CAFETRAN ESPRESSO - FILE FORMATS

Updated for CafeTran Espresso 2019 - Croissant 10.0 (20190117)

This document aims to cover natively supported file formats in CafeTran, provide compatibility information for external bilingual projects (files and packages), and discuss solutions to handle not yet supported file types.

To keep track of the latest CafeTran developments, review (and subscribe to) the official update annoucements or the aggregated ChangeLog.

Curated by Jean Dimitriadis (EN-FR/EL-FR translator).

1. SUPPORTED FILE FORMATS AND WORKFLOWS

Be it office suite documents, web, DTP, AutoCAD or localization file types, CafeTran supports natively a variety of different file formats with its default "**Translate document**" workflow.

It also offers a "**Multiple document folder**" option that allows you to translate and filter multiple files and file types within a folder (which may include nested folders).

With its "**Translate external projects**" workflow, CafeTran lets you further leverage its advantages as well as your own resources by handling a host of bilingual files and packages that originate from other CAT tools, all without leaving your favorite program.

This wide compatibility with external bilingual files may come in handy if you wish to prepare files on a different CAT tool which handles **file types not yet supported by CafeTran** or offers **additional filter options** for already supported formats, so that you can widen your net even more while still performing the bulk of the translation in CafeTran.

Finally, additional workflows make it possible to translate images and scanned/paper documents ("**Translate Paper document**"), work with PDFs and source document edited directly at the application designed for this document type (**Translate through Clipboard**), edit and perform maintenance tasks on TMX memories (**Edit Translation Memory**) or quickly align short documents to produce reference TMs (**Align two documents**).

File formats overview

OFFICE	WEB	EXTERNAL PROJECTS
MS Word .docx, .docm	.html	.xliff, .xlf
OO/LO Writer .odt	.xhtml	.sdlxliff, .sdlppx (SDL Trados)
iWork .pages	.xml	.ttx (Trados TagEditor)
AbiWord .abw	Tagged .xlsx	.mqxliff, .mqxlz (MemoQ)
MS Excel .xlsx, .xlsm	**DTP**	.txml (Wordfast Pro 3)
OO/LO Calc .ods	FrameMaker .mif	.txlf (Wordfast Pro 4/5)
iWork .numbers	InDesign .inx	.xlf (Déjà Vu External View)
MS PowerPoint .pptx	InDesign .idml	.mxliff (Memsource)
OO/LO Impress .odp	QX Tagged DTP	.ppf & Transit XML (Star Transit)*
iWork .key(note)	**AUTOCAD**	.xlf (MateCat), .xliff (Crowdin, Smartcat)
OO/LO Draw .odg	.dfx	Unclean .docx, .odt*
Plain text .txt	**LOCALIZATION**	**Paper document workflow**
OTHER	.properties	Images (.jpg, gif, .bmp, .png)
Vector Image .svg	.resx	**Edit TM workflow**
SUBTITLES	.xml	.tmx
SubRip Text .srt	.strings	**Clipboard workflow**
iTunes Timed Text .itt	.loc	.pdf, external documents

CT uses a native filter for external formats marked with an asterisk.

Note: The external projects compatibility list does not reflect an official statement.

CafeTran's own file formats

CafeTran embraces easy formats and open standards for its own internal operations. It uses:

- **Projects**: industry standard **XLIFF** (XML Localization Interchange File Format) files
- **Translation Memories** and fragment bases: **TMX** (Translation Memory Exchange) files
- **Segmentation rules**: **SRX** (Segmentation Rules Exchange) file format
- **Glossaries**: simple **tab-delimited TXT** files

It also features a powerful **TMX editor** and can import **TBX** (TermBase Exchange) glossaries, among many other resource types.

CafeTran can export any native project to a compressed **CTP package** Project > Export and exchange > To Package. This zipped file contains the project's XLIFF, the source files, the TMX and glossary files, as well as any other document contained in the project folder (exported target files, bilingual review documents, etc.) for easy sharing.

Last but not least, CafeTran offers an excellent **Export bilingual document (.docx)** (with or without notes) feature for native projects, which allows you to review the translation in a DOCX file and import the changes and additional notes back into CafeTran. External projects can also benefit from the same feature (minus the import functionality), and all projects can be Exported to HTML (along with other formats), which produces a bilingual table with notes in an HTML file that can be viewed in MS Word or OpenOffice/LibreOffice Writer.

CafeTran's File type filters

File type: Ms Word (*.docx/xml)

Here is the list of native file type filters CafeTran offers when creating a new project ("Translate document"):

- Text (.txt)
- OpenOffice (.odt/ods/odp/odg)
- MS Word (.docx/xml)
- MS Word OCR (.docx/xml)
- MS Excel (.xlsx)
- MS PowerPoint (.pptx)
- iWork (.pages/key/numbers)
- XML
- HTML
- XML/HTML with tags
- AbiWord (.abw)
- MS Word Uncleaned (.docx/xml)
- OpenOffice Uncleaned (.odt)
- AutoCAD (.dfx)
- InDesign (.inx)
- InDesign (.idml)
- FrameMaker (.mif)
- Subtitles (.srt)
- Subtitles (.itt)
- Java Properties (.properties)
- Windows .NET Resources (.resx)
- Mac String Resources (.strings)
- String resource header (.loc)
- DTP Tagged Text
- Tagged MS Excel (.xlsx)
- Transit XML File
- Vector image (.svg)
- Image (.jpg/gif/bmp/png)

File types in detail

Supported file types are discussed in more detail in this section. Additional information will be added later on, as needed.

Microsoft Office

Only files from MS Office 2003 and later (.docx, .xlsx, .pptx) are natively supported by CafeTran.

For .doc, .xls and .ppt formats, check the Unsupported formats section below.

Password-protected files are not imported. You need to remove the protection before importing them into CafeTran.

MS Word (.docx)

Notes:

- MS Word files containing macros **.docm** are also handled via this filter (although the file type might need to be selected manually).
- Hidden content is not imported into CafeTran, which can be an effective solution for filtering out content that you don't wish to translate. Just remember to unmask hidden content when done. You can also use the TransTools Word add-in to [Hide/Unhide Text](#).
- CafeTran filters out of the deleted text in Ms Word source documents with the track changes enabled.

MS Word OCR (.docx/xml)

CafeTran offers a special filter to handle MS Word documents after OCR, which clears the source text of unnecessary formatting tags.

Note: You can also use this filter anytime the MS Word document produces many unnecessary tags.

Alternatives for tag cleaning:
[CodeZapper](#) (MS Word add-in)
[TransTools- Tag Cleaner](#) (MS Word add-in). Tag Cleaner is part of the [TransTools](#) suite featuring many useful utilities for preparing different Microsoft Office documents.

Related links:
[Handling Tag Soup](#)
[Word Documents After Optical Character Recognition](#)

MS Word Uncleaned (.docx/xml)

Use this filter for Trados RTF/DOC Uncleaned documents (converted to DOCX). Uncleaned documents from other tools should work as well, please report if otherwise.

MS Excel (.xlsx)

Notes:

- MS Excel files containing macros **.xlsm** are also handled via this filter (although the file type might need to be selected manually).

Document settings	Memory & glossary	Project properties	Filter options

Import source column (A, B, C or A:B etc.): _____

Import target column (A, B, C etc.): _____

Import notes column (A, B, C etc.): _____

Export column (A, B, C etc.): _____

- When you choose the **Excel filter**, you have access to the following Filter options:

  - **Import source column**: allows you to specify from which column(s) to import source text
  - **Import target column**: allows you to specify from which column(s) you wish to import content in the translated segments. Note: If you define two Excel columns for import separately (in two Import fields of the filter), the second column is to be imported for the review of the existing translation. Only existing rows (not empty cells) will be imported if you define an Import target column. If you specify two import columns separately, leave the export column empty.
  - **Import notes column:** allows you to define a column that will be imported in the form of Notes in CafeTran.
  - **Export column**: allows you to specify to which column(s) you want to export your translation.

- CafeTran imports only visible (not hidden) columns, rows or sheets. Hiding these in Excel before starting a project allows you to filter out content that you don't wish to translate.

- [Custom target segment text formatting](#) is not supported.

Tagged MS Excel (.xlsx)

The Tagged MS Excel filter is designed for handling certain GUI resource file types using XML/HTML code, where strings are surrounded by tags.

Note: Final file should be checked against any missing parts in code, as some HTML codes can be interpreted or deleted.

MS PowerPoint (.pptx)

OpenOffice/LibreOffice

Writer (.odt)

Notes: CafeTran filters out of the deleted text in LibreOffice source documents with the track changes enabled.

OpenOffice Uncleaned (.odt)

Use this filter for Trados RTF/DOC Uncleaned documents (converted to ODT). Uncleaned documents from other tools should work as well, please report if otherwise.

Calc (.ods)

Impress (.odp)

Draw (.odg)

AbiWord (.abw)

Apple iWork

To translate Apple iWork documents, you need to save them in the older Pages/Numbers/Keynote '09 file format:

Upon creating a CafeTran project using the iWork filter, you will get a dialog. Just choose the *index.xml* file and press OK to load the segments.
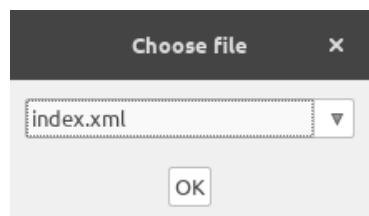
Related links:

[Open older iWork files in newer versions of Pages, Numbers, and Keynote](#)

iWork Pages (.pages)

[Import and export different file formats with Pages](#)

iWork Numbers (.numbers)

[Import and export different file formats with Numbers](#)

iWork Keynote (.key)

[Import and export different file formats with Keynote](#)

Web localization/XML files

XML

Suggestions:

- Enabling Preferences > Segment at all tags along with your segmentation rules might be a good solution for some XML files.
- Since this file type does not yet support filter options, if the text to be translated is not displayed or filtered in CafeTran, you might need to use an external tool (such as Memsource, memoQ, Okapi Framework, etc.) to prepare the XML file (see tools in Solutions section).

HTML

This filter also handles XHTML files.

Notes:

- CafeTran does not recognize the alt tag.

Suggestions:

- Since this file type does not yet support filter options, if the text to be translated is not displayed or filtered in CafeTran, you might need to use an external tool (such as Memsource, memoQ, Okapi Framework, etc.) to prepare the HTML file.

XML/HTML with tags

Choose this filter to include the text inside the XML/HTML tags to the displayed segments.

Transit XML File

See PPF and Transit XML files under External projects section.

Tagged MS Excel (.xlsx)

The Tagged MS Excel filter is designed for handling certain GUI resource file types using XML/HTML code, where strings are surrounded by tags.

Note: Final file should be checked against any missing parts in code, as some HTML codes can be interpreted or deleted.

DTP

InDesign (.inx)

InDesign Interchange (INX) is an older format, primarily used for saving documents back to a previous version of InDesign (CS3).

Note: For handling Adobe InDesign **INDD** files directly, see the Unsupported formats section.

InDesign (.idml)

The Adobe InDesign Markup Language (IDML) format replaces the Interchange (INX) format used for saving backwards in previous versions.

