

# **Content of Pre-treatment packages for Monolingual Documents and General Instructions for External Translators (with SDL Trados Studio)**

**v15**

Directorate for Support and Technological Services for Translation

PreTrad - Euramis Pre-Translation Unit

## Contents

1. Content of pre-treatment zip packages.....	3
1.1 Content of pre-treatment zip packages for MS Office source documents.....	3
1.2 Content of pre-treatment zip packages for different XML source documents.....	5
1.3 Translatable files .....	7
1.3.1 Word originals .....	7
1.3.2 XML4EP originals (monolingual documents without two-column amendments only) .....	8
1.4 XML4EP-specific subfolders in the project folder .....	10
1.5 Translation memories .....	10
1.6 Studio project file .....	12
1.7 Tmx subfolder .....	12
1.8 Support files in the zip package .....	13
2. General instructions for external translators .....	14
2.1 Use of DocEP.....	14
2.1.1 Word originals .....	14
2.1.2 XML4EP originals .....	17
2.2 Pre-translation of the document .....	17
2.3 Machine translation .....	17
2.4 Delivery of the TL document .....	18

# 1. Content of pre-treatment zip packages

## 1.1 Content of pre-treatment zip packages for MS Office source documents

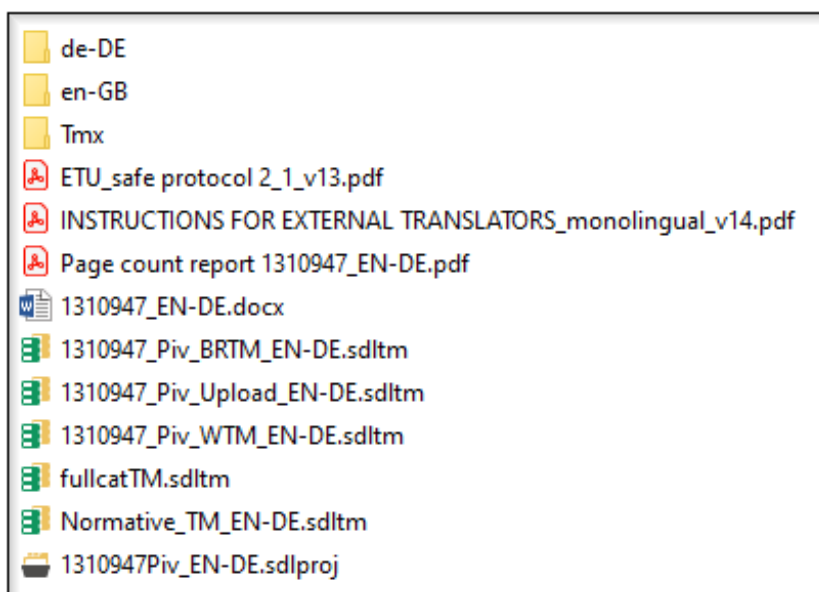


Figure 1: Content of a zip package for a document with two-column amendments (Word original).

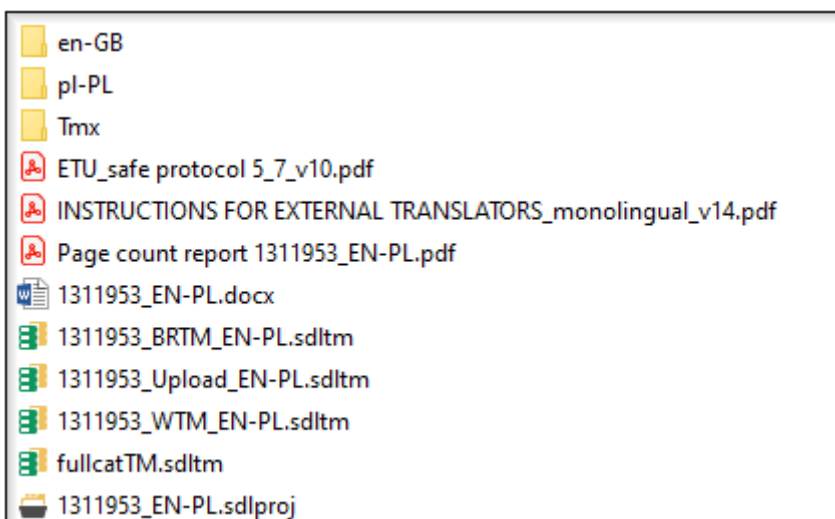


Figure 2: Content of a zip package for a document without two-column amendments (Word original).

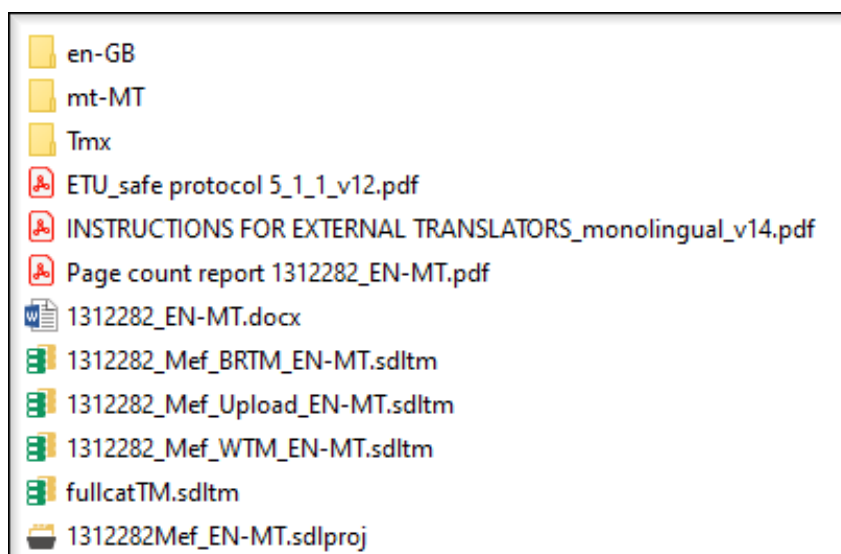


Figure 3: Content of a zip package for a *mise-en-forme* (MeF) version of a document (Word original without two-column amendments)

