

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

v16

Directorate for Language Technology, Support and Innovation

PreTrad - Pre-Processing and Translation Memories Management Unit

Contents

1. Content of pre-treatment zip packages.....	3
1.1 Content of a pre-treatment zip package for a multilingual (XM) document.....	3
1.2 Content of the individual source language (SL) to target language (TL) zip packages.....	3
1.2.1 Translatable files in the individual SL - TL zip packages.....	5
1.2.2 Translation memories in each individual SL - TL zip package.....	6
1.2.3 Studio project file in each individual SL - TL zip package.....	8
1.2.4 TMXs subfolder in each individual SL - TL zip package.....	8
1.2.5 Support files.....	9
2. General instructions for external translators.....	10
2.1 Setting-up the TL document with DocEP.....	10
2.2 Handling DocEP tags in the SDL Trados Studio Editor.....	11
2.3 Multilingual documents without two-column amendments.....	12
2.4 Multilingual documents with two-column amendments.....	13
2.5 Handling of documents with two-column amendments which contain non-contractual source languages (for which a Pivot is required).....	14
2.5.1 For the translation using the original zip package and multilingual document.....	14
2.5.2 For the translation using the Pivot document and package.....	16
2.5.3 Generation of complete TL document.....	17
2.6 Machine Translation.....	17
2.7 Delivery of the TL document.....	18

1. Content of pre-treatment zip packages

1.1 Content of a pre-treatment zip package for a multilingual (XM) document

The pre-treatment package for a multilingual document contains individually zipped packages for each source-to-target language combination, as well as support files.

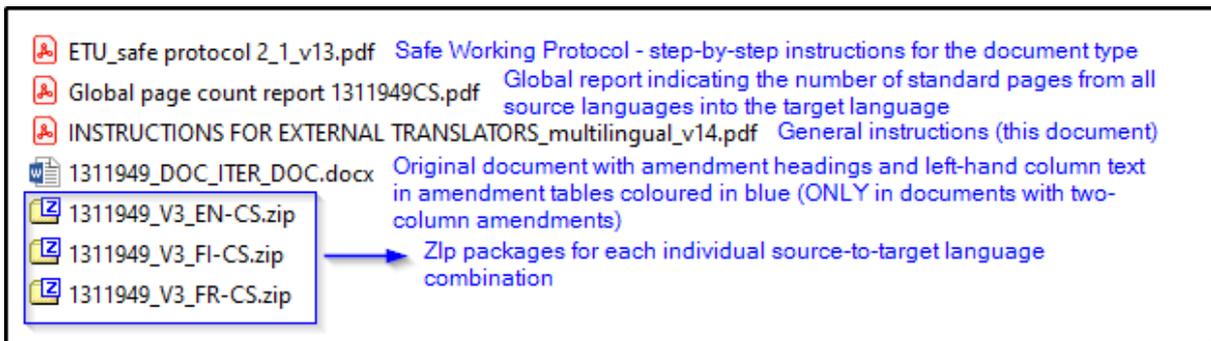


Figure 1: Content of zip package for a multilingual (XM) document.

1.2 Content of the individual source language (SL) to target language (TL) zip packages

The content of the individual SL-TL zip packages is different, depending on whether:

- the source language is EN or any other language, and
- the document contains two-column amendments or not.

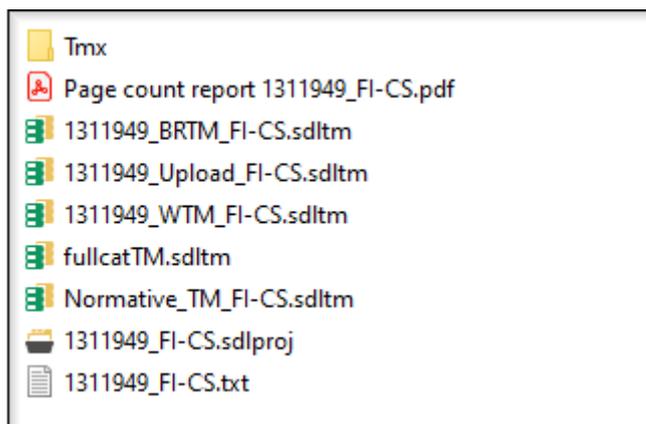


Figure 2: Content of a source language (other than EN) to target language zip package for documents with two-column amendments (general case).

