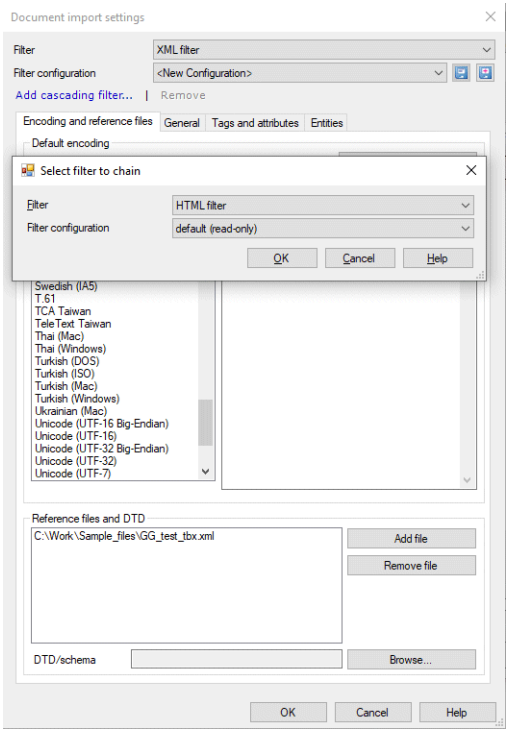


Figure 243: *memoQ's* cascading filters

Trados Studio (starting with version *2014 SP2*) also made it easier to process embedded content—though it's still more complicated than in other tools.

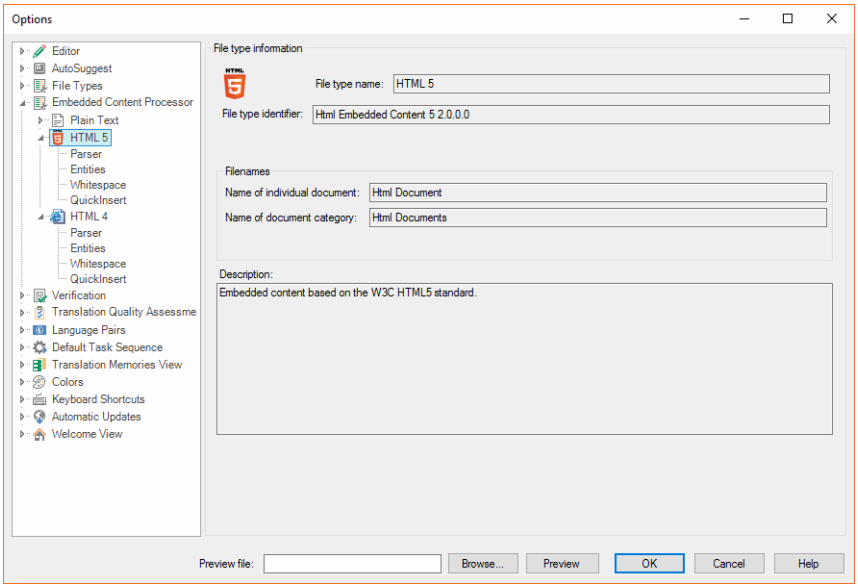


Figure 244: The first step of configuring processing embedded HTML in *Trados Studio 2017*

In *Memsource* you also have to specify which XML elements are supposed to be processed as HTML:

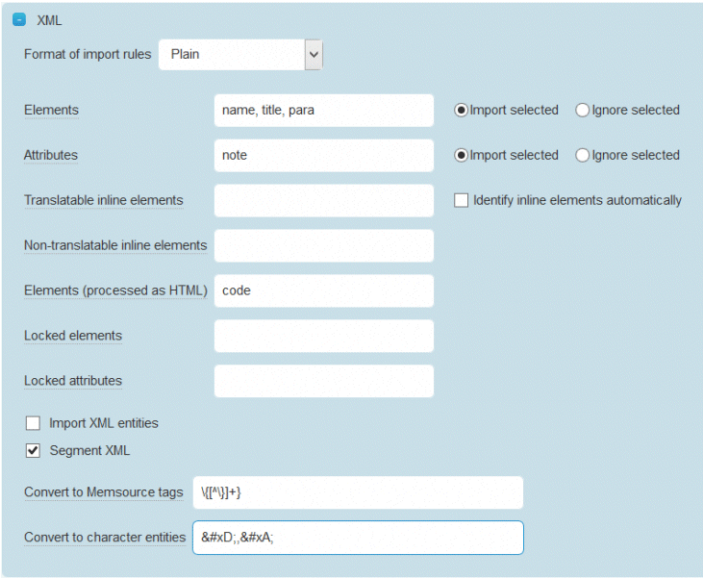


Figure 245: Setting which elements are processed as HTML

Across supports embedded HTML by providing customizable check-in rules.

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 a ny 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 résumés, invoices, file sharing and printing/publishing.

PDF File Types

To be able to utilize PDF files in the 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 **Edit PDF** (or: **Text Recognition**) 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 recognizes 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 364).

In addition to the free *Reader* version, the *Adobe Acrobat* product family also includes *Adobe Acrobat Standard* and *Adobe Acrobat Pro* versions.



Note that here the name "Adobe Acrobat" refers to these three paid versions and "Adobe Reader" to the free Reader version.

You should be 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 product comparison, see acrobat.adobe.com/us/en/acrobat/pricing/compare-versions.html.

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* (gonitro.com), *Foxit PDF Tools* (foxitsoftware.com), *Solid PDF Tools* (soliddocuments.com), and many others.



It's also possible to use the PDF features within other tools. For the past year or so, I have been using ABBYY FineReader (see abbyy.com/finereader) not only for its scanning capabilities but also for its advanced PDF reading and editing features.

I will concentrate on *Adobe Acrobat* here to the exclusion of the other tools.

Editing and Commenting on PDF Files

Commenting Tools

One of the most common tasks that translators need to do with PDF files is commenting on 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 364).

The **Comment & Markup** tools can be accessed through **Tools> Comment & Markup** (before *Acrobat X*) or **Comment** (*Acrobat X* and higher). 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 various text edit tools are particular 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** 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.

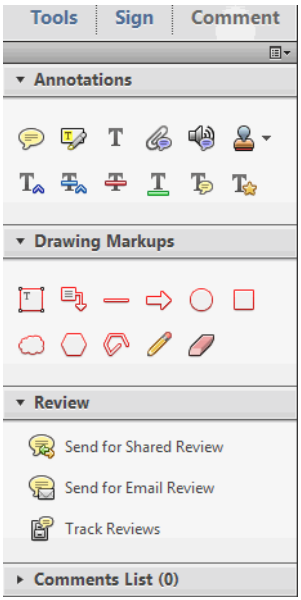


Figure 246: Commenting tools in the **Comment** pane *Adobe Acrobat X* and higher



Figure 247: Tools for commenting and markup in *Adobe Acrobat DC*

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 organize and manage comments. Clicking the **Show** button in the **Comment & Markup** toolbar opens a menu that includes several options for showing or hiding all comments (**Show/Hide Comments**) or only certain

comments based on comment type, reviewer or comment status. In *Adobe Acrobat X* and higher you can access this under **Comment> Comment List** (in *Acrobat DC* the list of comments is automatically displayed once you select **Comment**).

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.

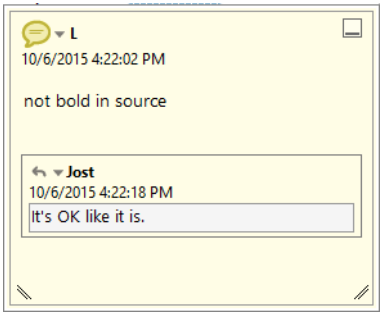


Figure 248: A comment with a reply

Other Editing Options

The *Standard* and *Pro* versions of *Adobe Acrobat* also have text editing tools (*Acrobat 9* and earlier: **Tools> Advanced Editing> TouchUp Text Tool**; *Acrobat X* and *XI*: **Tools> Content Editing> Edit Text & Images**; *Acrobat DC*: **Edit PDF**) 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.

Other editing features available in *Adobe Acrobat* include inserting, extracting, deleting and rearranging pages and document splitting (*Acrobat 9* and earlier in the **Document** menu; *Acrobat X* and *XI* under **Tools> Pages**; *Acrobat DC*: **Organize Pages**). 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.

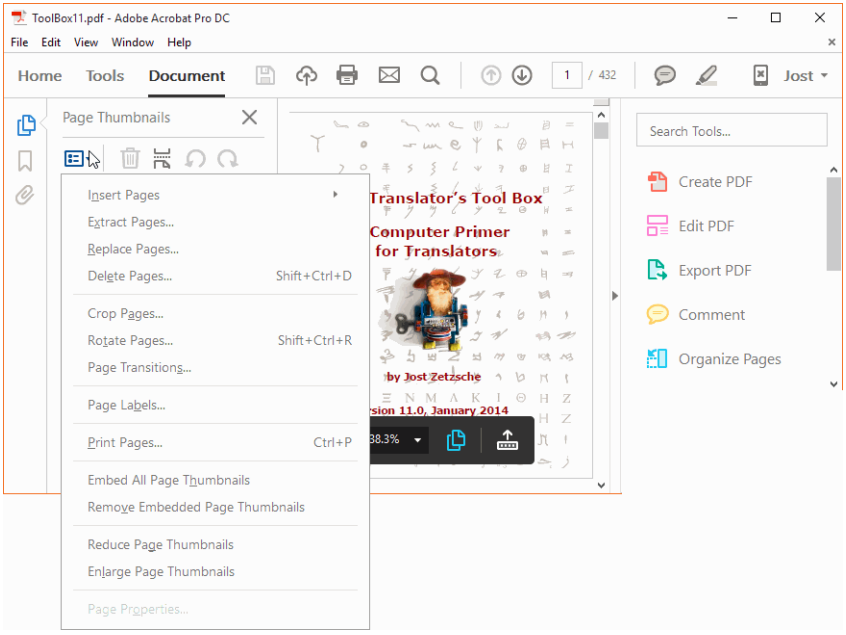


Figure 249: The **Pages** navigation pane

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** (*Acrobat X* and above: **File> Save as Other> Reader Extended PDF**).

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 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.

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** or **File> Save as Other> Text**). With both methods, each line ends with a hard return (paragraph mark), so they are practical only for a small amount of text.

What·is·backup·withholding?· Persons·making·certain·payments¶
to·you·must·under·certain·conditions·withhold·and·pay·to·the¶
IRS·28%·of·such·payments.·This·is·called·"backup·withholding."¶
Payments·that·may·be·subject·to·backup·withholding· include¶
interest,·tax-exempt·interest,·dividends,·broker·and·barter¶
exchange·transactions,·rents,·royalties,·nonemployee·pay,·and¶
certain·payments·from·fishing·boat·operators.·Real·estate¶
transactions·are·not·subject·to·backup·withholding.¶

Figure 250: Text copied using *Adobe Reader*, showing paragraph marks at the end of each line



AutoUnbreak is a handy little utility that allows you to delete those unnecessary hard returns in a text file but retain those that truly separate paragraphs. You can only paste up to 65,500 characters, but for the purposes of quickly copying some material out of the PDF file for research or alignment purposes, this is an extremely welcome utility with a price tag you can't beat. You can presently download it at download.cnet.com/AutoUnbreak/3000-2079_4-10504900.html.

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 key 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 to (File)**, 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**.

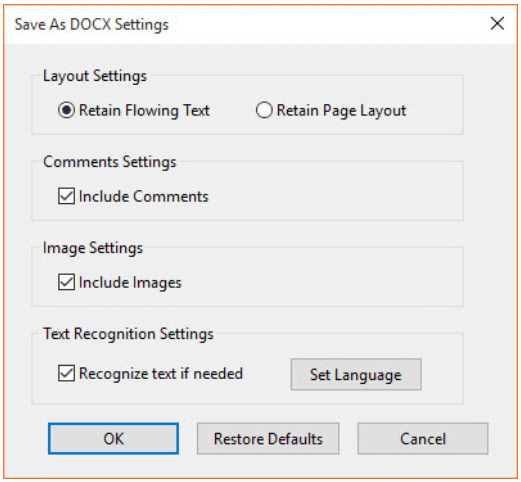


Figure 251: Available settings for converting PDF files to DOCX format in *Adobe Acrobat*

Using *Word 2013* and above to convert PDF Files

Starting with version 2013, *Microsoft Word* offers the *PDF Reflow* feature, an *MS Word*-internal PDF conversion program. It's a nice conversion utility that does fine with simple PDF files, but it shares the same problems common to almost any other program that converts PDFs to truly editable and therefore translatable files. Laudably, Microsoft is not too shy about talking about the limitations. In fact, when you open a PDF file in *Word*, you are shown the following dialog:

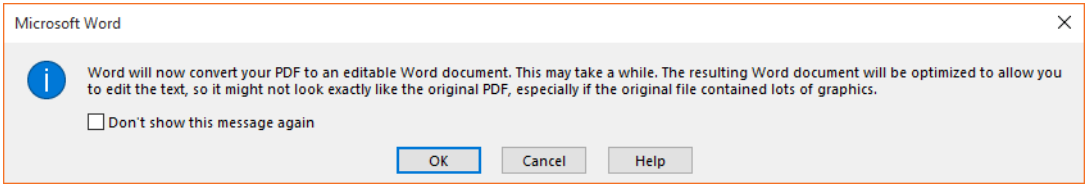


Figure 252: *MS Word's* warning that PDF conversion might only be partially successful.

Using *Apache OpenOffice/LibreOffice* to Work with PDF Files

Apache OpenOffice/LibreOffice offer the option of creating a PDF that has a *LibreOffice/Apache OpenOffice* file embedded, making this PDF completely editable within its originating application. In fact, when you open this PDF within *LibreOffice/Apache OpenOffice*, it automatically opens in the word processing/spreadsheet/presentation component it was created in just like a normal document. If it's "only" a normal PDF that is not directly editable, it opens in the graphics component.

These kinds of PDF files are called hybrid PDFs, and you can create them by selecting **File> Export as PDF> Embed this document inside the PDF** (*Apache OpenOffice*) or **Hybrid PDF** (*LibreOffice*).

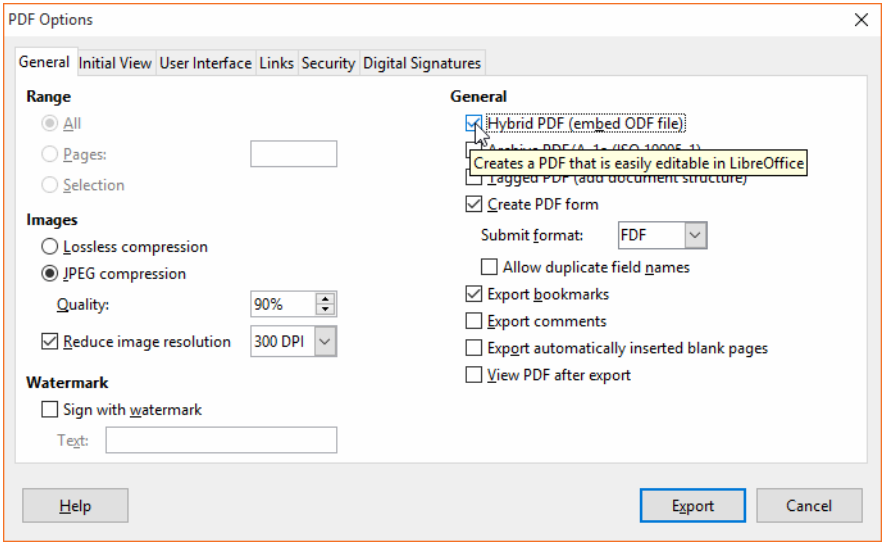


Figure 253: **PDF Options** in *LibreOffice*

OCR (Optical Character Recognition) Tools for PDF Conversion

The only way for image-based PDFs (also, and accurately, called "dead" PDFs) to be converted into editable files is to apply an optical character recognition process, i.e., the conversion of images of text into computer readable (and not just displayable) text. Image-based PDFs can come from a variety of sources, including scanned documents, images that were converted into a

PDF, or *Word* files that mainly contained graphics. And some PDF documents are protected in such a way that even though they are not technically image-based, they are for our purposes since there is no other way to digitalize and extract the text.

Adobe Acrobat comes with an internal OCR reader. You can access that feature by selecting **OCR Document> OCR Text Recognition** (before version XI) or the **Edit PDF** command in the **Tools** bar. The result will be a PDF file that is text-based and can therefore be searched and edited (within the limited possibilities that *Acrobat* offers for editing).



Acrobat DC offers a much-improved way of converting image-based documents into editable PDFs. The system essentially cheats by faking the look of fonts that it does not have installed or by characters that in reality it does not recognize—but it works surprisingly well. And it even does that when you enter new text for which it will also try to emulate the look of the previous text. To enable this feature, you'll need to select ***Edit> Preferences> Content Editing> Enable Artificial Bold/Italic Font Styles****.*

Unfortunately, this has no effect on any text that you might try to export—here the results are similar(ly poor) to previous versions.

You can also use third-party programs to convert image-based PDF files into translatable files, and even a number of translation environment tools now also offer integrated OCR-based PDF converters (see *Using OCR Features for PDF Conversion in Translation Environment Tools* on page 378). The most commonly used standalone OCR tool for PDF file conversion can be found as part of the ABBYY *FineReader* program (see abby.com/finereader).

ABBYY FineReader is a full-scale OCR program that 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.

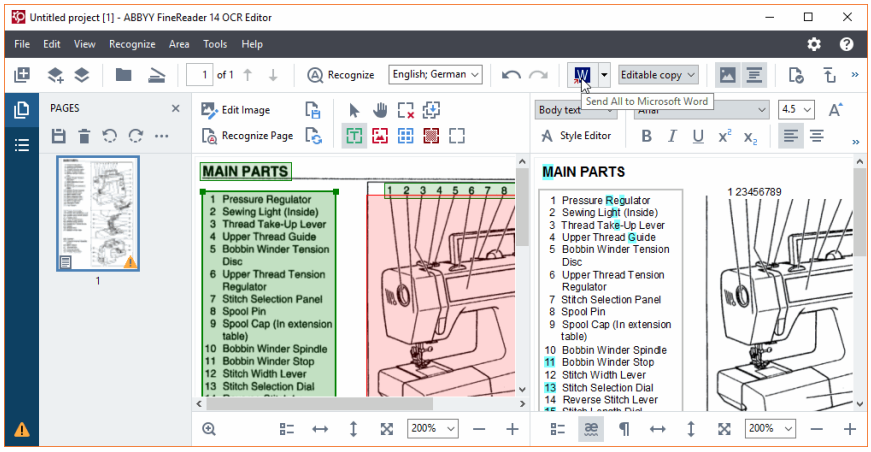


Figure 254: Converting a PDF file to *Word* format in *ABBYY FineReader*

FineReader offers several general and file-specific options for perfecting the conversion process and the output (**Tools> Options**). When converting PDF files to editable file formats, it's important to select the **Format Settings** 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.

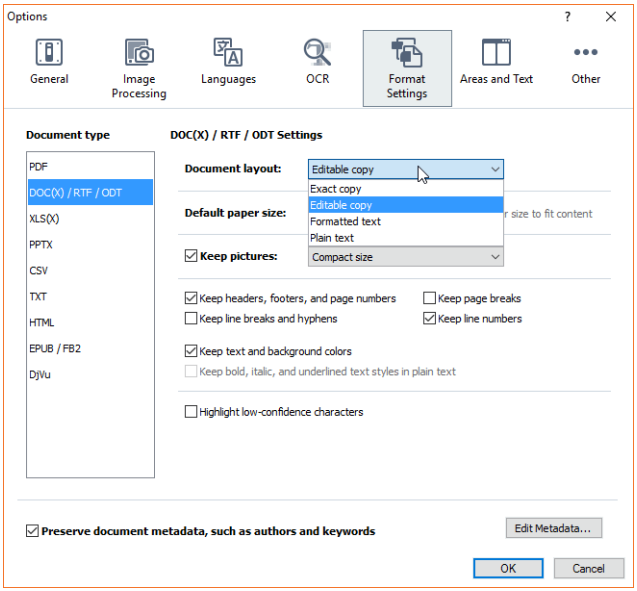


Figure 255: **Document Layout** options in *FineReader*

As you can see above, 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.

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 to 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 clean up converted Word files. You can obtain it from asap-traduction.com/CodeZapper.

A comparable set of features is offered as part of the TransTools suite of tools (see translatortools.net/products/transtools#word_formatting)

Translating PDFs Inside the PDF (Sort Of)

Iceni Technology specializes in preparing PDFs so they can be edited almost like a normal document in their own *Infix PDF Editor* (see iceni.com). At its core, it's a tool that opens any PDF and lets you format, edit, or add text—which is displayed in resizable text boxes—and manipulate and replace graphics within the PDF without performing any conversion to another format.

Theoretically it's possible to also translate within *Infix PDF Editor*, but since you want to use a translation environment tool for text of any appreciable length, the tool also offers the options **Export as XLIFF** and **Import translated XLIFF** in the **Translate** menu.

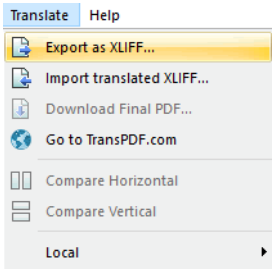


Figure 256: *Iceni PDF Editor's* **Translate** menu

The XLIFF file will be converted at Iceni's cloud-based part of the offering (see iceni.com/transpdf.htm). The reason why it's important that part of this process happens within the cloud and partly within the desktop-based tool has to do with its licensing model.

Infix PDF Editor comes with subscription-based licensing. It costs $9.95 a month, or about $100 a year. This might be a good deal if you work with PDFs all the time and have to process dozens a month. But if you only run into a PDF file for translation maybe a couple of times a month, this would be a steep price to pay. Here the free edition of the *Infix PDF Editor* in combination with 50 cents a page for the conversion into XLIFF and back into PDF is a much better deal.

The quality of the actual conversion is good albeit not fantastic, but you'll need to keep in mind that we are dealing with a "pretty darn frustrating" format. Sample files that I ran through as a test were relatively heavily formatted, and though I would have had to spend some time in Infix to fix them in the direct PDF-to-PDF conversion, it was certainly better than expected. Much better.

And if your PDF is image-based, the tool will recognize that and automatically send it to ABBYY's OCR (optical character recognition) server (see page 368).



One thing that you need to consider before using this tool is whether your client actually wants a PDF back or would prefer a word processing format (RTF or DOCX) so that he can do some editing, proofreading, or reformatting with the file as well. In that case this is not the right kind of solution.

If you use Trados Studio, memoQ, or Memsource as your translation environment tool, you can also use this solution right from within those environments. For more information, see appstore.sdl.com/app/transpdf/718 (Trados Studio), memoq.com/en/translate-pdf-documents and iceni.com/blog/memsource-integrates-with-transpdf-com (Memsource).

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 now a number of tools (*Trados Studio*, *Wordfast Pro*, *Wordfast Anywhere*, *Smartcat*, *memoQ*, *Alchemy Publisher*, *XTM Cloud*, *Déjà Vu* and *Fluency* among others) 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 a *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*, *Fluency* and *Wordfast Pro* use *BCL*; and *memoQ* employs Aspose.PDF for .Net (by Aspose) for conversion into DOCX and Xpdf (by Glyph & Cog) for conversion into TXT files.

Interestingly, the strategies of the different TEnTs differ.

In addition to conversion via TransPDF (see *Translating PDFs Inside the PDF (Sort Of)* on page 372), *memoQ* offers two different ways of processing PDF files. With the first and possibly more efficient option, it does not even pretend to save the formatting of the file. Instead, it "only" converts 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 end 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.