While INDD represents the extension given to all InDesign files created in the standard way in any version of the design software, IDML (InDesign Markup) is an XML format which allows users of different versions of InDesign to open up the file without a problem.

Typically, IDML files are smaller than INDD files, since non-embedded images aren't included. IDML is the standard format for translating InDesign files.

Note: For handling Adobe InDesign **INDD** files directly, see the Unsupported formats section.

FrameMaker (.mif)

Adobe FrameMaker is an authoring and publishing solution for XML and Maker Interchange Format (MIF) is the markup language Adobe FrameMaker uses for technical document.

DTP Tagged Text

CafeTran handles Quark Express DTP Tagged Text files.

Vector image (.svg)

Scalable Vector Graphics (SVG) is an open standard XML-based vector image format for two-dimensional graphics with support for interactivity and animation.

SVG images and their behaviors are defined in XML text files. They can be produced by Inkscape [free and open source software], Adobe Illustrator, Adobe Flash Professional, or CorelDRAW, and rendered to common raster image formats such as PNG using the same software.

NOTE: Since Adobe Illustrator can import and export SVG files, an SVG filter can help handle Adobe Illustrator files (*.ai or *.eps) as well as DWG or CAD files, exported to SVG by Adobe Illustrator. For handling AI files using the SVG file filter, see the Unsupported formats section.

AutoCAD

AutoCAD, developed by Autodesk, Inc. and first released in December 1982, is a software application for computer-aided design (CAD) and drafting that supports both 2D and 3D formats.

AutoCAD (.dxf)

A DXF (Drawing eXchange Format) file is a type of drawing interchange file used to transfer data between various applications. It is either a binary or an ASCII representation of a drawing file, often used to share drawing data between other CAD programs.

In AutoCAD, you can convert a DXF file to DWG format by opening the file and saving it in DWG format. You can then work with the resulting drawing file as you would with any other drawing file.

Exporting a DXF file in AutoCAD.

Be sure to visit the Filter options tab in the New document creation dialog, in order to check which of the available layers you wish to filter.
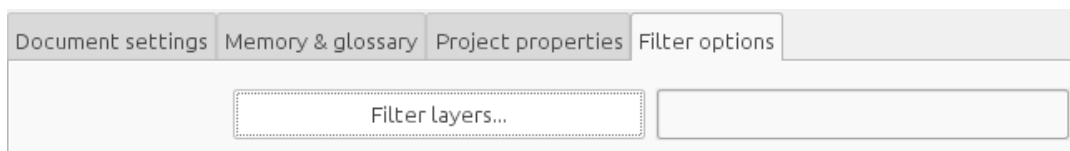
NOTE: For handling the binary **DWG** files, see the Unsupported formats section.

Software localization

CafeTran provides some filters that allow you to localize various mobile and desktop software applications.

Java Properties (.properties)

Java Properties files are simple text files that are used in Java applications.

You can localize these within CafeTran with the help of this file filter.

Windows .NET Resources (.resx)

The file filter for NET XML Resource Template (.resx) files can be used for localizing .NET and Windows Phone applications.

Mac String Resources (.strings)

The Mac String Resources filter can be used to localize iOS or OS X applications.

The '/* comments */' in .stringd are displayed as Notes in CafeTran.

Android (and other) XML files (.xml)

The XML filter can be used to localize Android applications, among other content (since XML can have various applications).

String resource header (.loc)

String resource header files are used for localization of programs written in C or C++ language.

Subtitles

Subtitling offline or online software is often used not only to create, edit, and time code, but also to translate subtitles, including SubRip files, because it integrates video and audio context and offers various features specific to subtitling.

For those wishing to translate subtitling files in CafeTran Espresso, the developer has recently added support for SRT and ITT files. The program can also open videos (Resources > Open movie) and offers access to YouTube as a Web resource (Resources > Web > YouTube.)

Subtitles (.srt)

SubRip Text files.

Subtitles (.itt)

iTunes Timed Text files.

Paper document workflow

Image (.jpg/gif/bmp/png)

Images and scanned documents can be translated with the help of the "Translate Paper document" workflow.

For an OCR approach to images, see the PDF > OCR section in Unsupported formats.

Related links:

[Translation of Scanned Images](#)
[Translation of Paper Documents](#)

2. EXTERNAL PROJECTS

Intercompatibility (both in and out) is an important CafeTran Espresso feature. "Translate external project" workflow lets you handle a host of bilingual files and packages that originate from other CAT tools, all without leaving your favorite program. This section offers details on different external bilingual project types.

GENERAL NOTES ON BILINGUAL XLIFF FILES

"Translate external project" workflow presents some differences when compared to the native "Translate document" workflow. Here are some of these differences you may want to consider:

- **Export target file format**. For external projects, CafeTran does not export to the original target language document file format. Instead, it Finalizes the edited bilingual file(s). To final target file export is generally done by the client who created the translation project in their tool (or, in some cases, by the translator, if they use the same tool). This is a general limitation and is not inherent to CafeTran.

- **Export target file Preview**. For the same reason as above, CafeTran does not export a target file Preview for external projects.

- **Segmentation**. For external projects, CafeTran does not handle the segmentation phase as this takes place at the bilingual file creation. If you open an unsegmented file, CafeTran will show a warning message. Ask for properly prepared (segmented) files.

- **Split/Merge** segments. CafeTran only supports **virtual joining of segments** in external projects. The feature allows to join segments in segments' editor and rearrange their contents comfortably. As soon as the joined segments leave the editor, their number is the same as before joining but their respective text contents is changed (rearranged), hence the virtual aspect of this feature.

- **Project creation**. For native projects, CafeTran creates a Project folder, copies the source documents and generates a segmented native XLIFF, which it opens for translation. For external projects, **CafeTran edits the external bilingual files directly at their own directory location**. Make sure you have a copy if needed.

- **CafeTran does not remember/reopen the previously used resources for external projects**. Use [Project Templates](#) to achieve this.

- **Project memory/Project glossary**. If Project memory and/or Project glossary are checked in the Dashboard when opening an external project, the ProjectTM.tmx and ProjectTerms.txt files are conveniently created and placed at the same working folder.

- **Packages**. Supported packages (.sdlppx, .mqxlz) are expanded and edited at the working folder location. To reopen an already expanded package, just drag and drop the package file once again. You can export a package (or return package) via Project > Export and exchange > To Package.

- Export button becomes **Finalize button** in CafeTran's segments grid. This sets the Checked status (Translated in SDL Trados/Confirmed in memoQ) to all target segments and saves the project. Generally, this step should be performed prior to delivering the files. Note: If you need to keep a different status for some segments than the translated status, DON'T use the Finalize button. Instead, filter the segments you want to set as Translated, and apply the Task > Set translated status for target segments.

- **QA > Tags check** should be performed prior to finalizing the project, to ensure no tag issues are present in the delivered files. It is also recommended to run the **QA > At tags spaces check**.

- **TM match marks**. The translation memory match values are not stored in non-CafeTran project. They are available for the current session only, unless the projects provide their own values.

- **Segment statuses**. CafeTran attempts to map its own additional segment statuses (Checked, Approved, Locked) to the statuses applied by the external tool. More information can be found in each tool's section below.

- **Segment Notes (Comments).** CafeTran generally maps its Notes feature to the Comments supported by external tools. More information can be found in each tool's section below.

- **Extra formating tags** cannot be added (contrarily to native projects). Additional formatting will need to be done at final document export.

- **Export bilingual DOCX for review**. You can export bilingual DOCX files (with or without notes) for external review via Project > Export & Exchange menu. However, importing the changes directly back into CafeTran via Projects > Export & Exchange > Import bilingual file is reserved to native projects.

- **Auto-propagation** is supported (also across multiple files, if they are virtually glued together). However, the P and nP marks for auto-propagation and no propagation of individual segments are displayed **for the current session only**.

- **Review** jobs that rely on Track changes feature cannot be performed in CafeTran (CafeTran's own simple Version tracking cannot be used for this purpose). However, an solution such as the online [TQAuditor](#) can be used to produce a changes report which highlights the document corrections. CafeTran does map Translator Approved and Translator Rejected SDL Trados statuses.

- Related links:
  [Project File Formats Created in Other Tools](#)

SDL Trados Studio

SDL Trados Studio files and packages are very well supported in CafeTran.

CafeTran Espresso can also import SDLTM translation memories and SDLTB termbases, which are converted into the TMX format.

In addition to the general notes above, please consider the following **known limitations**:

- Tags at the start of the source segment cannot be moved around in the target segment.
- Trados Comments are shown as Notes in CT and CT Notes are imported as Comments in SDL Trados Studio (but do not get the levels FYI, Warning, and Error). In CT, segments with Notes/Comments have

a leading and an ending tag. This is normal.

- Alternative translations work for the current session only in Studio projects. If they interfere with SDL Trados, you can try removing them via Task > Remove alternative translations in CafeTran.

Segment status:

In the screenshot below, you see how the various SDL Trados statuses mentioned in the green Notes are displayed in CafeTran.

Exemples de questionnaires	Exemples de questionnaires	1
Translated		N
Des bases solides pour réussir votre étude de marché	Des bases solides pour réussir votre étude de marché	2
No status		N
Exemples de questionnaires	Exemples de questionnaires	3 100%
100% match		N
Envie de partir d'une base solide pour structurer votre questionnaire ?	Envie de partir d'une base solide pour structurer votre questionnaire ?	4
Translation approved		N
Nos exemples de questionnaires sont là pour ça !	Nos exemples de questionnaires sont là pour ça !	5
Signed-off		N
Tests de concept & mesure du potentiel de votre produit	Tests de concept & mesure du potentiel de votre produit	6 L
Locked		N
Mesurez l'attractivité et le potentiel de votre produit en amont de vos développements.	Mesurez l'attractivité et le potentiel de votre produit en amont de vos développements.	7
Draft		N
Obtenez rapidement une estimation de la pénétration de votre cible et de la taille de votre marché.	Obtenez rapidement une estimation de la pénétration de votre cible et de la taille de votre marché.	8
Translation rejected		N
Apprenez-en plus sur le proSl et les motivations des personnes intéressées par votre produit.	Apprenez-en plus sur le proSl et les motivations des personnes intéressées par votre produit.	9
Sign-offrejected		N

- "Translated" segments in SDL Trados get the "Checked" status in CafeTran (gray background in segments, plus simple gray background in segment number) and vice versa. This is the status segments get when you press the Finalize button in CafeTran.
- "Translation approved" segments in SDL Trados appear as "Approved" in CafeTran (gray segment background, plus gray segment number background with green frame).
- "Translation rejected" segments in SDL Trados appear as "Rejected" in CafeTran (no segment background, plus gray segment number background with red frame).
- "Signed-off" segments in SDL Trados do not have a specific status in CafeTran.
- "Sign-off rejected" segments do not have a specific status in CafeTran.
- "Locked"/"Complete Sign-off" (marked with PM and a lock in SDL Trados) have the Locked status in CafeTran.
- Context Matches (CM) and 100% matches have the 100% match status in CafeTran.
- Fuzzy matches in SDL Trados have the same match mark in CafeTran (but this is not so the other way around).