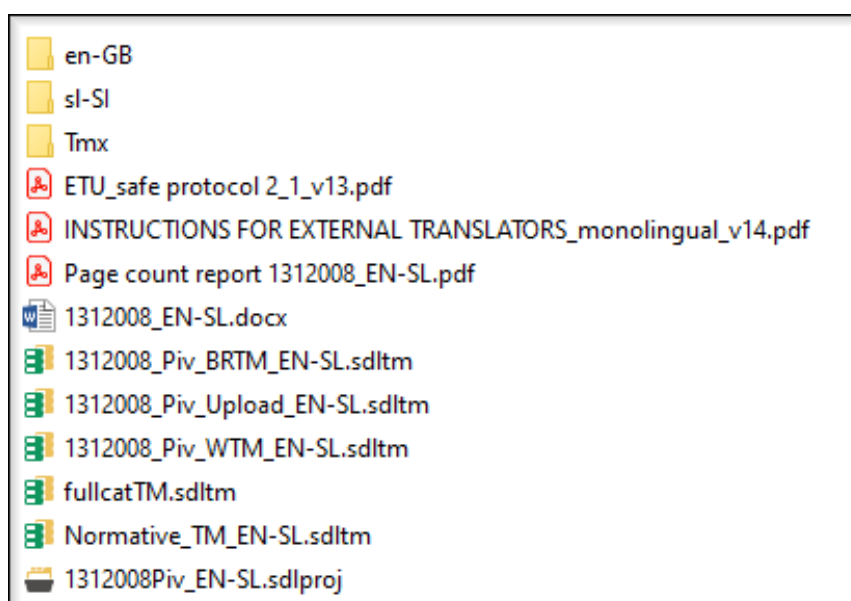


Figure 4: Content of a zip package for a Relay (Pivot) version of a document (Word original with two-column amendments)

## 1.2 Content of pre-treatment zip packages for different XML source documents

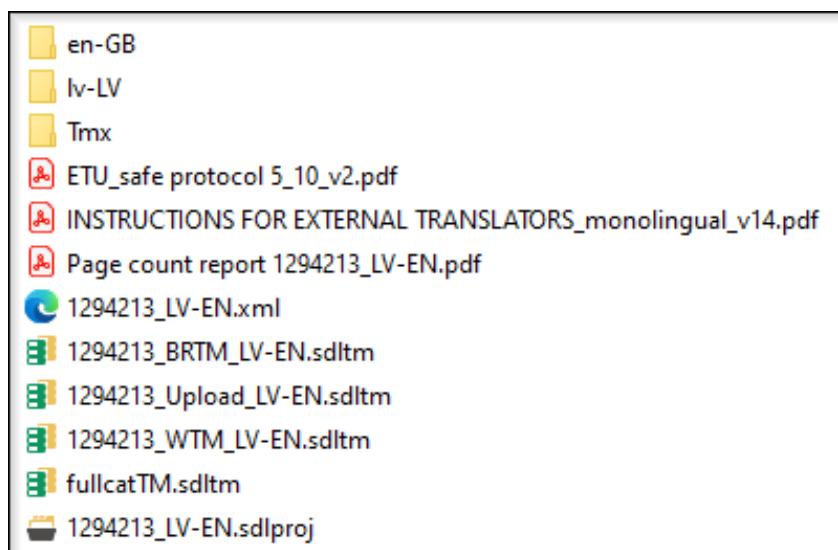


Figure 5: Content of a zip package for a simple XML original document without two-column amendments (document type CV).

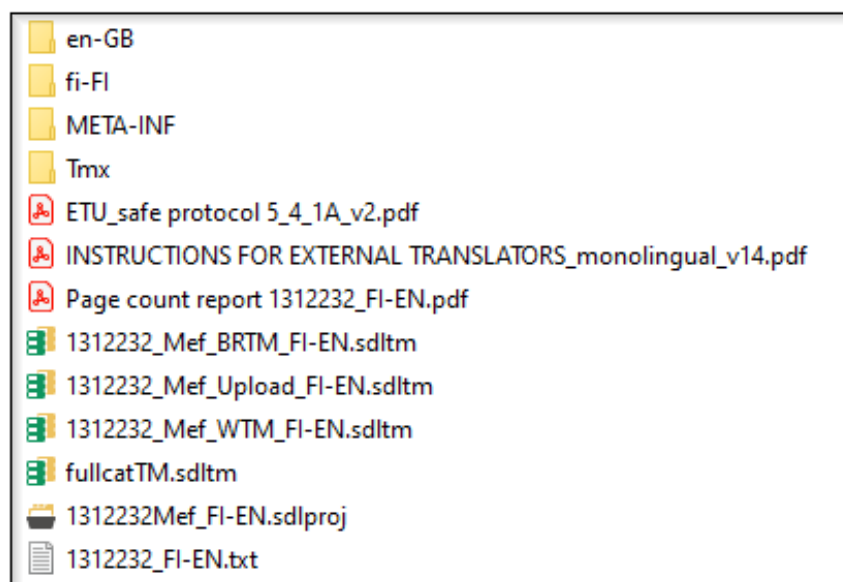


Figure 6: Content of a zip package for a document without two-column amendments - XML4EP original (parliamentary questions: QE, QO).