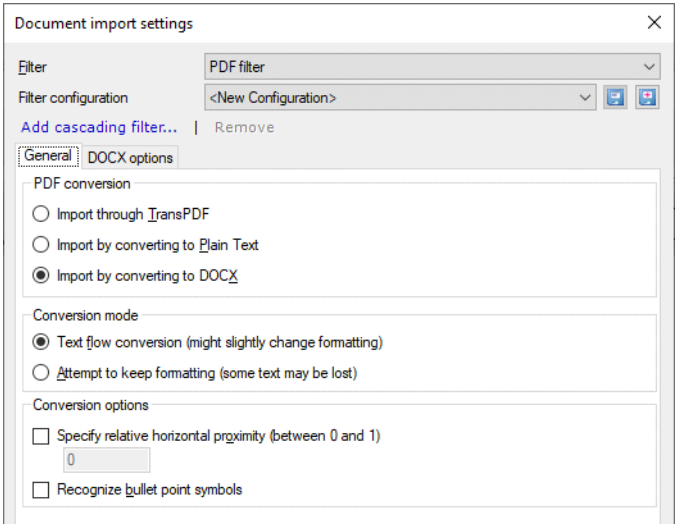


Figure 257: PDF import options in *memoQ*

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.

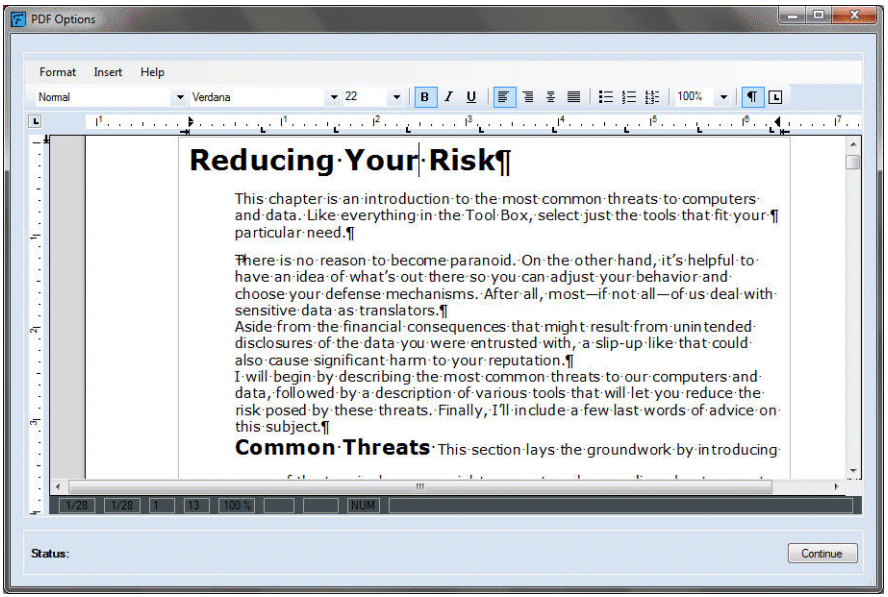


Figure 258: *Fluency's* interim editing interface

Déjà Vu X2/3 converts the PDF into a *Word* file and uses an integrated version of the *CodeZapper* tool (see page 371) to eliminate most unnecessary codes. Alternatively, *Déjà Vu X3* also offers an automatic conversion through the integrated PDF conversion that Microsoft *Word 2013* and above offers (see page 367)—with often better results.

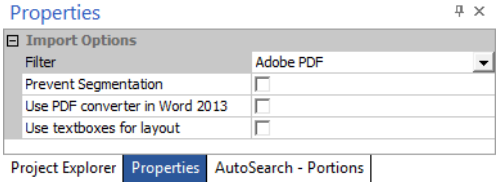


Figure 259: PDF conversion options in *Déjà Vu X3*.

Trados Studio offers a number of options to direct your PDF conversion.

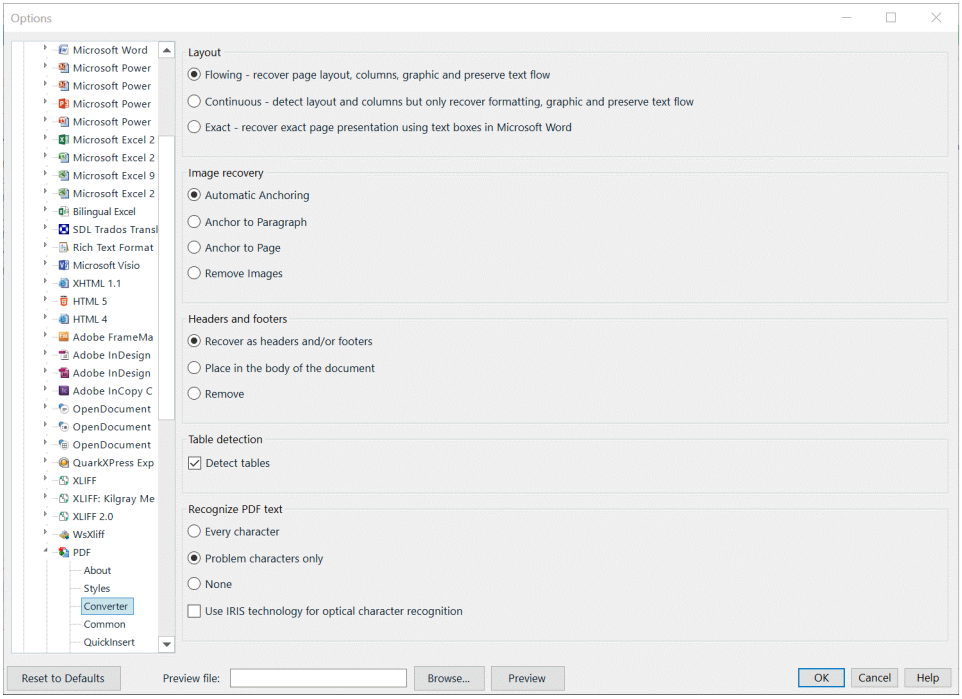


Figure 260: *Trados Studio's* PDF conversion options



*Trados Studio contains a setting under the **Common** option entitled **Skip advanced font formatting**, which also helps with a smoother PDF import.*

Other tools try to do everything "behind the scenes." When you open a PDF file, those tools attempt to keep the layout, presenting it 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:

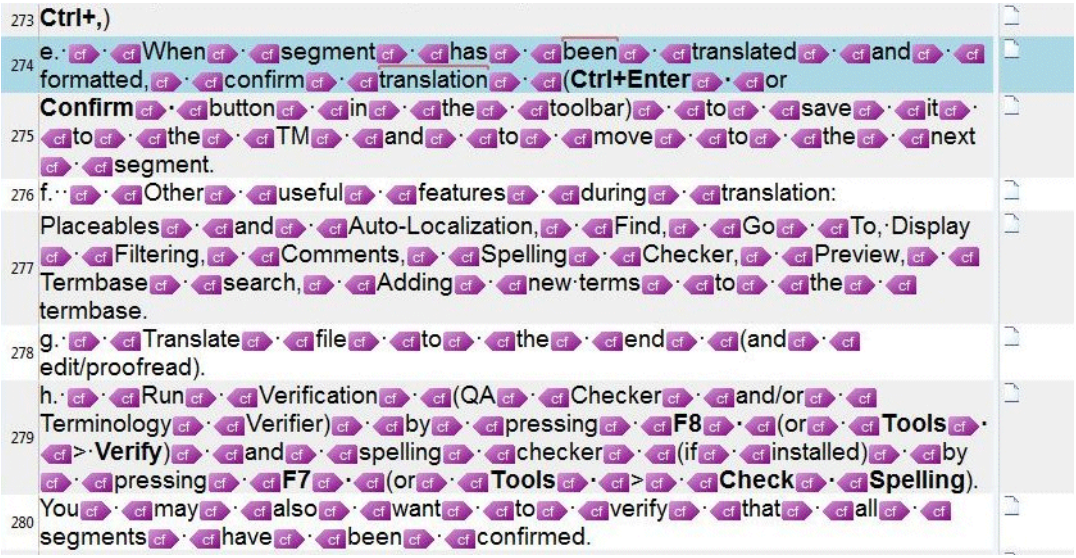


Figure 261: A PDF file that has been opened directly in an early, non-customizable version of *Trados Studio*. An overabundance of tags makes translating the file virtually impossible to translate.

You won't 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—erroneous hard returns at the end of lines—is also handled relatively well with the ABBYY and Nuance tools. You may occasionally find a rogue hard return in converted files from these solutions, but they are few and far between.



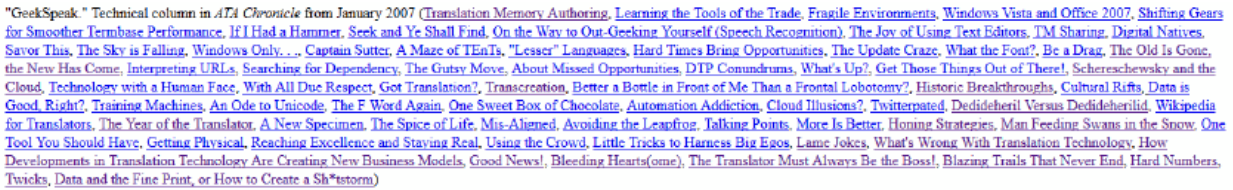
For Trados Studio OCR conversion of PDF files, see Using OCR Features for PDF Conversion in Translation Environment Tools on page 403.

Using OCR Features for PDF Conversion in Translation Environment Tools

A number of translation environment tools have started to use OCR features to decipher image-based PDFs. Presently the list includes *Trados Studio 2015*, *Wordfast Anywhere* and *Smartcat*.

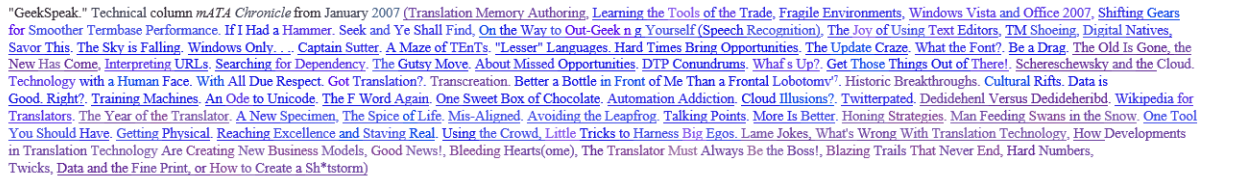
I did some testing with an admittedly difficult document—my own bibliography—which contains a number of different languages and a number of made-up words. I converted the first page into a graphic and then had the different tools have a go at it.

This is a section of the original:



"GeekSpeak." Technical column in *ATA Chronicle* from January 2007 (Translation Memory Authoring, Learning the Tools of the Trade, Fragile Environments, Windows Vista and Office 2007, Shifting Gears for Smoother Termbase Performance, If I Had a Hammer, Seek and Ye Shall Find, On the Way to Out-Geeking Yourself (Speech Recognition), The Joy of Using Text Editors, TM Sharing, Digital Natives, Savor This, The Sky is Falling, Windows Only. . ., Captain Sutter, A Maze of TEnTs, "Lesser" Languages, Hard Times Bring Opportunities, The Update Craze, What the Font?, Be a Drag, The Old Is Gone, the New Has Come, Interpreting URLs, Searching for Dependency, The Gutsy Move, About Missed Opportunities, DTP Conundrums, What's Up?, Get Those Things Out of There!, Schereschewsky and the Cloud, Technology with a Human Face, With All Due Respect, Got Translation?, Transcreation, Better a Bottle in Front of Me Than a Frontal Lobotomy?, Historic Breakthroughs, Cultural Rifts, Data is Good, Right?, Training Machines, An Ode to Unicode, The F Word Again, One Sweet Box of Chocolate, Automation Addiction, Cloud Illusions?, Twitterpated, Dedideheril Versus Dedideherilid, Wikipedia for Translators, The Year of the Translator, A New Specimen, The Spice of Life, Mis-Aligned, Avoiding the Leapfrog, Talking Points, More Is Better, Honing Strategies, Man Feeding Swans in the Snow, One Tool You Should Have, Getting Physical, Reaching Excellence and Staying Real, Using the Crowd, Little Tricks to Harness Big Egos, Lame Jokes, What's Wrong With Translation Technology, How Developments in Translation Technology Are Creating New Business Models, Good News!, Bleeding Hearts(ome), The Translator Must Always Be the Boss!, Blazing Trails That Never End, Hard Numbers, Twicks, Data and the Fine Print, or How to Create a Sh*tstorm)

Here is what the *ABBYY FineReader* made of it:



"GeekSpeak." Technical column *mATA Chronicle* from January 2007 (Translation Memory Authoring, Learning the Tools of the Trade, Fragile Environments, Windows Vista and Office 2007, Shifting Gears for Smoother Termbase Performance. If I Had a Hammer. Seek and Ye Shall Find, On the Way to Out-Geek n g Yourself (Speech Recognition), The Joy of Using Text Editors, TM Shoeing, Digital Natives, Savor This. The Sky is Falling. Windows Only. . .. Captain Sutter. A Maze of TEnTs. "Lesser" Languages. Hard Times Bring Opportunities. The Update Craze. What the Font?. Be a Drag. The Old Is Gone, the New Has Come, Interpreting URLs. Searching for Dependency. The Gutsy Move. About Missed Opportunities. DTP Conundrums. Whaf s Up?. Get Those Things Out of There!. Schereschewsky and the Cloud. Technology with a Human Face. With All Due Respect. Got Translation?. Transcreation. Better a Bottle in Front of Me Than a Frontal Lobotomv?. Historic Breakthroughs. Cultural Rifts. Data is Good. Right?. Training Machines. An Ode to Unicode. The F Word Again. One Sweet Box of Chocolate. Automation Addiction. Cloud Illusions?. Twitterpated. Dedideheni Versus Dedideheribd. Wikipedia for Translators. The Year of the Translator. A New Specimen. The Spice of Life. Mis-Aligned. Avoiding the Leapfrog. Talking Points. More Is Better. Honing Strategies. Man Feeding Swans in the Snow. One Tool You Should Have. Getting Physical. Reaching Excellence and Staving Real. Using the Crowd, Little Tricks to Harness Big Egos. Lame Jokes, What's Wrong With Translation Technology, How Developments in Translation Technology Are Creating New Business Models, Good News!, Bleeding Hearts(ome), The Translator Must Always Be the Boss!, Blazing Trails That Never End, Hard Numbers, Twicks, Data and the Fine Print, or How to Create a Sh*tstorm)

You can see that in general it did a good job. It did not like words like "lobotomy" and other strange words; it had a difficult time when formats were switched (see the "*mATA*" rather than "in *ATA*" in the first line); and it had a hard time knowing where to use commas and where to use periods.

Here is what we get from *Smartcat*:

"GeekSpeaL" Technical column *mATA Chronicle* from January 2007 (Translation Memory Authoring. Learning the Tools of the Trade. Fragile Environments. Windows Vista and Office 2007. Shifting Gears for Smoother Termbase Performance. If I Had a Hammer. Seek and Ye Shall Find. On the Wav to Qut-Geeking Yourself (Speech Recognition). The Joy of Using Text Editors. TM Sharing. Digital Natives. Savor This, The Sky is Falling. Windows Only... . Captain Sutter. A Maze of TEnTs. "Lesser" Languages, Hard Times Bring Opportunities. The Update Craze. What the Font?, Be a Drag. The Old Is Gone, the New Has Come. Interpreting URLs. Searching for Dependency.. The Gutsy Move. About Missed Opportunities. DTP Conundrums. What's Up?. Get Those Things Out of There!. Schereschewsky and the Cloud. Technology with a Human Face, With All Due Respect. Got Translation?, Transcreation.. Better a Bottle in Front of Me Than a Frontal Lobotomy?. Historic Breakthroughs. Cultural Rifts, Data is Good. Right?, Training Machines. An Ode to Unicode. The F Word Again One Sweet Box of Chocolate, Automation Addiction. Cloud Illusions?, Twitterpated. Dedideheril Versus Dedideherilid, Wikipedia for Translators. The Year of the Translator, A New Specimen. The Spice of Life. Mis-Aligned. Avoiding the Leapfrog. Talking Points. More Is Better. Honing Strategies. Man Feeding Swans in die Snow. One Tool You Should Have, Getting Physical. Reaching Excellence and Staving Real. Using the Crowd, Little Tricks to Harness Big Egos. Lame Jokes, What's Wrong With Translation Technology, How Developments in Translation Technology Are Creating New Business Models, Good News!. Bleeding Hearts(ome). The Translator Must Always Be the Boss!, Blazing Trails That Never End. Hard Numbers. Twieks- Data and the Fine Print, or How to Create a Sh*tstorm)

Overall it's about the same, with slightly different errors but generally quite acceptable.

Trados Studio 2019 essentially fails with the same document:

"GeekSpeak " Technical column *in ATAChronicle* from January 2007 (Translation Mrmory Authorine, . Leamine the Tools of the Trade. Fra e ile Environm ents. Wind ov.-.s Vist a and O ffic e ')007+ . Shi ft ine G earn for Smoot her Term base Perform anc e, If I H a d a Hamm er Se ek and Ye Shal l Find . On the Way to Ou-t Geekine Yourself (Speech Reco enit ion). Th e Joy of Usina Text E ditor s .TM S hann.a Digital Natives

_ SaYOI"This, The Skv is Falline . Windows Onyl . Ca ptain Sutt e r; A Maz e of TE nT s. "Lesse r' Lan zuaz e s. H ard Times Brin11 Opportunitie s. The Update Cra z.t> \Vb at the Font . B e a lli a v;, Th e Old Is Gone . the Ne w H as C om t>. In terpretinv: URL s., S earc hin ..- fo f"Dt"Jen d en c .:v Th e Gut sy MoY e., About Missed Opportunitise . :DTP Conun dru l 1J5 \Vbat's Up., G et Those Thinas Out of There '. Schet""es cb. t>ws k y and the Clo u d, .Tt"Chnol o 11y w ith a H uman Fa c e., Wit h A U Du t> Re s J\(.c t G ot Transl at io n?., Traru cJ"eation. . B e n et"" a B o n le in fr on t o f \-l:c Th an a Fm nta l L o bot o mv . Hi st OJ"iC". B r t>akthr ouah ., 5: C u ltu ra l Ri ft s Data is.

_ G ood Riaht . T rainin a M ach int>s. ANO de t o U n ico d.e. Tut> F Wo rd A v:ai n On.e weet Box of Chocola te Auto ma tion 4..ddiction . C loud IUusions Twin etl)ated . :Dedideh el"il Versus Dedidehcrili d. \Vikipedia fol"Translatof S:. . Tut> Year of the Trans lato r; A New Specim t>n . 'fhe Spice of Lif e, Mi-s Aliaued, AYoidin iz: the L ea p fm 11 T alk in.- P o i n t s , Mo I"t' I c: B en er., Honina Stra te ics. Man F ecdin .,- S·wans in th e now. On e Tool You Shou ld HaY e; G t>n in iz: P hysical R eac h in - E i:cel lt>n C' t" and Stayin - R t>a l Us in e- the Crowd Little T ricks to H arn e s B i..- E e-os L am e Jo kes Wha t's Wrnn e: With Translation Tt"Chnolo v:y. How D c elopm e nt in Trnnslation Tt'<:hnolo e-y Arc Cr e atin" N e w Busin ess Modd s., JOod New sl . B leedin ..- H cart s(om c), Th e Trans la to r Mu t Always Be tht> Bos si . B lazin Trails That Ney er En.d; HM d Numbe rs , Twick 5:. D at a and tb t> Fine Pl"int o r H ow to Cl"ea te a Sh " tstonn)

Maybe most surprisingly, the best result comes from *Wordfast Anywhere*:

"GeekSpeak." Technical column *mAIA Chronicle* from January 2007 (Translation Memory Authoring, Learning the Tools of the Trade, Fragile Environments, Windows Vista and Office 2007, Shifting Gears for Smoother Termbase Performance. If I Had a Hammer. Seek and Ye Shall Find. On the Wav to Out-Geeking Yourself (Speech Recognition). The Jov of Using Text Editors. TM Sharing. Digital Natives. Savor This The Sky is Falling. Windows Only... . Captain Sutter. A Ma?e of ThwTs. "Lesser" Languages. Hard Times Bring Opportunities. The Update Cm?e. What the Font?. Be a Drag. The Old Is Gone, the New Has Come. Interpreting URLs. Searching for Dependency. The Gutsy Move. About Missed Opportunities. DTP Conundrums. What's L'p?. Get Those Things Out of There!. Schereschewsky and the Cloud. Technology with a Human Face. With All Due Respect. Got Translation'?. Transcreation. Better a Bottle in Front of Me Than a Frontal Lobotomv?, Historic Breakthroughs. Cultural Rifts. Data is Good, Right?. Training Machines. An Ode to Unicode. The F Word Again. One Sweet Box of Chocolate. Automation Addiction. Cloud Illusions?. Twitterpated. Dedideheril Versus Dedideherilid. Wikipedia for Translators. The Year of the Translator. A New Specimen. The Spice of Life. Mis-Aligned, Avoiding the Leapfrog. Talking Points. More Is Better. Honing Strategies, Man Feeding Swans in the Snow, One Tool You Should Have. Getting Physical. Reaching Excellence and Staving Real, Using the Crowd, Little Tricks to Harness Big Egos. Lame Jokes. What's Wrong With Translation Technology, How Developments in Translation Technology Are Creating New Business Models, Good News!. Bleeding Hearts(ome), The Translator Must Always Be the Boss!. Blazing Trails That Never End, Hard Numbers, Twicks, Data and the Fine Print, or How to Create a Sh*tstorm)

While it capitulated with terms like "TEnT" or "Craze," it gives the best overall result.



According to Wordfast representatives, the reason for the pristine rendering is a third-party ABBYY OmniPage server that Wordfast Anywhere uses in the background.

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 *Nuance OmniPage* (or *Power PDF*).

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).

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 *MS Office* 2007 and higher (**Save As> Adobe PDF**) and *Apache OpenOffice*/*LibreOffice* (**File> Export as PDF**).

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 Resources 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 page 250).

Another way to use PDFs as a translation resource is to use *memoQ's* LiveDocs feature within that translation environment (see page 210).

Multimedia and Audiovisual Translation



This chapter was written by Carolina Alfaro de Carvalho. Carolina has been a full-time professional translator for 20 years, 15 of which she has spent specializing in the translation of audiovisual materials, particularly film subtitling. She regularly works closely with video producers and directors, often following a film from its early treatment stages all the way to the big screen. She also regularly provides audiovisual translation and editing services to many clients in the corporate and technical world. Carolina teaches subtitling courses in Rio de Janeiro, Brazil, and Toronto, Canada, as well as online courses.

Feel free to contact her for inquiries or to suggest new content to this chapter.

Email: carolalfarotranslations@gmail.com; websites: carolalfarotranslations.com (Canada), scribatraducoes.com.br (Brazil); blog: multimediatranslation.org; LinkedIn: ca.linkedin.com/in/carolalfaro.

Multimedia or audiovisual materials have become a common way to convey ideas and messages in all different fields, far exceeding the realm of entertainment such as films and shows for cinema and TV. In any area, audio and video files are being used as a quick and engaging way to train, sell, motivate and much more; they are often integrated into presentations or websites, as multimedia tools and online resources make it easier and cheaper to produce and share these kinds of files.

If you are a technical translator, even if you've never had a particular interest in audiovisual (AV) translation, chances are that multimedia files will soon find their way to you through your regular clients if they haven't done so already. They are becoming ubiquitous, and for many technical translators they will soon have be as much a part of their tool arsenal as translation environment tools. You might not become an AV translation expert (there are so many different forms of AV translation that it would be hard to master all of them), but the sooner you familiarize yourself with some of the basic tools and processes, the better you can serve and retain your clients.