- You have the option to omit setting the Translated status for target segments during the export to .sdlrpx packages and apply a statuses to filtered segments.

Note: If you mark Studio segments as Locked in CafeTran, this status is active for the current CT session only. The CafeTran status does not interfere with the locked segment status (Complete sign-off) that is set in Studio by the project manager.

Known Workarounds:

- If Trados files that have been Finalized in CafeTran do not get the Translated status back in Studio (they show as untranslated in the statistics, although their segments have the "translated" icon), just run Batch Tasks > Translation Count on the project to get the Translated status to 100%. This information can be shared with the PM if needed, as it requires virtually no effort.

SDLXLIFF

SDLXLIFF is the SDL Trados version of the bilingual XLIFF format. Please refer to the above general notes on bilingual XLIFF files and read more specific details concerning SDL Trados files in CafeTran.

SDLPPX

- **Opening SDLPPX packages**
  When you drag and drop an SDLPPX package on CafeTran's Dashboard, the program unpacks it automatically and loads the included SDLXLIFF files, and imports any attached SDLTM translation memory and SDLTB termbase in the form of a (read-only) TMX memory.

You need to check the unpacked package to see if it contains any additional material that could be useful to you (reference material, project analysis, etc.).

- **Exporting SDLPPX packages**
  After you finish the translation of SDLPPX package, finalize it and export it back to package via Project > Export > To package. Then, you have two options for the export: Project package or Return package.

The translated return package (SDLRPX) should be opened in SDL Trados Studio which created the normal (SDLPPX) package. If you did not create this project yourself, export to as a normal project package (SDLPPX) to review it in your installation of SDL Trados Studio.

SDLTM translation memories

CafeTran handles SDLTM translation memories, which are imported as read-only TMX files in CafeTran. The process is very quick and painless.

You can do so manually via Memory > Import > Import SDLTM memory.

SDLTM files present in Trados packages (SDLPPX) are automatically imported as a TMX in CafeTran when you open a package. The TM options dialog is shown, allowing you to fine-tune their options or apply additional filtering. All SDLTM files present in the package are joined together during the import. To open them separately, use the manual process above.

In the SDLTB section below, you can find external utilities that also handle SDTM conversion.

SDLTB termbases

CafeTran handles SDLTB terminology files, which are imported as TMX termbases into CafeTran.

You can do so manually via Memory > Import > Import SDLTB termbase.

SDLTB files present in Trados packages (SDLPPX) are automatically imported as a TMX in CafeTran when you open a package. The TM options dialog is shown, allowing you to fine-tune their options or apply additional filtering. All SDLTB files present in the package are joined together during the import. To open them separately, use the manual process above.

If you prefer to convert the imported SDLTB termbase to the glossary format, proceed as follows: After importing the SDLTB, select the glossary in which you wish to import the converted entries, then go to Memory > Export > Export segments to glossary. You will be asked to select the memory you wish to export from in the next dialog.

External SDLTM and SDLTB conversion utilities

Although their use is now redundant, since CafeTran already handles SDLTM and SDLTB files on its own, here are some additional utilities that offer SDLTM and SDLTB conversion capabilities outside CafeTran:

- Van Nellen's Trados Studio Resource Converter is a free cross-platform Java app that offers conversion from both SDLTM and SDLTB files.
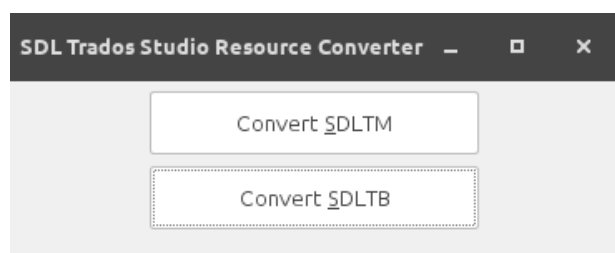
- WfConverter by Jean-Philippe Odent.
- Glossary Converter, available as an SDL Trados plugin and as a standalone Windows program. It can serve to convert term bases from and to SDLTB.
- XBench QA and Terminology Management tool version 3.x (paid, yearly licencing plan) can open SDLTB files and convert them to either TMX or tab -delimited TXT format.

TRADOS UNCLEAN RTF/DOC

These older SDL Trados files should be handled with the MS Word Unclean file filter, once you save them as DOCX files.

If you own a SDL Trados licence, you can also use the SDLXLIFF to Legacy Converter to deal with those using a round trip.

TRADOS WORKBENCH - TTX

TRADOStag files (TTX) are bilingual files used in Trados TagEditor, part of SDL's legacy tool called Trados Translator's Workbench.

These files need to be pre-segmented (Tools > Translate > Segment unknown sentences in Trados) before you can translate them in CafeTran.

The process is described in detail in this [article.](#)

TRADOS WORKBENCH - RTF

Trados Workbench unclean RTF files should be handled with the MS Word Unclean file filter, once you save them as DOCX files. These should have been pre-segmented first in Trados Workbench.

IMPORTING CAFETRAN'S TMX TO STUDIO

You can import CafeTran's and other TMX translation memories to SDL Trados Studio using the Import wizard. Imported data contains translation units with their associated fields.

In order to import data into a translation memory, the translation memory must already exist and be displayed in the navigation pane.

To display this wizard for the active translation memory, select File > Import from the menu bar in the Translation Memories view.

Related links:
[Overview: Importing and Exporting Translation Memory Data](#)
[Overview: Import Wizard](#)

CONVERTING CAFETRAN GLOSSARIES TO SDLTB

[Glossary Converter](#), available as an SDL Trados plugin and as a standalone Windows program can be used to convert CafeTran glossaries into SDLTB term bases. It needs to be set up correctly before the conversion. Instructions to follow.

memoQ

memoQ's MQXLIFF and MQXLZ files are well supported in CafeTran, although MQOUT packages require additional care.

Known limitations:

Segment status:

- Edited and/or Pre-translated segments in memoQ appear as Translated in CafeTran.
- (Translator) Confirmed segments in memoQ are marked as Checked in CafeTran. Finalizing in CafeTran will set the status of all segments to Translator Confirmed in MemoQ.
- Warning: 100% segments in memoQ are invisible in CafeTran. Have the original file and/or a bilingual RTF table at hand for context.
- Warning: Locked segments in memoQ are invisible in CafeTran. Have the original file and/or a bilingual RTF table at hand for context.

Known workarounds:

MQXLIFF

MQXLIFF is a memoQ-flavored XLIFF file format. It is well handled by CafeTran (please refer to the above general notes on bilingual XLIFF files and read more specific details concerning memoQ).

Note: CafeTran now displays the match percentage for the target segments included in .mqxliff translation projects.

MQXLZ

MQXLZ are memoQ's zip packages containing MQXLIFF files. MQXLZ files are unpacked for translation and zipped back by CafeTran via Project > Export and exchange > To Package (without overwriting the original [unpacked] .mqxlz package file).

MQOUT

MQOUT are memoQ's project packages.

There are several ways to receive MQOUT packages. By e-mail or another electronic mean, in memoQ via Project > Check Out Online Project, or on a web browser, via WebTrans.

NOTE: As described below, handling MQOUT requires some additional steps, which may represent a hassle. If possible, ask to receive MQXLZ (or MQXLIFF) files and TM/TB separately, and to simply send back the translated MQXLZ or MQXLIFF files.

*The following is based on a memoQ Package Guide. *

A memoQ handoff (.mqout) package is a zip in disguise. Unzipping it reveals at least three subfolders:

- Translation documents = Contains the documents to be translated, in a zipped XLIFF format (MQXLZ) which includes the MQXLIFF files. You do not need to unpack the MQXLZ file, CafeTran does it for you. Just drag and drop it on the Dashboard.
- TM = Contains the translation memories used in the project in TMX format, which can be imported and opened in CafeTran.
- TB = Contains the term bases used in the project, in CSV format. These require a small preparation before being able to import them in CafeTran.

How to handle MQOUT packages:

- Unpack the contents of the .mqout file. Depending on your software, you may need to rename .mqout to .zip for that.
- Drag and drop the MQXLZ file (containing the MQXLIFF files) in CafeTran's Dashboard, to start working on the project.
- Open the TMX files in CafeTran. Optionally set them as read-only, just for reference.
- Add the converted CSV term base as a CafeTran glossary (tab-delimited .txt file).
- Translate the project in CafeTran. Run QA.
- Save the translation with the Finalize button.
- Export .mqxlz package via Project > Export and exchange > To Package (without overwriting the original [unpacked] .mqxlz package file).

- You now need to create the handback file (.mqback), containing only the translated files, no TM or TB. This is described below.

Packing up the translated XLIFF files in a memoQ Handback (MQBACK) package:

- When you have finished translating the XLIFF files, be sure to save them in a new, empty folder. Preferably, name this folder ProjectName.YourName. Then follow these steps:
    - In the original handoff package that you unpacked in Section 2.2, there is a file called HandoffPackageInfo.xml. Copy this file to the new folder, and rename the second copy to DeliveryPackageInfo.xml.
    - Open the new DeliveryPackageInfo.xml file in a plain text editor such as Notepad.
    - Delete almost all lines from the file, but do not delete the lines going from <?xml version="1.0"?> to the closing tag of HandoffPackageInfo.
    - In the file, change <HandoffPackageInfo to <HandoffDeliveryPackageInfo. Change <UserName to <User.
    - Save the file.
    - Create a zip file from the contents of this folder. (It should now contain the translated MQXLIFF files and the DeliveryPackageInfo.xml file.)
    - Compress the folder to a .zip. Rename the .zip file so that it is called Project-Name.YourName.mqback.
    - Send the file to your project manager. The project manager will be able to update the master project the same way as he/she would when he/she received a handback package directly from memoQ.

memoQ online projects

Information taken from the [old wiki](#):

- When handling online projects, many aspects depend on the export/import rights the PM assigns to you on the server settings.
- In some cases server files cannot be exported or cannot be re-imported. There is a simple trick to export and re-import server files: Simply click on Create view on the Document tab in memoQ. This view can be exported and re-imported.
- Please bear in mind that in most cases TMs and/or TBs cannot be exported from memoQ server projects. And the PM won't see any progress until you re-import your file.
- Note that memoQ translator free (the edition you get after your trial version ends) cannot be used in memoQ online projects, although it can be used for accepting jobs from a memoQ server.

memoQ bilingual RTF files

MQXLIFF, and MQXLZ files are the recommended way to receive memoQ projects in CafeTran. Please ask for one of these formats instead.

For instructions on how to handle bilingual RTF files exported from memoQ, refer to this [old wiki article](#).

Also see the RTF entry in Unsupported formats section below.

memoQ Translation Memories

memoQ Translation memories can be exported to TMX from a memoQ installation. TMX can also be used to import TMs into memoQ.

TMs included in MQOUT packages are already in the TMX format.

memoQ Term Bases

To prepare Term Bases exported via memoQ or received as CSV via a MQOUT package:

- Open the CSV file with a spreadsheet program (MS Excel, LibreOffice Calc, etc.).
- Look for the two columns that match your source and target language and copy them to a new worksheet, arranging them in that correct order (source terms - target terms).
- You can then copy-paste the two columns into a plain .txt file with UTF-8 encoding and open the file as a glossary in CafeTran (the first line of each column should include the language beginning with a #. Example: #en-US, #en-GB or #fr-FR, to match the language pair you are working on).