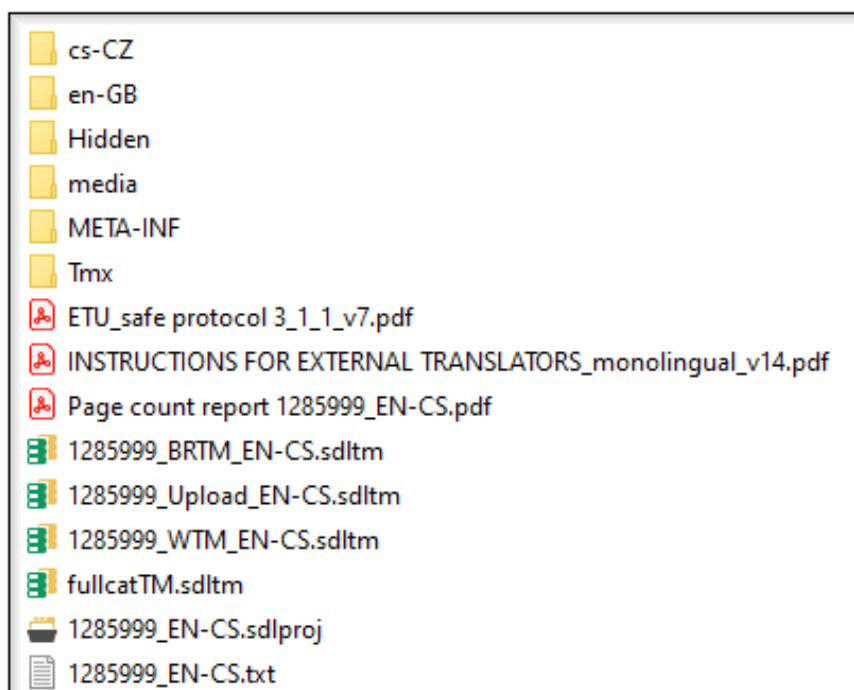


Figure 7: Content of a zip package for a document without two-column amendments - XML4EP original (document types PR and RR).

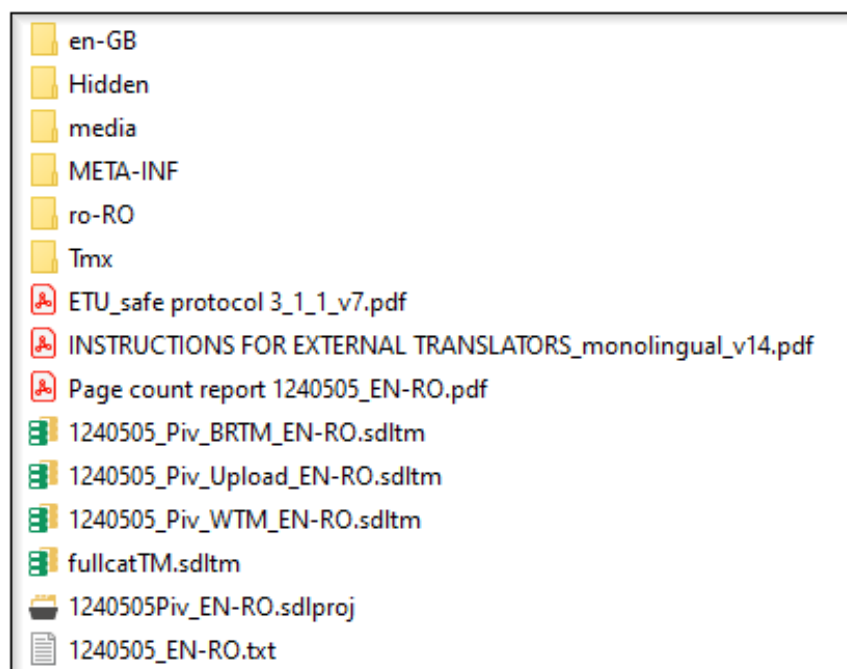


Figure 8: Content of a zip package for a Relay (Pivot) version of a document without two-column amendments - XML4EP original (document types PR and RR)