Leaving aside the well-established and more traditional entertainment segment of AV translation and focusing more on the technical and institutional market, there are a few clear trends in this scenario:

- The end clients, who are not from the audiovisual field and are using AV materials as a means to an end—just like any other text—tend to hire a translator or an agency to translate those materials—just like they would do with any other text. However, these translators or agencies more often than not lack AV translation expertise.
- This can lead to poor AV translation (even if the translation itself is excellent, its integration with the multimedia file can be far from ideal), which the end client is often unable to assess. This tends to reinforce common negative views related to AV translation and makes it harder for expert AV translators to prove their worth and be rewarded for it.
- Many AV translators from the entertainment sector who know the processes and tools well are moving to the technical/institutional market to fill this gap. However, the technical segment is a lot more varied than the entertainment segment in terms of the formats used, which are constantly evolving. It often requires the translator to take care of the entire process, from translation to final editing, whereas in entertainment the video producers are in charge of the more technical editing tasks. Also, many of these translators are not experienced technical translators and might lack knowledge of the different fields or terminology and research resources.
- There is high demand for capable technical translators who know what to do when presented with audiovisual tasks—or AV translation experts who are also competent technical translators. Everyone will benefit from this convergence.

This chapter provides some information and suggests tools for many common AV translation-related tasks. Bear in mind that most forms of AV translation, such as subtitling and dubbing, are fields of expertise that take time to master and are a lot more complex and difficult than most people think. The best way to learn them is through specialized courses. It would be impossible to teach something like subtitling or dubbing briefly and without hands-on practice, and this is not the purpose of this chapter.

Some Definitions and Clarifications

There are so many different forms of audiovisual translation—and we are usually presented with the final products—that it can be a bit overwhelming or confusing when we have to navigate all the possibilities or assist a client who is also often confused. So here are some basic definitions and a few brief explanations to help you navigate any potential confusion.

Common types of AV translation:

- **Dubbing**: The audio tracks of dialogs are entirely replaced by audio tracks in the translated language, interpreted by professional actors and with lip-sync. The need for lip-sync accounts for often significant editing.
- **Voiceover**: Audio tracks in the translated language, interpreted by professional voice talents and without lip-sync, are added to the original film without deleting the original track, whose volume is lower. Although there is no lip-sync, there are time constraints imposed by the pace of the original language.
- **Narration**: Often confused with voiceover, professionally recorded off-screen narration in the translated language replaces the original off-screen narration. It is technically different both from dubbing (because there is no lip-sync) and from voiceover (because the original audio track is removed).
- **Subtitling**: The written translation of oral or written text on the video is added to the screen in text segments that remain on the screen for the duration of that segment in the original language. Because we require more time to decode written information than oral information, subtitles have to be concise enough to be read in a very few seconds and simple enough to be understood at first sight, leaving time for the viewer to look at the images as well. This can require a great deal of paraphrasing and synthesis.
- **Closed captions**: For hearing-impaired people, these include not only oral texts but also meaningful sounds (phone ringing, shot gun, sometimes voice intonation, etc.).

- **Teletext**: For hearing-impaired people, this is scrolling text (also known as "running captions") included in TV programming, often live, usually done with voice recognition software.

Working with multimedia involves many different processes and tasks carried out by many different professionals. It's important to know which tasks are the translator's job and which are carried out by other professionals.

In oral forms of AV translation (such as dubbing or voiceover), the translator works on the script, which may require cues, phonetic transcription and other conventions to help the voice talents. In written forms of AVT (such as subtitling or closed captions), the translator prepares the text file in a specific format that includes time codes. Time codes are often prepared by video producers, but translators may be asked to time the subtitles as well (some accept this task, some don't).

On the other hand, recording the voiceover or dubbing is done by professional voice talents or actors, and editing the final video with subtitles is the task of video editors or producers. This is not part of the core job of a translator, although there are many translators who are also voice talents or video editors. Even when we are able to provide all of these services, they are still distinct jobs, with different requirements and different fees. Not all clients are aware of this and we might need to explain exactly which task(s) we do.

There are different ways to display a film with subtitles. Dynamic or soft-coded subtitles can be selected by the viewer, as required. This is the case of DVDs with language options, digital cable TV (in most cases) or YouTube's closed caption feature. On the other hand, embedded or hard-coded subtitles are permanently "printed" onto the video images.

Different programs are used for different tasks. While some AV translators don't use specialized tools for their translations, many professionals use specific software to prepare a dubbing script or a subtitle file with more efficiency and quality. Other tools are used specifically for audio editing or video editing, as shown in the following sections.

Codecs and Media Players

Audio and video files can be codified and compressed in countless different ways. "Codecs" (portmanteau from "coder-decoder") are programs that help media players and other software to encode and decode the files. All computers come equipped with a number of basic codecs, but never all, especially the newest codecs.

If your usual media player or media-related software won't play a video correctly (for instance, you may hear the audio but see no image, or vice versa), you can install a codec pack with more and newer codecs.

K-Lite Codec Pack

K-Lite (see codecguide.com/download_kl.htm) is one of the best and most widely used codec packs. It is completely free. There are four different packs, with *Standard* being more than enough for most common uses and *Full* recommended for more advanced video editing. Select your preferred type, then choose one of the mirror locations for a download.

Simply install it and any media-related software will use the codecs as needed.

There are also other media players, some of which include their own codecs and plug-ins and perform much better than media players that you might already have installed.

VLC Media Player

VLC Media Player (VLC) (see videolan.org) is one of the most widely used video players. Simply download and install it. It's free and fully functional and will play just about any audio and video file.

It will also play many different subtitle files with a video, and even allow hard-coding subtitles. (See page 400.)

Media Player Classic

Media Player Classic (MPC) (see mpc-hc.org) is a free, open-source, light and reliable media player for *Windows*. It will also play just about any audio and video type, and will play several different subtitle files (also known as "soft-coding"—see page 388).

AVS4YOU Media Player

AVS4YOU is a suite of multimedia software, which also includes a great media player (see avs4you.com/AVS-Media-Player.aspx). You can download each program separately or the entire suite. The license costs around $60 for all programs, which might be a good investment as a number of the programs can be used professionally. (More on other *AVS4YOU* programs on page 396.)



There are dozens of different file types and extensions, because each of the many video editing programs has its own proprietary subtitle file format. However, two types of subtitle files have become virtually universal—i.e., they are imported by most media players, subtitling, and video editing software. These are SRT (SubRip Text) and SSA (SubStation Alpha) or ASS (Advanced SubStation Alpha).

SRT files contain only simple text and time codes, while SSA/ASS files also include coding for text positioning on the screen, font type, size, color, and style. This means that when an SRT file is encoded on a video, the subtitles will be displayed using the default configuration of the media player or software used, and they need to be configured by the video editor and/or user.

SSA/ASS files, on the other hand, carry all the formatting information in the subtitle file, so they can be configured by whoever prepares the subtitle file (the client, translator, or editor) and displayed in the same way by the media players and programs that accept this file type.

To run videos with a separate subtitle file ("soft-coding")

Most video players, and certainly the three mentioned above, can play a subtitle file (a separate, offline file, such as SRT, SSA or SUB) together with the video file, superimposing them in real-time. This is useful when, for any reason, you don't have a video file with embedded ("hard-coded") subtitles but have a separate subtitle file for that video. In my case, I can send the subtitle file for the client to review before I actually edit the video with embedded subtitles.

A simple way to do it is to put both files (the video file in any digital format and the subtitle file) in the same folder, both with the same name except for the extension. If you've received them with different names, just rename them. For instance:

`myfavoritefilm.avi` (video file)

`myfavoritefilm.srt` (subtitle file)

Then just run the film file with your favorite media player and the subtitles should be played together with the film. For instance, if I want the client to review the subtitles I prepared, I will send them a subtitle file with exactly the same file name as the video they sent me, and I will instruct them to save the subtitle file in the same folder as the video.

Bear in mind that with most subtitle files, the media player will use its default settings for the font, size, and color of the subtitles. Also, it's important to note that the subtitle file has the time codes determining when each subtitle will appear and disappear on the screen. If you notice that the subtitles displayed on the video are not in sync with the film, you might have a video file (or subtitle file) with different frame rates or some other editing difference.

Using the three media players listed in the previous section, you can change the look of soft-coded subtitles in their settings—such as the font type, size, and color. This doesn't change the subtitle file itself, just how it is displayed on the videos you play with that player.

While the *AVS Media Player* will not run SSA/ASS files, *VLC Media Player* and *Media Player Classic* will. This is an advantage if you want to also embed the subtitles and are working with this type of file, because the client can have a preview of the final look of the video. You can send them the SSA or ASS file and instruct them to view it with the video using *VLC* or *MPC*, enabling them to review not only the translation, but also the font size, positioning, and so forth.

Other options and settings with VLC and MPC

VLC Media Player (*VLC*) and *Media Player Classic* (*MPC*) also allow several different types of subtitle files to be added from a different folder or with a different file name. An easy way to do this is to open the video on one of these media player and then simply drag and drop the subtitle file onto it.

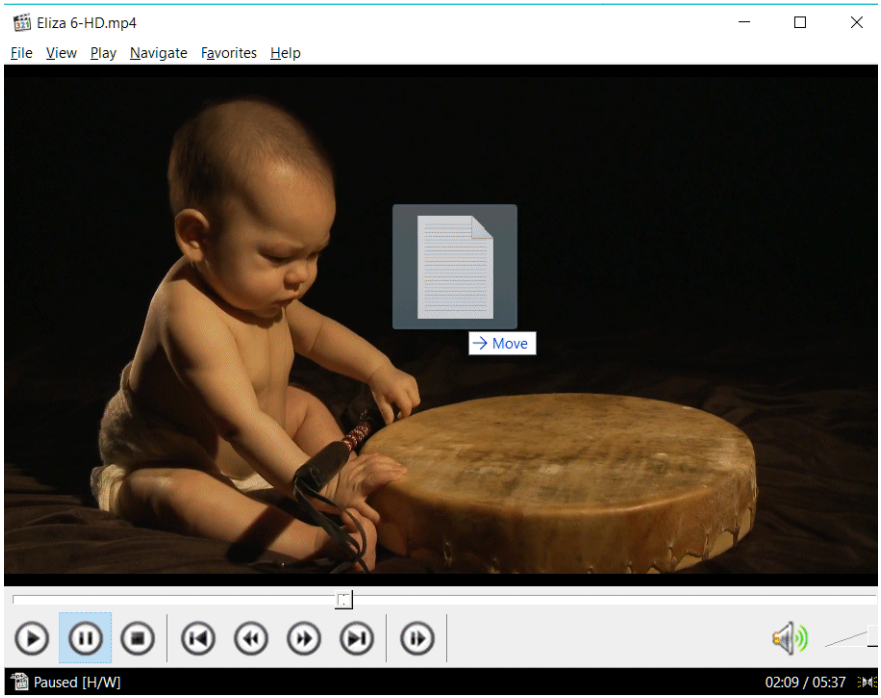


Figure 262: Dropping a file into *MPC*

In *MPC*, you can also select a subtitle file with **File> Subtitles> Load Subtitles**.

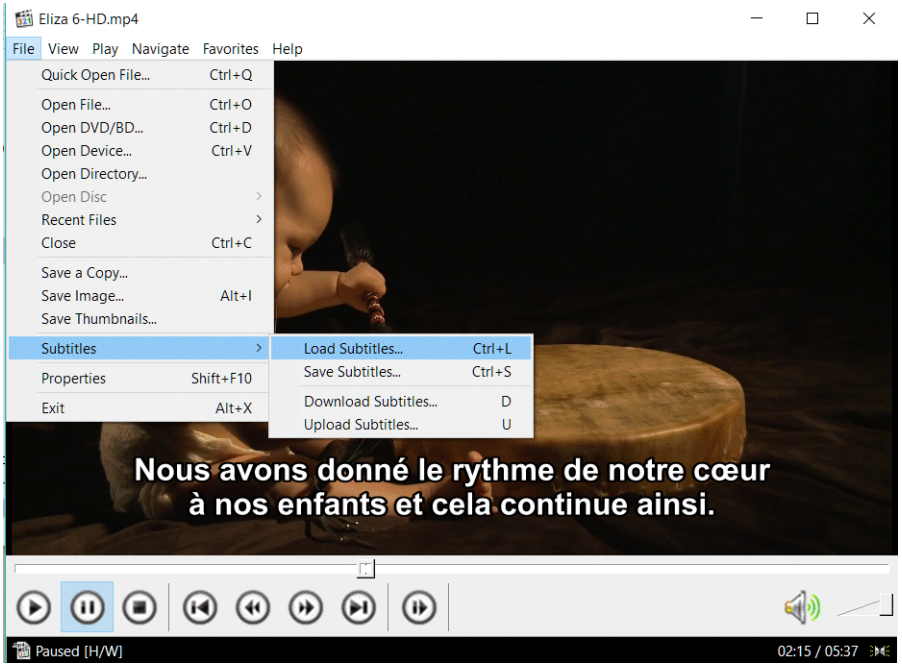


Figure 263: Opening a file in *MPC* through the menu

You can configure the style and other settings under **Play> Subtitle Track**.

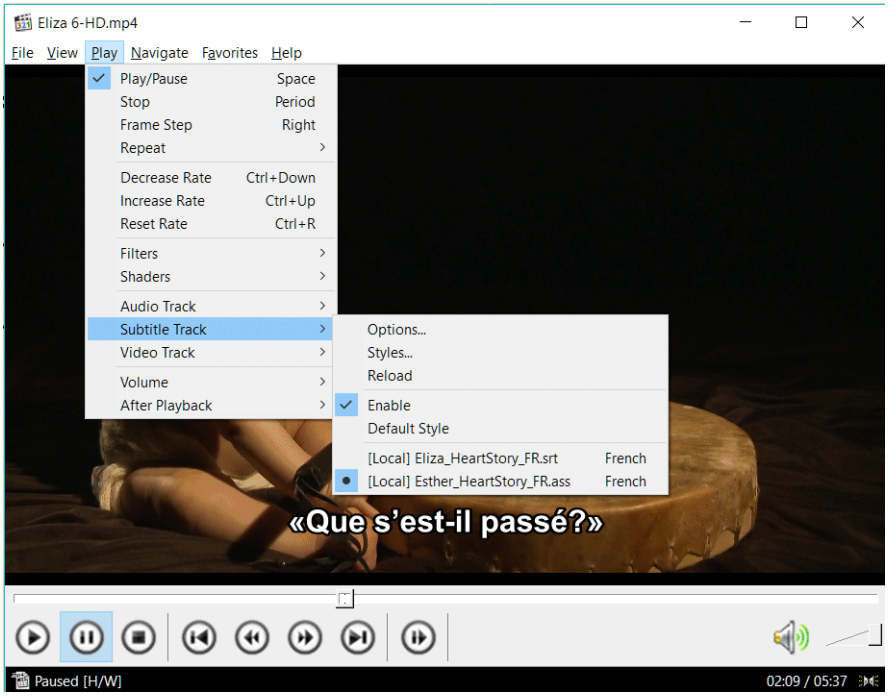


Figure 264: Configuring the settings of the subtitle track in *MPC*

Under **Options** you find general options and the default style, which you can change.

Under **Styles** you can edit and manage more than one preferred style, as well as the default one. In fact, if you change the look of your subtitles using these options in *MPC*, you also have the option of saving or even converting your subtitles under **File> Subtitles> Save Subtitles**. For instance, if you have an SRT file, you can configure the font and the positioning of the entire subtitle file and then save it as ASS, thus retaining these visual settings in the resulting subtitle file.

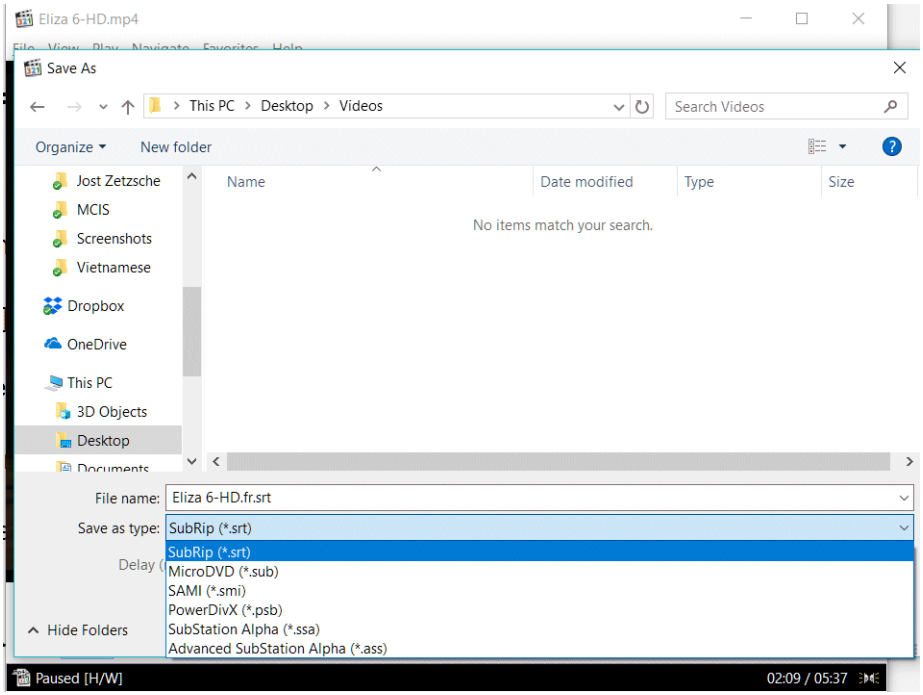


Figure 265: Saving the subtitle file into a different format in *MPC*

In *VLC*, apart from drag & drop, you can also select a subtitle file from the **Subtitle** menu and then select **Add Subtitle File**.



Figure 266: Adding a subtitle in *VLC*

General settings for subtitles in *VLC* are under **Tools> Preferences**> **Subtitles/OSD**.

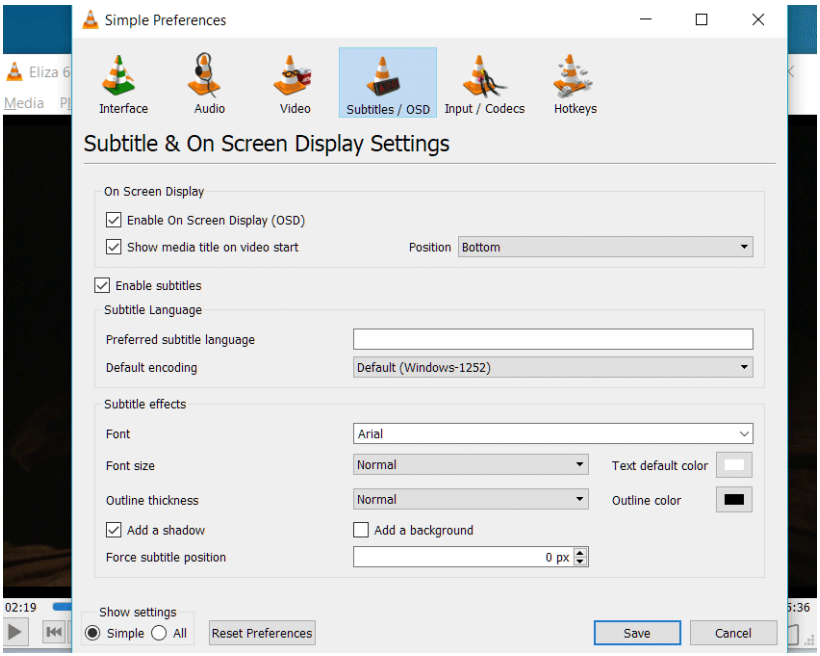


Figure 267: **Subtitle & On Screen Settings** in *VLC*

VLC's subtitle settings are more limited than *MPC's* and it doesn't support converting or saving subtitle files. On the other hand, it supports "hard-coding" (see page 400).

DVD Ripping, Audio/Video Conversion, and Embedding Subtitles ("hard-coding")

A number of programs are able to perform many types of conversions for audio and video files. Conversion may be necessary for a number of reasons. Commercial DVDs or Blu-Rays, for instance, cannot be opened directly into editing, dubbing or subtitling software and are composed by several very large files. In order to use them in a program, edit, share with a team, etc., you must extract them from their original disk into a compressed digital file (aka "ripping").

Likewise, some programs you may have to use accept only certain types of video files, so you may need to convert a work file to be able to use it.

The programs below perform many of these conversions and also allow adding a subtitle file, so they allow for simple ways to embed (hard-code) subtitles into a film. These are not high-end, professional tools, but they get the job done quite well and are free (or almost free) and easy to learn.

Audiovisual files are complex, and occasionally your favorite converter might not be able to render a file correctly and you can try a different one. That's why it might be a good idea to have access to more than one.



Each time a video is converted, some quality is lost. So if you use any of the following programs to embed subtitles while converting a video file, the resulting file will have a slightly lower video and audio resolution. That's why it's a good idea to first convert the video to a very high-quality one—preferably "lossless" or "uncompressed," if your computer can handle the huge resulting file that can easily reach several gigabytes, or at least with a higher video and audio bit rate—and then use this file to embed the subtitles. The final subtitled file can be converted again, now reproducing the same audio and video encoding as the original.

AVS4YOU Video Converter

I use the *AVS4YOU Video Converter* (see avs4you.com/AVS-Video-Converter.aspx) on a daily basis. It can read any type of file, from MPG to Blu-Ray, and convert it into any other type. The tool also gives a great deal of control in customizing the presets, with a very user-friendly interface.

All you need to do is select the file or DVD from the **File** menu or drag the video file into the **Input File Name** box and select the output format, which are arranged into **Formats**, **Devices** and **Web**.

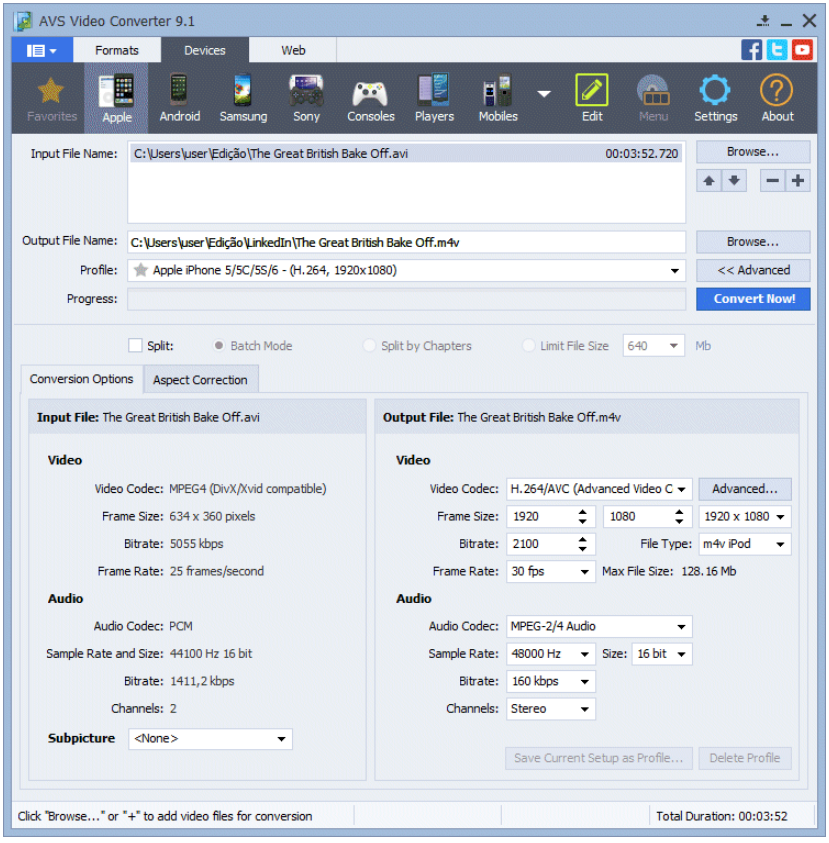


Figure 268: Advanced conversion screen of the *AVS4YOU Video Converter*

For instance, if you are adding subtitles that have been timed based on the input file, which has a frame rate of 25 frames per second, you would want to change the frame rate of the output file to 25 frames per second, so the subtitles would not be out of sync with the audio. This can easily be done through the controls on the right side of the screen.

