

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

v13

Directorate for Support and Technological Services for Translation

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
1.1 Translatable files.....	7
1.1.1 Word originals.....	7
1.1.2 XML4EP originals (monolingual documents without two-column amendments only).....	8
1.2 XML4EP-specific subfolders in the project folder.....	10
1.3 Translation memories.....	10
1.4 Studio project file.....	11
1.5 TMXs subfolder.....	12
1.6 Support files in the zip package.....	12
2. General instructions for external translators.....	13
2.1 Use of DocEP.....	13
2.1.1 Word originals.....	13
2.1.2 XML4EP originals.....	16
2.2 Pre-translation of the document.....	16
2.3 Delivery of the TL document.....	16

1. Content of pre-treatment zip packages

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

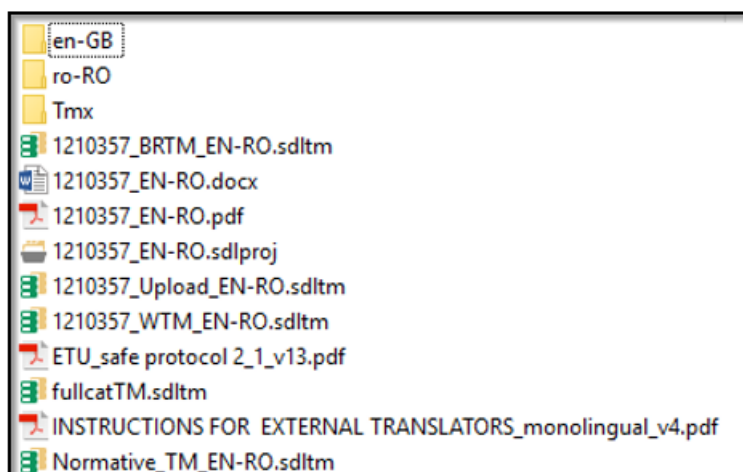


Figure 1: Content of a zip package for a document with two-column amendments (Word original).

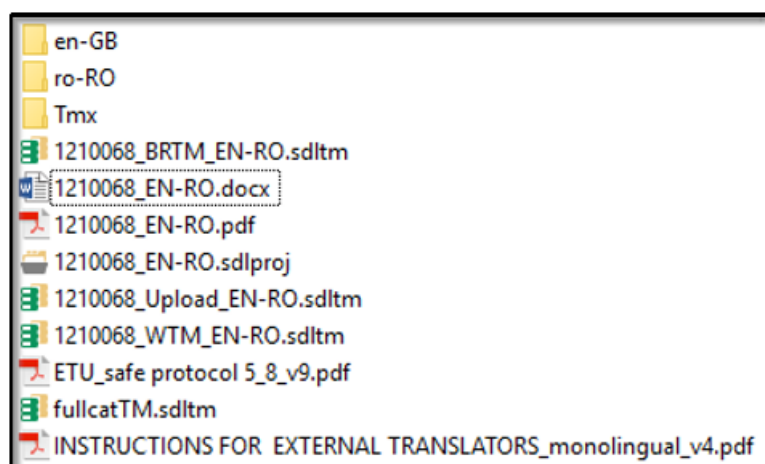


Figure 2: Content of a zip package for a document without two-column amendments (Word original).

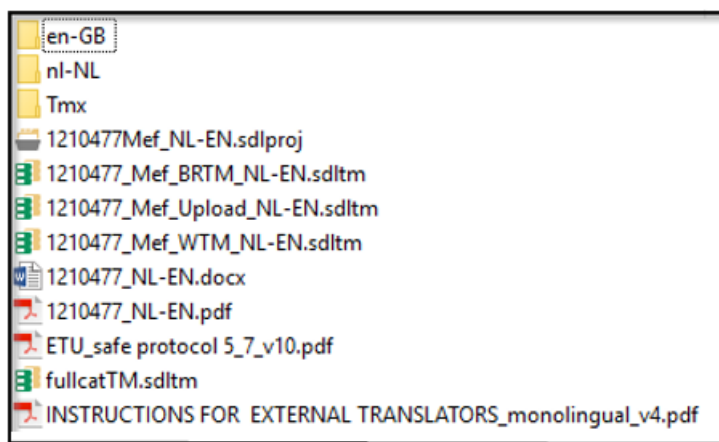


Figure 3: Content of a zip package for a *mise-en-forme* (MEF) version of a document (Word original without two-column amendments)