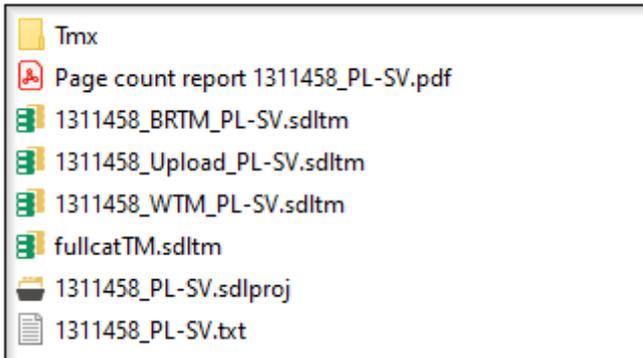


Figure 3: Content of a source language (other than EN) - target language zip package for documents without two-column amendments (general case).

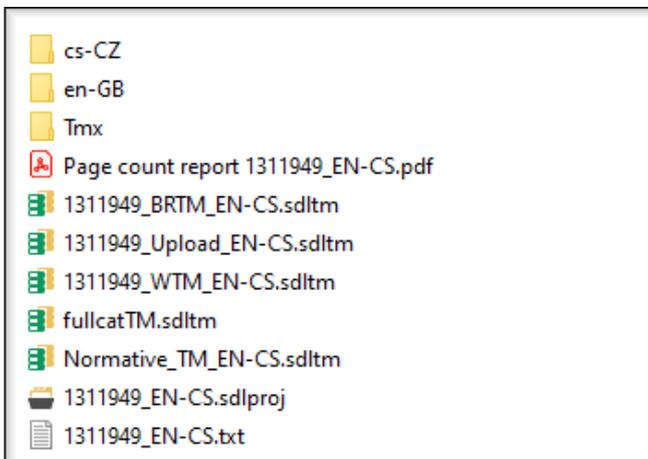


Figure 4: Content of an EN - target language zip package for documents with two-column amendments (exception).

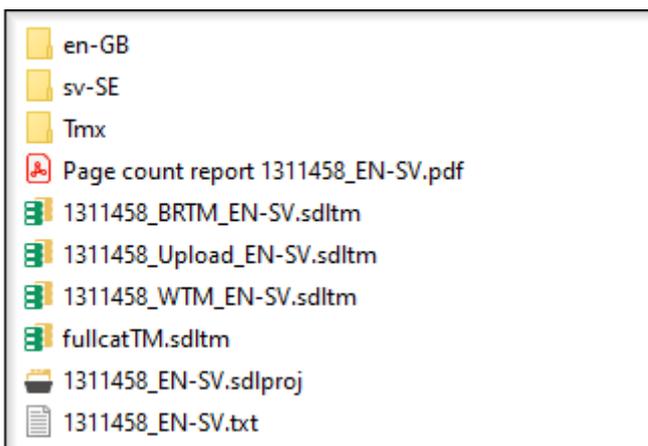


Figure 5: Content of an EN - target language zip package for documents without two-column amendments (exception).

1.2.1 Translatable files in the individual SL - TL zip packages

General rule (Figures 2 and 3) - any source language other than EN

There are NO translatable files in the Studio projects. Therefore, there are no SL or TL folders either.

Exception (Figures 4 and 5)

The only exception concerns the **EN** language - target language zip package, which contains several files, as described below.

Case 1: documents with two-column amendments:

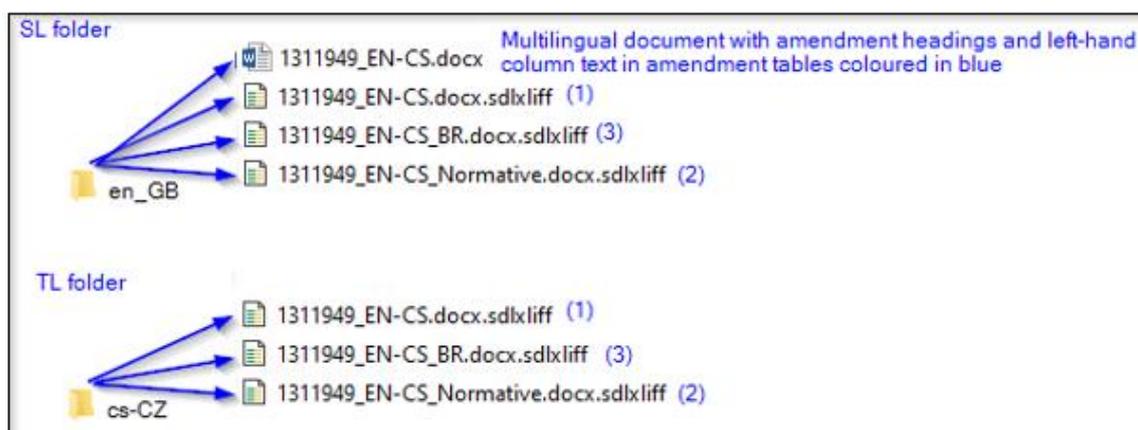


Figure 6: Translatable files in the SL and TL folders - documents with two-column amendments

The **multilingual 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, this is 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.

The following translatable SDLXLIFF files are included:

- The SDLXLIFF conversion (1234567_EN-TL.**docx.sdlxliff**) of the original document to be translated, with the text of amendment headings and of the left-hand column in the amendment tables coloured in blue **(1)**.
- An SDLXLIFF file (1234567_EN-TL_**Normative.docx.sdlxliff**) resulting from the pre-translation at 98% with the Normative_TM **(2)**.
- An SDLXLIFF file (1234567_EN-TL_**BR.docx.sdlxliff**) resulting from the pre-translation at 98% with the Normative_TM and a consecutive pre-translation at 100% with the BRTM **(3)**.

Case 2: documents without two-column amendments:



Figure 7: Translatable files in the SL and TL folders - documents without two-column amendments

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

The following translatable SDLXLIFF files are included:

- An SDLXLIFF conversion (1234567_EN-TL.**docx.sdlxliff**) of the original document to be translated **(1)**.
- An SDLXLIFF file (1234567_EN-TL_**BR.docx.sdlxliff**) resulting from the pre-translation at 100% with the BRTM **(2)**.

1.2.2 Translation memories in each individual SL - TL zip package

Depending on the document type, each individual SL-TL package may contain:

 1197349_BR2TM_EN-RO.sdltm	a Basic Reference 2 Memory
 1197349_BRTM_EN-RO.sdltm	a Basic Reference Memory
 1197349_Upload_EN-RO.sdltm	an Upload Memory (can be disregarded in your workflow)
 1197349_WTM_EN-RO.sdltm	a Working Memory
 fullcatTM.sdltm	a fullcat Memory (can be disregarded in your workflow)
 Normative_TM_EN-RO.sdltm	a Normative Memory

Figure 8: Translation Memories

The **Normative_TM** contains the tmx file of the standard segments 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 from 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 from each individual SL-TL package as a resource for interactive translation in Studio or a Studio-compatible translation tool.

1.2.3 Studio project file in each individual SL - TL zip package

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

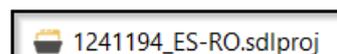
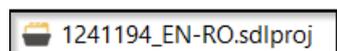


Figure 9: Studio Projects

1.2.4 TMXs subfolder in each individual SL - TL zip package

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.