For information on how to migrate Term Bases from memoQ or convert them for use in CafeTran, you can refer to this [old wiki article](#).

Wordfast Pro

WFP 3 files are well supported in CafeTran, while WFP 4/5 currently less so.

WFP 3 - TXML

TXML is the bilingual file format used by Wordfast Pro 3.

It is supported by CafeTran, as part of the "Translate external project" workflow.

Notes:

- CafeTran supports comments in .txml files.
- CafeTran displays the match score information for the target segments included om .txml translation projects.

WFP 4/5 - TXLF

Wordfast Pro 4 and Wordfast Pro 5 use TXLF bilingual file format.

This file format is handled by CafeTran, as part of the "Translate external project" workflow, with the following limitations.

Segment status, known limitations:

- Segments with confirmed/Unconfirmed and Locked status all show as simply Translated in CafeTran.
- WFP segment Notes are not mapped to CafeTran Notes
- Finalizing the TXLF file in CafeTran does not change the status in WFP 4/5.

WFP 4/5 - GLP

Wordfast Pro 4/5 zipped packages have the .glp file extension.

They are not handled directly by CafeTran and need to be expanded.

Translate the TXLF files contained in the folder bearing the same title.

RTF/DOC/DOCX Bilingual Review

These Wordfast Pro exported files are usually meant to be used for external review or for exchange with Wordfast Classic or for external review in MS Word. Whenever possible, ask for a TXML/TXLF file instead.

You may want to try CafeTran's MS Word Uncleaned file filter if you wish to translate these files.

Related links:
[Wordfast Pro - Bilingual Export](#)

Wordfast TXT TMs

Memory > Import tab delimited memory menu option allows you to import segments from a Wordfast tab-delimited memory file to the selected translation memory in CafeTran.

Note: Wordfast Pro can also export its memory to the TMX format that CafeTran can open or import directly.

Related links:
[Exchanging TMX TMs with Non-Wordfast Users](#)

Wordfast Glossaries

Wordfast Pro can export Glossaries as tab-delimited TXT files or TBX files. Both file formats are handled by CafeTran. Tab-delimited TXT might need some preparation, but is otherwise used as CafeTran's native Glossary format. TBX can be imported via TMX in Memory > Import TBX terminology.

Déjà Vu

[Atril's Déjà Vu](#) CAT tool offers the [External View](#) (EV), a format specifically created for the purposes of proofing and outsourcing translation work and one of the first implementations of bilingual review files.

Déjà Vu X3 Professional can export External View files to RTF and XLIFF. Proofed or outsourced files can then be imported back into Déjà Vu.

CafeTran can handle External View files, with some limitations, although it cannot be used to translate Satellite files and Pack & Go packages, as explained below.

Related links:
[Déjà Vu - External View](#) (Contains seemingly outdated information about XLIFF availability in Déjà Vu X3 Professional, as opposed to Déjà Vu X3 Workgroup edition).

Known limitations:

- Déjà Vu Comments are not mapped to CafeTran Notes.

- Déjà Vu segment statuses are not differentiated in CafeTran and CafeTran does not change their status back in Déjà Vu, even after the Finalize button is clicked (see below). You will need to change the segment status to Translated in Déjà Vu.

Segment status:

- Segments marked as Translated, Proofread, Approved, Pending, Do not send (or simply Edited, with no status) all appear as simply translated in CafeTran.
- Warning: Locked segments in Déjà Vu are invisible in CafeTran (there is also an option to Suppress Locked segments, etc. when exporting an External View file). Have the original file and/or a bilingual RTF table at hand for context.
- Segments with 100% matches are not marked as such in CafeTran.
- Finalizing segments in CafeTran does not set a specific status in Déjà Vu. Segments are shown as Edited (no status). Furthermore, previous status (Translated, Proofread, Approved, Pending, Do not send, 100% match) is kept, even if the segments have been edited in CafeTran. You will need to change the segment status to Translated in Déjà Vu.

Known workarounds:

- According to this [old wiki article](), you can update comments to show CafeTran's Notes in Déjà Vu by using an SQL command in Déjà Vu.
- According to the same article, If a project contains more than one file, check Project > Documents in CafeTran and glue all files as needed. By default the last file is selected.

External View XLF

CafeTran handles External View XLIFF files (and their tags), with the segment status limitations explained above.

External View bilingual RTF tables

RTF files are mostly meant for users without a translation tool. XLIFF is the preferred format for users of another translation tool. Please ask for a XLIFF file instead.

For instructions on how to handle External View bilingual RTF files from Déjà Vu, refer to this [old wiki article]().

Déjà Vu Satellite files (.dvsat)

After a Déjà Vu X3 Workgroup project has been created and the external files have been imported, it is possible to generate satellite work files. The free version of Déjà Vu (after the Déjà Vu Professional 30-day trial ends) can be used to view and work on satellite projects sent by translation firms).

A satellite file contains only one source and one target language (as opposed to a full project file, which typically consists of the source language plus a number of target languages) with all the information needed by the translator. This information can exclude the formatting and image code, thus significantly reducing the file size and making it easier to transmit.

A satellite file contains a number of automatic restrictions. It is not possible to delete files, split/join rows, edit the source, or lock/unlock rows.

If the security options for the originating project file are enabled, users of the satellite files will have to log in with a user name and password. This will give them access to a predefined security level, which in turn will either permit or prohibit the overwriting of translation entered by other users or the project owner.

Once the translation of a satellite is finished, it can be sent back to the project owner who can import it into its parent project file.

Note: As per the above, DVSAT files can only be handled from within Déjà Vu. Ask your client whether they can send you XLIFF External View files instead.

Related links:
[Déjà Vu - Documentation - User manual](#)

Déjà Vu Pack & Go packages (.dvpng)

Pack & Go packages are highly compressed exports from project files that are ideal for the transmission of Déjà Vu X3 Professional data. The Pack & Go feature is available as a standalone feature for the transfer of complete projects, or as part of the satellite creation for the transfer of even smaller bilingual satellite projects.

Opening a Pack & Go package with Déjà Vu X2 or Déjà Vu X3 will start the Pack & Go Wizard, which will allow you to define where you want to have your new project stored. When the translation of the project is finished, you can once again export it into the Pack & Go format to transmit it back to the project owner.

Note: As per the above, DVPNG files can only be handled from within Déjà Vu. Ask your client whether they can send you XLIFF External View files instead.

[Déjà Vu - Documentation - User manual](#)

Déjà Vu TMs (.dvmdb)

Déjà Vu uses DBMDB access database files for its native TMs and can import or export TMX files (which can be used by CafeTran), among other formats.

Related links:
[Déjà Vu - How To Export A Translation Memory To TMX](#)
[Déjà Vu - How To Use A TMX File In Déjà Vu](#)

Déjà Vu TBs (.dvtdb)

Déjà Vu uses DVTDB access database files for its native TBs and can import or export tab-delimited TXT glossary files (which can be used by CafeTran), among other formats.

Related links:
[Déjà Vu - How To Export A TB To CSV Format](#)
[Déjà Vu - How To Import A TXT Or CSV File Into A TB](#)

STAR Transit NXT Professional+

CafeTran offers a native Transit XML file filter for handling [STAR Transit NXT](#) translation projects.

PPF AND TRANSIT XML files

*The following procedure has been copied verbatim from CafeTran's Knowledge base, see link below. *

PPF packages are translation projects created by Transit CAT tool. They are zipped files which need to be unpacked before the translation in CafeTran. Follow these steps to prepare and translate a PPF package:

1. Rename the received .ppf package to the .zip package and unzip it to a folder.
2. Create a separate working folder outside the unpacked package.
3. Locate the source language files in the unpacked package. They end with the three-letter language code for your source language.
4. Copy the source language files to the working folder you created in step 2.
5. Run CafeTran and create a new project choosing a single source language file in the working folder or the whole working folder with several files. If you choose a folder instead of a single file, please select "Multiple document project" option in the Dashboard menu.
6. Choose Transit XML filter from the "File of type" list in the Project Configuration dialog and click the OK button.
7. Translate the project and perform QA > Tags check making sure that tags are correct.
8. Export the translation and go to CafeTran project's folder where the exported file(s) are located.
9. Change the three-letter source language code in the name of exported file(s) to the three-letter language code of your target language.
10. Copy the translated file(s) with the three-letter target language code to the unzipped PPF package folder.
11. Zip back the folder and change the name (extension) of the zipped package from .zip back to .ppf.
12. Send the translated .ppf package to your client.

Note: You can find three-letter language codes for your language pair here.

If PPF package contains *.TXE terminology files to use in the translation, please perform the following:

1. Choose Memory > Import TBX terminology... from CafeTran's menu.
2. Navigate to the folder with the *.TXE terminology file and choose it.
3. In Memory options, make sure "Fragments memory" option is selected.
4. Click the OK button to import *.TXE terminology to CafeTran memory.

Related links:
[Translation of PPF packages](#)

Transit TermStar terminology files - TXE

To import a TXE terminology file in CafeTran:

1. Choose Memory > Import TBX terminology... from CafeTran's menu.
2. Navigate to the folder with the *.TXE terminology file and choose it.
3. In Memory options, make sure "Fragments memory" option is selected.
4. Click the OK button to import *.TXE terminology to CafeTran memory.

Transit XV packages - PXF

The older version, Transit XV used the PXF extension for packages. You should be able to ask for a PPF file instead, in the unlikely event you receive a PXF package.

Memsource Cloud

CafeTran can handle MXLIFF files that you download from your own [Memsource](#) translator account (both Personal Edition and 1+ Freelancer Edition) or those you are allowed to download from an account enabled by a translation agency on a project basis (some clients/accounts seem to disable the download feature).

Translation status:

- Checked segments in CafeTran appear as Confirmed in Memsource and vice versa.

MXLIFF

Follow these steps to translate MXLIFF files in CafeTran:

1. Download the MXLIFF files from the Memsource Cloud account (Go to the project's dashboard, check the files you wish to download and click on the Download button).
2. Create a working folder and copy the bilingual MXLIFF files.
3. Translate the files in CafeTran and perform QA > Tags check making sure that tags are correct.
4. Run Task > Set checked status for target segments. CafeTran's "Checked" status is mapped to Memsource's "Confirm" status.
5. Launch Memsource Editor and connect with the same Memsource credentials (click [here](#) to download and install it on your computer), and open one MXLIFF file at a time.
6. (Recommended) Run Memsource's QA checks and/or run bilingual file export/import operation as needed.
7. Upload the MXLIFF file to the Memsource server via Document > Upload to server in Memsource Editor. Repeat for each file.
8. Finalize the project in Memsource Cloud.

This last step generally requires the following actions:

- Check that all files are 100% confirmed.
- if you need to upload files to a server upon project completion, use Download > Completed file.
- Click on Edit and change the Status to "Completed by Linguist".