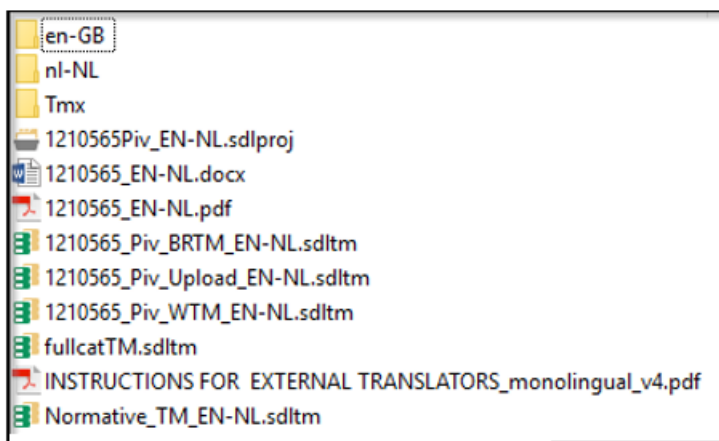


Figure 4: Content of a zip package for a Relay (Pivot) version of a document (Word original with two-column amendments)

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

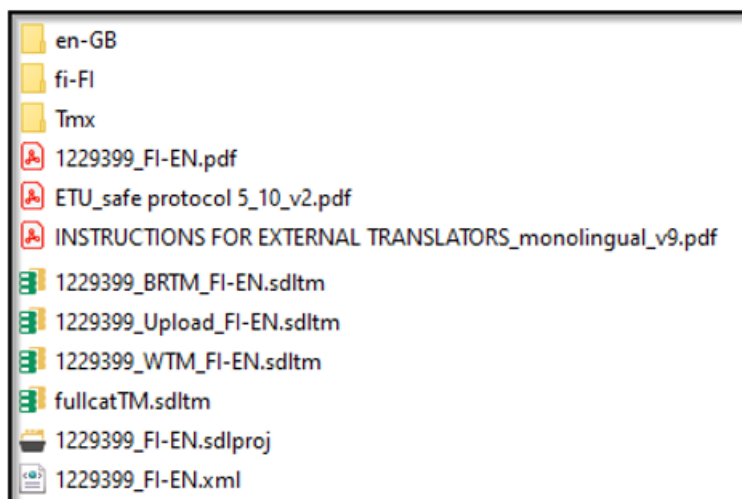


Figure 5: Content of a zip package for a simple XML original document without two-column amendments (document types CV, HH).

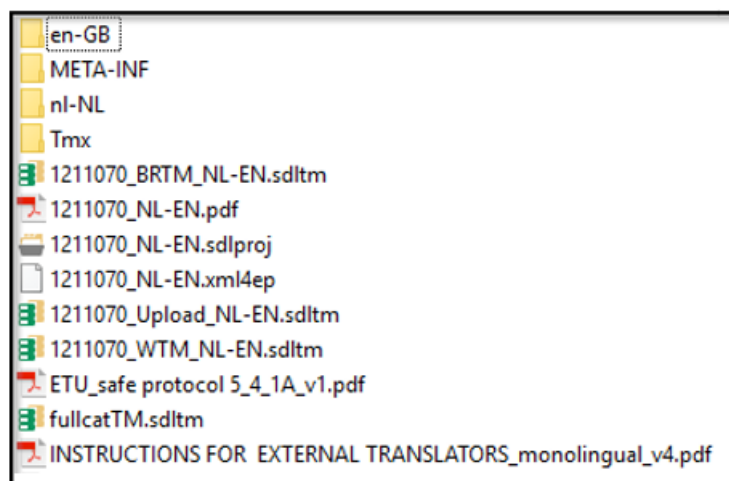


Figure 6: Content of a zip package for a document without two-column amendments - XML4EP original (parliamentary questions: QE, QO).