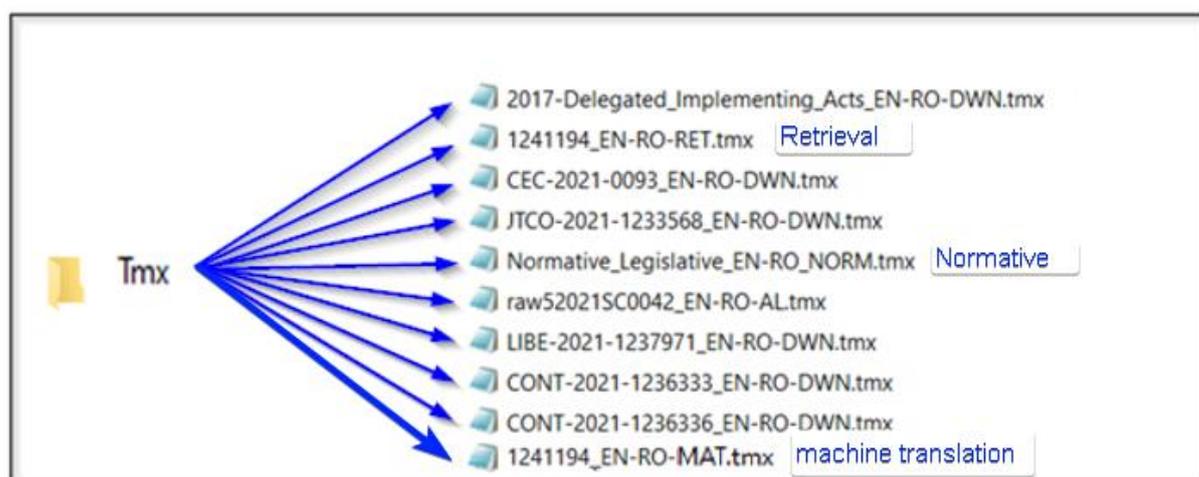


Figure 10: Tmx files in the Tmx subfolder

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

1.2.5 Support files

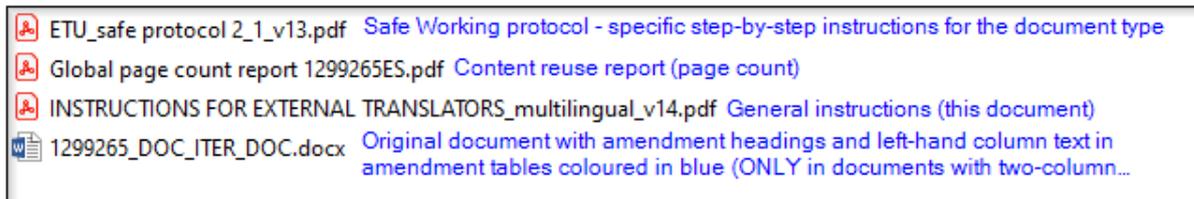


Figure 11: Support files in the zip package for the multilingual document.



Figure 12: Additional support files in the individual SL - TL zip packages.

2. General instructions for external translators

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

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

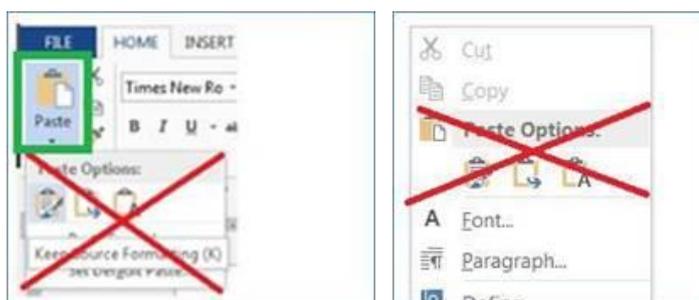
2.1 Setting-up the TL document with DocEP



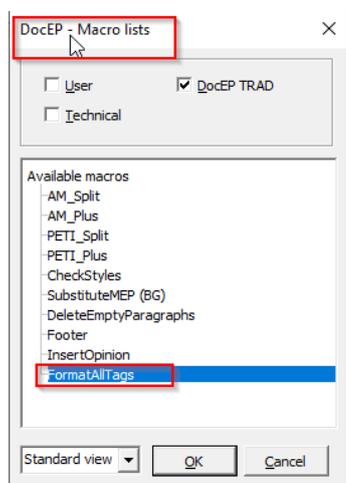
docep is a document creation tool (Word macro) used in the European Parliament. It allows the creation of documents which contain the official standard phrases encoded in the Recueil de Modèles (RdM) already in the target language.

If the SWP recommends setting up the target language (TL) document to be used as working document in Studio with DocEP 'Create Document', when 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:

- ✓ 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 AM-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.