## 1.3 Translatable files

### 1.3.1 Word originals

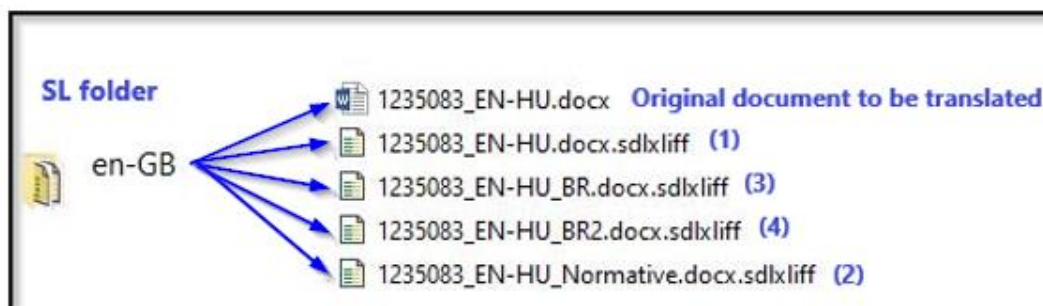


Figure 9: Translatable files - source language folder view

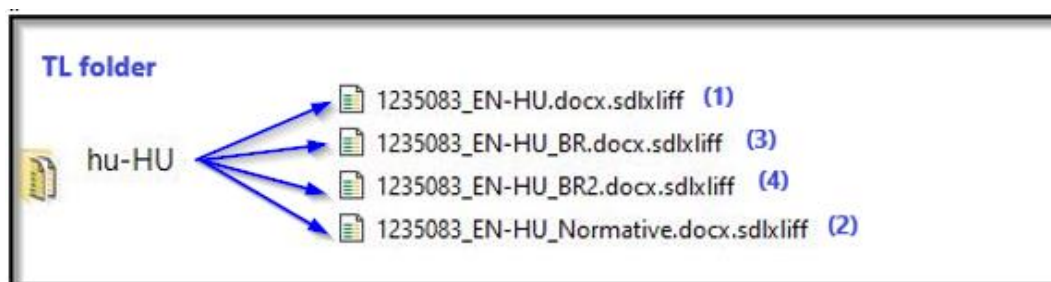


Figure 10: Translatable files - target language folder view

The **original** document in its original format used to create the pre-treatment package is included in the source language (SL) folder.

**N.B:** For documents containing **two-column amendments**, the SL folder contains the so-called **Coloured version** in which the amendment headings and the left-hand column text in the amendment tables have been coloured in blue in order to facilitate the identification of the corresponding segments in the Studio Editor.

Depending on the Safe Working Protocol to be followed, the following translatable SDLXLIFF files may be included:

- An SDLXLIFF conversion (1234567\_SL-TL.docx.sdlxliff) of the original file format of the document to be translated **(1)**.
- An SDLXLIFF file (1234567\_ST-TL\_Normative.docx.sdlxliff<sup>1</sup>) resulting from the pre-translation at 98% with the Normative\_TM **(2)**.

<sup>1</sup> Concerns only documents with two-column amendments.

- An SDLXLIFF file (1234567\_SL-TL\_BR.docx.sdlxliff) resulting EITHER from the pre-translation at 98% with the Normative\_TM, followed by the pre-translation at 100% with the BRTM (in the case of two-column documents), OR from the pre-translation at 100% against the BRTM only if the Safe Working Protocol does not call for a Normative\_TM (in the case of documents without two-column amendments) **(3)**.
- An SDLXLIFF file (1234567\_SL-TL\_BR2.docx.sdlxliff) resulting from the sequential pre-translation with the Normative\_TM (at 98%), then with the BRTM (at 100%) and finally with the BR2TM (again at 100%) **(4)**.

### 1.3.2 XML4EP<sup>2</sup> originals (monolingual documents without two-column amendments only)

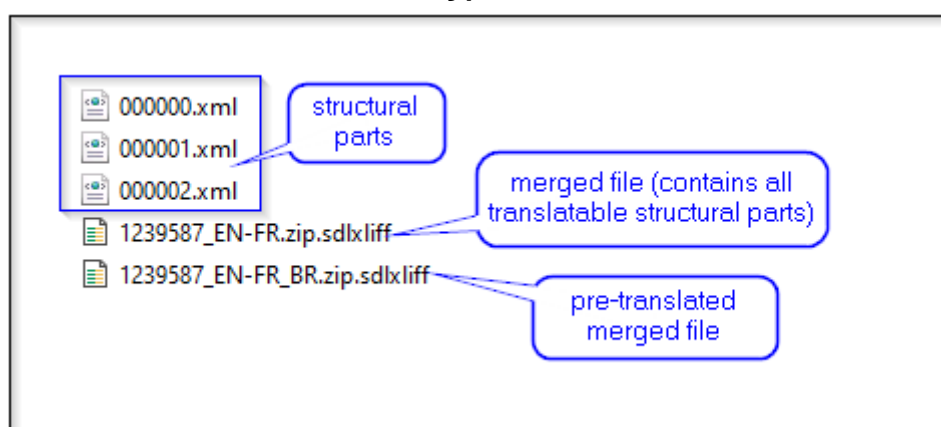


Figure 11: Project SL/TL folder view in Studio - document types PR and RR

The complete original document consists of several XML files (in XML4EP format)<sup>3</sup> corresponding to defined structural parts: Cover page, Motion for a European Parliament resolution, Explanatory statement, Annex.

To see the complete original text, you can open the PDF/Word file delivered with the package, named after the FdR number and the source language (e.g. 1239587EN.pdf or 1239587EN.docx).

<sup>2</sup> XML4EP is a special XML flavour used by the European Parliament in their e-Legislate chain (i.e. digitalisation of legislative documents).

<sup>3</sup> For document types PR and RR. Parliamentary questions consist of only one XML file.



The **translatable structural parts** are **visually merged into one translatable master SDLXLIFF file** in the Studio project. Some structural parts may not need translation. In this case, they are placed in the Hidden subfolder in the project folder and they are not visible in the Studio Editor.

After the delivery, the different translatable and hidden structural parts will be assembled into a complete document and saved to the central repository by the EP Language Units in the finalisation phase.