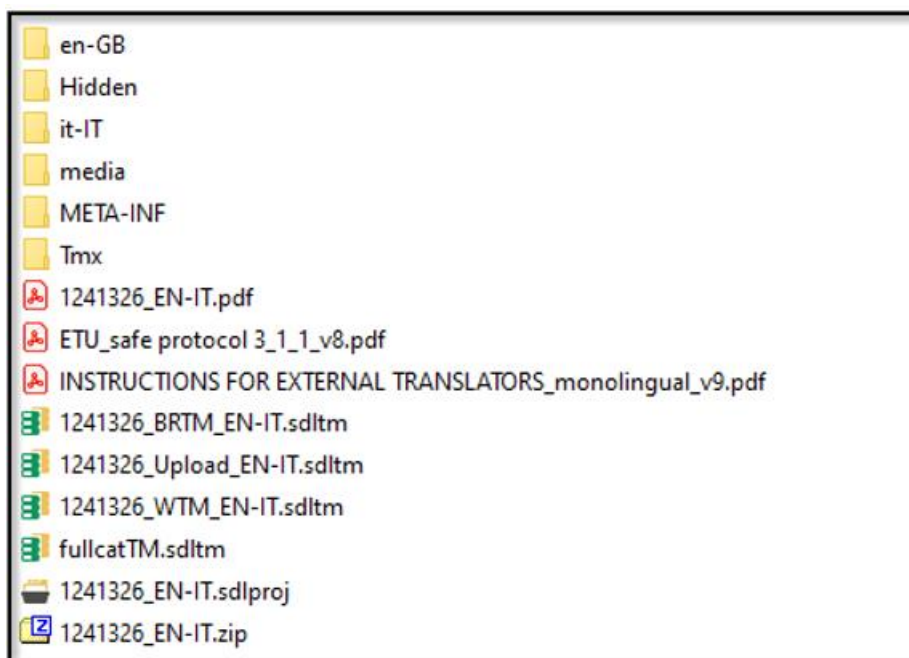


Figure 7: Content of a zip package for a document without two-column amendments - XML4EP original (document types PR, PA).