Note: Access to Memsource's Translation Memories ™ and Termbases (TB) can only be achieved via Memsource Cloud or Memsource Editor, unless you are allowed to download them. In that case, TMs can be downloaded as TMX or XLSX and TBs as TBX or XLSX.

Handling Memsource tags:
Tags used in Memource MXLIFF files appear to be some internal markings unrelated to the XLIFF format itself, but rather to the source document being translated. They are displayed in the form of {1> <1} or {1} and are not recognized as tags in CafeTran.

To make them easier to handle and insert in your target segments, you can add the following Regular Expression (regex) as a non-translatable fragment (Resources > Non-translatable fragments > Add selection to non-translatable fragments):

```
|[{<]\d+[>}]
```

Then, you will be able to easily place these tags with the F4 keyboard shortcut for inserting non-translatables.

To exclude these tags from memory matching (so that they don't hurt the TM fuzzy matching algorithm), you can also insert the above regex in the "Do not match" section of Preferences (Options) > Memory.

Related resources, for your reference:
[Memsource Editor User Manual](#)
[Memsource Cloud User Manual](#)

CROWDIN

Crowdin is an online localization project management platfom and translation tool for developers and website owners.

XLIFF

XLIFF files downloaded from Crowdin are properly handled using CafeTran's "Translate external project" workflow.

Crowdin's Context information is displayed as CafeTran Notes.

Note: Access to Crowdin's Translation Memories ™ and Termbases (TB) can only be achieved via Crowdin.

XTM Cloud

XTM Cloud is an online Translation Management Software.

XLIFF files downloaded from XTM Cloud are properly handled (including tags) by CafeTran's "Translate external project" workflow. They can be imported back into XTM Cloud by a Project Manager.

If TIPP packages containing multiple XLIFF files are used, ask for a Multi-XLIFF file instead (a single XLIFF file which basically joins all the XLIFF files inside), so that you can translate and upload translation for all files at once.

Note: Access to XTM Translation Memories ™ and Termbases (TB) can only be achieved via XTM Cloud.

XLIFF 2.0

XLIFF 2 is being developed by the OASIS XLIFF TC. The latest published version of XLIFF 2 can be found [here](#).

CafeTran uses an experimental filter for XLIFF 2.0 files. Report any issues to the developer.

Note: [Okapi XLIFF Toolkit](#) is a library that implements a set of Java components to create, read and manipulate XLIFF 2 documents. It can be used to validate a XLIFF 2.0 file (which can also be done online via this [XLIFF checker](#)).

Interoperability Now! TIPP and XLIFF:DOC

The TIPP package format (along with the XLIFF:DOC bilingual XML files it contains) is part of the Interoperability Now! initiative. It aspires to make it easier to pass projects and assignment packages among different translation tools. TIPP also has the objective to make resources and other settings interoperable among tools.

TIPP zipped packages cannot be simply unzipped, translated and zipped back.

However, XLIFF:doc files can be edited in CafeTran, so you can ask for those instead.

See the Unsupported section for more details.

Related links:
[Interoperability Now! project](#)

3. UNSUPPORTED FORMATS

You may come across translation projects which require handling various yet unsupported formats. This section lists and discusses different such formats (the list is bound to expand as needs arise) along with possible ways to deal with them. The next section (Solution) catalogs various tools you can use to prepare unsupported file formats and export them back to their original format after translating them in CafeTran, so be sure to check it as well.

PDF

PDF (Portable Document Format) represents a difficult format to deal with. You should ask your client for the editable file that generated the PDF file instead. If this is not possible, different strategies and solutions exist for editable and non-editable PDF documents, some of which are presented below.

- **CafeTran Espresso**

CafeTran does not offer OCR, but can handle editable PDFs with its **Translate through Clipboard** workflow (either exporting to MS Word, OO/LO Writer or directly pasting the translation into a dedicated PDF editor) and short non-editable (scanned) PDFs with its **Translate paper document** workflow.

Related links:
[Basic Clipboard Workflow](#)
[Translation of PDF Documents](#)
[Translation of Paper Documents](#)

- **OCR**

Especially for scanned (non-editable) PDFs and images, an obvious solution is to use an OCR program to scan the file and prepare an editable form of the source document (generally in MS Word or OpenOffice/LibreOffice format) that you then translate in CafeTran.

Note: OCR conversion, source file preparation and final layout editing require additional time and are generally billed as an extra service.

Because scanned PDF documents are routinely sent in for translation, obtaining an efficient OCR solution can often be a worthy investment. [ABBYY FineReader](#) (Windows, Mac) is the most popular choice (and, in version 14, also integrates the PDF editing capabilities of PDF Transformer+), but it's not the only one. Other offline paid solutions include [Adobe Acrobat Pro DC](#) (Windows, Mac, OCR, PDF editor, OCR, PDF to Word, Word, Excel or PowerPoint to PDF), [Nuance OmniPage](#) (Windows), [Readiris](#) (Windows, Mac) and [Wondershare PDFelement](#) (Windows, Mac, PDF editor, OCR, PDF to Word).

Free offline OCR solutions which make use of Tesseract, an open source OCR engine (which produces accurate results, but is less efficient with complex layouts), include: [FreeOCR](#) (Windows), [gImageReader](#) (GNU/Linux, Windows, also create searchable PDFs), VietOCR (Windows, Mac, GNU/Linux).

Online solutions: [FineReaderOnline](#) (10 page free trial, yearly paid plans, multiple languages supported, excellent results), [Convertio](#) (10 page free trial, API available, prepaid and API packages available, multiple languages supported, excellent results, identical to FineReaderOnline in my test), [Free Online OCR](#) (max. size 15 MB, good OCR recognition for basic layouts, uses Tesseract), [TM-Town](#) (online, uses Tesseract).

On the CAT tools front, [MateCat](#) (online, uses Nuance OmniPage SDK) and [Wordfast Anywhere](#) (online) offer PDF OCR conversion. [SDL Trados Studio](#) also offers OCR, powered by Solid Documents Technology, or via the IRIS PDF OCR plugin (powered by Readiris).

Inceni's [TransPDF](#) (free or pay as you go) online service tackles the PDF problem in a different way, via conversion to XLIFF and back, offering an OCR feature paid with credits along the way (see below for more details).

Note: Many PDF editors and OCR solutions also offer the possibility to run OCR on the PDF itself, in order to make it editable and/or searchable. Making the PDF editable allows you to further edit or convert the document. Making it searchable is interesting if you wish to use the scanned PDF for reference purposes.

Handling OCR tags

CafeTran offers a special filter to handle MS Word documents after OCR, which clears the source text of unnecessary formatting tags.

Alternatives for tag cleaning:
[CodeZapper](#) (MS Word add-in)
[TransTools- Tag Cleaner](#) (MS Word add-in). Tag Cleaner is part of the [TransTools](#) suite featuring many useful utilities for preparing various types of Microsoft Office documents.

- **EDITABLE PDF TO TEXT CONVERSION**

While OCR can still be used on editable PDFs, there are generally other effective methods for converting this type of PDF to text (DOCX, etc.).

Typically, the resulting file still needs source preparation before translation, or at least some final layout editing after file export.

Offline tools with PDF conversion capabilities include: [WPS PDF to Word Converter](#) (paid, free for 5 page documents, Windows, tested with excellent results), [Foxit's PhantomPDF](#) (paid, Windows, PDF editor, PDF to

Word), [Adobe Acrobat Pro DC](#) (paid, Windows, Mac, PDF editor, OCR, PDF to Word, Word, Excel, or PowerPoint to PDF), [Wondershare PDFelement](#) (paid, Windows, Mac, PDF editor, OCR, PDF to Word), [PDF Architect](#) (paid, Windows, PDF editor, OCR, PDF to Word), [Solid Converter](#) (paid, Windows and Mac), [Nitro](#) (paid, Windows, PDF editor, PDF to Word), [PDFSam](#) (paid for PDF to word, OCR and other stuff on Windows / free cross-platform utility for basic editing tasks, such as splitting and merging PDFs), [Ableword](#) (free, Windows, Word and PDF editor, PDF to Word), [Sejda Desktop](#) (free/paid, Windows, Mac, GNU/Linux, PDF editor, PDF to Word, Word to PDF, also online version) and [ABBY PDF Transformer+](#) (still available for a free 30-day /100 pages trial, but not offered for purchase as a separate product. Now its functionality is included in [ABBYY FineReader](#)).

[Microsoft Word](#) (2013 and later, on Windows) also offers the ability to turn a PDF into an editable document. [LibreOffice](#) does this as well, but text appears within text boxes, which only makes it usable for PDF editing, not document preparation.

For simple text extraction, you can also use the free [Adobe Reader](#) and other PDF readers.

Online:

- [Foxit PDF to word converter](#) produces excellent results.
- [CloudConvert](#) does a very nice job as well, with the added bonus that it also handles a host of different file conversion types and offers an API. It is used by MateCat.
- Other online converters, with less impressive results include: [Smallpdf](#) (PDF to word is based on Solid Documents, other PDF related actions and conversions are supported), [Convertio](#), [Zamzar](#) (many different conversion types), [PDFtoWord](#) (also offline as Nitro).

For command line enthusiasts, programs such as pdftotext, pdfreflow (pdftohtml), Calibre's e-book-convert, pdf2htmlEX, pdfbox can be added to the mix for PDF text extraction.

On the CAT tools front, [MateCat](#) (online) and [Wordfast Anywhere](#) (online) offer PDF conversion. [SDL Trados Studio](#) also handles editable PDFs directly.

Inceni's [TransPDF](#) (free or pay as you go) online service tackles the PDF problem in a different way, via conversion to XLIFF and back (see below for more details).

- **PDF TO XLIFF TO PDF - INCENI INFIX & TRANSPDF**

If you need to translate a PDF document with tricky layout that you need to replicate and deliver back into PDF (not DOCX) format, the following solution can be very efficient.

Iceni Infix is a PDF editor (with a monthly subscription or one-off payment) that offers the possibility to export text strings from editable PDFs as XLIFF files, then import them back and perform final adjustments. Infix 7 now uses a online TransPDF service or an offline legacy solution (with some drawbacks).

Related links:
Infix 7 - [Translating PDF](#)
Infix - [Pricing](#)

TransPDF is on online (free/pay as you go) service that aims to end the frustration of translating PDFs by converting them to good quality XLIFF for use with your own translation tools.

Simply upload your PDF to TransPDF and translate the XLIFF you get back using your own CAT tool (connectors exist for SDL Trados, Memsource and memoQ). Upload your translated XLIFF and you will get a

fully-formatted, translated PDF. Any post-editing can be done for free using Infix 7 PDF editor. This solution can also perform OCR (paid through credits) on scanned PDFs.

Note: By default, text is segmented per paragraph, although some tools, such as SDL Trados and Memsource, apply additional segmentation, which makes the resulting XLIFF far more usable. I would recommend performing a round trip via one of these two tools to take advantage of the improved segmentation. You can also try to resegment the XLIFF file from paragraph to sentence via the Okapi Framework.

On the plus side, the original layout is preserved and should only require limited PDF editing (although in language pairs with high expansion rate, keeping the same layout can be troublesome), greatly reducing the time needed for layout finalization/DTP work.