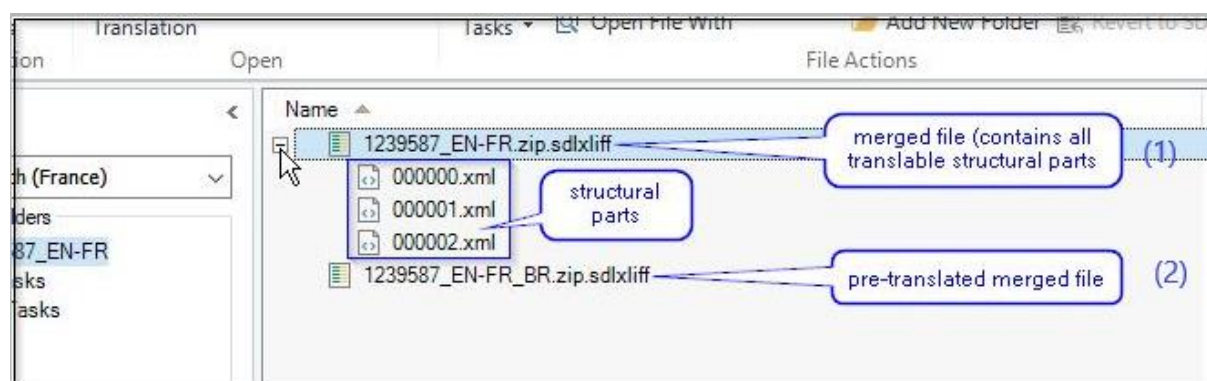


Figure 12: Target language Files view of an XML4EP Studio project - document types PR and RR

The following translatable SDLXLIFF files are included in the project:

- An SDLXLIFF conversion (**1234567\_SL-TL.zip.sdlxliff**) of the translatable original xml files (structural parts), physically merged for ease of use **(1)**. The structural parts are visible in the 'Files' view when clicking on the “+” sign. In the Studio Editor, the beginning and the end of each such structural part are marked with delimiters (orange tags).
- An SDLXLIFF file (**1234567\_SL-TL\_BR.zip.sdlxliff**) resulting from the pre-translation of the translatable merged file at 100% with the BRTM **(2)**.

## 1.4 XML4EP-specific subfolders in the project folder

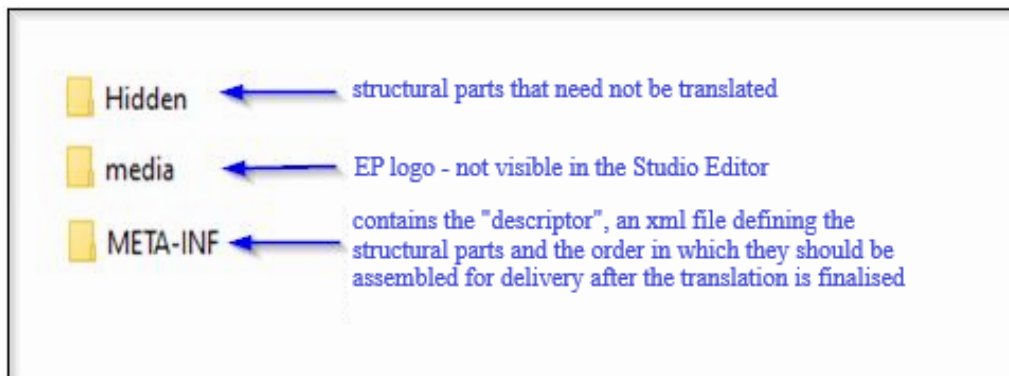


Figure 13: XML4EP-specific subfolders

These subfolders relate to the structure of the XML4EP document and have no relevance for the translation proper. The files contained therein are not part of the files visible in the Studio Editor and it is crucial that they are not accessed directly or changed in any way.

## 1.5 Translation memories

Depending on the document type, the package may contain:







 Normative_TM_EN-HU.sdltm	a Normative Memory
 fullcatTM.sdltm	a fullcat Memory (can be disregarded in your workflow)
 1235083_WTM_EN-HU.sdltm	a Working memory
 1235083_Upload_EN-HU.sdltm	an Upload memory (can be disregarded in your workflow)
 1235083_BRTM_EN-HU.sdltm	a Basic Reference Memory
 1235083_BR2TM_EN-HU.sdltm	a Basic Reference 2 Memory

Figure 14: Translation Memories

The **Normative\_TM** contains the tmx file of the standard segments used in the amendment headings preceding each amendment. This TM is only created for documents with two-column amendments.

The content of the **BRTM** varies, depending on whether the document contains two-column amendments or not.

- In the case of documents with two-column amendments, the BRTM contains the Basic Reference tmx file(s) corresponding to the document(s) being amended and the Normative tmx file compiling the source-target translation of general standard text applicable to all document types (*Normative\_Main*). For example, for protocol 2.1, the BRTM contains the tmx of the COM document and the Normative\_Main.

- For documents without two-column amendments, the BRTM contains the Basic Reference tmx file(s), the Normative tmx file compiling the source-target translation of general standard text applicable to all document types, and a second Normative tmx file containing source-target standard text specific for that particular document type. For example, for documents treated under protocol 3.1.1, the BRTM contains the tmx file corresponding to the previous document from the same procedure, the Normative\_Main and the Normative\_General\_NonLeg.

This memory is used for pre-translating the original document at 100% in order to get the Basic Reference document text (in documents with two-column amendments this text is present in the left-hand column) and the translation of the standard text.

The **BR2TM** contains the Secondary Basic Reference tmx file(s) for specific two-column amendment document types (legislative final opinions and final reports). For protocol 2.6.1, for example, the BR2TM contains the tmx files corresponding to the draft report from the same procedure and the amendments thereto. This memory is used for pre-translating the BR.docx.sdlxliff file at 100% in order to get the text from the previous, secondary basic reference documents into the amendment right-hand column.

The **Working TM (WTM)** contains all the available resources, i.e. the tmx files for the Basic Reference and Secondary Basic Reference document(s), the Retrieval, the Normative tmx file(s), all the other reference documents belonging to the document's procedure, and any other reference deemed useful for the translation of the document.

The **Upload TM** and the **fullcat TM** are for EP internal use only and should be disregarded in your workflow. Disable them before you start working on the project.