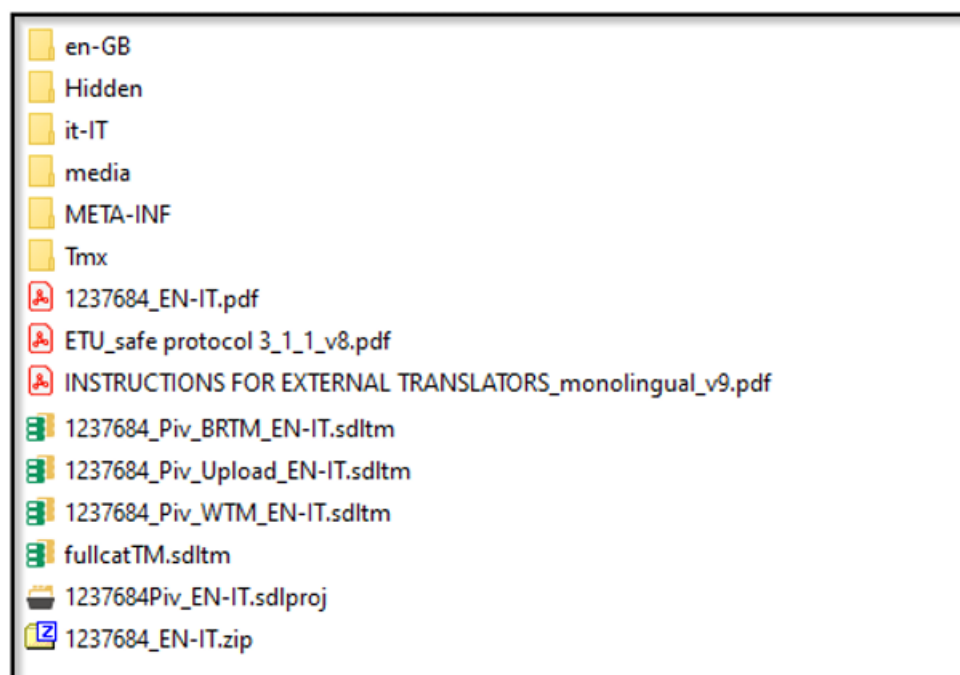


Figure 8: Content of a zip package for a Relay (Pivot) version of a document without two-column amendments - XML4EP original (document types PR, PA)

1.1 Translatable files

1.1.1 Word originals

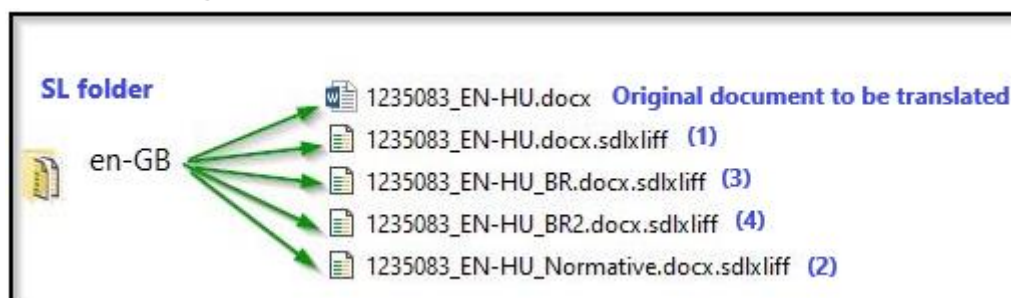


Figure 9: Translatable files - source language folder view

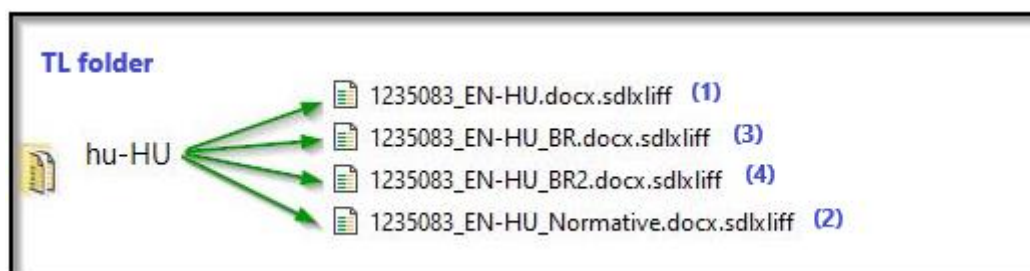


Figure 10: Translatable files - target language folder view

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 are 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**¹) resulting from the pre-translation at 98% against the Normative TM **(2)**.
- 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

¹ Concerns only documents with two-column amendments.

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)**.