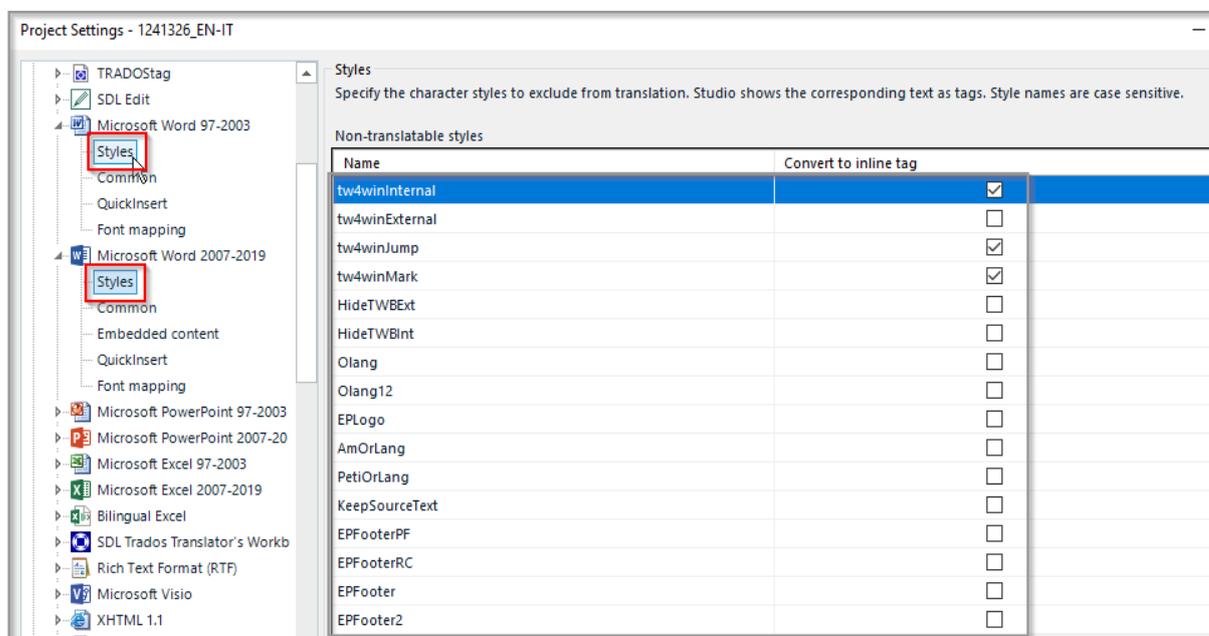


- ✓ Before adding the set-up document to the Studio project, always check that all DocEP tags have the two styles mentioned above. Should it not be the case (the tag style is *Normal*), use the macro 'FormatAllTags' from the DocEP Macro Lists menu. It will re-apply the correct styles throughout the document.



2.2 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.

2.3 Multilingual documents without two-column amendments

If the document contains text in non-contractual source languages (for a which a Pivot is required), the general rules are as follows:

- Set up the target language document with the document specific DocEP template (if recommended by the SWP) and copy the body of the original multilingual document into the new TL document.
- Split or divide the multilingual TL document by source languages using DocEP 'AM_Split'. Leave the splits corresponding to non-contractual source languages aside.
- For the contractual languages: add each SL-TL split document to the corresponding SL-TL Studio project. Perform the following steps for each SL-TL document split:
 1. Pre-translate with the BRTM at 99% in order to get the unchanged text from the basic reference document and/or any relevant standard text from the Normative TMX file(s).
 2. Open the pre-translated file. Since the target language document was initially set up with DocEP 'Create Document', 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.
 3. Go through the document, checking that the pre-translated target language segments come from the appropriate TM and the correct reference document(s). Use the WTM to complete the translation.
 4. Generate the translated TL document splits and a bilingual SDLXLIFF file (or a standard bilingual XLIFF file) for each contractual SL-TL combination.
- When the Pivot package arrives: split the Pivot document source language-wise and save the parts corresponding to the source languages that you do not cover in a separate document.

- Add this split to the PIV-TL project (you can also use the EN-TL project from the initial zip package).
- Perform steps 1 - 3 described above. When the translation has been completed, generate the TL translation and a bilingual PIV-TL SDLXLIFF file (or a standard bilingual XLIFF file).
- Once you have all the translated TL document splits in MS Word, assemble them to get the final complete TL document.
- Run DocEP 'Change document ID'.

2.4 Multilingual documents with two-column amendments

For multilingual documents with two-column amendments, it is recommended to:

- either work with the Studio project from the EN - target language package, which contains the coloured complete multilingual original, pre-translated with the Normative_TM (Normative.docx.sdxliff);
- or pre-translate yourself the complete coloured multilingual original¹ at 98% using the Normative_TM from the EN - target language Studio project.

This will allow you to translate the standard text of the **amendment headings**, which are **always in English**. After that, you can split the original document source language-wise using the DocEP 'AM_Split' macro.

Follow the steps described below, under point 2.5.1.

Once you have all the translated TL document splits, assemble them to get the final complete TL document with the DocEP 'AM_Plus' macro.