Although it's not a full video editor, it includes some editing functions. You can change the aspect ratio of the film in the **Aspect Correction** tab (see image above), or you can select the **Edit** mode to trim the film, add various effects, extract the audio file, etc.

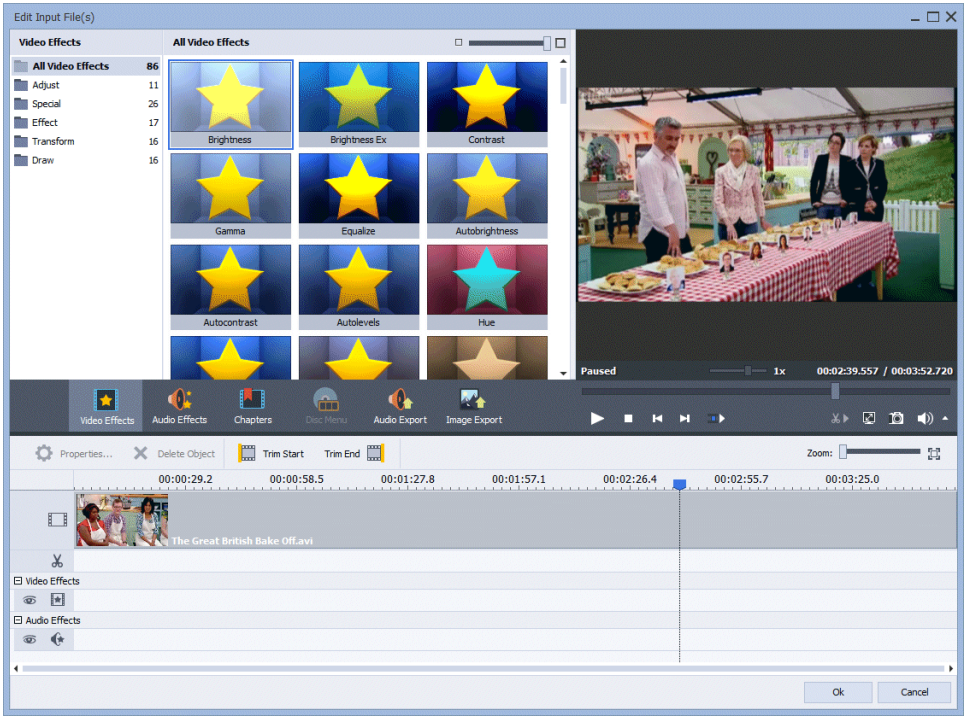


Figure 269: Edit mode of *AVS4YOU Video Converter*

If you have a previously prepared subtitle file (SRT or SSA), it must have the same name as the video file, except for the extension. Then, on the advanced screen of the video editor, select the subtitle track under **Subpicture**.

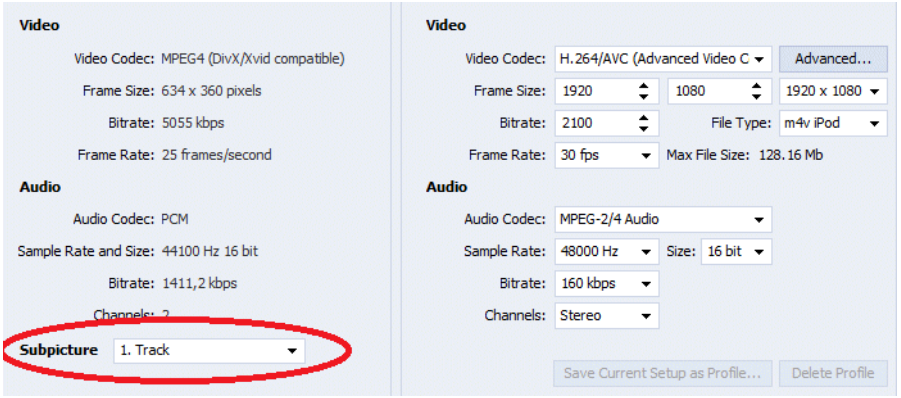


Figure 270: Selection of subtitle file in *AVS4YOU Video Converter*

The appearance of the subtitles can be adjusted in **Settings> Subtitles**, where you can select the font type and size, color, border, and top or bottom alignment.

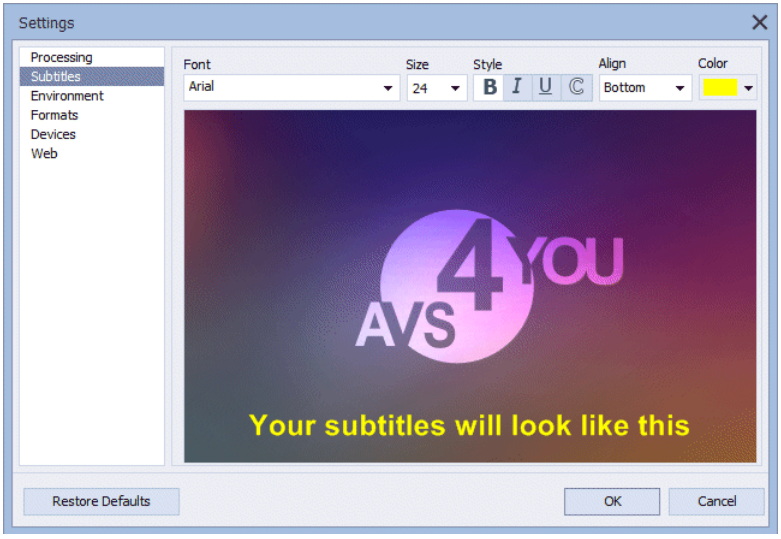


Figure 271: Subtitle formatting in **Settings** of *AVS4YOU Video Converter*

When you're all set, select the output file name and location and click the **Convert Now** button.



The AVS4YOU support site (avs4you.com/Guides) is a great resource with clear user guides and video tutorials.

VLC Media Player

The VLC Media Player (see videolan.org) also supports video file conversion, which includes embedding subtitles. If your subtitle file does not include visual formatting, such as SRT, be sure to first configure how the subtitles will look over the video under **Tools> Preferences**. If it does include visual formatting, such as ASS, *VLC* will render them according to the subtitle file information. It's always good to first check whether everything looks good using soft-coding (see page 388).

When you're ready to embed the subtitles, start by opening *VLC* (without opening the video file), and select **Media> Convert/Save> File> Add** to select the video file. Check **Use a subtitle file** and click **Browse** to select your subtitle file. Then click the **Convert / Save** button.

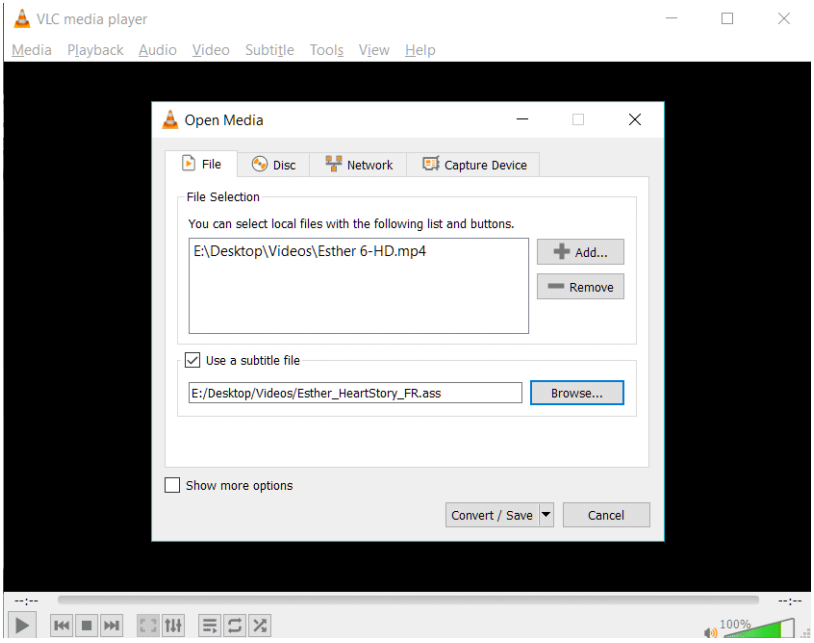


Figure 272: Selecting the subtitle file

On the following screen, select the output video profile, i.e., the file extension, its audio and video codecs, and click the wrench icon to access the advanced settings for the profile.

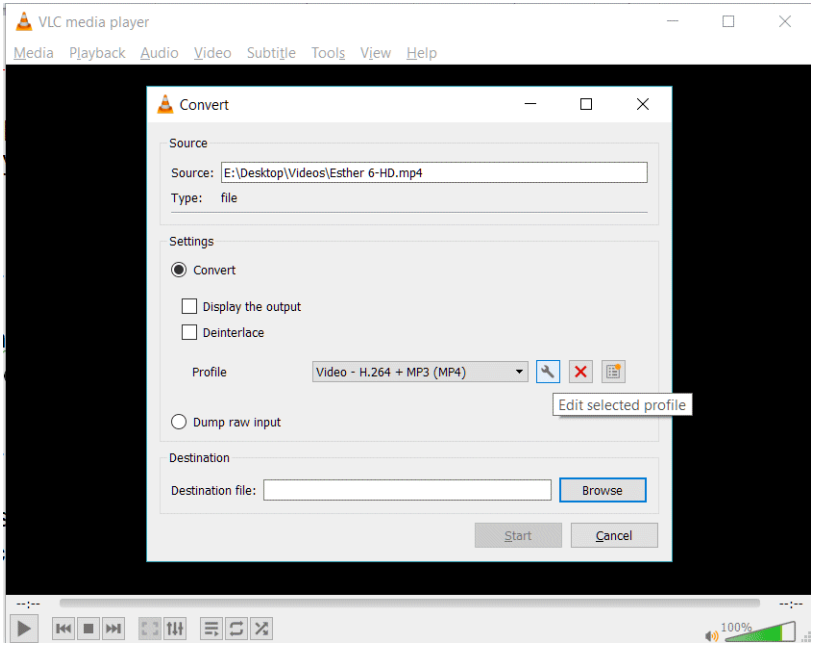


Figure 273: Conversion settings in *VLC*

In the following dialog are four tabs where you can input more specific file, audio, video, and subtitle settings. In the **Subtitles** tab, ensure that both check boxes are checked. For codecs I use "T.140," but you can try both if for some reason it doesn't work for you. Then click **Save** to start converting/embedding.

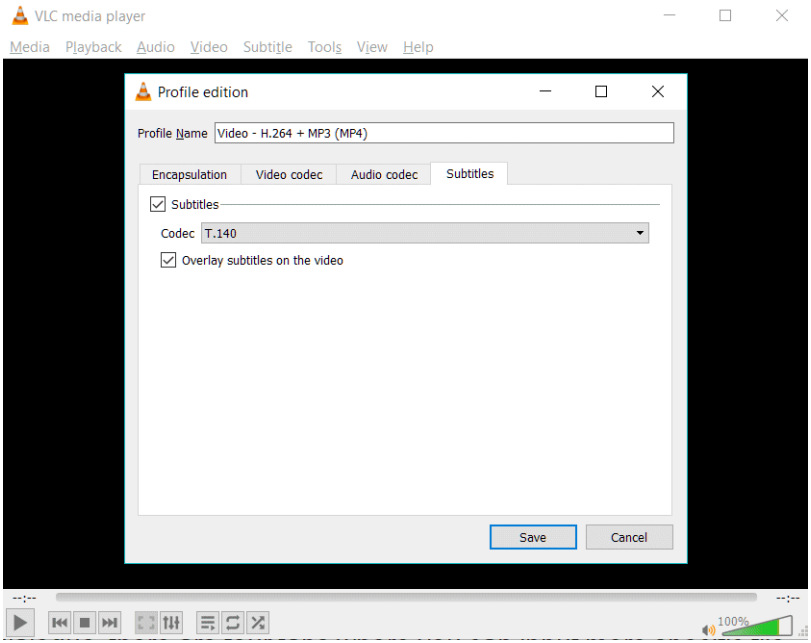


Figure 274: Selecting a codec

Go back to the previous screen, browse to the location, and name the output file. Then click **Start**.

Xvid4PSP 5.0

Xvid4PSP is a video converter with some editing functions. It is currently at version 7.0 (see winnydows.com); however, for some mysterious reason, subtitle embedding is only supported up to version 5.0, when the developers did a major overhaul. This is why many translators who work with subtitles stick with "good old" *Xvid4PSP 5.0* (see xvid4psp.en.softonic.com, or for more information, see videohelp.com/software/XviD4PSP).

Technically speaking, *Xvid4PSP 5.0* uses AviSynth code (as do the *Media Player Classic* and *Aegisub*), which is used to add multiple types of subtitle files as layers to video files. Because they share this code, subtitles configured using *Aegisub* (such as an ASS file with a customized style for the font, positioning, and so forth) will be accurately rendered in these other converters and media players.

The interface of *Xvid4PSP 5.0* may not be as simple and friendly as that of Format Factory or Any Video Converter, but it gives the user more control and complex options.

To simply convert a video, all you have to do is open the video file and then select its output format.

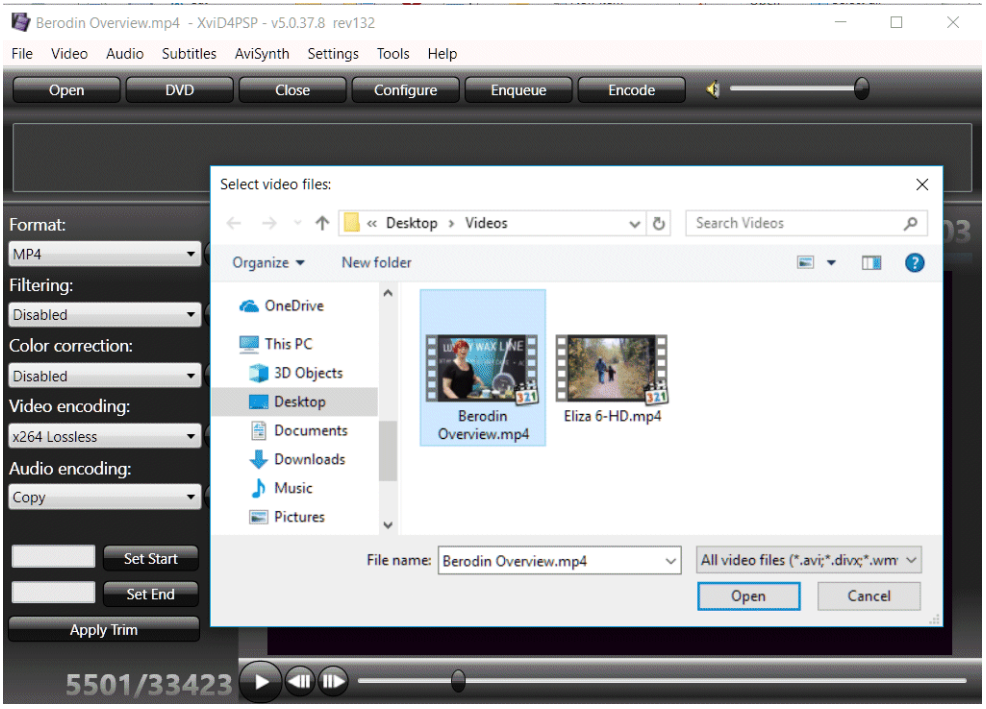


Figure 275: Opening a video file in *Xvid4PSP*

There are many video formats to choose from. When I am embedding subtitles, I never use filters or color correction, and I choose "lossless video encoding" (this decompresses the video to preserve full quality) and "copy audio encoding" (to preserve the original audio settings). As mentioned above, I usually decompress and then recompress the video file to avoid losing video/audio quality. All the options can be further configured by clicking the gear icon.

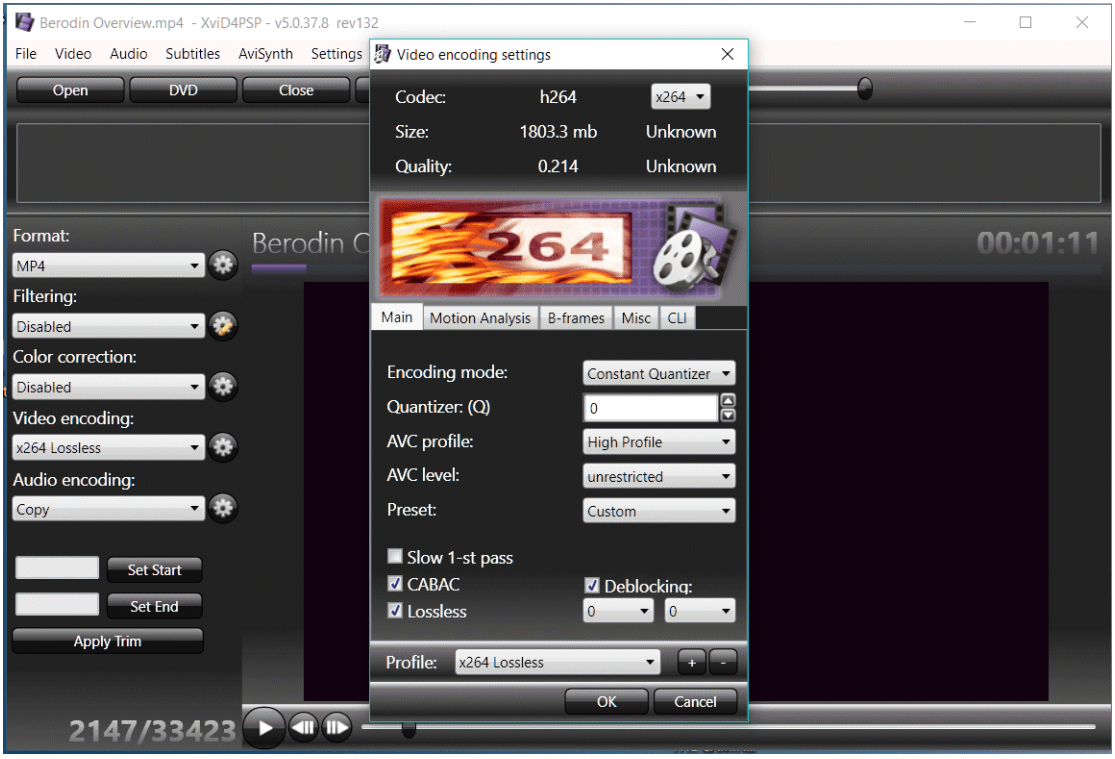


Figure 276: Video encoding settings in *Xvid4PSP*

To embed subtitles, select the subtitle file using **Subtitles> Add**. Many file types will work, including ASS/SSA, which will contain style information in the converted video.

Finally, to encode just click **Encode** below the menu bar and select the new video's file name and location.

Format Factory

Format Factory (see format-factory.en.softonic.com) is a free and simple DVD ripper and video and audio converter with very simple editing tools.

Simply drag the video file onto the **Input File** box (or select the file or DVD from the **File** menu) and select the output format in the pop-up dialog.

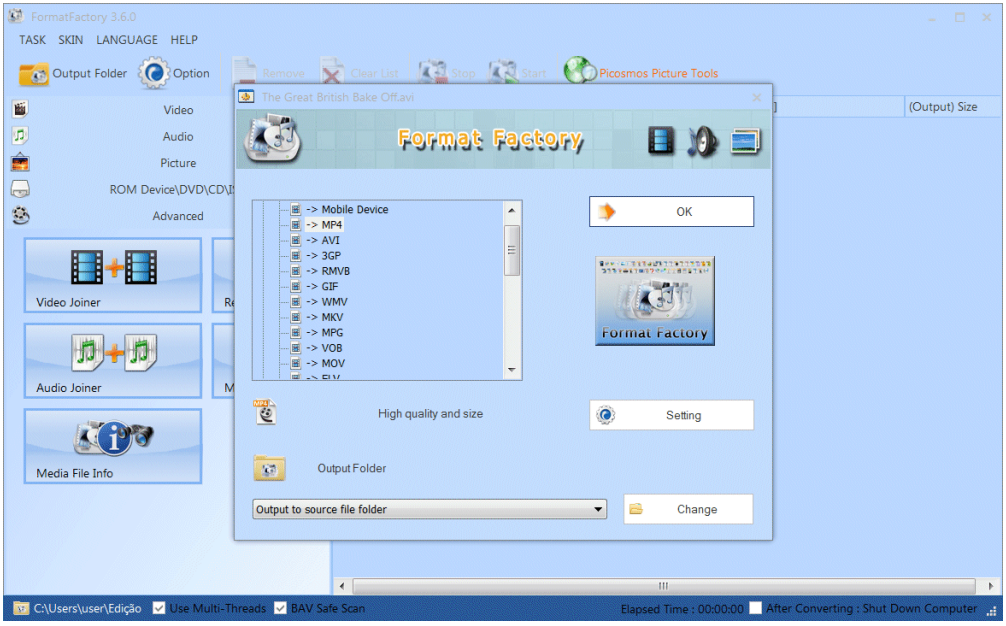


Figure 277: *Format Factory* output selection dialog

After the file has been added, you can right-click it and select **Output Settings** for more advanced settings. There you can adjust video and audio settings, as well as select a subtitle file. With this tool the subtitle file doesn't have to have the same name as the video file.

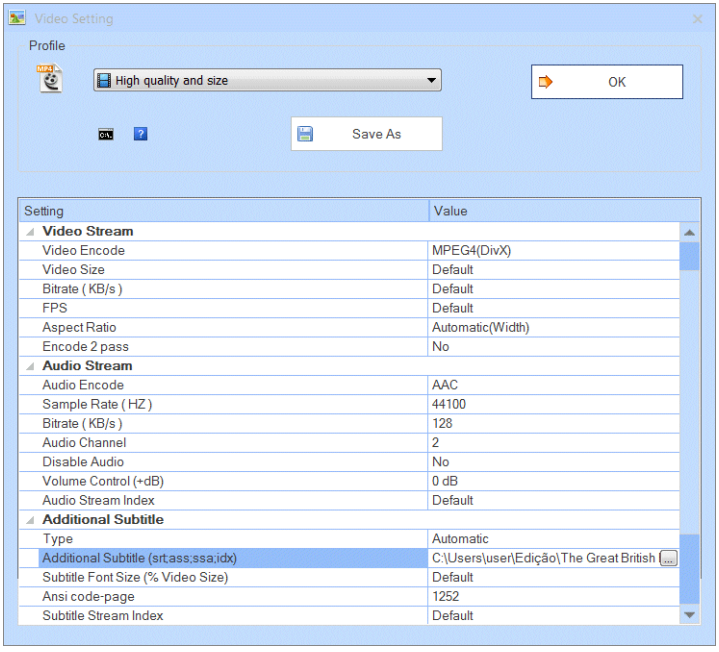


Figure 278: Output settings (advanced options) in *Format Factory*

Under **Option** (from the main menu) you can select the subtitle font and color. Subtitle positioning cannot be changed.

Once you're done with your settings, click **Start** to encode the file.

Any Video Converter

Any Video Converter (see any-video-converter.com/products/for_video_free) is very similar to the other programs in this section. You can open a DVD, drag a video file into the program, or select a URL to download a video directly from the Internet.