1.1.2 XML4EP² originals (monolingual documents without two-column amendments only)

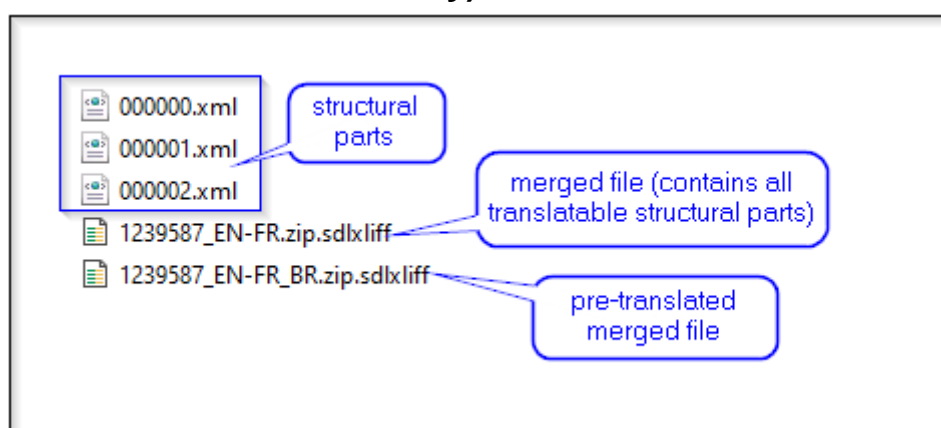


Figure 11: Project SL/TL folder view in Studio - document types PR and PA

The complete original document is saved as a zip file in the project folder. It consists of several XML files (in XML4EP format)³ corresponding to defined structural parts:

- for a PA: the Cover page, Suggestions, Annex
- for a PR: the Cover page, Motion for an EP 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).

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

³ For document types PR and PA. Parliamentary questions consist of only one XML file.

The **translatable structural parts** are **visually merged into one translatable master sdxliff file** in the Studio project. Some structural parts (e.g. Annexes) 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.

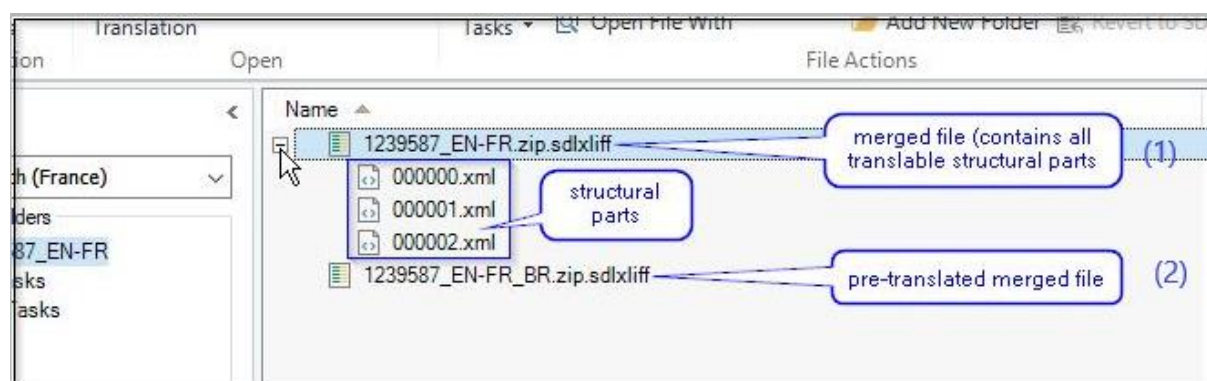


Figure 12: Target language Files view of an XML4EP Studio project - document types PR and PA

The following translatable SDLXLIFF files are included in the project:

- An SDLXLIFF conversion (**1234567_SL-TL.zip.sdxliff**) 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.sdxliff**) resulting from the pre-translation of the translatable merged file at 100% with the BRTM **(2)**.