❖ **How to handle amendments when the amended text in the source language does not affect the translation in the target language**

If the document contains amendments where the changes in the source language do not affect the target language translation, follow these steps:

1. Check that the pre-translation of the left-hand column (the 'blue text') is correct and that it comes from the relevant reference document.
2. Confirm the left-hand column text, including the footnotes, if any.
3. In the first segment of the amendment right-hand column ('black text') – whether it is a 'num' (e.g. a paragraph number) or a content segment – insert the standard "Does not affect TL" phrase. Lock the segment without confirming the translation.

¹ 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.

You can find the target language translations of “Does not affect TL” in Annex 1 below.

4. Add a space in the remaining right-hand column segments, including any footnote segments, then lock them without confirming.
5. If the amendment has a Justification, translate the heading “Justification” into your target language and confirm the segment.
6. Insert the standard “Does not affect TL” phrase in the first segment of the Justification text. Lock the segment without confirming the translation.
7. If the Justification contains more than one segment, add a space in the remaining segments, then lock them without confirming.

130	Amendment		Amendement
131	5		
132	Proposal for a regulation		Proposition de règlement
133	Recital 10		Considérant 10
134	<i>Text proposed by the Commission</i>		<i>Texte proposé par la Commission</i>
135	<i>Amendment</i>		<i>Amendement</i>
136	(10)		
137	In this context, the processing of special categories of personal data which is necessary for the purposes of carrying out the obligations and exercising specific rights of the controller or of the data subject in the field of employment and social security and social protection law, as referred to in Article 9(2), point (b), of Regulation (EU) 2016/679, should not as such require that records of processing be <i>maintained</i> .		Dans ce contexte, le traitement de catégories particulières de données à caractère personnel qui est nécessaire aux fins de l'exécution des obligations et de l'exercice des droits propres au responsable du traitement ou à la personne concernée en matière de droit du travail, de la sécurité sociale et de la protection sociale, tel que visé à l'article 9, paragraphe 2, point b), du règlement (UE) 2016/679, ne devrait pas, en tant que tel, exiger la tenue de registres pour les activités de traitement.
138	(10)		<i>(Ne concerne pas la version française.)</i>
139	In this context, the processing of special categories of personal data which is necessary for the purposes of carrying out the obligations and exercising specific rights of the controller or of the data subject in the field of employment and social security and social protection law, as referred to in Article 9(2), point (b), of Regulation (EU) 2016/679, should not as such require that records of processing be <i>kept</i> .		
140	<i>Justification</i>		<i>Justification</i>
141	<i>To ensure terminological consistency.</i>		<i>(Ne concerne pas la version française.)</i>
142	<i>To clarify that this refers to the record-keeping obligation stipulated under recital 9.</i>		

8. Ignore the errors or warnings related to these locked and draft segments (such as empty content or missing tags), displayed when running the QA Checker.

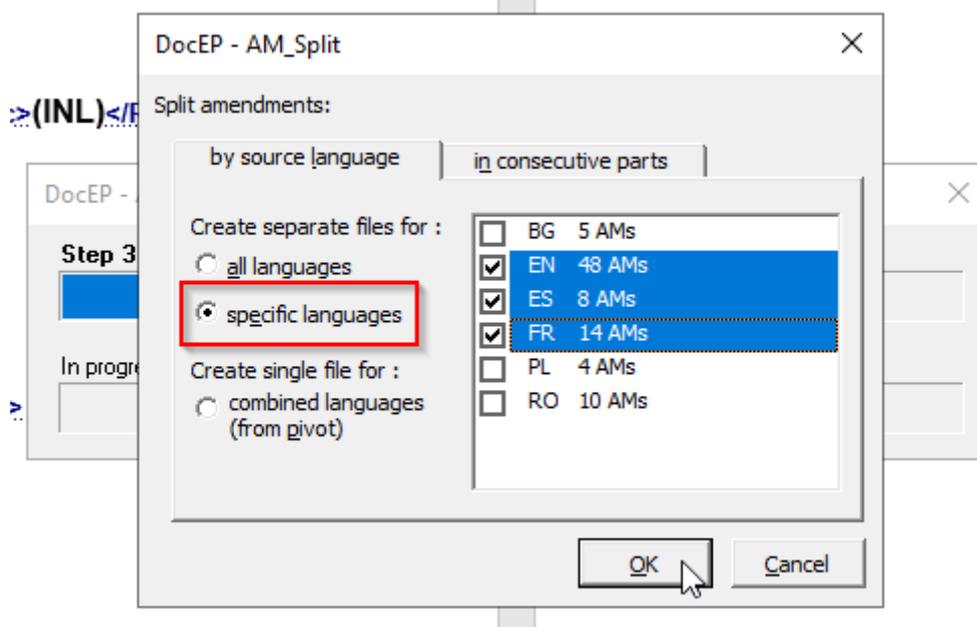
2.5 Handling of documents with two-column amendments which contain non-contractual source languages (for which a Pivot is required)