On the down side, the service has some price attached to it which, although quite low, represents an additional cost. Furthermore, even with the improved segmentation if you go for a roundrip through SDL Trados or Memsource, the file might present a few less than optimal segmentation situations that you cannot fix as you would in a native CafeTran project.

Still, TransPDF represents an interesting option you should definitely try!

Related links:
TransPDF - [Home page](#)
TransPDF - [Pricing](#)
TransPDF - [Help and support](#)
TransPDF - [PDF Translation, Step-by-step](#)
TransPDF - [Using Trados Studio for PDF translation](#)
TransPDF - [Using Memsource for PDF translation](#)

- **EXTRACT PDF TABLES - TABULA**

If you want to extract tables from editable PDFs (not only for translation, but also for terminology files), Tabula ([Windows, OS X, GNU/Linux](#)) is a great open source and free solution. Output is exported to CSV or Microsoft Excel file formats.

Note: There is an ongoing discussion for integrating Tesseract OCR capabilities into Tabula, allowing it to handle scanned PDF tables, stay tuned.

- **PDF EDITING**

PDF editing solutions are also useful to keep around for various tasks, including CafeTran's "Translate through Clipboard" workflow.

[Adobe Acrobat Pro DC](#) (Windows, Mac), [Nitro Pro](#), [Nuance Power PDF](#) (Windows, PDF editor, PDF to Word), [Foxit Phantom PDF](#) (Windows, PDF editor, PDF to Word), [Wondershare PDFelement](#) (paid, Windows, Mac, PDF editor, OCR, PDF to Word), [PDF Architect](#) (paid, Windows, PDF editor, OCR, PDF to Word), [Soda PDF](#) (Windows, online, PDF editor, PDF to Word), [iSkysoft PDF Editor](#) (paid, Windows, Mac), [FlexiPDF](#) (paid, Windows, PDF Editor, doc to PDF converter) and [Master PDF Editor](#) (free with watermarking/paid, Windows, Mac, and GNU/Linux // Linux free edition is 100% usable without PDF watermarking) are some of the available options. [ABBYY FineReader](#) (Windows, Mac), mostly known for OCR, now also integrates the PDF editing capabilities of PDF Transformer+.

On the (mostly) free front: Sejda Desktop (free/paid, Windows, Mac, GNU/Linux, PDF editor, PDF to Word, Word to PDF, also free/paid online version) AbleWord (free, Windows, Word and PDF editor), PDF-XChange Editor (Windows, Free and Plus versions available), LibreOffice Writer or Draw (free, Windows, Mac, GNU/Linux, text appears within text boxes), Inkscape (free, Windows, Mac, GNU/Linux, SVG editor, can also serve to edit PDFs - each page separately - and translate images).

Online PDF editors: Sejda PDF Editor.

- **CONVERT TO PDF**

Sometimes, you may need to deliver documents to PDF format as well.

Office suite applications (Microsoft Office [Windows, Mac, online], OpenOffice/LibreOffice [cross-platform], WPS Office [Windows, GNU/Linux], SoftMaker Office [cross-platform], Google Docs [online]) and many software applications can export a document they handle directly to PDF.

Most PDF editing tools mentioned above should also handle document to PDF conversion, such as Adobe Acrobat Pro DC (Windows, Mac), Nuance Power PDF (Windows), Foxit Phantom PDF (Windows), Wondershare PDFelement (Windows, Mac), PDF Architect (Windows), Soda PDF (Windows) and iSkysoft PDF Editor (Windows, Mac), FlexiPDF (Windows, PDF Editor, Word to PDF converter).
There are also apps like PrimoPDF by Nitro (Windows, free), Icecream PDF converter (Windows, free/paid).

Online: Soda Word to PDF, Smallpdf, as well as general (batch) file converters such as: Zamzar, CloudConvert, Convertio.

DOC / XLS / PPT

Microsoft Office 1997–2003 files (.doc, .xls, .ppt) are not directly supported in CafeTran.

You can either convert these files to DOCX, XLSX, PPTX before translating them in CafeTran or, if compatibility/presentation issues might arise, use a round trip solution (MateCat, Wordfast Pro 3 & 5, Okapi Framework, Memsource, SDL Trados Studio, memoQ, etc.).

RTF

Short for Rich Text Format, RTF is a text file format with only some additional basic formatting.

You will need to convert RTF to DOCX or ODT before translating it in CafeTran. After export, you can convert it back to RTF using MS Word or OpenOffice/LibreOffice. Conversion should be lossless.

CAT tools that allow you to handle RTF via a round trip: MateCat, Wordfast Pro 3, SDL Trados, MemoQ.

Bilingual RTF/DOC files / Multi-column RTF tables

RTF (or DOC) is being sometimes used as a bilingual document format by a number of CAT tools, such as: some versions of Trados ("External Review"), Wordfast Pro ("Bilingual Review"), Déjà Vu ("External View") and memoQ ("Two-column RTF" or Multi-column RTF tables).

In general, you will be better off asking for a bilingual XLIFF-flavored file.

Otherwise, please note additional file preparation and handling might be needed before proceeding to the translation in CafeTran.

For such Déjà Vu and memoQ files, you can access this old wiki article. See also below:

Multi-column RTF tables

Multi-column RTF tables are Rich Text Format documents that contain one or more memoQ documents.

	03_Termbase CAUTION: Do not change segment ID or source text MQ420511 49b761b7-0ca1-45a6-8ccf-2317c2b740a8			
ID	**English**	**Hungarian**	**Comment**	**Status**
1	Most translation jobs concern technical or other specialized text.	Most translation jobs concern technical or other specialized text.		Edited
2	In specialized texts, some words and expressions are very specific to the subject field: they refer to concepts and objects of the subject field.	In specialized texts, some words and expressions are very specific to the subject field: they refer to concepts and objects of the subject field.		Edited
3	Subject field experts tend to use these words and expressions consistently.			Not started
4	The set of such words and expressions is called the terminology.			Not started
5	The translator is seldom an expert of the subject field of the source text, so she is not familiar with the subject field terminology either.			Not started
6	Even if the translator recognizes a term in the text, it requires a lengthy research to find its translation.			Not started
7	Once found, the term should be written into a glossary, so that the same research won't be necessary again.			Not started
8	Translators often receive glossaries from the clients, containing previously unfamiliar words and expressions.			Not started
9	In this unit, you will learn how to build a glossary – a termbase – in memoQ, and how to use the glossaries received from your client.			Not started
10	It is important to clarify the meaning of the word 'terminology' because the definition tells you how to look for words or expressions that might need to be added to the glossary. memoQ stores the glossaries in databases called termbases: it is worth getting acquainted with their structure, too.			Not started

They are also used in Déjà Vu External View.

Related links:
memoQ - Multi-column RTF tables

PUB (Microsoft Publisher)

Microsoft Publisher is an entry-level desktop publishing application from Microsoft, differing from Microsoft Word in that the emphasis is placed on page layout and design rather than text composition and proofing.

This format is tricky to handle as it has few export options, Few CAT tools support it.

Fluency CAT tool (paid, Windows and OS X, offers a free 15-day trial) is able to handle .pub files and even export to xliff (File > Export Fluency XLIFF file), allowing you to work in CafeTran.

If you own the right edition of MS Publisher (It is part of the Microsoft Office suite), you can also use "Translate through Clipboard" workflow (explained above) to translate in CafeTran and paste the results in MS Publisher.

Finally, this interesting article @ Multifarious discusses a different solution for MS Publisher files.

VDX (Microsoft Visio)

Microsoft Visio is a diagramming and vector graphics application that is part of the Microsoft Office family.

If you have Microsoft Visio, you can use TransTools For Visio that allows you to extract translatable text from Visio drawings (all formats supported by Microsoft Visio) and to merge translations back into the original

Visio drawings.

CAT tools that allow you to handle Visio files via a round trip: Wordfast Pro 3, Okapi Framework, MemoQ, Deja Vu.

INDD (InDesign)

If you need to handle .indd instead of .idml files, [Language Terminal](#)'s free InDesign online converter allows you to work with InDesign in any CAT tool. Registration required.

Steps:

1. In Language terminal, drag and drop the INDD file in the InDesign conversion section.
2. Select source and target language, then press convert.
3. Once the conversion is completed, click on the file name to download a ZIP file.
4. Extracting the ZIP file produces an MQXLZ (memoQ ZIP) file and a reference PDF.
5. Change the MQXLZ extension to zip and unzip the file.
6. Translate the extracted MQXLIFF file in CafeTran.
7. When done, add the translated and finalized MQXLIFF file to Language Terminal.
8. The MQXLIFF file will be converted to a translated INDD file.

[Video](#) tutorial.

AI (ILLUSTRATOR)

Adobe Illustrator Artwork (AI) is a proprietary file format developed by Adobe Systems for representing single-page vector-based drawings in either the EPS or PDF formats. The .ai filename extension is used by Adobe Illustrator. The AI file format was originally a native format called PGF.

While AI files are not directly supported in CafeTran, since Adobe Illustrator can import and export SVG files, you can use CafeTran's SVG filter to handle Adobe Illustrator files (*.ai or *.eps).

To save the AI file as an SVG, you need to use a compatible application. There are several options:

Paid options:

- Purchase the Sysfilter pack or a monthly/annual Sysfilter subscription for Illustrator [here](#). Generates XLIFF files for import/export.
- Get a monthly subscription to Adobe Illustrator itself ([here](#)). Use that to save as .SVG.

Free option: Use the free/libre software [Inkscape](#). For more information, review the related documentation for [Inkscape for Adobe Illustrator users](#) and an [additional explanation/workaround](#).

Adobe Illustrator instructions:

- Before you export your document, you need to change the text from single line text boxes to full paragraph text boxes in Illustrator. This ensures correct segmentation.