1.2 XML4EP-specific subfolders in the project folder

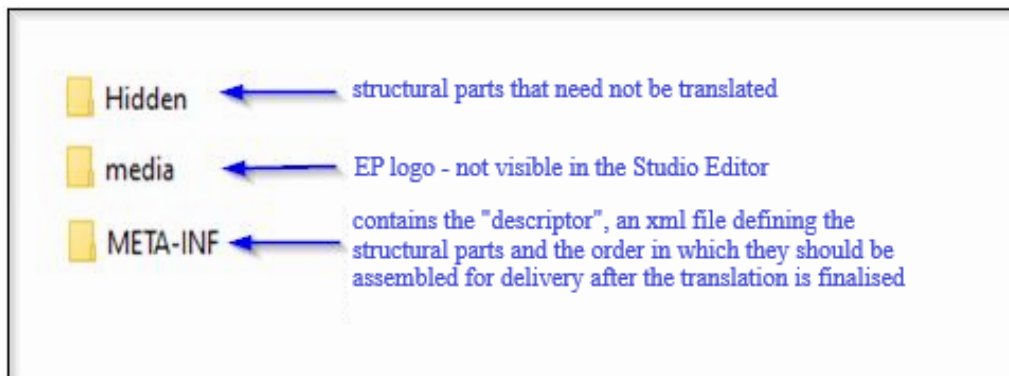


Figure 13: XML4EP-specific subfolders

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.

1.3 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 in the amendment headings preceding each amendment. This TM is created for documents with two-column amendments only.

The **BRTM** contains

- either the Basic Reference tmx file only, in the case of documents with two-column amendments (e.g.: for protocol 2.1 the BRTM contains the .tmx of the COM document);
- or the Basic Reference tmx file and the Normative tmx file containing standard text specific for that particular document type, for documents without two-column amendments.

This memory is used for pre-translating the original document at 100% in order to get a) the Basic Reference document text into the left-hand column or b) the Basic Reference document text and the translation of the standard text.

The **BR2TM** contains the Secondary Basic Reference tmx file(s) only (for example: for protocol 2.6.1 the BR2TM contains the .tmx of the draft report and its amendments). This memory is used for pre-translating the .BR.docx.sdlxliff file at 100% in order to get the text from previous versions into the right-hand column.

The **Working TM (WTM)** contains all the material, i.e. all the tmx files for the Basic Reference document(s), the Retrieval, the Normative tmx (in case of documents without two-column amendments) and the other reference documents belonging to the procedure.

The **Upload TM** and the **fullcat TM** are for EP internal use only, they should be disregarded in your workflow.

Each zip package contains the General Instructions and the ETU Safe Working Protocol (i.e. 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, please use the WTM as a resource for interactive translation in Studio or a Studio-compatible translation tool⁴.

1.4 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).

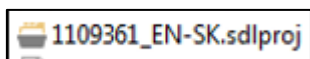


Figure 15: Studio Project

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

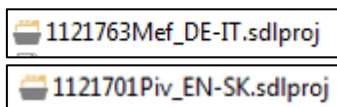


Figure 16: Studio projects for MeF / Pivot versions

⁴ For Word originals only

1.5 TMXs subfolder

The folder contains the tmx files that have already been imported into the TMs according to the Safe Working Protocol.