Each zip package contains the General Instructions and the ETU Safe Working Protocol (a table in Word format, describing step-by-step how to work with the memories) to be used for that document. If no protocol is available, use the WTM as a resource for interactive translation in Studio or a Studio-compatible translation tool<sup>4</sup>.

---

<sup>4</sup> For Word originals only

## 1.6 Studio project file

The project is named after the FdR number of the original document, plus the source and the target languages (**1234567\_SL-TL.sdlproj**).

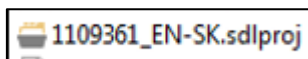


Figure 15: Studio Project

In case of a *mise-en-forme* or Relay (Pivot) version of the original document, the project is identified with the 'MeF' or 'Piv' label respectively.

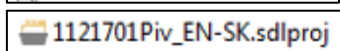
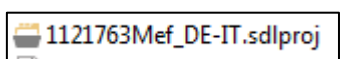


Figure 16: Studio projects for MeF / Pivot version

## 1.7 Tmx subfolder

The folder contains the Retrieval, the Normative and document-relevant references in the form of tmx files that have been imported into the project TMs according to the Safe Working Protocol.

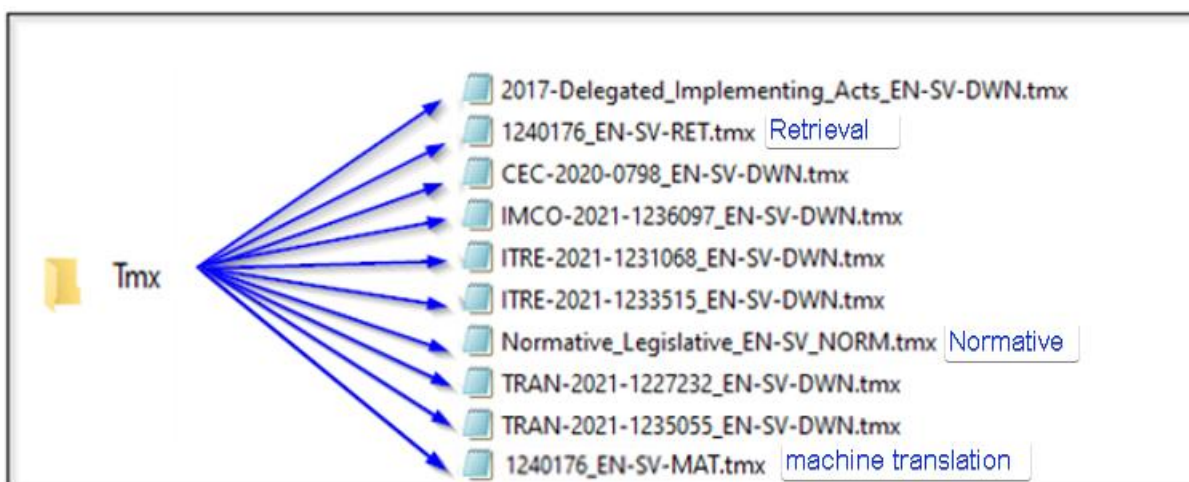


Figure 17: Tmx files in the Tmx subfolder

You will also find here a tmx file which contains the machine translation into the target language, obtained from the interinstitutional machine translation provider eTranslation. The naming pattern for this tmx file is 1234567\_SL-TL-MAT.tmx.

Section 2.3 below outlines general considerations about machine translation and provides guidelines for its safe use.

## 1.8 Support files in the zip package





	ETU_safe protocol 2_1_v13.pdf	Safe Working Protocol - specific step-by-step instructions for the document type
	INSTRUCTIONS FOR EXTERNAL TRANSLATORS_monolingual_v14.pdf	General instructions (this document)
	Page count report 1312008_EN-SL.pdf	Content reuse report (page count)
	1312008_EN-SL.docx	Original document with amendment headings and left-hand column text in amendment tables coloured in blue (ONLY in documents containing two-column amendments)

Figure 18: for Word originals (see Figure 1)






	ETU_safe protocol 3_1_1_v7.pdf	Safe Working Protocol - specific step-by-step instructions for the document type
	INSTRUCTIONS FOR EXTERNAL TRANSLATORS_monolingual_v14.pdf	General instructions (this document)
	Page count report 1285999_EN-CS.pdf	Content reuse report (page count)
	1285999_EN-CS.txt	Plain text document containing the translatable text extracted from the various structural parts of the original XML4EP document

Figure 19: for XML4EP originals (see Figure 7)

## 2. General instructions for external translators

### 2.1 Use of DocEP

 **docep** is a document creation tool (Word macro) used in the European Parliament. It enables the creation of documents which contain the official standard phrases encoded in the Recueil de Modèles (RdM) already in the target language.

#### 2.1.1 Word originals

Always refer to the specific Safe Protocol for the type of document concerned before you start to work on the document.

❖ **If the SWP recommends setting up the target language (TL) document to be used as working document in Studio with DocEP 'Create Document', the general rules are as follows:**

- Set up the target language document with the document specific DocEP template and copy the body of the source language<sup>5</sup>.
- Add the TL document to the project.
- Pre-translate first against the Normative\_TM at 98%, then with the BRTM and the BR2TM (if it exists) at 100%.
- Open the pre-translated file. The source language (SL) column in SDL Trados Studio will display the DocEP text in the target language (TL). Change the status of those segments to 'locked' and leave the corresponding TL cells empty. If you are using a different CAT tool, copy the source (i.e. DocEP text already in the TL) to the target cells.
- Go through the document, checking that the pre-translated target language segments come from the appropriate TM(s) and the correct reference document(s). Use the WTM to complete the translation.