2.5.1 For the translation using the original zip package and multilingual document

Open the Studio project from the individual EN-TL package. It contains the original multilingual document pre-translated with the Normative_TM in order to get the translation of the standard segments of each amendment heading and the DocEP text from the cover page. Use this pre-translated SDLXLIFF file (Normative.docx.sdlxiff) for all further steps. Check that all the pre-translated standard segments are correct and complete the translation of the amendment headings using the Normative_TM.

Convert the Normative.docx.sdlxiff back to Word and split the document source language-wise (with the DocEP 'AM_Split' macro) in order to get the amendment

batches from the contractual source languages (EN, FR, DE, ES, IT). Use the option 'Create separate files for specific languages'.



The steps below have to be repeated for each contractual SL-TL split.



Add each split document to the corresponding SL-TL Studio project. At this stage, the source language column in the Studio Editor already contains the DocEP text and the amendment headings in the target language. Lock these segments.

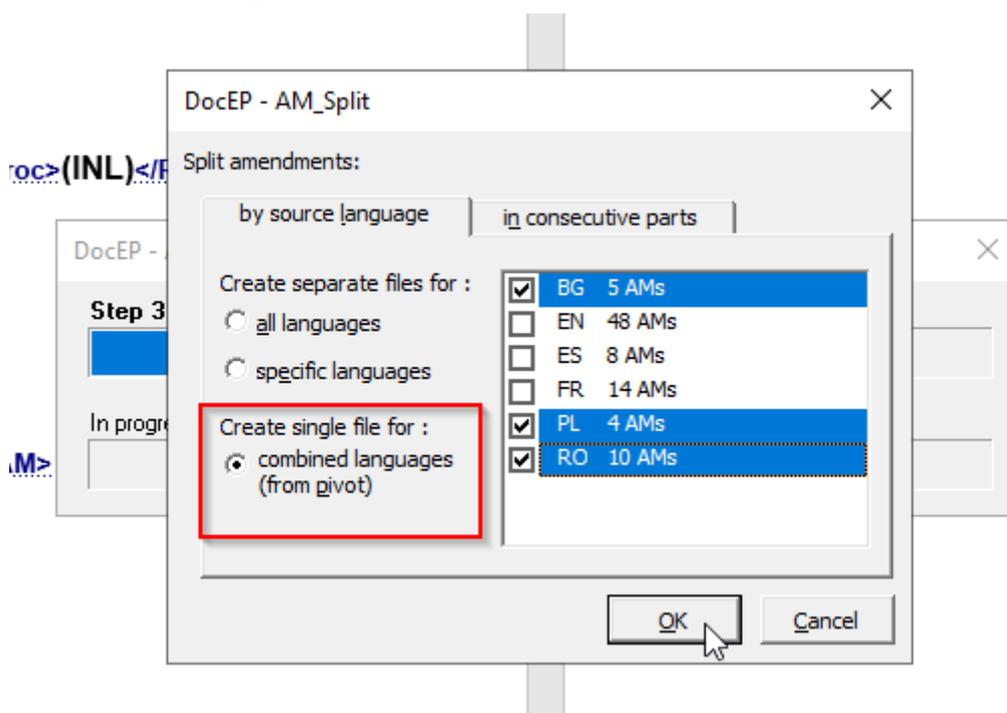
Pre-translate the TL document using the BRTM at 100% in order to get the reference document text into the left-hand column. Check that all the pre-translated segments (including those marked as 100%) are correct and complete the translation of the left-hand column using the BRTM.

Enable the WTM and use it to complete the translation of the document. Save the file frequently. While interactively translating in Studio, mark the differences between the left-hand column and the right-hand column in ***Bold and Italic***.

When the translation has been completed, generate a bilingual SDLXLIFF file (or a standard bilingual XLIFF file) for each contractual SL-TL combination.

2.5.2 For the translation using the Pivot document and package

Split the pivot document source language-wise (with the DocEP 'AM_Split' macro) in order to get the amendment batches from the source languages you do not cover. Use the option 'Create single file for combined languages (from pivot)'. You will obtain a single document named 'FdRnumberpivot2TL'.