- To prevent too many tags and ensure correct segmentation, use the following Illustrator export settings (make sure you do not use "Preserve Illustrator Editing Capabilities"):
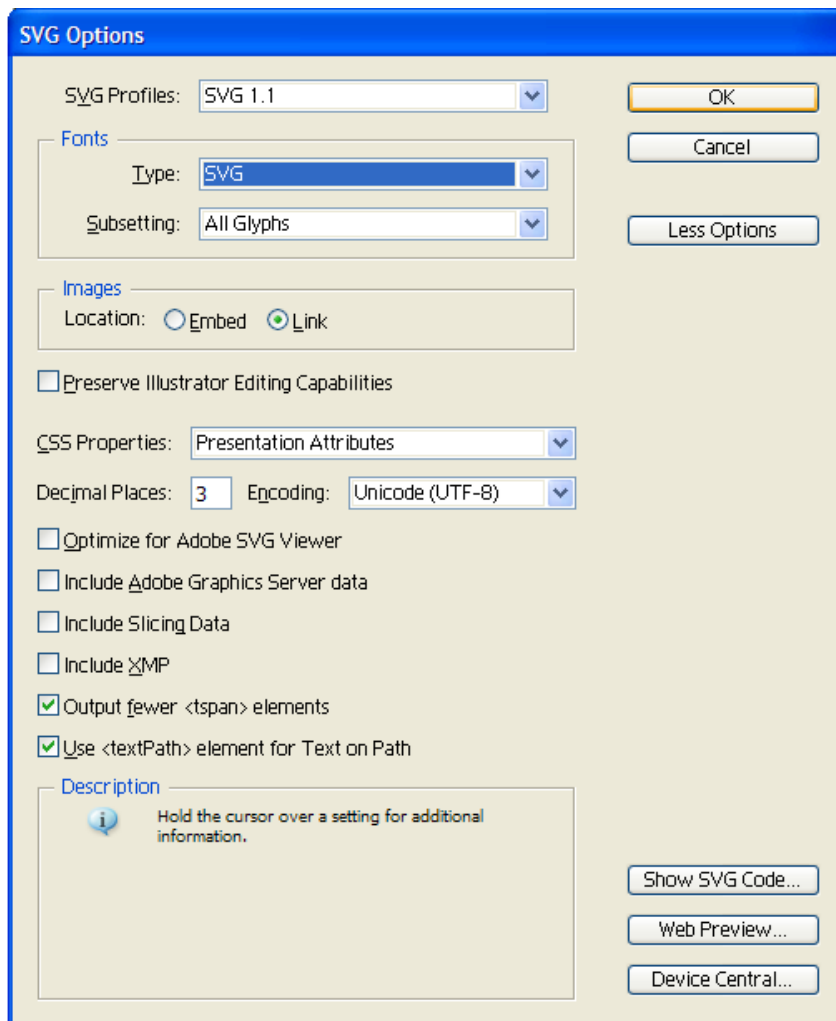
- Saving the document as SVG separates out the images. Therefore, save the SVG to its own separate folder.

- Import the SVG into CafeTran. After the translation, export the file back into the original folder so that the images can be linked correctly.

DWG (AutoCAD)

Generally, prior to translation, the AutoCAD DWG proprietary binary format needs to be converted into DXF, an exchange plain text format, which is handled by CafeTran (see Supported filters section above).

Ideally, this should be done by the client in AutoCAD. Various (often expensive) software solutions also offer DWG to DFX conversion.

If the DWG is exported as an SVG file (in Adobe Illustrator), if can be translated using the SVG file filter.

Free conversion solutions include: [Teigha File Converter](#) (free, cross-platform)

Free DWG/DFX viewers include: [DraftSight](#) (Windows, OS X, GNU/Linux)

PO

GNU Gettext Portable Object (.po) format used in localization is not yet directly supported by CafeTran.

Tools that allow you to handle PO files via a round trip: MateCat, Wordfast Pro 3, Okapi Framework, Memsource, Swordfish, Heartsome, SDL Trados Studio, memoQ.

JSON

JavaScript Object Notation (JSON) files are not directly supported by CafeTran.

CAT tools that allow you to handle JSON files via a round trip: MateCat, Wordfast Pro, memoQ, Swordfish, SDL Trados Studio (via an Open Exchange application). Okapi Framework can also help prepare files for translation and then back to the original format.

EPUB / MOBI / AZW / e-book

Report any issues as described in the Feedback section of this document

EPUB

EPUB (or ePub, short for electronic publication) is an e-book file format with the extension .epub that can be downloaded and read on devices like smartphones, tablets, computers, or e-readers.

An EPUB file is a ZIP archive that contains, in effect, a website—including HTML files, images, CSS style sheets, and other assets. It also contains metadata. EPUB 3 is the latest version. By using HTML5, publications can contain video, audio, and interactivity, just like websites in web browsers.

So, while the EPUB format itself is not directly handled by CafeTran, since the text to translate is contained in HTML files, one method of translating EPUBs is be the following:

1. Extract the EPUB to a folder: rename the EPUB to ZIP and extract as usual. Alternatively, 7-Zip file archiver (among others) can uncompress EPUBs directly. Another method: Convert the ePub to ZIP via Calibre e-book management application.
2. Translate the subfolder (typically called "text") containing the .html files in CafeTran (enable the "Multiple documents project" option in the Dashboard menu).
3. Rebuild the .ePub file. This may include the following steps:
   - Replace the .html files in the "text folder" with the translated ones.
   - Regenerate/Edit the TOC.NCX (Table of Contents) file. TOC generation and Editing can be done via the Sigil free E-book editor, as described here. Editing can also be done with Calibre's e-book editor (accessible via the "Edit book" button or context menu action).
   - ZIP the file and rename it to EPUB.
   - View it in Calibre and run additional validation checks (see below).

Additional validation checks can be done in Sigil (via its FlightCrew ePub validation plugin). You should also run a validation test in EpubCheck (online/offline) or the GUI version Pagina ePub checker. EpubCheck is an open source tool that validates EPUB books, so that you can make sure that your book complies with the EPUB format's specifications. This may be a requirement for some publishing platforms. Final checks can of course be conducted by the author or the publisher.

More details on the external tools mentioned:

- [Calibre](#) is a complete cross-platform e-book management suite. It lets you manage and access your digital library, view and edit e-books, convert them from one format to another, and much more. You can read its [user manual](#).
- [Sigil](#) is an excellent cross-platform E-book editor.

MEMSOURCE ROUND TRIP METHOD

Since [Memsource](#) handles EPUBs directly, you can use it to import the EPUB file, export the bilingual MXLIFF, translate it in CafeTran, upload the MXLIFF file in Memsource and finalize the translation. This is described in detail in the External projects > Memsource Cloud section of this document.

CALIBRE CONVERSION METHOD

You can convert any e-book to DOCX or RTF (and other formats) with Calibre, translate it and convert it back (to any format). This method is nice for producing a usable text file, but the back-conversion is likely to produce problems. If you need to deliver an EPUB file (and not a DOCX or other text type), you're better off using the methods described above.

MOBI / AZW / AZW3

The **Mobipocket** e-book format (.mobi) is based on the Open e-book standard using XHTML and can include JavaScript and frames. It also supports native SQL queries to be used with embedded databases. Mobipocket can be read in Amazon Kindle ad well as other e-Book devices and readers.

The **Amazon Kindle's AZW** format is basically just the Mobipocket format with some non standard variations. The Kindle format is available on a variety of platforms, such as through the Kindle app for the various mobile device platforms.

Older Kindle e-readers use the proprietary format, **AZW**. It is based on the Mobipocket standard, with a slightly different serial number scheme (it uses an asterisk instead of a dollar sign) and its own DRM formatting.

Since, [Amazon.com](#) has released Kindle Format 8, also known as **.AZW3**. The .azw3 file format supports a subset of HTML5 and CSS3 features, with some additional non-standard features; the new data is stored within a container which can also be used to store a Mobi content document, allowing limited backwards compatibility.

Translating MOBI / AZW3 E-books would require the following additional steps, with regards to the EPUB translation process described above:

1. Preparation: Convert the azw/mobi file to an ePub file with Kindleunpack or Calibre (import the book in Calibre, then Convert books > EPUB).
2. Finalization: Convert the updated .ePub file to an .azw3 file. You can do so via Calibre or online converters (such as Zamzar and CloudConvert).

https://www.mobileread.com/forums/showthread.php?t=243216

IBOOKS

The .ibooks format is created with the free iBooks Author E-book layout software from Apple Inc... This proprietary format is based on the EPUB standard, with some differences in the CSS tags used in an ibooks

format file, thus making it incompatible with the EPUB specification

An option to handle iBooks would be to use iBooks Author (OS X only) to export from IBOOKS to EPUB format, perform the EPUB translation as described above, then create a new iBooks Author file and import the translated .ePub into it for finalizing and publishing (since iBooks Author only opens iBooks Author projects, .i-books).

Related links:
Apple - iBooks Author official support page

TIPP (Packages)

The TIPP package format is part of the Interoperability Now! initiative and contains **XLIFF:doc** documents. It aspires to make it easier to pass projects and assignment packages among different translation tools. TIPP also has the objective to make resources and other settings interoperable among tools.

TIPP zipped packages cannot be simply unzipped, translated and zipped back. You should ask for or use the included XLIFF:doc files instead.

memoQ has full support for TIPP packages and XLIFF:doc documents, so you also can use it for handling these packages.

Related links:
Interoperability Now! project

4. SOLUTIONS

Since CafeTran offers a wide compatibility with external bilingual files, you can take advantage of this to prepare files (into a format CafeTran can accept) using a different software tool which handles **file types not yet supported by CafeTran** or offers **additional filter options** for already supported formats, then export the files back to their original format using the same tool once you've completed the translation in CafeTran (what we will call a "round trip").

Note: Working on External projects presents some differences and/or limitations when compared to the native "Translate document" workflow in CafeTran. These differences and limitations are discussed in detail in the General notes of the External projects section above.

CafeTran Espresso - Clipboard workflow

Before reviewing external options, it's good to know that CafeTran offers a Clipboard workflow which is not limited to any file format and can be used in various translation scenarios, allowing you to work in collaboration with many different software programs.

In the context we discuss here, these scenarios might include:

- translating content from dedicated software programs/editors that use file formats which aren't supported by CafeTran, while still viewing the content in full context;
- translating content from CAT tools that don't support appropriate exchange formats;

- translating content which can be copied but not modified, as is the case for many non-scanned PDF files.

In this workflow, View > Detached windows can help you arrange your screen to include both CafeTran and an external program, even without using two displays.

Related links:
[Basic Clipboard Workflow](#).

MateCat

MateCat is a free online CAT tool* which handles no less than 70 different formats and offers an excellent and easy way to translate file types for which CafeTran does not yet offer native support, acting as an additional file filter.

MateCat round trip procedure:

1. (Optional) Create a MateCat account or login with your Google credentials, for easier access and management of your projects.
2. Create a project in [MateCat](#) by giving it a name, selecting the language pair (and the subject if desired) and uploading the files.
3. Once you click Analyze, you will see a breakdown of the word count. You can access the job itself by clicking Translate in the bottom right, then Open job.
4. On the Translation Editor page, click the arrow next to the Preview button in the top-right corner of the screen and select the option Export XLIFF from the drop-down menu to download the translatable XLIFF file, which you can save on your PC. The XLIFF will be pre-translated with suggestions from your private TMs, the public TM and machine translation.
5. Translate and Finalize the XLIFF file in CafeTran. (You may need to use Project > Documents to view/glue all documents, some text sections of each file being held in separate .xml documents within the XLIFF itself.)
6. Use this [MateCat tool](#) to convert the translated XLIFF back into the original file format (DOCX, if OCR or PDF conversion is used). Just drag and drop, the download will start!

Additional notes:

- Added bonus: the XLIFF will be pre-translated with suggestions from your private TMs (you can import them prior to or at project creation), the public MyMemory TM and machine translation (Google/Microsoft translate/DeepL). Other MT options are also available.
- If you wish so, you can disable MT (Settings > Machine Translation), although public MyMemory TM matches will still appear in the XLIFF file you download and target segments with no matches will be populated with source text. To start with a clean slate, you can use Task > Remove target segments in CafeTran. This will also set the progress statistics to 0%.
- Highly recommended steps if you intend to perform at least part of the translation in MateCat: Create a private TM resource:
    - In the Project creation page, click on Settings (Alternatively, in the TM and glossary field, expand the drop-down menu and select Create resource).
    - Click on + New resource button in the opened dialog. Give the TM an optional name. Hit Confirm. You will see that "MyMemory: Collaborative translation memory" resource is Enabled for Lookup,

but not set to be Updated anymore. That way, translated segments will only be stored in your private resources.