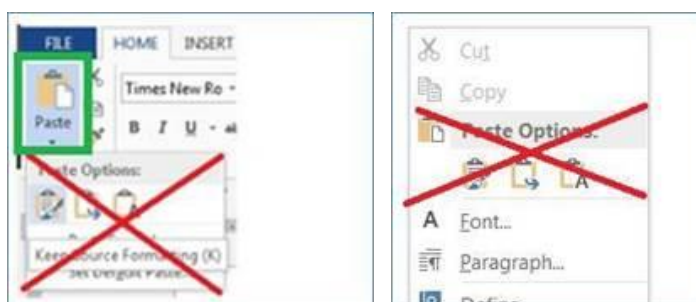
When setting up the document with DocEP and copying the body of the source language text, be careful not to change the DocEP tag-specific style definition

---

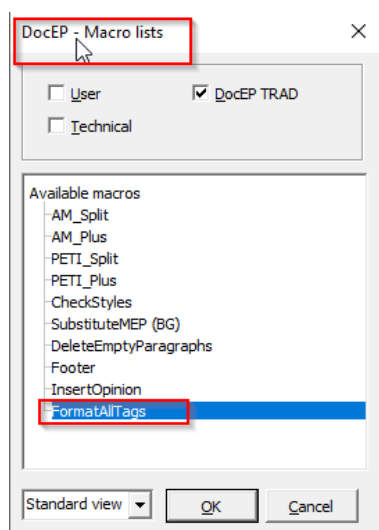
<sup>5</sup> N.B.: If the document contains two-column amendments, use the Coloured original document when copy-pasting. Having the amendment headings and left-hand column text of amendment tables in blue will facilitate the identification of the corresponding segments in the Studio Editor and the use of the cascading translation memories provided.

(*HideTWBExt* and *HideTWBInt* in dark Blue). Please take the following into consideration:

- ✓ Use the standard MS Word Paste – i.e. either the 'Paste' button in the first screenshot below, or *Ctrl+V* – styles will be kept.
- ✓ For amendments, use the AM-specific functions in DocEP ('Copy AM', 'Copy all AMs') which allow you to safely copy text with styles.
- ✗ If you use the additional Paste options in MS Word, the styles will be lost. This is also the case if using the option 'Keep Source Formatting': the text looks exactly the same, but all the styles are *Normal*. In the context menu (right click), the regular Paste is not even present and all the available options will result in a faulty Paste.



- ✓ Before adding the set-up document to the Studio project, always check that all DocEP tags have the two styles mentioned above. Should it not be the case (the tag style is *Normal*), use the macro 'FormatAllTags' from the DocEP Macro Lists menu. It will re-apply the correct styles throughout the document.



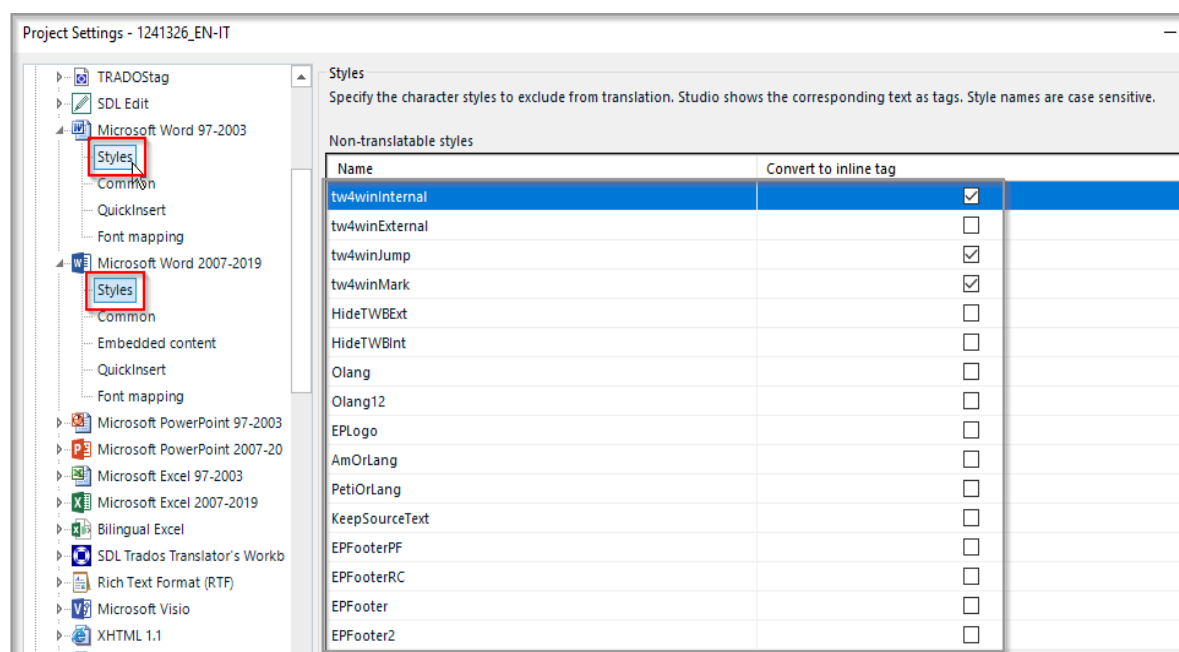


❖ **If the SWP recommends the use of the pre-translated SDLXLIFF file(s) from the pre-treatment package**, the standard workflow is described below:

- Open the project.
- Depending on the document type and SWP, it may contain several pre-translated SDLXLIFF files. If there is a pre-translation which has been produced using the BR2TM (1234567\_SL-TL\_BR2.docx.sdlxliff), use it. Otherwise, use the file resulting from the pre-translation with the BRTM (1234567\_SL-TL\_BR.docx.sdlxliff).
- Any standard phrases from the official Parliament models needed for the translation of the original document are included in the Normative\_TM or BRTM. Go through the document, checking that the pre-translated target language segments come from the appropriate TM(s) and the correct reference document(s). Use the WTM to complete the translation.
- DocEP 'Change document ID' function will be performed by the EP Language Units in DG TRAD at the end of the workflow.