Add this document to the PIV-TL project (you can also use the EN-TL project from the initial package).

Pre-translate the TL document using the Normative_TM at 98% in order to get the standard segments of each amendment heading. Check that all the pre-translated standard segments are correct and complete the translation of the amendment headings using the Normative_TM.

Pre-translate the TL document using the BRTM at 100% in order to get the reference document text into the left-hand column. Check that all the pre-translated segments (including those marked as 100%) are correct and complete the translation of the left-hand column using the BRTM.

Enable the WTM and use it to complete the translation of the document. Save the file frequently. While interactively translating in Studio, mark the differences between the left-hand column and the right-hand column in *Bold and Italic*.

When the translation has been completed, generate a Pivot-TL bilingual SDLXLIFF file (or a standard bilingual XLIFF file).

2.5.3 Generation of complete TL document

Generate all split source language-specific TL documents treated according to points 2.3.1 and 2.3.2 above.

Assemble the final complete TL document with the DocEP 'AM_Plus' macro.

Run DocEP 'Change Document ID'.

2.6 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 available in the individual SL-TL zip packages. 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 files, you need to always keep an eye on the source text to prevent the risk of serious mistranslations.

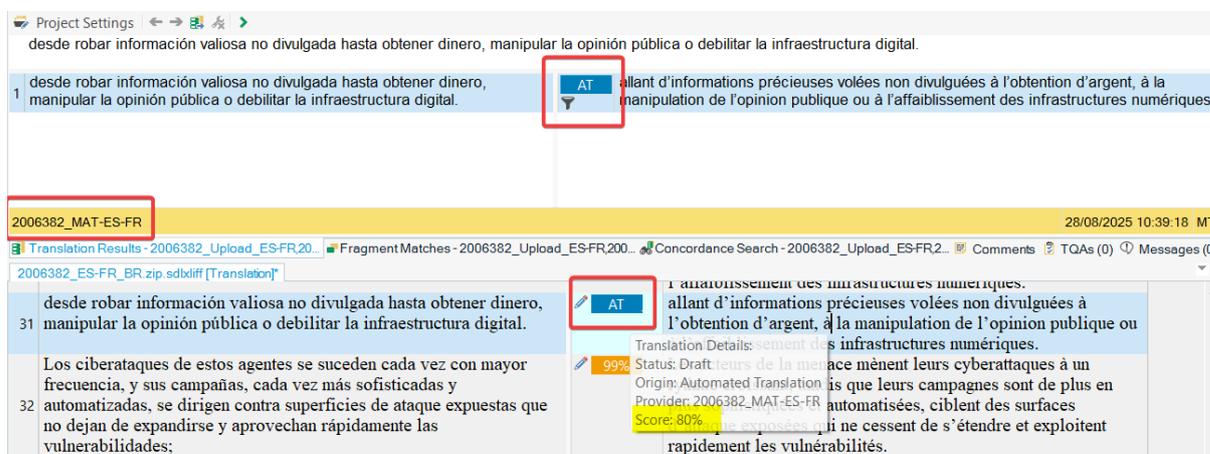
To safely use the machine translation tmx provided for each contractual SL-TL combination, you will need to import it into a separate Machine Translation_TM. This is the current practice in DG Trad.

The steps to follow for each contractual SL-TL combination:

- Create a separate SL-TL Machine Translation_TM
- Add the Machine Translation_TM to the SL-TL 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.



2.7 Delivery of the TL document

Deliver

1) a monolingual MS Word document corresponding to your translation of the complete multilingual document.

If the multilingual file has been split into monolingual parts during the translation process, reassemble it using the DocEP function 'AM_Plus' (for document type AM) or do it manually (in case of all other types of documents) before the delivery.

All tagging and formatting from the original document must be kept or replicated in the MS Word file that you will deliver. Check that the DocEP tags have the correct *HideTWBExt* and *HideTWBInt* styles. If they take the *Normal* style instead, use the 'FormatAllTags' macro on the DocEP Macro Lists menu.;

AND

2) a bilingual SDLXLIFF file (or a standard bilingual XLIFF file) for each SL-TL combination, for the contractual languages and, if applicable, a bilingual SDLXLIFF PIV-TL file for the other languages. All tagging and formatting from original document must be kept or replicated in the bilingual (SDL)XLIFF files you will deliver.