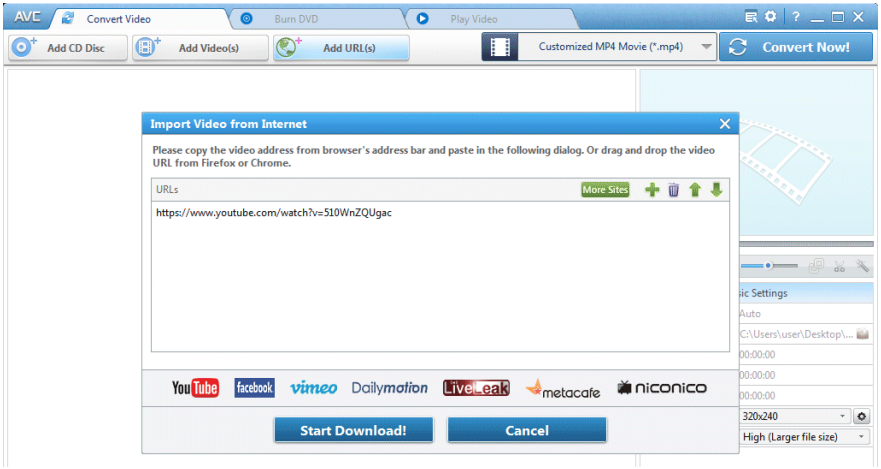


Figure 279: Using *Any Video Converter* to download a video file from a website, such as YouTube.

You can see the settings of the original video under the file name. On the top right, you can select the output format. To the right, below the video preview, you can select basic and advanced audio and video options. If you want to add a subtitle file, click the menu behind **No Subtitle** (under the added file) and select the subtitle file (which also doesn't need to have the same name as the video).

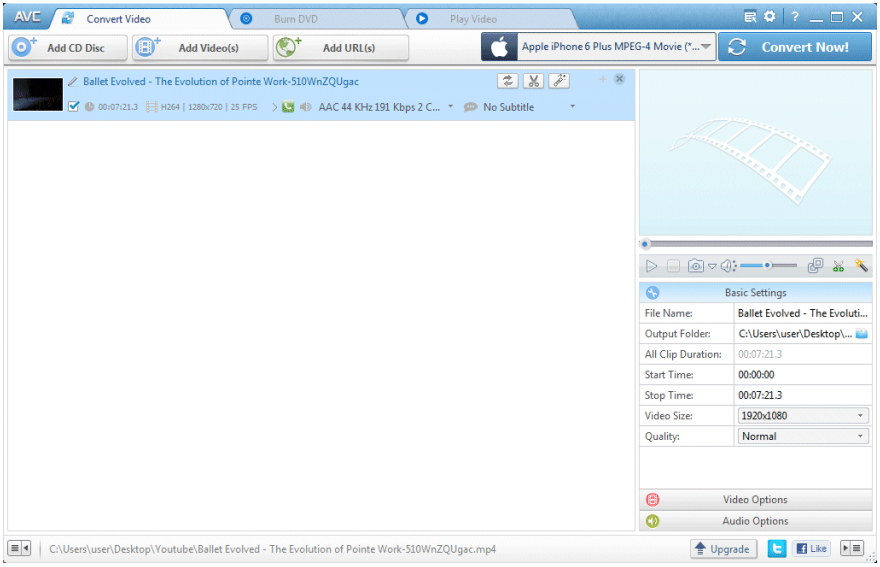


Figure 280: Main menu and options in *Any Video Converter*

Various options, including the appearance of subtitles, can be changed under **Options**, which is accessed through the tiny engine icon above the **Convert Now** button.

For the subtitles, you can select the encoding (for different writing systems), screen positioning from bottom to top of the screen using a percentage scale, subtitle size in a percentage scale, and the font. You cannot change the color of the subtitles.

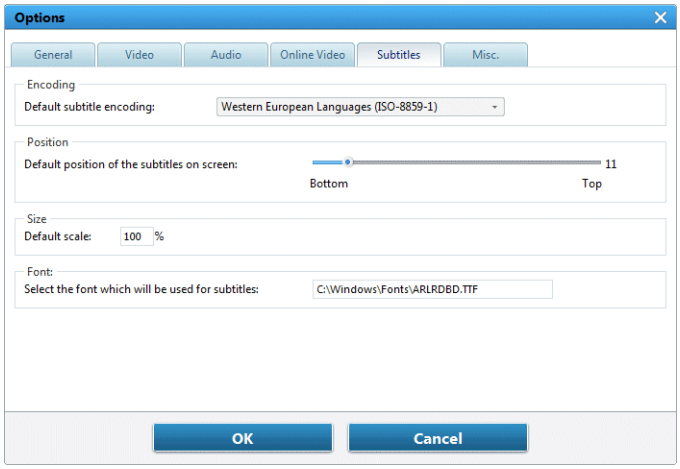


Figure 281: Subtitle formatting options in *Any Video Converter*

While this program might be less powerful for adding subtitles, its conversion is quite reliable, with the added advantage of downloading videos from the web. In fact, this option alone is worth installing this tool.

Subtitling Tools

There are many online and offline programs and tools for subtitling. Following are just two popular solutions.

Subtitle Workshop

The free and user-friendly *Subtitle Workshop* (see subworkshop.sourceforge.net) is meant for all types of operations involving subtitle files and is especially good for format conversions, since it supports over 60 different formats.

In *Subtitle Workshop* you always work with two files open at the same time: one is the video and one is the subtitle file. The video is opened and managed from the **Video** menu (or the file can be dragged into the program to be opened), while the subtitle file is created, opened, saved and edited using the **File** and **Edit** menus.

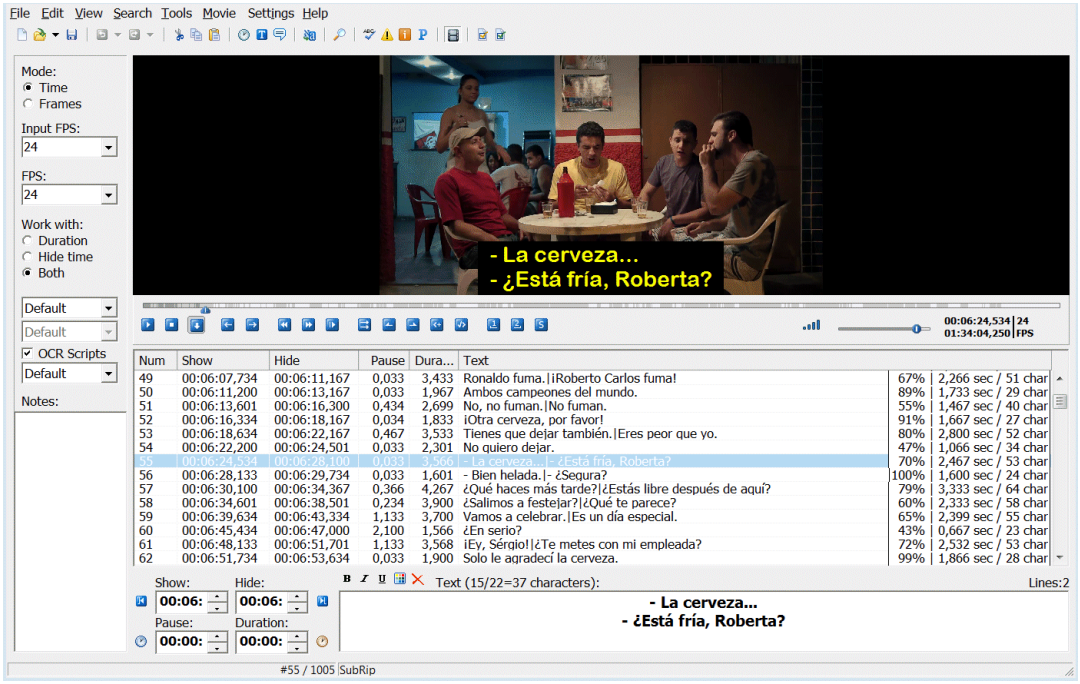


Figure 282: *Subtitle Workshop's* interface

To save a new subtitle file, select **Save** and double-click the desired format in the screen that is displayed. The file formats shown are the video editing programs that will import the corresponding subtitle file. This can be counter-intuitive, as often our clients tell us the file extension rather than the software they use. Commonly used formats are ASS, SSA, and SRT. When in doubt, use SRT, currently the most common format and imported by almost all editing software.

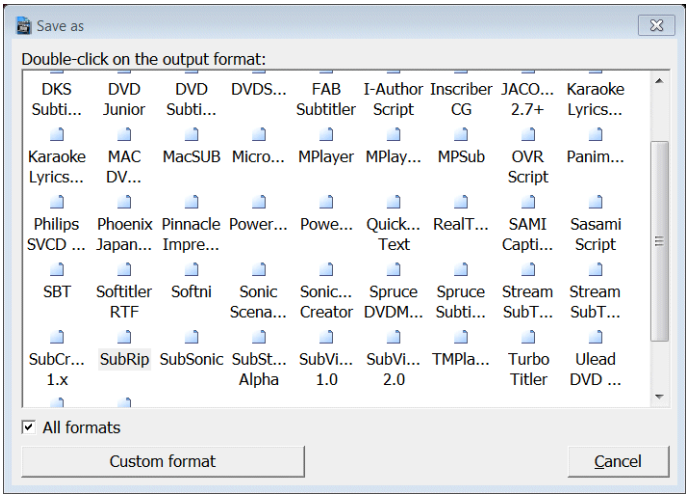


Figure 283: Format options when saving a subtitle file in *Subtitle Workshop*

To create subtitles from scratch, add the subtitles one by one by pressing the INSERT key and typing the text of each subtitle in the box at the very bottom of *Subtitle Workshop*. The program shows the number of characters per line and in total for each subtitle, as well as its show (start) and hide (end) times.

In the subtitle list below the video, you can also see the pause (i.e., the interval between subtitles) and the duration of each subtitle, as well as the reading speed to the right of the text. All of this and many more options can be customized in the general settings and under **Information** and **Errors** (the two last icons just above the video).

The best way to sync the subtitles with the video is to use shortcuts while the video is playing:

- Video controls:

CTRL+SPACE: Toggles Play/Pause

CTRL or ALT + < or >: Rewinds or forwards two different time increments

- With the video running:

  ALT+C: Sets the moment when the selected subtitle should appear (the software sets the current time as that subtitle's initial "show" time)

  ALT+V: Sets the moment when the selected subtitle should disappear (the software sets the current time as that subtitle's final "hide" time, then moves to the next subtitle and sets its initial time immediately after the previous one)

- After timing all subtitles:

  Edit> Timings> Set delay (or CTRL+D): Sets the timing to be subtracted from our natural delay when setting the times (on average, between 200 and 500 milliseconds)



*The software offers many text formatting options, such as italics, bold, underline, as well as font, size, color, shadow, etc. (found in **Settings**). Note that most subtitle formats accept only very limited and specific types of tags, which are not necessarily the ones used by Subtitle Workshop, so do not simply rely on these tags without checking which kinds of tags are accepted by the subtitle format you're using or asking your client about it.*

Also, font type, size, color, etc. are not transferred to the subtitle file and will not affect the final edited version. They are only for your own comfort when working. The same applies to the subtitle's position on the screen. Subtitle Workshop superimposes your subtitles onto the video file just for simulation purposes. In most cases, these visual configurations are determined in the video editing software. The main purpose of Subtitle Workshop is to create subtitles with adequate line lengths, reading speed and time codes.

Subtitle Workshop also has a "Translator mode" (Under **Edit> Translation> Translator mode**), which is useful when you have a pre-existing subtitle file in another language and want to use it as template. It opens a new column for the translated version, based on the spotting (subtitle segmentation) and timing of the template. When saving in Translator mode, there is an option to save the project (i.e., the project in translator mode, with both versions), only the original, or only the translated version.

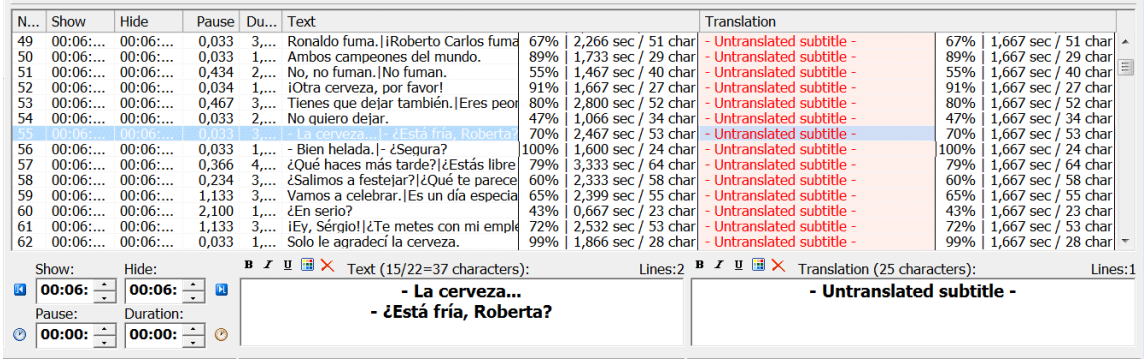


Figure 284: Translator mode in *Subtitle Workshop*

This program offers a large number of resources and settings which would be impossible to cover here. There is certainly a learning curve, and even the most tech-savvy require many hours of training to achieve professional results. Check the many available manuals and tutorials for more information, and if you're interested in learning the most efficient and quality-oriented processes for film subtitling, it would be advisable to take a course.



It's important to bear in mind that this program deals only with subtitle files—meaning text files. It does not edit video, nor does it embed the subtitles onto the videos. This is done with video editing tools (see DVD Ripping, Audio/Video Conversion, and Embedding Subtitles ("hard-coding") on page 395).

Aegisub

Aegisub (see aegisub.org) is a free, cross-platform open source tool for creating and modifying subtitles.

It is a complex tool, with the ability to do much more in terms of editing than *Subtitle Workshop* (see page 410). On the other hand, it's considerably less user friendly, and I find it less comfortable and efficient for typing and timing, which is what audiovisual translators do the most. The text box is rather small, and it uses tags even for simple things such as line breaks. However, many translators prefer working with the audio wave, which *Aegisub* includes.

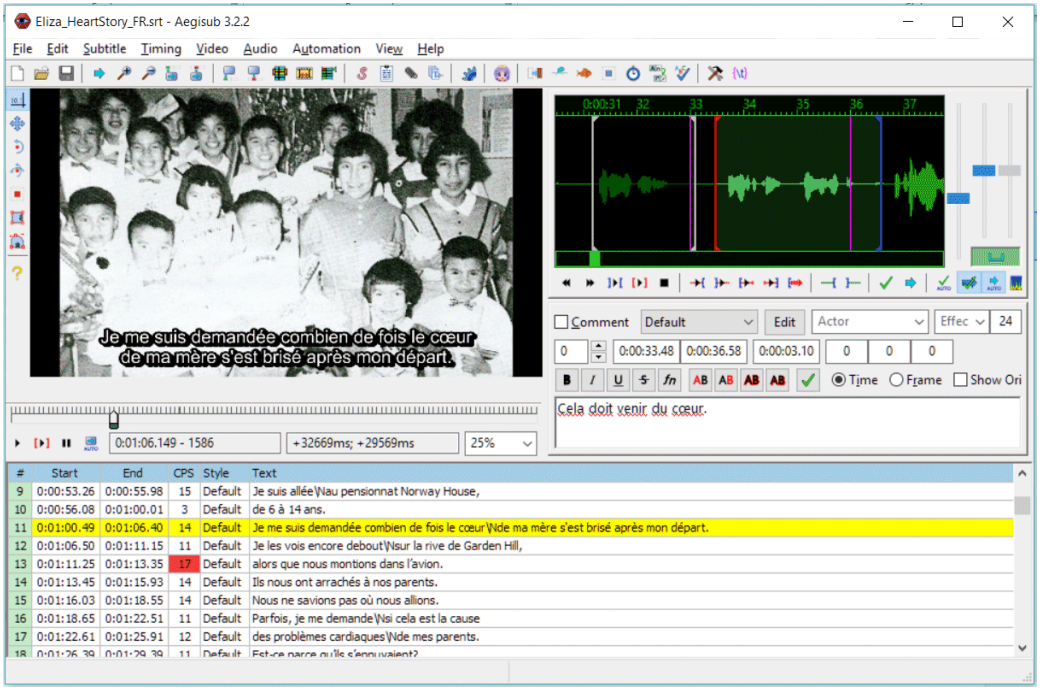


Figure 285: *Aegisub's* interface

Video files and existing subtitle files can be opened by dragging them into the program. The **File** menu manages subtitle files: **New**, **Open**, **Save,** and **Export**, among others. The **Video** menu opens, closes, and manages video files. Then there are the usual text and subtitle editing tools, such as **Cut**, **Copy & Paste**, **Find & Replace**, **Split**, and **Join**. The most common of these are also available by right-clicking the subtitles on the list below the video.

The timing can be set by either selecting each subtitle in the list under the video preview and then dragging the mouse cursor on the audio wave to determine its duration, or using shortcuts C and V (making sure the cursor is not within the text editing box) and then pressing Enter. The in and out times are respectively called "lead in time" and "lead out time" in *Aegisub*. In the **Timing** menu, **Shift Times** can be used to discount or add time to the subtitles, which is usually done to compensate for a manual delay while timing.

To translate an existing subtitle file into another language, *Aegisub* provides a **Translation Assistant** via the **Subtitles** menu.

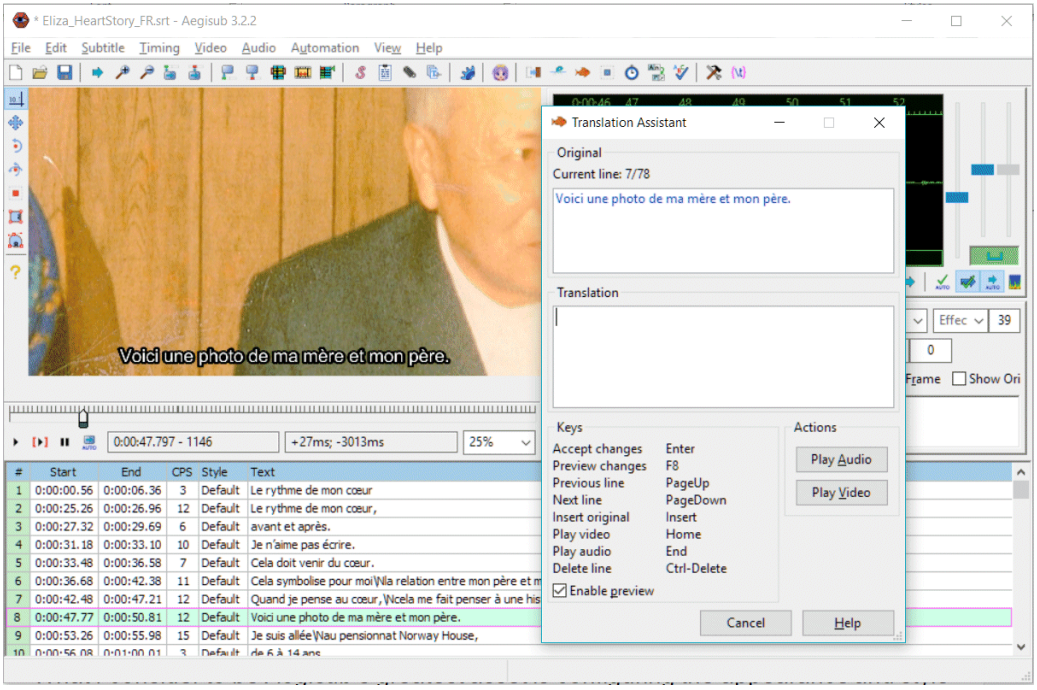


Figure 286: *Aegisub's* Translation Assistant

You can type the translation in the bottom box, play the audio or video for that segment, then move up or down the list using the PAGEUP and PAGEDOWN key.

Aegisub saves subtitles as ASS files. A few other formats are available through **File> Export Subtitles**, but not as many as *Subtitle Workshop* offers.

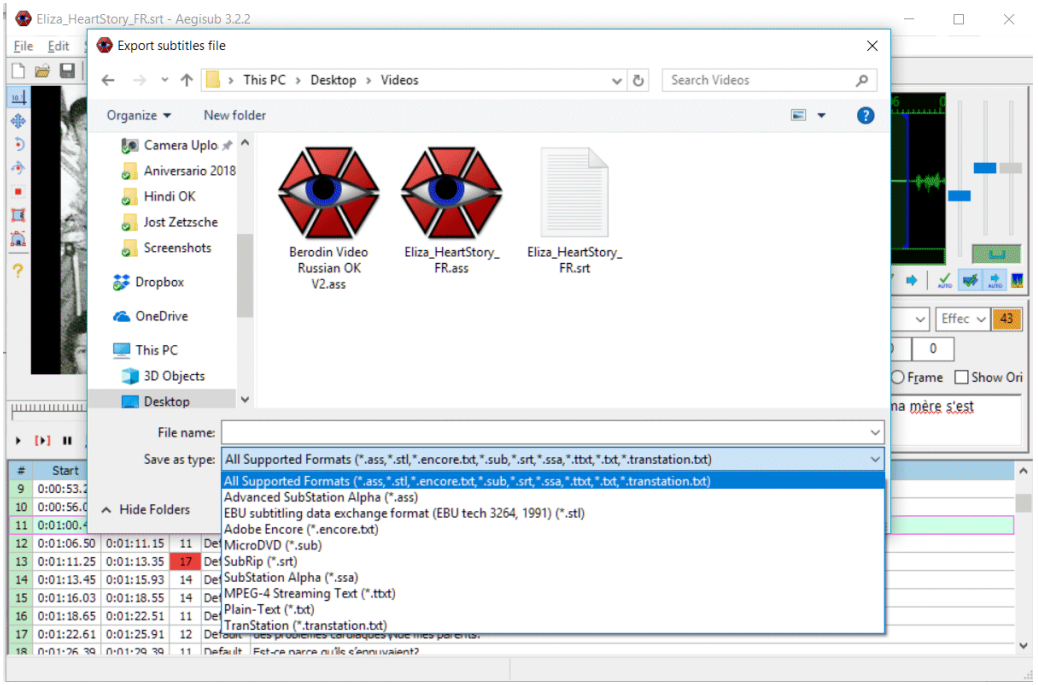


Figure 287: Exporting a subtitle file in *Aegisub*

What I consider to be *Aegisub's* greatest asset is configuring the appearance and style of the ASS subtitles. In **Subtitles> Style Manager**, you can modify the default style and create as many customized ones as you like (according to each client's preference, for instance). These are then stored by the program, and can be moved to the current project as needed.

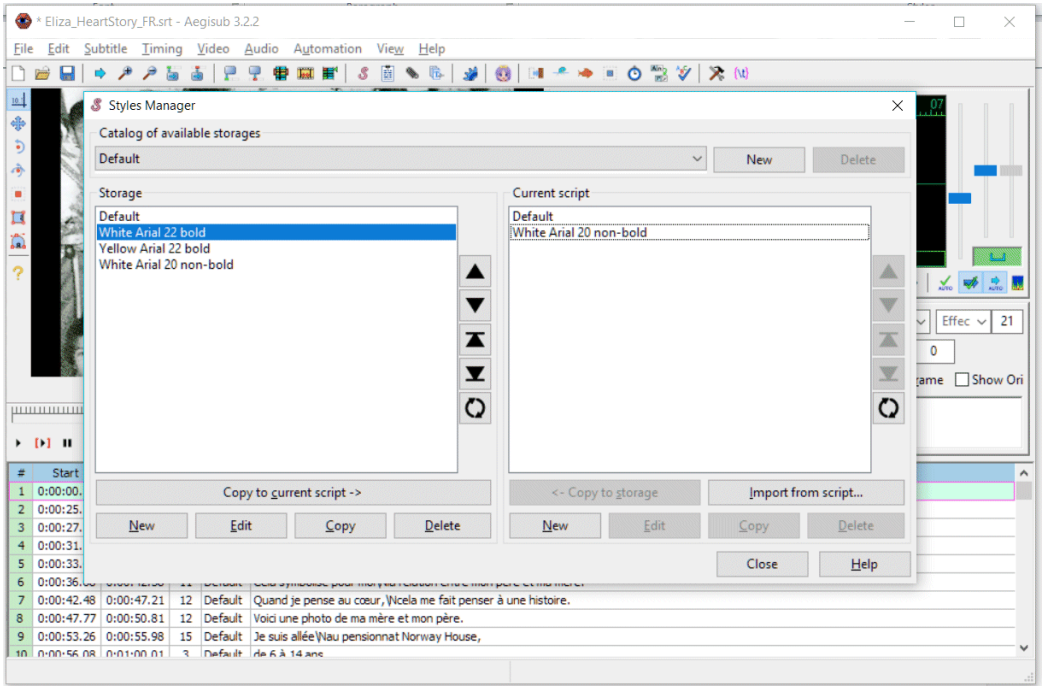


Figure 288: *Aegisub's* Styles Manager

By double-clicking a style or selecting it and clicking **Edit**, you can customize the font, positioning of the subtitles, encoding and many other details, which are all encoded in the ASS file.