❖ **Handling DocEP tags in the SDL Trados Studio Editor**

If you use SDL Trados Studio, we recommend that you replicate the settings below in the 'File Type' definitions for MS Word 97-2003 and 2007-2019 (under SDL Trados Studio – Project Settings – File Types). DocEP tags will then be treated as structure tags which, by definition, remain outside the segments. This eliminates the risk of damaging them while translating. The settings shown on the images below are also the EP's default in-house settings.







If these settings are not used in SDL Trados Studio, the DocEP tags may be present and visible in the segments in the Editor. Should this be the case, they must be copied from the SL to the TL segments.

If you use another CAT tool, we recommend that you look at the possibility of introducing similar settings, so that the DocEP tags do not need to be handled by the translator. If such a possibility does not exist, then the standard XLIFF files delivered must contain the relevant tags in the appropriate location in the TL.

### 2.1.2 XML4EP originals

No DocEP set-up is needed for XML4EP originals. Always refer to the specific Safe Working Protocol for the type of document concerned.

As a general rule, you have to use the pre-translated SDLXLIFF file (**BR.zip.sdlxliff**) from the package. The basic reference document(s) and the standard phrases needed for the translation of the original document are included in the BRTM. Go through the document, checking that the pre-translated target language segments come from the appropriate TM(s) and the correct reference document(s). Use the WTM to complete the translation.

## 2.2 *Pre-translation of the document*

Follow the steps described in the Safe Working Protocol to (pre)translate the document, if needed.

If no protocol is available, use the Working Memory (WTM) to translate the BR.sdlxliff file or, in its absence, the SDLXLIFF conversion of the original file.

## 2.3 *Machine translation*

Machine translation should be used only to supplement the matches provided by the reference and Retrieval tmx files imported into the various project translation memories. It should never be your first choice, but rather a last resort. Especially for procedural documents, the reuse of previously translated text from the same procedure or related ones is of the utmost importance.

The quality of the machine translation output is still below the level of human translation, as shown by recent EP internal tests, and it can vary greatly between target languages. Therefore, if you do make use of the provided machine translation tmx file, you need to always keep an eye on the source text to prevent the risk of serious mistranslations.

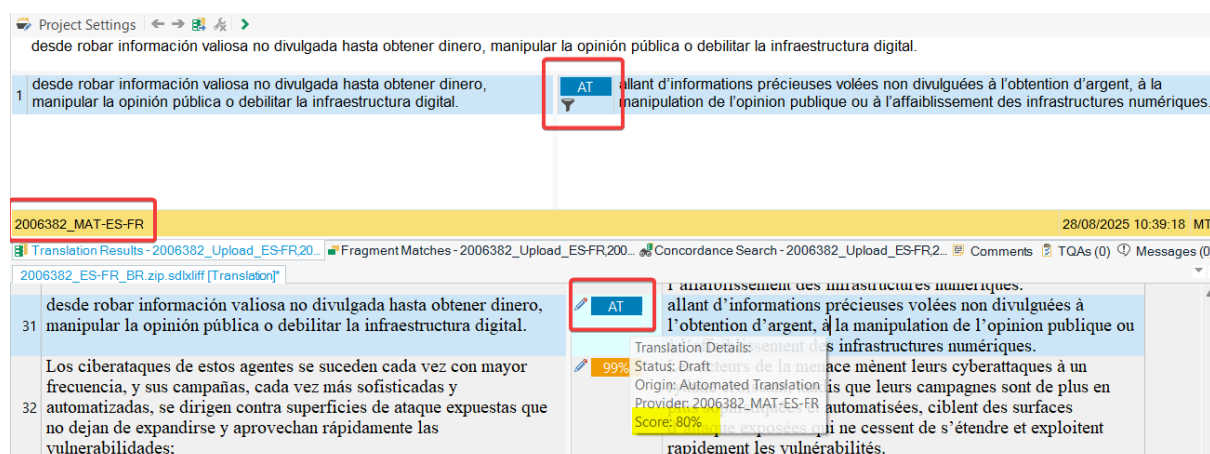
To safely use the machine translation tmx, you will need to import it into a separate Machine Translation\_TM. This is the current practice in DG Trad.

The steps to follow:

- Create a separate Machine Translation\_TM
- Add the Machine Translation\_TM to the project
- Make sure that it is placed last on the list of project Translation Memories
- Import the MAT.tmx file into the Machine Translation\_TM
- Enable the Machine Translation\_TM for lookup and concordance

You can easily identify the matches from the Machine Translation\_TM / MAT.tmx by the AT label (automatic translation). It is visible in the translation results window, as well as on the Segment Status Column in the Studio Editor.

A 20% machine translation filter penalty is applied by default on machine translation matches.



## 2.4 Delivery of the TL document

Deliver a bilingual SDLXLIFF file or a standard bilingual XLIFF file in the case of Word originals.

For XML4EP content, you must deliver only a bilingual SDLXLIFF file in order to guarantee that the EP Language Units can finalise the document successfully.

All tagging and formatting from the original document must be kept or replicated in the bilingual (SDL)XLIFF file that you deliver.