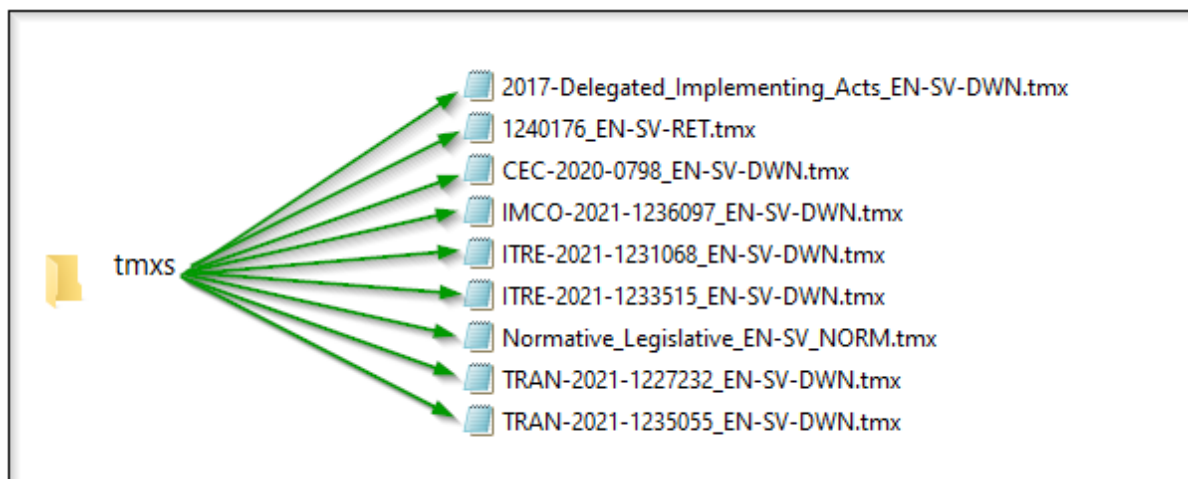


Figure 17: Tmx files in the tmxs subfolder

1.6 Support files in the zip package

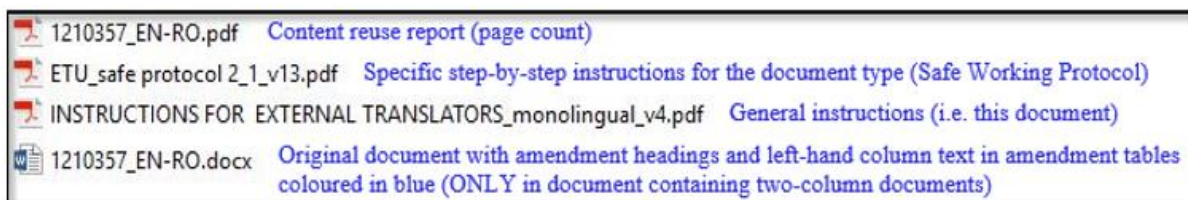


Figure 18: for Word original (see Figure 1)

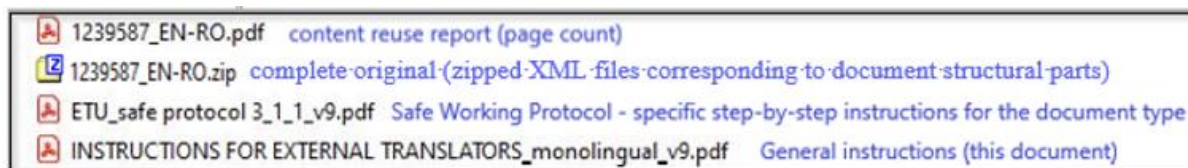



Figure 19: for XML4EP original (see Figure 7)

2. General instructions for external translators

2.1 Use of DocEP

 **docep** is a document creation tool used in the European Parliament. It allows to create 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 starting 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 language⁵.
- 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 Studio will display the DocEP text already available in the target language. 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 target language segments in the pre-translation come from the appropriate TM(s) and the correct reference document(s). Use the WTM to complete the translation.



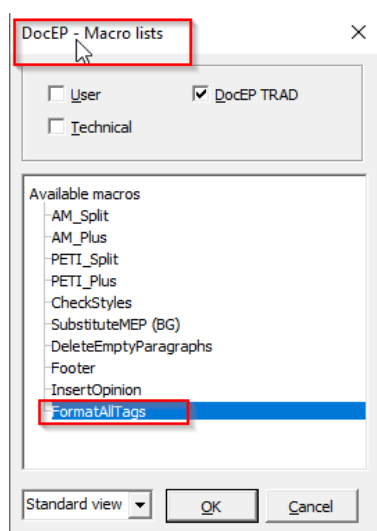
When setting up the document with DocEP and copying the body of the source language text be careful not to change the DocEP tag-specific style definition (*HideTWBExt* and *HideTWBInt* in dark Blue). Please take the following into consideration:

⁵ 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.

- ✓ Use the normal 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 specific functions in DocEP (*Copy AM*, *Copy all AMs*) which allow you to safely copy text with styles.
- ✗ When you use the additional Paste options in MS Word, the styles will be lost. This is also the case with 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 project always check that all DocEP tags have the two mentioned styles. Should it not be the case (the tag style is *Normal*), use the macro *FormatAllTags* from the DocEP Macro Lists menu. It re-applies the correct styles throughout the document.