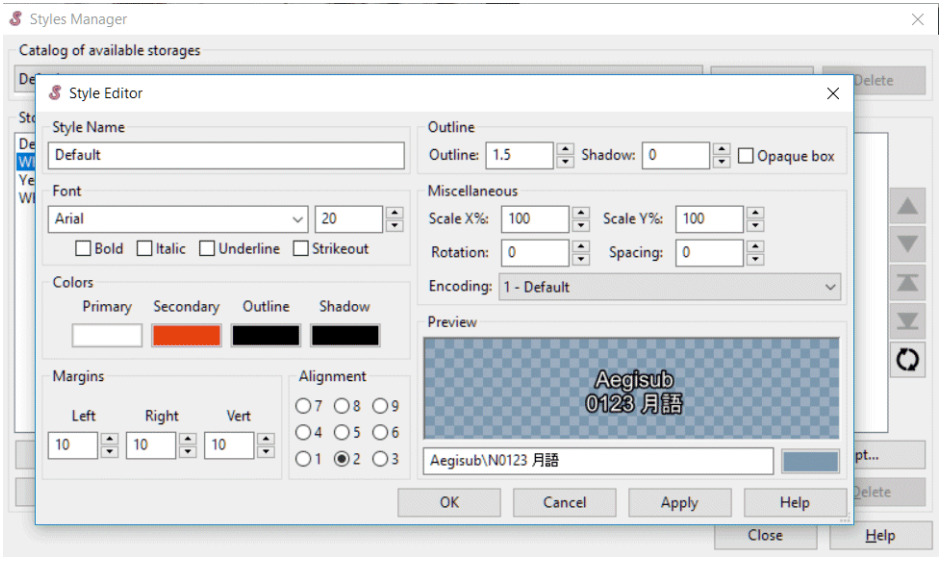


Figure 289: *Aegisub's* Style Editor

Personally, I prefer to create my SRT files elsewhere and then open them in *Aegisub* to save them as ASS to hard-code the subtitles.

FluencyTranscription

If you are using the translation environment tool *Fluency*, you already have an advanced video transcription tool installed: *Fluency Transcription* (see westernstandard.com/Fluency/Transcription.aspx) You can either start it from within *Fluency Now* (**Tools> Transcription / OCR**) or as a standalone application. Once launched, you can open any image-based PDF, most graphic formats (JPG, GIF, BMP, PNG, TIF), many sound formats (WMV, MP3, WAV, WMA) or video formats (AVI, MPG or WMV), have them displayed or played on one side of the screen and transcribe on the other side (time codes are entered with the keyboard shortcut CTRL+T).

Once the file is readily transcribed, you can open the file in *Fluency Now* for translation or—in the case of subtitles for a video—save it as an SRT file.

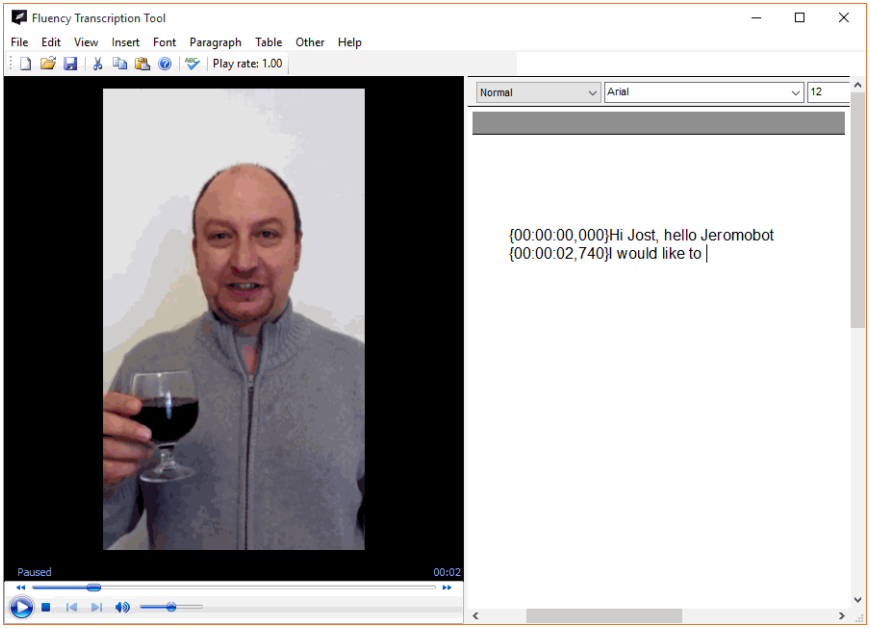


Figure 290: Creating subtitles in *Fluency Transcription*

Amara

Amara (see amara.org) is a non-profit, online collaborative platform/community for video subtitling. This is the platform used by *Ted Talks* for volunteers to translate the talks into dozens of languages (see ted.com/participate/translate).

As a user, you can volunteer to translate videos uploaded to *Amara*, or you can upload your own. If you want to upload videos to subtitle them in *Amara*, you can make the videos publicly available or work in a private space with restricted access. The first option is free, while the second is paid. Needless to mention, *Amara* is not a good choice to use as a subtitling tool for your clients that require confidentiality.

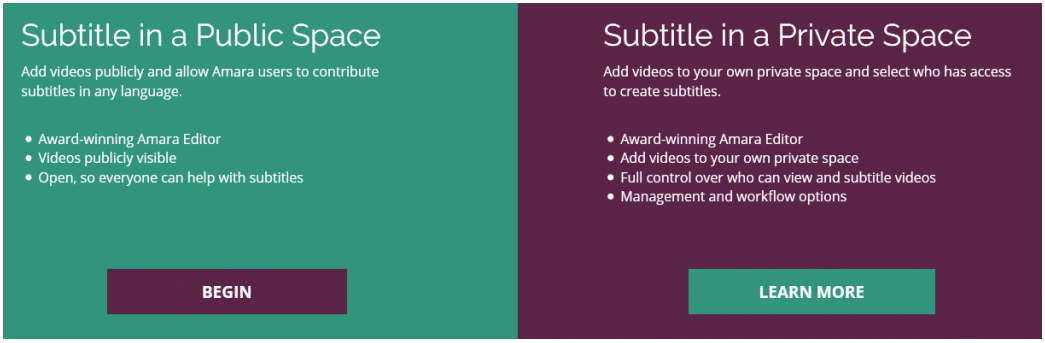


Figure 291: Uploading files in *Amara*

Amara also offers two subtitling platforms: one that is public, collaborative, and free; and a professional one. This is done through their *Amara On Demand* service (see amara.org/en/purchase-subtitles), and freelance transcriptionists and translators can apply to be service providers (see amara.org/en/recruitment).

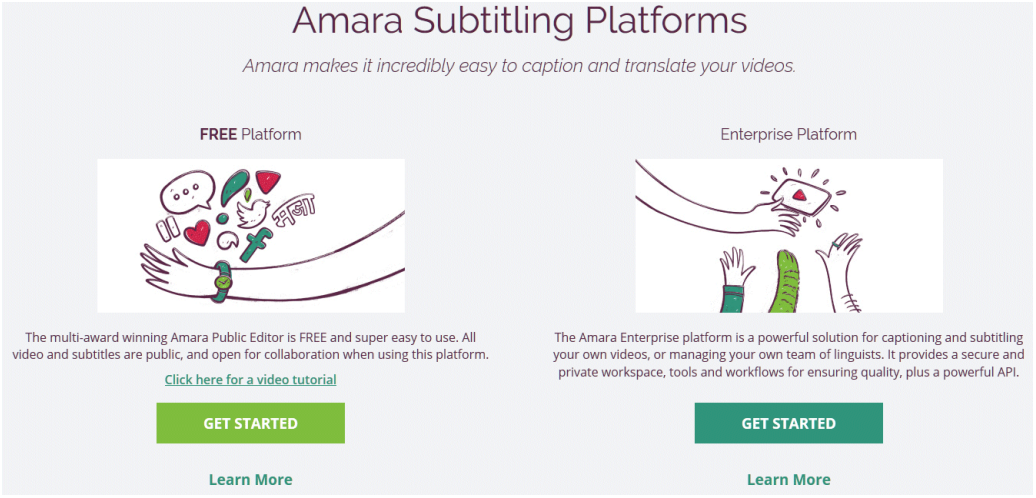


Figure 292: *Amara's* subtitling platforms

The online platform is very user friendly. The subtitles are typed into text boxes, which can be manually dragged over a timeline to be synced to the video. A tutorial is provided in the "Subtitling Platform" section. You can learn to use it and get some practice by volunteering to translate videos uploaded by the several organizations that use *Amara*.

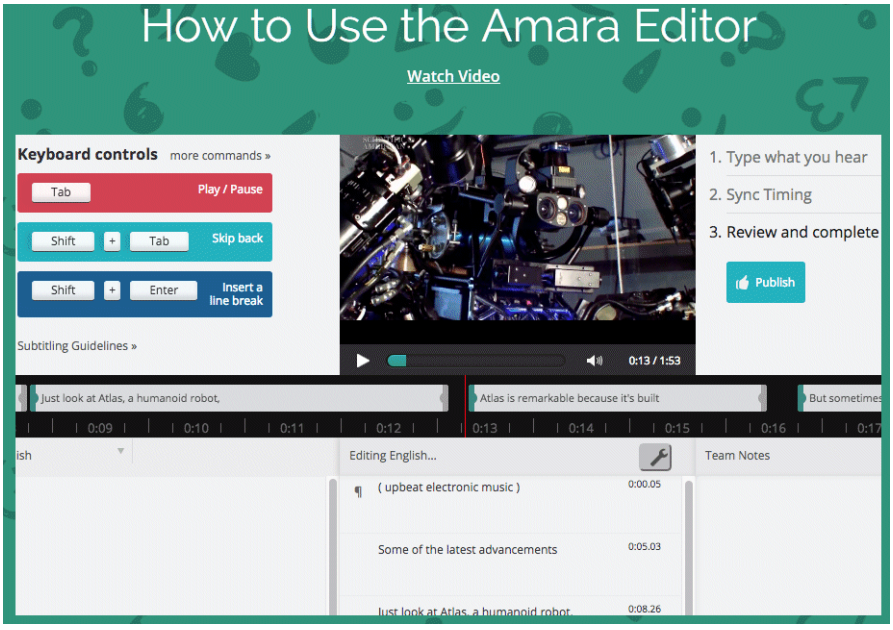


Figure 293: Tutorial for the *Amara Editor*

The nicest thing about *Amara* is its ability to create a very simple and friendly subtitling environment and to encourage the idea of a community of volunteer translators. It's a lovely idea for non-profits or for independent, low-budget projects. On the other hand, from a professional point of view, it will not necessarily prepare you to work with more sophisticated tools or give you all the skills required by quality-focused clients. Still, it's a great place for translators who have never tried their hand at subtitling to gain a feel for how it's done.

SubtitleNEXT

Another too, SubtitleNEXT (see profuzdigital.com) was the subject of a helpful write-up by Fernanda Brandao-Galea in the Resources section of the *ATA Chronicle*. You can find it here: ata-chronicle.online/highlights/audiovisual-translation-preparing-for-this-market.

Translating Subtitle Files with Translation Environment Tools

In the last few years, a number of translation environment tools have started to support the translation of subtitle files. Here is a (most likely incomplete) list of TEnTs that support subtitle files and the supported format(s):

- *Fluency* supports SRT and ASS files
- *OmegaT*, *Wordfast Pro,* and *Lilt* support SRT files
- *Smartcat* supports SRT and TTML files
- *Cafetran* supports ITT and SRT files
- *Déjà Vu X* and above supports STL files
- Via its *Studio Subtitling* app (see appstore.sdl.com/language/app/studio-subtitling/978/), *Trados Studio* supports SRT, webVTT, STL, SBV, and ASS/SSA files and a synchronized display of the video that is being subtitled

- *Star Transit NXT*, *Wordbee* and *memoQ* support the translation of SRT files with a synchronized display of the video that is being subtitled

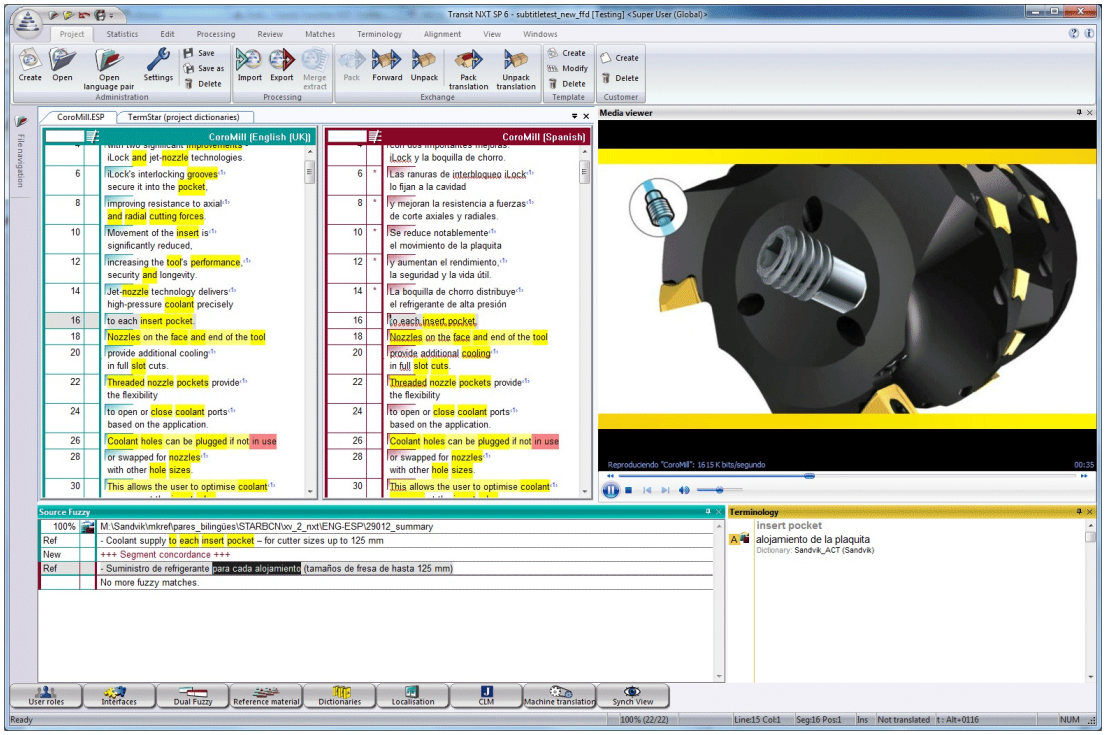


Figure 294: *Star Transit* with STL file and corresponding video (image source: star-spain.com)

Sound Editing

Audacity

Audacity (see audacityteam.org) is a free, open-source and cross-platform software for sound recording and editing. The interface is very user-friendly and self-explanatory, though it has a lot more resources than you will ever need. This is not surprising since this is a tool that is widely used for tutorials, podcasts and semi-professional voiceover tasks.

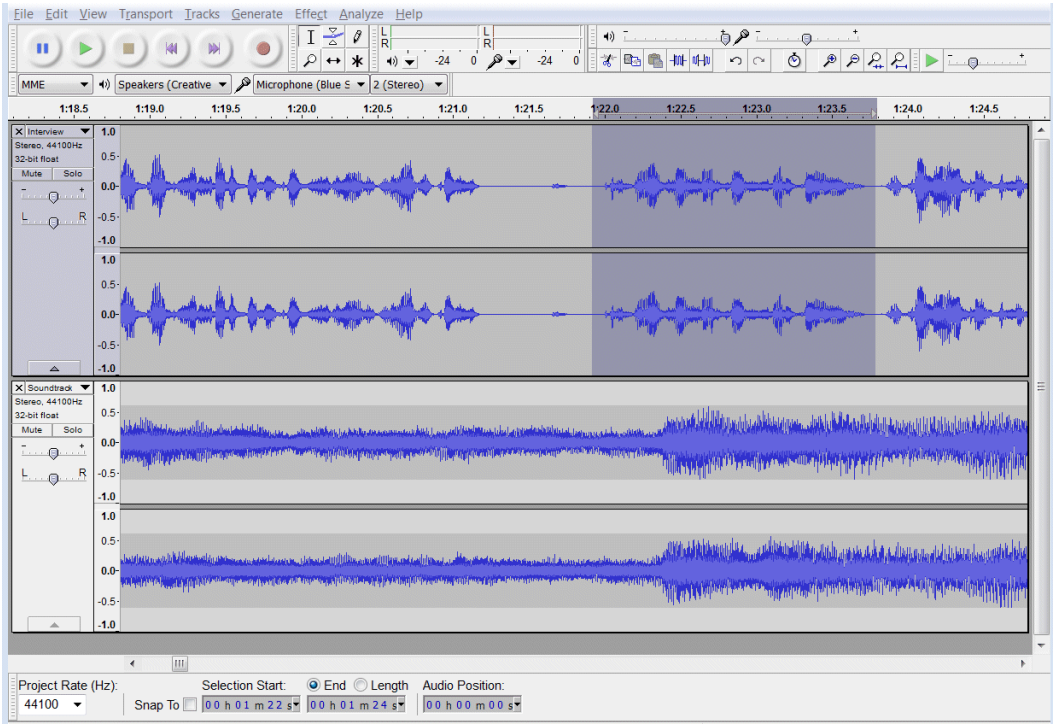


Figure 295: *Audacity* interface

Existing audio files can be imported, edited and exported in various different formats. New recordings can be created with great quality and many adjustments, such as noise reduction.

The software allows creating multiple tracks. In fact, each time the **Record** button is pressed, a new track will be created for the new recording, separating each part for easier editing. For single-track recording, press **Record** only once and then use the **Pause** button to pause and resume.

Each track can be equalized, edited (parts can be deleted, cut, copied, pasted, etc.), and multiple effects can be added.

When the final file (or portion) is exported, the tracks are collapsed into the output file.

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*, available at nuance.com/dragon.html—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.)

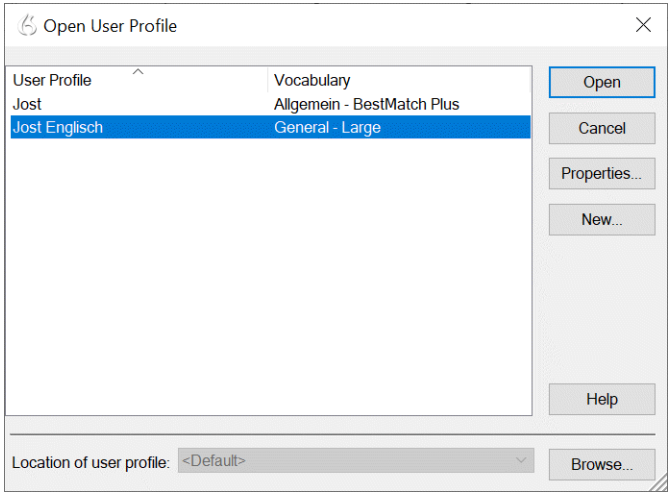


Figure 296: Selecting one of two possible languages in a German version of *Dragon*

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. *Dragon NaturallySpeaking* comes with all basic formatting in environments like *MS Word* or its own editor, *DragonPad*. But even when you use a translation environment tool, you are able to do everything with voice commands if you are willing to invest some more money into the Professional edition of the program and some effort in creating customized macros that then present virtually unlimited possibilities.

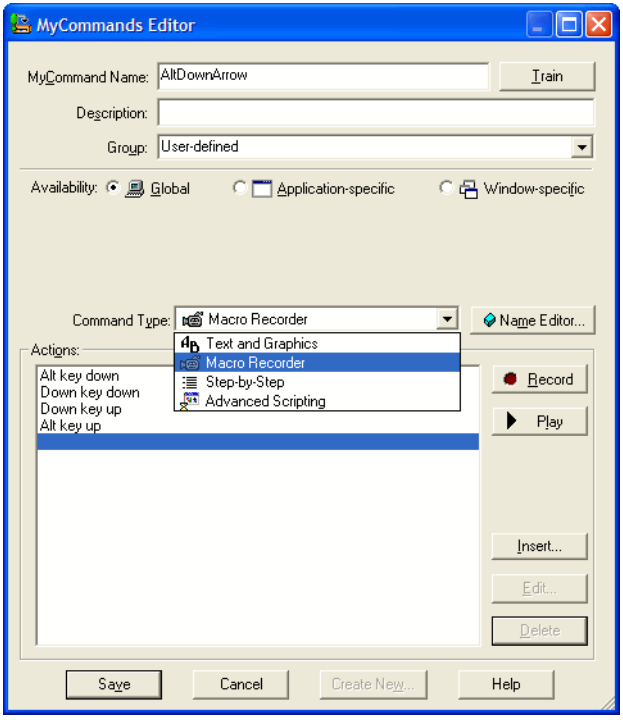


Figure 297: Recorded Macro in *Dragon NaturallySpeaking*

Windows Vista and later also contain an internal voice recognition program.

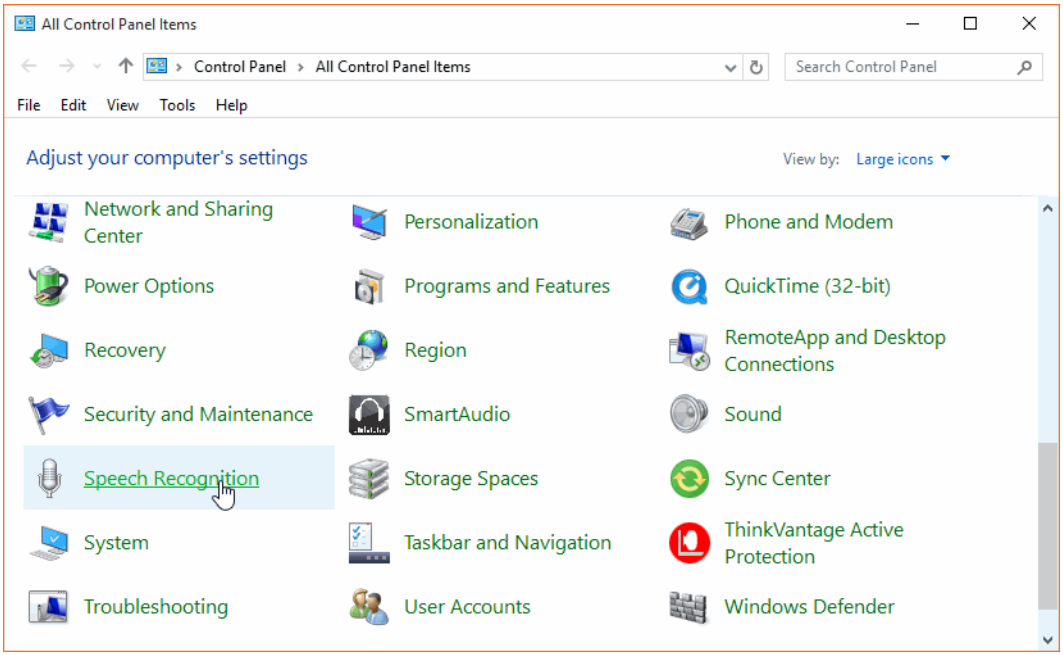


Figure 298: Starting *Windows Speech Recognition* in the *Control Panel*



In Windows 7 and above, 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 *Windows* than in *Dragon* (96% vs. 98%).