- Be sure to check MateCat's Privacy policy to confirm this solution is appropriate for you. If you prefer offline solutions, you'll be better off using other options, such as the Wordfast Pro 3 demo version or, if you feel comfortable, Okapi Framework (see below).
- The same goes if you wish to configure additional filter options, since MateCat does not seem to offer any for its supported file formats.

Related links:
[MateCat - online support documentation](#)
[MateCat - Translating offline](#)
[MateCat - Philosophy and Terms of service](#) (includes Privacy policy)

MateCat file formats

Below you will find the full list of supported file formats in MateCat (in bold, those not natively supported by CafeTran): Microsoft Office (DOCX, XLSX, PPTX, **DOC**, **DOT**, DOCM, **DOTX**, **DOTM**, **XLS**, **XLT**, XLSM, **XLTX**, **XLTM**, **PPT**, **PPS**, **PPTM**, **PPSX**, **PPSM**, **POT**, **POTM**), OpenOffice (ODT, **OTT**, ODS, **OTS**, ODP, OTP), Web (HTM, HTML, XHTML, **CSV**, **TSV**, XML, **DTD**, **JSON**, **YAML**, **YML**) Localization (SDLXLIFF, XLIFF, XLF, **PO**, TTX, TXML, Android XML), Desktop publishing (MIF, IDML, **ICML**, **DITA**), Others (TXT, PROPERTIES, RESX, STRINGS, **SRT**, **WIX**), OCR [via Nuance OCR SDK] (BMP, GIF, PNG, JPEG, TIFF, Scanned PDFs). Others (**RTF**, Regular PDFs [converted to DOCX via CloudConvert]).

Note: CafeTran does not offer OCR, but can handle image translation (except for TIFFs) via the "Translate paper document" workflow, editable PDFs with its Clipboard workflow and short non-editable PDFs with its Paper document workflow.

Related links:
List of [MateCat - Supported file formats and languages](#)
List of [MateCat filters - Supported file formats](#)

*Note:
An open source version of Matecat's client-server web application can also be installed and used offline, although the list of supported file types does not include those requiring commercial licenses for its filters.

MateCat filters

Online API, offline open source

This section is added for the sake of completeness. It is geared towards more advanced users.

Online

MateCat filters allows you to easily extract all the translatable contents from a wide range of file format into a convenient XLIFF file. Translate it and use Filters again to get back a completely translated file with perfectly preserved formatting.

Written in Java using Jetty and Okapi Framework, MateCat filters is offered as a hosted API (at [Mashape.com](#)) and includes commercial licenses to support the full range of file formats MateCat does.

There is a Mashape plan for 250 free conversions per month, as well as other paid plans.

For more information:- 20180831 Updated for CafeTran Espresso 2018 - Akua Update 11

- [MateCat Filters](#) (home page).
  [Supported file formats](#)

Offline

You can build and use your own instance of MateCat filters (open source), although you will optionally need to install the MateCat Win Converter to support legacy MS Office formats, and buy licenses for the commercial software MateCat Filter uses to support OCR for images, scanned files (Nuance OCR SDK) and editable PDFs (CloudConvert).

Related links:
[Matecat Filters](#) (git page)
[Matecat Win Converter](#)

Wordfast Pro 3

Offline, cross-platform, free trial version, no time limit

You can use the cross-platform Wordfast Pro 3 to perform a round trip for different file types.

While WFP demo version has a cap on the size of resources (TMs and glossaries) that can be imported, it can import a source document of any size.

WFP roudtrip procedure:

1. Import the source document into WFP and save it.
2. Open the WFP work file (.txml) in CT and translate it.
3. Open the WFP work file in WFP and choose "Save Translated File."

Additional supported file formats include:

- Adobe InCopy ICML, PO, RTF, DOC, XLS, PPT, ASP, JPS, SRT, VDX (Microsoft Visio)

Also, some file formats already supported by CafeTran can offer additional filter options in WFP (in Edit > Preferences > Formats), such as the ability to choose whether to include slide notes, hidden slides etc. in PowerPoint presentations.

Related links:
Download [Wordfast Pro 3](#)

Wordfast Pro 5

Offline, cross-platform, free trial version, no time limit

If you own a Wordfast Pro 5 license, you can use this CAT tool to create projects which support additional file formats or filter options, and perform the translation in CafeTran, since CT handles TXLF files.

You can use the cross-platform Wordfast Pro 5 to perform a round trip for different file types. While WFP demo version has a cap on the size of resources (TMs and glossaries) that can be imported, it can import a source document of any size.

Supported file formats | Download (Specifications tab)

Memsource

Online/offline, cross-platform, Personal Edition or Freelancer edition

Memsource free Personal Edition account offers max. 2 files for translation (10 MB each). Old files will need to be deleted before adding new ones. Memsource paid Freelancer Edition (monthly subscription) offers unlimited file use.

Related links:
Memsource Cloud - Translator editions

Memsource handles a wide range of file types and offers various filters.

Additional supported file filters include: Multilingual MS Excel, Multilingual XML, Adobe Photoshop - PSD, JSON, CSV, YAML, DITA (dita, .ditamap), DocBook, PO Gettext, PHP, Markdown (.md, .markdown), Quark Tag, CATKEYS, EPUB, INI, LANG, PLIST, RC, SRT, SUB, TS, WIKI, YAML, TAG, XTG.

Smartcat

Free, online

Smartcat supports exporting and importing back a translated XLIFF. This makes it excellent for roundtrip scenarios. See also their best practices document.

Additional supported file formats include: DOC, XLS, PPT, POT/POTX, RTF, PHP, PO, MIF, DITA XML, HELP+MANUAL XML, CSV, TSV, DTD, JSON, TJSON, YML, INC, SRT, various image formats, powered by OCR (JPG/JPEG, TIF/TIFF, BMP, PNG, GIF, DJVU/DJV, DCX, PCX, JP2, JPC, JFIF, JB2), WSXZ packages.

Related links:
Smartcat Help Center

Swordfish

Offline, cross-platform, 30-day trial, license request needed

Additional supported file filters include: DocBook 3.x, 4.x and 5.x, JavaScript, PHP Arrays, PO (Portable Objects), RC (Windows C/C++ Resources), TS (Qt Linguist translation source), DITA

Related links:
Download Swordfish.

Heartsome

Offline, cross-flatform, free and open source

Additional supported file filters include: DITA, PO, RC.

Related links:
[Heartsome Translation Studio 8.0](#)

Okapi Framework - Rainbow

Offline, cross-platform, free and open source

This solution is for more advanced users.

The Okapi Framework is a cross-platform and free open-source set of components and applications that offer extensive support for localizing and translating documentation and software.

Rainbow is an Okapi application which allows you to launch different utilities to help you perform various localization-related tasks. Like other Okapi applications, it uses the components called "filters" that take input files in a given native format and extract its translatable content into common resources that the Okapi tools can manipulate and modify. You can create and execute [Pipelines](#), a series of [Steps](#) that produce a given outcome. For example, you can create a Pipeline that uses [Rainbow Translation Kit Creation Step](#) to create a (XLIFF or other type of) translation package from a set of input documents, translate the XLIFF, and use [Rainbow Translation Kit Merging Step](#) to merge the files of the package back into their original format.

[Filters](#) are the configurable components that convert input documents from their native file format into a common internal set of resources that all Okapi components use. The extracted content can be re-written into the original file format. When using the steps, the extraction is done by the [Raw Document to Filter Events Step](#) and the rewriting by the [Filter Events to Raw Document Step](#).

Related links:
[Okapi Framework - Filters](#) (to get an idea of supported file types)
[Okapi Framework - Knowledge Base](#)

Note: Open source MateCat filters mentioned above are partly built using Okapi filters.

Online/Offline file converters

As mentioned in the Unsupported formats subsection when considering options for handling PDFs, various online or offline file converters might help you perform a conversion from an unsupported file type to another which CafeTran supports.

Not all conversions are lossless though, and caution/testing is required.

- **Online**

Hundreds of online file converters exist, here are few good ones:
[CloudConvert](#), [Zamzar](#), [Convertio](#).

- **Offline**

These generally depend on the type of file conversion you need to make, it is difficult to make a specific recommendation. Using specialized software is in general more appropriate offline. For example, converting a DOC to DOCX is just a matter of using MS Word, OpenOffice/LibreOffice, WPS Office, etc.

SDL Trados Studio

Paid, offline | Windows only

If you own an SDL Trados Studio license, you can use this CAT tool to create projects which support additional file formats or filter options, and perform the translation in CafeTran, since CT handles SDLXLIFF files.

An SDL Trados license also helps perform operations such as Document Preview, Bilingual document import/export and, of course, Export translated files function and importing TMX to SDLTM, that can't be done on Trados files in CafeTran.

SDL Trados supports additional plugins that further enhance its features, some of them adding support for additional file formats.

Related links:
[Supported file types](#) in SDL Trados Studio.
[Additional information](#) on supported file types.

memoQ

Paid, offline | Windows only

If you own a memoQ license, you can use this CAT tool to create projects which support additional file formats or filter options, and perform the translation in CafeTran, since CT handles MQXLIFF files.

memoQ offers a 45-day free trial version of [memoQ translator pro](#) edition. The trial is valid for new customers only and can be later activated if you purchase a full license. You also need to download this version if you are working for customers using a [memoQ server](#).

[memoQ translator free](#) is the basic software your trial version turns into at the end of the trial period if you choose not to purchase a license. It has only very basic translation memory and term base functionality and does not help you in large projects (only one document can be translated at a time, so not a great solution for free round trips after the trial period ends). It's still suitable for personal use on small translation projects though. It is not possible to use memoQ translator free in memoQ online projects (although, as stated above, it is used for accepting jobs from memoQ server).

[Supported file formats](#) in memoQ

Déjà Vu

Paid, offline | Windows only

If you own a Déjà Vu license, you can use this CAT tool to create projects which support additional file formats or filter options, and perform the translation in CafeTran, since CT handles Déjà Vu External View

XLIFF files.

Related links:
[Déjà Vu - Supported File Formats](#)
[Déjà Vu - Supported Formats And How To Use Them](#)

LIST OF CHANGES

- 20171127 File creation
- 20171203 Added section for e-books (EPUB, MOBI, AZW), and information about Memsource tags.
- 20180217 Updated for CafeTran Espresso 2018 - Akua, Migrated from [Classeur.io](#) to [GitHub.com](#)
- 20180621 Updated for CafeTran Espresso 2018 - Akua Update 10
- 20180830 Updated for CafeTran Espresso 2018 - Akua Update 11
- 20180831 Updated for CafeTran Espresso 2018 - Akua Update 11 - Added section for SVG filter in Supported file formats, Adobe Illustrator (AI) in Unsupported formats, and Smartcat in Solutions
- 20190121 Updated for CafeTran Espresso 2019 - Croissant 10.0

Feedback

You can send me feedback via a [ProZ message](#). Please start your subject with the name of the document: "CafeTran - File formats".