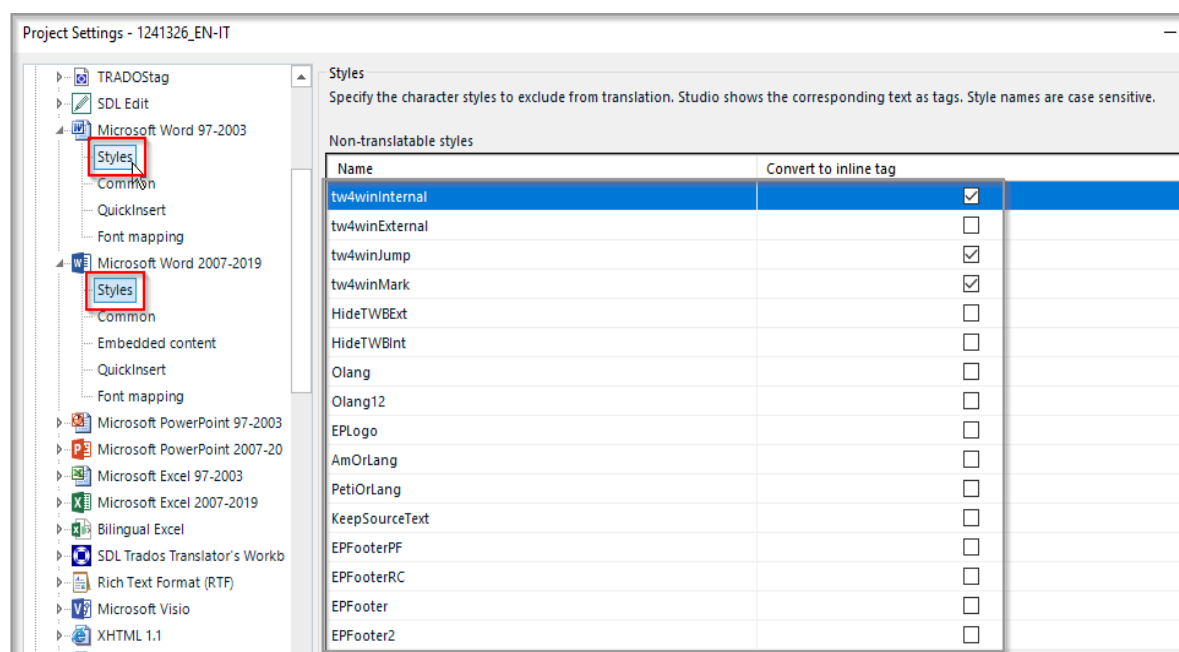


❖ **If the SWP recommends the use of the pre-translated SDLXLIFF file(s) from the pre-treatment package**, the standard workflow is described below:

- Open the project.
- Depending on the document type and SWP, it might contain several pre-translated SDLXLIFF files. If there is a pre-translation which has been produced using the BRT2TM (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 DocEP standard phrases needed for the translation of the original document are included in the Normative_TM or BRTM. Go through the document, checking that the target language segments in the pre-translation 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 run by the EP Language Units in DG TRAD at the end of the workflow.

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

If you use SDL Studio, we recommend that you replicate the settings below in the *File Type* definitions for MS Word 97-2003 and 2007-2019 (under SDL Studio – Project Settings – File Types). DocEP tags will then be treated as structure tags, which by definition remain outside the segments, thus eliminating 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 Studio, the DocEP tags may be contained within the segments. 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.sdlxliiff**) in the package. The standard RdM text has been automatically prefilled, i.e. retrieved and inserted in the file in the target language (e.g. the standard text of the cover page)⁶.

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 sdlxliiff file or, in its absence, the SDLXLIFF conversion of the original file.

2.3 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 you will deliver.

⁶ N.B. It is not possible to see the source text for the prefilled parts in the Studio Editor or in the Preview since the target text replaces the source text. If you want to see the complete original text, you can open the PDF or the Word file provided with the pre-treatment package