All of this leaves you high and dry if you don't work in the relatively few languages that are covered by *Dragon* or *Windows*. Here is the good news, though: GBoard, Google's keyboard for mobile devices, allows you to either type or dictate into your phone. And unlike Dragon, it's available in an insane number languages (see support.google.com/websearch/answer/

6380730#languages). If you follow translator Tiago Neto's advice on how to "virtually connect" your cell phone with your computer, you can dictate into your cell phone and see the results simultaneously show up in your preferred translation environment tool on your Mac or Windows PC. You can find Tiago's tutorial right here: ata-chronicle.online/none/speech-recognition-is-in-your-back-pocket-or-wherever-you-keep-your-mobile-phone.



In late 2018, memoQ introduced "hey memoQ," an iOS app that uses Apple's speech-to-text service to allow dictation into memoQ's translation editor in 30+ languages and commonly used voice commands. For more information, see docs.memoq.com/current/en/hey-memoQ/hey-memoq.html and apps.apple.com/app/hey-memoq/id1440587736.

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 115. 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.

The same is true for translation environment that are supported only by a single person. (The clear drawback there is, of course, what happens if that person changes careers or is otherwise not available for an extended period of time.)

Larger companies, including those behind translation environment tools have various levels of support for their products. While some products are known for outstanding support from their development communities, other products are—rightly or not—not particularly well-known for this. Either way, the best way to get support for translation environment tools is through user groups, which are usually located at groups.io, Google Groups, or Facebook. Most of these groups are extremely supportive and are very forgiving of even the most basic questions.

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. You don't have to become quite as extreme as me, 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 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 monthly *Tool Box Journal* (see internationalwriters.com/toolkit), please send me an email at toolbox@internationalwriters.com, or reach me on Twitter at twitter.com/jeromobot. (While you're on Twitter, be sure to follow twitter.com/translationtalk, a rotation curation account that highlights translators and interpreters from around the world.)

Index

How to . . .

access Search and Replace 117
access system configuration 32
access the registry 36
access zip functions 126
add shortcuts to the SendTo folder 18
add to the Startup folder 32
adding a subtitle in VLC or MPC 390
ask for help 52
assemble dictionary resources with Wordfinder 190
author on the basis of the translation memory 300
automatically detect language in Word 167
avoid animation in Windows 34
avoid word counts 134
back up the system 51
backup the system in Windows 8 48
batch import and export stories from Quark, PageMaker and InDesign 335, 340
batch rename file 116
be cautious with the registry 36
be creative with text editors 111
block traffic with router 83
change spelling language in PowerPoint 184
change the code page in Word 166
change the date and time of a file 116
change the spell-checking language in PowerPoint 184
check and replace fonts in PowerPoint 184
check for graphic in PowerPoint 183
check terminology 280
choose the right version of Windows 5
clean the registry 36
clean Word files with a macro 371
comment PDF files 361
compare files 106
compress files 125
connecting to FTP servers via Windows/File Explorer 72
control services 33
control which icons are displayed on the taskbar 31
control which programs are started automatically 31
convert code pages from Mac to Windows and vice versa 342
convert measurements 130
convert PDFs 365
convert PDFs with FineReader 370
convert time or date formats in Word 172
convert video files 396
copy files or folders 23
copy texts from screenshots 146
copy with the Shift and Ctrl keys 23
count words 133
create a new library in Windows 16
create editable PDFs in Apache OpenOffice/LibreOffice 368
create PDF files 380
create translation resources from PDFs 381
creating subtitles 410
define delimiters for word counts 136
define macros in AllChars 64
defragment your computer 42
delete cookies 40
delete from the Startup folder 32
delete remains of programs manually 36
delete temporary files 38
determine the format of a software development format file 318, 319, 320, 322, 323, 329
disable templates in Word 175
disable the Aero and Transparency features 34
display Windows in a different language 7
display Windows in different languages 7
distinguish between DTP programs 332
distinguish between the translatables vs. untranslatables 331

distinguishing between different types of subtitle files 388
downloading videos from the Internet 410
drag files into open programs 19
duplicate an item within a document 23
edit PSD files 343
edit psd files 342
efficiently browse on the Internet 67
enable extended features for Acrobat Reader 364
enable templates in Word 176
encrypt individual files 88
error check hard drives 42
exchange data through XML 158
export web pages into Excel 177
extract terminology 304
find large unneeded files 41
find out what's new in the Tool Box 3
find support for utilities 435
guard against phishing 93
guard against pop-ups 95
guard against spam 98
guard against spyware 92
guard against viruses 91
guard your computer 90
guard your computer screen 89
identify file types 110
identify zip files 111
install additional keyboard on Windows 8 60
install an additional keyboard 57
install an MSKLC keyboard on Windows 10 63
installing many utilities simultaneously 155
installing 500+ different keyboards 62
inverse the color in MS Word 164
keep the computer clean 35
list search results in Excel 179
make a decision for a TEnT 290
make QuarkXPress Passport files compatible with English version 340
make sure that your computer starts 43
manage comments in a PDF 362
manage graphic files 115
manage pages in PDF files 363
manage the Windows clipboard 142
memorize keyboard shortcuts 28
merge files 147
name and rename files 31
not to overwrite a file when restoring previous version 47
not use Word for HTML 113
open the Windows/File Explorer 14
organize files and folders 31
password protect zip files 127
perform backups 45
perform TM maintenance 301
pin files in Windows/File Explorer 28
prepare PowerPoint documents for translation 182
process vector-based graphic formats for translation memory 348
quickly navigate within Windows Explorer 18
read this book 2
record your screen with the Game bar 53
reduce risk with tools 83
reduce your risk 75
remove programs 35
rename several files 116
rent DTP programs 331
reorder lists in Word 171
restore previous version in Vista and 7 45
restore the system 51
restore your computer 44
run videos with a separate subtitle file 388
run Windows in different languages 7
save HTML files from the Internet 114
save PDF files with Acrobat Reader into editable format 365
save PDF files with Adobe Acrobat into editable format 366
save websites from the Internet 114
search by country with a search engine 68
search by file type with a search engine 68

search by language with search engine 68
search for definitions with search engine 69
search in dictionaries on your computer 119
search in Internet glossaries 119
search intelligently with Bing 67
search intelligently with Google 67
search through files 117
search through several text files 104
select a version of Microsoft Office 158
select an HTML editor for double-byte languages 112
select an HTML editor for right-to-left languages 112
select an office suite 157
select HTML editors 112
select multiple files in Windows 8 and above 15
send files to other drives or programs 18
set up a router 84
setting the display language in Windows 10 9
setting the preferred language in Windows 10 9
setting the primary language in Windows 8
split zip files 127
start into Safe Mode 43
Start with the desktop in Windows 8.1 10
store search-and-replace processes for later reuse 172
subscribe to the Tool Box Journal 437
switch between keyboards in Windows 8 62
switch between open applications 24
switch languages in voice recognition 429
take off all formatting in Word 163
take screenshots 144, 343
track changes in Word 168
track time 137
transfer files 71
translate desktop publishing formats 331
translate embedded documents 185
translate FrameMaker files 334
translate help systems 147
translate HTMLHelp 151
translate INDD files 336
translate InDesign IDML files 337
translate Java Properties 318
translate PageMaker files 338
translate PageMaker with Trados 339
translate PDFs with a TEnT 373
translate PDFs within the PDF 372
translate Photoshop file 344
translate Photoshop files in GIMP 344
translate pixel-based graphic formats 343
translate RC files 317
translate tagged formats in Déjà Vu 351
translate vector-based graphic formats 347
translate Ventura files 334
translate WinHelp 148
translate XML files with embedded HTML 353
understanding MT and privacy 273
uninstall programs 35
unzip several zip files simultaneously 127
upgrade to Windows 10 or not 12
use a mask when searching 118
use AllChars for special characters 64
use File History 48
use libraries in Windows 16
use magic numbers 111
use MT in TEnTs 278
use ScanDisk 42
use System Configuration 32
use text editors 104
use the Format Painter in Word 163
use the Research feature in Word 169
use the shell command 18
use the Windows/File Explorer 14
use URLs to find translation data 69
use voice recognition in Windows 432
use Wikipedia as a language resource 70
use Word for words counts 135
use Word 2013 for PDF conversion 367

using the Fill feature in Excel 180
using the Flash Fill feature in Excel 181
voice recognition on mobile devices 432
wisely use voice recognition 430
work in complex file formats with translation environment tools 192
work securely 75
work with Jump Lists 27
work with the registry 36
Toggl
time tracking software 140
Tortoise Tagger 355
Total Recall
CafeTran Espresso 228

A

ABBYY Screenshot Reader 146
ABBYY Transformer+ 378
Acrolinx IQ 297
Across
overview 212
support of binary localization formats 317
translation environment tool 202
adaptive machine translation
MateCat 241
AdaptiveMT 203
Ad-aware 93
Adblock Plus 98
additional keyboard
on Windows 8 60
Adobe AEM
support by Beebox 323
Adobe products
SaaS offering 331
Adobe Reader
enabling extended features for 364
advance-fee fraud 80
Adware 77
Aegisub
create and modify subtitles in 414
AEM
support by iLangL 325
Aero
disable 34
interface in Vista 34
ai
vector-based graphic extension 347
Alchemy Publisher
translate FM files 334
translation environment tool 202, 214
Aliado 194
AlignFactory 381
alignment 253
alignment
AlignFactory 253
caution against 250
reversal in Trados Studio 2015 252
AllChars
defining custom input 64
entering special characters 64
input utility 64
Amara
online collaborative platform 420
translate subtitles in 420
Amazon Translate 273
Amplexor MT
in TEnTs 276
Amptran 194
Anaphraseus 195
Andrä 196
Android localization 320
animated environment
preventing 34
anti-phishing 93
anti-spam 98
anti-spyware 92
anti-virus software 91
Any Video Converter 408
AnyCount 135
AnyLexic 309
Apache OpenOffice 157
creating editable PDFs 368
XML-based 158
Apertium
in TEnTs 275
app localization 320
Apple iWork
CafeTran Espresso 228
applications
switching between 24
ApSIC Comparator 296
ApSIC Xbench 309
search MS glossaries with 104

Apsic Xbench 302
Arabic
  QuarkXPress 341
ArabicXT 341
Araxis Merge
  tool to compare files 106
Araya 217
Archivarius 3000 118
artificial intelligence
  in Memsource 239
ASAP Utilities
  add-on for Excel 181
ASCII code
  enter characters 55
ASS files 388
  support via TEnTs 424
assemble feature 259
  Déjà Vu 206
attacks 78
Audacity 426
audiovisual translation 383
AutoCAD
  Star Transit 208
autocomplete 258
  in Wordfast Classic 198
Auto-completer
  OmegaT 226
AutoCorrect
  in TEnTs 270
automatic adjustment of MT matches
  CafeTran Espresso 228
automatic assemble
  CafeTran Espresso 228
automatic defragmentation
  in Windows 7 and above 43
automatic language detection
  in Word 167
AutoSuggest
  Trados 203
AutoUnbreak
  for dealing with short PDF excerpts 361, 366
AutoWrite
  Déjà Vu X2 206
AV translation 383
  benefits 383
AVG Antivirus 92
AVS4YOU
  suite of multimedia software 388
AVS4YOU Media Player 388
AVS4YOU Video Converter 396

B

Böblingen
  place of origin for TEnTs 193
backdoor 77
backing up the system 51
Baidu
  in TEnTs 276
Baidu Translate 273
Beebox
  CMS connector 232
  CMS connector by Wordbee 323
BeGlobal
  in TEnTs 276
behavior-based detection
  in anti-virus software 92
Belarc Advisor
  computer inventory 54
Belazar
  in TEnTs 275
Belkin Linksys
  router 84
bi-directional languages
  PageMaker support 339
binary software localization 313
Bing 67
  tips for 67
Bitdefender Antivirus 92
BlueGriffon
  HTML editor 112
bmp
  graphic extension 343
book
  FrameMaker extension 334
  mission 1
  purpose 1
botnet 81
breadcrumb trail 14
browsers 67
Browsing Tips 67
browsing tips 67
brute force

password cracking 129
built-in encryption 88
bulk email 82

C

cable lock 90
Cafetran
support of subtitle files 424
CafeTran Espresso
overview 228
translation environment tool 202, 228
Carbonite
backup program 45
CAT tools
alignment tools 192
categories 190
conversion and maintenance tools 191
project management tools 190
quality assurance tools 192
resource lookup tools 190
software localization tools 191
term extraction tools 191
terminology management tools 190
text extraction tools 191
word count tools 191
Catalyst
localization tool 315
purchased by translations.com 316
translate CHM files 153
categories
of CAT tools 190
categorization
of CAT tools 189
CCleaner
clean cookies and browsing data with 40
registry maintenance 37
Censhare
support by Beebox 323
changing
date/time stamp of file 116, 118
Character Map 55
character set
definition 166
CheckMate 291
Chinese
language in Linguee 121
chm
HTMLHelp extension 151
Chrome
anti-phishing 95
browser 67
Classic Shell 11
Clear-CAT 194
click fraud 81
clipboard
definition 140
translation with Wordfast Anywhere 235
clipboard management utilities 140
clipboard translation
CafeTran Espresso 228
clipboard utility 142
ClipMate
clipboard utility 142
closed captions
definition 385
CMS-based websites 322
cnt
WinHelp extension 149
code page
changing in Word 166
definition 166
switching in Word 166
code pages
differences between Mac and Windows 342
Codecs
definition 387
codepage conversion
in text editors 109
CodeZapper
macro to clean Word files 371
collaboration features
in TEnTs 284
collaborative localization 320
collaborative localization platforms 320
combination of linguistic resources
in TEnTs 259
comma-delimited files
vs. Excel files 179
Common threats to computers 75
Comodo Firewall 91
comparing
files 106

compression utilities 125
computer
  keeping it clean 35
computer-assisted translation tools. See CAT tools
Congree 300
Contentful
  support by iLangL 325
content-oriented DTP programs
  automation 333
  definition 333
  publishing in different formats 333
  suitability for translation 333
Contentstack
  support by iLangL 325
Control Panel
  Windows 10 9
  Windows 8 12
Convert
  conversion utility 131
converting
  codepages 109
  measurements 130
  PDF files to editable text 365
converting measurements
  in Windows 133
  in Windows 10 132
Convertus
  in TEnTs 276
Cookies
  deleting in different browers 40
cookies
  definition 40
  deleting 40
CopyFlow
  and QuarkXPress 340
  not supporting newer versions of Quark 335
  support for Quark 7 and above 340
copying and pasting
  in MS Word 144
copying files
  Windows Explorer 23
Corel Draw
  translate in 348
count
  words 133
creating PDF files 380
Creative Cloud 331
Crosslang Gateway
  in TEnTs 275
crossMarket 213
Crowdin
  collaborative localization 321
crowdsourcing
  Wordbee 233
csv files
  vs. Excel files 179
customized keyboard shortcuts in Word 55

D

Déjà Vu
  Déjà Vu 206
  first Windows-based TEnT 193
  first Windows-based translation environment tool 206
  overview 206
  translate embedded documents 186
  translate tagged formats 351
  translation environment tool 201
Déjà Vu X
  support of subtitle files 424
Déjà Vu X2
  deep integration of machine translation 206
  support of XML with embedded HTML 355
Déjà Vu X2/3
  PDF support 375
date/time stamp of file
  changing 116, 118
dead key 57
dead keys
  on US-International keyboard 57
DeepL 273
  and privacy 273
  in TEnTs 276
Deeplingo
  in TEnTs 275
DeepMiner
  Déjà Vu X2 206
definition

HTML 112
defragmenting
definition 42
deleting
cookies 40
programs 35
deleting programs 36
after uninstall 36
deleting temporary files 38
Denial-of-Service (DoS) 80
design-oriented DTP programs
definition 332
post-translation layout 336
text flow 335
Desktop Publishing. See DTP
dictionary resources
assemble with Wordfinder 190
Disk Cleanup 38
disk failure
precaution 51
Distributed-Denial-of-Service 81
D-Link
router 84
routers 84
DNN
support by iLangL 325
Document Type Definition
SGML 350
document versioning
memoQ 210
DoS 80
double-byte languages
PageMaker support 339
QuarkXPress 341
DQF 282
Dragon NaturallySpeaking 429
voice recognition 431
Dreamweaver
HTML editor 112
drive-by download 80
drivers
definition 36
drives
defragmenting 42
Dropbox
backup program 45
Drupal
support by Beebox 323
support by iLangL 325
dtd
SGML extension 350
DTP programs
classification 332
common denominator 331
content-oriented 332
design-oriented 332
Mac vs. Windows 341
reasons to buy 336
workaround to obstacles 332
working in translation environment
tools 333
dtSearch 118
dual fuzzy
Star Transit 208
dubbing
definition 385
duplicating
an item within a document 23
Dutch
language in Linguee 121
DVD ripping 395
Dynamic Quality Framework 282

E

Eastern European languages
and QuarkXPress 340
editable PDF
Fluency 219
editing time
count in PractiCount 136
ElcomSoft 128
email attachments
restrictions 71
email spam 82
embedded documents
translating 185
embedded files
memoQ 211
embedding subtitles 395
EmEditor 104
EpiServer
support by Beebox 323
support by iLangL 325

eps
  vector-based graphic extension 347
error checking
  hard drive 42
Error Spy 291
EU translation corpora 256
EuroTermBank
  Tilde Terminology 306
Everything
  search utility 21
ex TranslationFilter 340
  Quark filter 335
ExactSpent 139
ExamDiff
  file comparison 106
Excel
  list search results 179
  tips and tricks 177
  tracking hours and pricing 326
exchange formats 285
export to Microsoft Excel
  Internet Explorer command 177
Extensible Stylesheet Language
  XML 350
Extract and Reinject Embedded Objects
  Trados app 188
extract terminology
  SynchroTerm 304

F

Farsi
  QuarkXPress 341
Felix
  open-source 219
file association memory
  benefits and drawbacks 208
file comparison
  with text editors 106
File Explorer 14
  access 14
  copying files in the same directory 23
  file management 14
  Ribbon menu 16
file formats
  support of 289
File History 45
  in Windows 8 and above 48
  libraries 16
file management
  File Explorer 14
  Windows Explorer 14
File Merger 147
file structure 31
file transfer 71
File Transfer Protocol 72
files
  naming conventions 31
  renaming 116
FileZilla 72
Fill feature in Excel 180
Firefox 67
firewall software 90
flame wars
  Lantra-L 206
Flare
  RoboHelp's successor 221
  support by MadCap Lingo 221
Flash Fill
  in Excel 181
Fluency
  MQM support 284
  overview 219
  PDF support 374, 375
  support of subtitle files 424
  transcription in 419
  translation environment tool 202, 219
Fluency Transcription 420
FM files
  FrameMaker files 334
  translate with Alchemy Publisher 334
Focus
  tool in MS Word 164
folder structure 31
fonts
  different naming conventions on Mac vs. Windows 341
ForeignDesk 227
Format Factory 406
Format Painter
  using in Word 163
FormatCheckers for Word and FrameMaker 297
Foxit PDF Tools 361

fragmentation
  definition 42
FrameMaker 332
  save fm files as mif 334
  translate files 334
  translation memory support 334
free tools
  using 219
French
  language in Linguee 121
F-Secure Internet Security 91
FTP 72
  use Windows/File Explorer 72
fuzzy match repair
  Déjà Vu 206

G

GBoard
  voice recognition 432
Gengo 196
German
  language in Linguee 121
gettext
  translation 319
ghost imaging
  definition 51
  use of 51
GIF
  graphic extension 343
GIF89a
  GIF header 110
GIMP 344
  free graphic application 343
Globalese 275
Globalsight Ambassador 196
GNU
  translation 319
GNU gettext
  support by translation environment tools 319
Google
  tips for 67
Google Chrome
  browser 67
Google Docs 157
Google Drive 157
Google Translate 273
  and privacy 273
  in TEnTs 274
Google Translator Toolkit 231
  historical release 194
graphic applications
  common denominator 342
  double-byte languages 343
  right-to-left languages 343
  working with clients 343
graphic files
  pixel-based formats 343
  Smartcat 241
graphic formats
  translate 343
graphic management applications 115
Greenshot 146
GTCOM
  in TEnTs 275
GT4T
  access to many machine translation engines 277
  machine translation connector 277
  wholesale machine tranbslation vendor 277
Guess Tags
  MateCat feature 244

H

Hansards corpus 256
hard disk
  with built-in encryption 88
hard-coding
  media files 395
Heartsome
  open-source 217
  overview 216
  translation environment tool 202
Hebrew
  QuarkXPress 341
Helium files
  quality assurance for 291
help systems
  translation 147
HelpDeco

WinHelp 149
helpful shortcuts
Windows 18
hhc
HTMLHelp 151
hhk
HTMLHelp extension 152
hhp
HTMLHelp extension 151
highlighting
occurrences of a word 108
Hightail 73
history
of translation environment tools 193
hlp
WinHelp 150
WinHelp extension 149
Ho to
Install an additional display language in Vista 7
how to
Imitate the Start menu in Windows 8 11
HTML
definition 112
tagged formats 349
HTML editors 112
hands-on tools 112
HTML files
saving from the Internet 114
HTML Help Workshop
download 151
HTML mode
Text editors 106
HTMLHelp 148
translate 151
HTML5 112
support 112
hybrid machine translation
definition 272

I

IATE
Tilde Terminology 306
IBM
and privacy 273
IBM Watson Language Translator
in TEnTs 274, 275
Iconic
in TEnTs 276
Iconic IP Translator
in TEnTs 275
IDML files
translate 337
iLangL Cloud 325
Illustrator
translate in 348
Illustrator files
process with 3rd-party tools 348
INDD files
translate 336
InDesign 332
batch import and export stories 335, 340
translating in 336
InDesign CS4 and later
support 337
InDesign Markup Language
InDesign 337
indexing options
in Windows Vista and above 21
Infix PDF Editor 372
inline spell-checking
Déjà Vu 207
in TEnTs 270
inline tags 285
IntelliWebSearch 190
glossary search tool 119
Intento 276
in TEnTs 275
interactive machine translation
Lilt 247
Internet Explorer
anti-phishing 93
INX
exchange format for InDesign 336
iOS localization 320
IPTranslator
in TEnTs 275
Italian
language in Linguee 121
ITT files
support via TEnTs 424

J

Japanese
  language in Linguee 121
Java Properties
  software development format 318
  support by translation environment tools 318
  translate 318
Java-based
  CafeTran Espresso 228
  Fluency 221
Jawi
  QuarkXPress 341
Jeromobot 3
  on Twitter 437
JiveFusion 195
Journal
  Tool Box 437
Joust 194
JPEG
  identify 110
JPG
  graphic extension 343
Jump List
  Windows 7 and above 27
junk email 82

K

KantanMT
  in TEnTs 275
Kensington Security Slot 90
Kentico
  support by Beebox 323
keyboard
  United Kingdom Extended 56
keyboard languages 54
  adding 54
  setting 54
keyboard shortcuts
  memory aids 28
Keylogger 77
Keyman
  virtual keyboards 62
K-Lite codec pack 387
K-Slot 90

L

language
  automatic detection 167
Language Terminal
  memoQ 210
  process InDesign files 337
languages
  keyboard 54
  supported for SynchroTerm 304
LanguageWeaver 271
Learning Tools
  in MS Word 164
Learning Tools menu 165
LetsMT
  in TEnTs 276
LetsMT!
  in TEnTs 275
lexicon
  Déjà Vu 206
lexicon and ExtraTerm
  function of 303
lexicon-based alignment
  XTM 255
LF Aligner
  open-source aligner 254
libraries
  definition 16
  in Windows 16
library
  create in Windows 16
  for backup purposes 17
LibreOffice 157
  creating editable PDFs 368
  creating PDF files 380
  XML-based 158
Lilt
  neural morphology 249
  support of subtitle files 424
  translation environment tool 231, 247
Lingenio
  in TEnTs 276
Lingo 307
Lingofy 300
Lingohub
  collaborative app localization 321
Lingotek 196, 229

specializing in CMS 229
Lingua Custodia
in TEnTs 276
Linguee 121
Linksys
router 84
Linport
exchange standard for translation packages 285
Language Interoperability Portfolio 287
Linux
users on CafeTran Espresso 229
vs. Windows 5
Lionbridge
release of ForeignDesk 227
Translation Workspace 196
LISA QA Model 283
LiveDocs
memoQ 210
memoQ features to process PDF files 381
localization tools 313
history 314
LocStudio
Microsoft's localization tool 315
LogiTerm
features 310
search features 312
Logoport
basis of Translation Workspace 199
purchase by Lionbridge 193
lossless video 396
LSC
memoQ 258
LTC Organiser 327
LTC Worx
project management 327
l10n
example for a poorly chosen term 189

M

Mac
vs. Windows 5
machine translation
definition 271
integrated in TEnTs 271
Macintosh
Wordfast support 197
MadCap Lingo 202
overview 221
support for Flare 221
translating MadCap Flare help projects 154
magic numbers 111
maliciously crafted content 80
malvertising 78
Malware 75
Man-in-the-Middle 81
Manypedia 70
MARTIF 307
MateCat
automatically placing tags 244
business model 242
DQF support 284
translation environment tool 230, 241
measurement conversion utilities 130
measurements
converting 130
Media Player Classic
adding subtitle files to video 390
as media player 388
media players 387
memoQ 297
DQF support 284
overview 209
PDF support 374
QA features 282
support of PSD files 345
support of subtitle files 425
support of XML with embedded HTML 356
translate embedded documents 187
translation environment tool 202
user interface 270
memoQ cloud 210, 288
Memsource 236, 297
DQF support 284
processing embedded HTML 358
real-time collaboration 288
support of PSD files 345
translation environment tool 236
use of AI 239
Memsource Cloud
translation environment tool 230

Memsource Editor for Mobile 238
Memsource Translate
  automated selected MT engine 239
Memsource Web Editor 238
merging
  files 147
MetaTexis 195
Metro interface
  of Windows 8 10
MicroMATER 307
Microsoft Edge
  anti-phishing 94
  pop-up blocking 96
Microsoft Help Workshop
  download 149
Microsoft Keyboard Layout Creator 62
Microsoft Language Portal 104
  TBX download 104
Microsoft Office
  in connection with other programs 157
  office suite 157
  release versions 158
  working with clients 157
Microsoft Publisher
  support by Fluency 221
Microsoft Translator 273
  and privacy 273
  in TEnTs 275
MIF files
  FrameMaker extension 334
Mirai/COTOHA Translator
  Mirai/COTOHA Translator 275
mission
  for book 1
Mite
  tracking time 140
MITM 81
ModernMT 273
  in TEnTs 275
monolingual term extraction
  by TEnTs 303
Moravia MT
  in TEnTs 275
MoraviaMT
  in TEnTs 275
morphological recognition
  Across 212
  in OmegaT 226
morphological support
  Star Transit 208
morphology
  Across 212
  Lilt 249
  OmegaT 226
  Smartcat 241
  Star Transit 208
Moses 271
  in TEnTs 274
Mozilla Firefox
  anti-phishing 94
  pop-up blocking 97
Mozilla Thunderbird
  block junk 99
MQM 282
MQXLZ
  memoQ-specific XLIFF 338
msconfig
  command 32
MSKLC
  keyboard creator 63
  on Windows 10 63
MT
  definition 271
MTQE
  machine translation quality estimation in Memsource 239
MTX 307
Multidimensional Quality Metrics 282
multi-language versions
  of Windows 7
Multilingual Toolkit
  for app localization 320
Multilizer
  localization tool 316
Multimedia
  translation 383
Multiple Find and Replace
  TransTools+ 172
MultiTerm Extract
  principles 303
MultiTrans
  acquisition by SDL 193
MyDutchPal

in TEnTs 275, 276
MyMemory 276
in TEnTs 274, 275
MZ
EXE header 110

N

narration
definition 385
Netgear
router 84
routers 84
neural machine translation
definition 272
neural morphology
Lilt 249
n-grams
use in SMT 271
NICT
in TEnTs 275
Nigerian scams 80
Ninite
installing many utilities simultaneously 155
NMT
definition 272
NoBabel 194
non-translatables
detection in Memsource 239
Notepad 103
Notepad++ 104
Notifications Area
Windows 31
Nvu
HTML editor 112

O

obfuscation data
in IntelliWebSearch 121
OCR
graphic files 345
integrated into Fluency 220
OCR engine
Smartcat 241
OCR of PDFs
in TEnTs 378
OCR tools for PDF files 368
Office Clipboard 141
Office documents
with embedded documents 185
office suite
definition 157
Office 2013
integrated multilingual spellcheckers 159
Office 365 159
Okapi CheckMate 291
Okapi tools 226
Olifant 191
translation memory maintenance 301
OmegaT
Auto-completer 226
overview 225
real-time collaboration 288
realtime cooperation 289
support of subtitle files 424
supported formats 225
translation environment tool 202
Omniscien
in TEnTs 274, 275
One Hour Translation 196
OneDrive
backing up files with 45
online interface TEnTs
advantages and disadvantages 229
online translation environment tool
Wordfast Anywhere 235
ontram 196
Open Language Tools
open-source translation environment tool 227
OpenNMT
in TEnTs 276
OpenOffice
creating PDF files 380
open-source TEnTs 225
open-source translation environment tool
Open Language Tools 227
OpenTM2 227
Opera
anti-phishing 95
browser 67

pop-up blocking 97
operating systems
definition 5
optical character recognition
Fluency 220
OPUS corpora collection 257
Outlook
block junk 100
Overlay Editor
Text United 223

P

package standards 286
PageMaker 332
batch import and export stories 335, 340
translate 338
upgrade to InDesign 339
Paint Shop Pro
low-price graphic application 342
PangeaMT
in TEnTs 275
PangeanMT
in TEnTs 275
Passolo
localization tool 315
purchased by SDL 316
Passport edition of Quark 340
password cracking utilities 128
password protection
zip files 127
PaymentPractices 80
PDF
alignment 253
commenting 361
file types 359
image-based files 359
managing comments 362
page management 363
Pretty Darn Frustrating 359
searchable image-based files 359
text-based files 359
PDF files
counting words in 137
creating 380
PDF Nitro 361
PDF Reflow
feature in MS Word 367
PDF tools 360
PerfectIt 299
phishing 79, 93
Photoshop 344
Photoshop Elements 342, 343
physical keyboard
vs. virtual keyboard 56
pivot languages
XTM 234
pixel-based formats
graphic files 343
pixel-based graphic formats 343
PK
identify zip files 111
PKZIP
zipping utility 125
Plunet
project management 327
pmd
PageMaker extension 338
pm5
PageMaker extension 338
pm6
PageMaker extension 338
PO
GNU gettext extension 319
Poedit
support of PO and POT files 319
Polish
language in Linguee 121
Polmann Services
products for Photoshop files 345
pop-under ads 82
pop-up ads 82
pop-up blockers 95
Portuguese
language in Linguee 121
post-editing 274
converted PDFs 371
pot
GNU gettext extension 319
Power User menu 12
PowerPoint
check and replace fonts 184
counting words in 136

in translation tasks 182
preparing documents for translation 182
word counts 137
PractiCount & Invoice 133
word count utility 135
predictive typing 257
in TEnTs 258
preferred language vs. display language
in Windows 10 9
preprocessing
XTM 234
PrepTags 355
preview
Déjà Vu 207
Wordfast Pro 216
preview feature
in Windows Vista and above 14
primary language
setting in Windows 8
Privacy Badger
pop-up blocking 98
privacy filters 89
Problem Steps Recorder 53
programs
removing 35
programs starting automatically
preventing 32
Project
for managing workflows and
scheduling 326
project management
Plunet 327
project-open 327
Projetex 327
Protemos 327
QuaHill 327
Translation Office 3000 328
T.O.M. 327
XTRF 327
project-open
project management 327
Projetex
project management 327
Prompsit
in TEnTs 276
PROMT
in TEnTs 275
properties
Java Properties 318
Protemos
project management 327
pseudo-translation 286
psr
Problem Steps Recorder 53
Publisher
Fluency 221
PDF support 374
translate CHM files 153
purpose
of book 1
p65
PageMaker extension 338

Q

QA Distiller 291
QTLaunchPad 282
QuaHill
project management 327
quality assessment 282
quality assurance
in Wordfast Classic 198
source documents 297
quality assurance features
in TEnTs 279
quality assurance tools 291
quality tracking
through editing 296
Quark
batch import and export stories 335, 340
QuarkXPress 332
CopyFlow 340
double-byte support 340
foreign language support 340
support for Arabic 341
support for Farsi 341
support for Hebrew 341
support for Jawi 341
translate 331
translate files in 340
Quintillian 194
QWERTY

vs. QWERTZ keyboard 56
QWERTY keyboard layout 63
QXD
QuarkXPress extension 340

R

Rainbow
conversion utility for many file formats 319
TM management tool 226
utility to convert files into OmegaT projects 226
ransomware 76
RbMT
definition 271
RC files
support from translation environment tools 317
RC-WinTrans
localization tool 316
Read mode
in MS Word 164
reader
description of 1
real-time collaboration 287
reducing
risk 75
reference material
Star Transit 207
registry 36
accessing 36
cleaning utilities 36
definition 36
RemoteTM 217
renaming
files 116
renaming tool
1-4a Rename 117
renaming utilities 116
reordering lists
in Word 171
Replace Studio Pro
search utility 117
Research
feature in Word 169
resource files
in text editors 108
restore previous version
process not to overwrite 47
restoring
system 44
the system 51
RESW files 320
Retrofit
feature in Trados 203
Reverso
in TEnTs 274, 276
revision history
Wordbee 232
Rewrite
feature in Word 365 170
Ribbon menu
in File Explorer 16
right-to-left languages
PageMaker support 339
ripping
definition 395
Rname-it
renaming utility 116
RoboHelp 221
rogue security software 78
router
setup 84
to block traffic 83
rules-based machine translation
definition 271
Russian
language in Linguee 121

S

SaaS
offering for Adobe products 331
Translation Workspace 200
Wordbee 231
XTM Cloud 233
SAE J2450 283
Safaba
in TEnTs 275
Safari
browser 67
Safe Mode 43
in Windows 8 and above 43

saving
HTML files from the Internet 114
saving complete websites 114
saving HTML files
complete websites 114
purpose 114
screenshots 144
applications 144
ScreenTips
different language in Office 2010 160
SDL AppStore
marketplace for Trados 205
SDL Language Cloud
in TEnTs 276
SDL Language Cloud Terminology 264
SDL LiveContent
support by Beebox 323
SDL MT
in TEnTs 276
SDL MultiTrans 198
translation environment tool 196
SDL Trados 201
SDL Trados Business Manager 190, 206
SDLX 194
quality assurance features 280
search
right-click access 117
search engines
search by country 68
search by file type 68
search by language 68
search for definitions 69
search utilities 117
searching documents
text editors 105
segmentation correction
CafeTran Espresso 228
segmentation rules on the fly
memoQ 211
SendTo folder
adding shortcuts 18
Windows 18
sending files
by email 71
sequence checking 57
services
description 33
disabling 33
stopping 33
SGML
tagged formats 349
shadowing
restoring files in Vista and above 45
share translation memory data
in Wordfast Classic 198
Sharepoint
support by Beebox 323
shell command 18
shortcuts
Send to 18
Similis 194
Simple Markup
in Word 2013 168
simultaneous scrolling
Word 165
Sisulizer
localization tool 316
SiteCore
support by Beebox 323
support by iLangL 325
Slate Desktop
in TEnTs 274
Smart Lookup
feature in Word 365 169
Smartcat
free use 241
OCR of PDFs 378
support of subtitle files 424
translation environment tool 230, 240
working with graphic files 345
Smartling 196, 230
SmartMATE
in TEnTs 276
SMT
definition 271
SnagIt 145, 146
Snipping Tool
Windows 145
social engineering 79
soft-coding
media files 388
software
development process 313

software development formats
classification 317
determine file type 318, 319, 320, 322, 323, 329
translation 317
Sogou
in TEnTs 276
Solid PDF Tools 361
solid state drive
and defragmentation 43
sound editing 426
spam 82
spam email 98
Spanish
language in Linguee 121
special characters
AllChars 64
with ASCII code 55
spiders
saving websites 114
Spike
copying in Word without the clipboard 144
splitting
zip files 127
Spybot 93
Spyware 77
SpywareBlaster 93
SQL Server
database for Across 212
SRT files 388
support via TEnTs 424
SRX
lack of implementation 285
SSA files 388
S-Tagger
Trados 202
standard
missing 287
Star CLM 190
Star MindReader 300
Star Transit
overview 207
PageMaker support 338
support of QuarkXPress 340
terminology management 265
translate embedded documents 186
translation environment tool 202
Star Transit NXT
support of subtitle files 425
Start menu
Windows 10 13
start up
in Safe Mode 43
Startup folder
adding shortcuts 32
cleaning it out 32
using it 32
Start10
replacement for Start menu 13
static websites 321
statistical machine translation
definition 271
Steps Recorder 53
Stingray Document Aligner 217
STL files
support via TEnTs 424
stories
in design-oriented DTP programs 335
STRINGS file 320
Stuffit
compression tool 126
subsegment matching 257
in TEnTs 257
subtitle file
naming 399
separate from video 388
Subtitle Workshop 410
SubtitleNEXT
translate subtitles in 424
subtitling
definition 385
subtitling files
Fluency 221
subtitling formats
supported by Subtitle Workshop 412
subtitling tools 410
Sunda Translator
in TEnTs 276
support
utilities 435
support options 435
SVG
support 348
switching

between open applications 24
Swordfish
overview 216
Symantec's Norton Security 91
SynchroTerm
extract terminology 304
System Configuration Utility 32
stopping programs from starting 32
System Restore 44
Systran
in TEnTs 275

T

tab-delimited files
vs. Excel files 179
tagged formats
definition 349
translate 349
TansTools
for Excel 182
target text-based matching
Star Translt 209
Targoman
in TEnTs 276
Task View
button 25
Windows 10 25
taskbar
cleaning it out 31
Control icons 31
Windows 31
TAUS 282
TAUS corpus
Tilde Terminology 306
TAUS Data Cloud 122
tauyou
in TEnTs 275, 276
TBX
complexity 285
Ted Talks
translate with Amara 420
Teleport Pro
saving websites 114
teletext
definition 386
Tell me what you want to do feature
in SDL Trados Studio 2019 204
templates
disabling in Word 175
enabling in Word 176
temporary files
deleting 38
Tencent
in TEnTs 275
TEnT
abbreviation for translation environment tool 189
why this term 189
TEnTs
PDF support 373
term extraction
CafeTran Espresso 228
terminology
XTM Cloud 233
terminology checks
strategies 280
terminology extraction
Text United 224
terminology management
importance in translation environment programs 262
terminology mining 303
from existing translation 303
Termsoup
translation environment tool 230
TermStar 265
text
in content-oriented programs 333
text editor
definition 103
highlighting 108
vs. word processor 103
text editors 103
external 104
file comparison 106
HTML mode 106
searching documents 105
standalone applications 104
uses 104
Text United
desktop-based translation application 224

overview 222
translation environment tool 202, 222
TextBase translation memory
MultiTrans 198
text-based localization 317
TIFF
graphic extension 343
Tilde Terminology
integration with TEnTs 300, 302, 306
term extraction 305
Time Stamp 138
time tracking 137
TIPP 285
Translation Interoperability Protocol Package 287
TM-Container
Star Transit 208
TMs
create from PDFs 381
TMX
problems with 285
TMX Editor
by Heartsome 301
TMXEditor 217
Tmxmall
in TEnTs 275
tokenizers
OmegaT 226
tools
to reduce risk 83
TQAuditor 296
Track Changes
in different versions of Word 168
in TEnTs 270
track changes 168
Track Changes feature
in Trados 203
tracking 82
traditional desktop-based tools
working on a browser-based version 230
Trados
acquisition by SDL 193
DQF and MQM support 284
history of development 193
PageMaker support 338
PDF support 374, 376
quality assurance features 281
support for QuarkXPress 340
terminology management 266
Ventura support 334
Trados GroupShare 288
Trados Studio 297
overview 202
redesigned version of Trados 203
support of subtitle files 424
support of XML with embedded HTML 357
XML support 350
Trados Studio 2014
alignment 251
Trados Studio 2015
OCR of PDFs 378
Trados Studio 2019
better usability 204
Tradupedia 71
Training
requirement for using tools 291
Trans Suite 2000 194
TransTools
changing spelling language in PowerPoint 184
Document Cleaner 371
tools to optimize Office files for TEnTs 192
TransTools+
collection of tools for Word 172
transcription
Fluency 220
in Fluency 419
Transifex 196, 230
collaborative localization 321
Transistent
in TEnTs 276
Transit NXT
support of binary localization formats 317
translate
desktop publishing formats 331
translate5
MQM support 295
translation environment tool
CafeTran Espresso 228
Lilt 247
MateCat 241

memoQ 202
Memsource 236
Smartcat 240
why this term 189
translation environment tools
analysis 192
categories 195
definition 192
definition of users 194
productivity 192
quality assurance 192
terminology management 192
working in complex file formats 192
translation environment yool
relevance of terminology management 262
translation management systems
cloud-based 230
translation memory driven segmentation
memoQ 210
Translation Office 3000
project management 328
Translation Workspace
based on Logoport 199
translation environment tool 196
translationtalk
on Twitter 437
Translator mode
Subtitle Workshop 414
transparency
disable 34
in Windows 34
TransPDF
in memoQ 374
Trojan Horse 76
TSS/Joust 194
TTML files
support via TEnTs 424
T-Windows applications
Trados 202
Twitter
following Jeromobot 437
translationtalk 437
two-column Word file
Wordfast Pro 216
two-column Word files
for exchange 286
TXML
support by Wordfast Anywhere 235
Typo 3
support by Beebox 323
T.O.M.
project management 327

U

UBE 82
Ubiqus NMT 275
UCE 82
UltraEdit 104
Umbraco
support by iLangL 325
UN corpus 257
uncompressed video 396
Unicode
definition 6, 166
handling in Windows 6
Unicode support
Office2000 166
Uniscape 229
United Kingdom Extended
keyboard 56
updated version of Tool Box
what is new 3
UpLIFT 203
URLs
use to locate translation data 69
use
of MT in TEnTs 278
US-International keyboard 55
utilities
definition 115
support 435

V

vector-based graphic formats 347
Ventura 332
retired by Corel 334
translate 334
Verifika 291
video controls
for Subtitle Workshop 412
video files

convert 396
View Side by Side feature
Word 165
Virtaal 227
virtual desktop
Windows 10 25
virtual keyboard
vs. physical keyboard 56
Virus 75
Visio
suppport by XTM Cloud 233
Vista
Aero interface 34
installing an additional display language 7
restoring files 45
search capabilities 119
3D Flip 24
Visual Editor
XTM 234
VLC Media Player 387
adding subtitle files to video 390
as media player 387
VLTM
Wordfast Anywhere 235
voice recognition
support in Windows 432
voice recognition applications 429
voiceover
definition 385
VP
Ventura extension 334

W

Warning
Word files with tracked changes 171
web browsers 67
web bugs 82
Web Page, Filtered
Word command 113
website localization 321
websites
saving complete 114
Welsh
support in United Kingdom Extended keyboard 56
WEP 86
What's On
support by Beebox 323
Wikipedia
as a translation resource 70
wildcard
definition 20
wildcard searches
in Windows Explorer 20
wildcards 20
WinAlign
Trados 202
WinDirStat
finding space hogs 41
Windows
business and home products 5
focus 1
helpful shortcuts 18
integrated text editors 103
languages supported for voice recognition 432
multi-language editions 7
multi-language versions 7
preventing animated environment 34
SendTo folder 18
Unicode 6
voice recognition 432
vs. Mac and Linux 5
Windows app localization 320
Windows Clipboard 142
Windows Defender 92
Windows Explorer 14
access 14
copying files 23
copying files in the same directory 23
file management 14
wildcard searches 20
Windows key
symbol 14
Windows NT series
application processing 5
Windows registry 36
Windows 10
Control Panel 9
interface 12
Settings menu 9
Start menu 13

Windows 2000/XP
  benefits over previous versions 5
Windows 7
  restoring files 45
  search capabilities 119
Windows 8
  backing up 48
  Control Panel 12
  installing additional keyboard 60
  restoring files 45
  search capabilities 119
Windows 8 and above
  display different language 7
Windows 8.1
  start with desktop 10
Windows/File Explorer
  connect to FTP 72
WinHelp 148
  convert to HTMLHelp 153
  translation 148
WinKey
  symbol 14
WinKey shortcuts 29
WinZip
  zipping utility 125
WIPO Translate
  in TEnTs 275
Wireless B 86
Word
  disabling templates 175
  enabling templates 176
  keyboard shortcuts 55
  taking off formatting 163
  templates 174
  using the Format Painter 163
word count utilities 133
word counts
  differences 133
  in Word for Mac 135
  in Word XP 135
  PDF files in PractiCount 137
  PowerPoint 137
  problems 133
  reasons for differences 134
  strategies 134
  using Word 135
word processor
  definition 103
Word 2003
  Web Page, Filtered 113
Word 2007/2010
  realtime word count 135
Word 2011
  Wordfast support 197
Word 98
  for Mac 342
WordArt
  in word counts 135
Wordbee 297
  offline mode 233
  overview 231
  real-time collaboration 288
  support of PSD files 345
  support of subtitle files 425
  translation environment tool 230
  user interface 232
Wordfast
  quality assurance 279
  real-time collaboration 288
Wordfast Anywhere
  free 236
  OCR of PDFs 378
  overview 235
  translation environment tool 230
  using ABBYY OmniPage 379
Wordfast Classic
  history and overview 196
  realtime cooperation 288
  translation environment tool 196
  use of MT 278
  user interface 269
Wordfast Pro
  overview 215
  PDF support 374
  support of subtitle files 424
  translation environment tool 202
WordFinder 190
Wordfisher
  free program 219
WordPerfect 157

previously preferred file format 331
WordPress
localization 322
support by iLangL 325
using memoQ for translation 322
WordPress Multilingual Plugin 322
Word2000
Unicode support 166
work environment
of TEnTs 268
workspace
Wordfast Anywhere 236
WorldLingo
in TEnTs 276
Worm 76
Worx
project management 327
WYSIWG
What-You-See-Is-What-You-Get in TEnTs 269
WYSIWYG
environment in TEnTs 270

X

Xbench 302, 309
QA features 291, 294
support for Microsoft TBX 104
Xbox Game bar 53
Xcode 320
Xerox Terminology Suite
features 304
XLIFF
exchange format 286
limitations 286
XTM Cloud 233
XLIFF Editor
Translation Workspace 200
XLIFF 2.0 286
XLIFF2.0 286
XML
definition 158
tagged formats 349
with embedded HTML 353
XnView 115
XSL
XML extension 350
XTM 297
alignment 255
MQM support 284
XTM Cloud
overview 233
partnership with Easyling 324
spartan interface 233
support of PSD files 345
translation environment tool 230
XTRF
project management 327

Y

Yandex 273
in TEnTs 274, 275, 276
YouAlign
free alignment 254
Youdao
in TEnTs 276
YouSendIt 73

Z

zip files
protecting with passwords 127
zip utilities 125
zipping
different file formats 125
right-click access 126
split feature 127
zipping utilities
variety of 125
Zombie 81
ZoneAlarm 91

Symbols

1-4a Rename
renaming tool 117
3D Flip
Windows Vista and 7 24
7-Zip 126