How to prepare technical documentation files for localization using CAT software?

Contents

What is localization and why is it important?

Explanation of what location is.

According to the Globalization and Localization Association (GALA), localization is the whole process of adapting a product or content for the target region, location, or market.

Usually, localization is regarded only as a synonym for translation of the user interface and documentation, but it's often a substantially more complex issue. Localization also involves adapting other elements to a target market, including:

• Addressing local regulations and legal requirements
• Using proper formatting for details like dates, addresses, phone numbers, etc.
• Converting to local currencies and measurement units
• Changing content according to preferences
• Modifying graphics, design, symbols, icons, and colors to display translated text correctly. It should also be noted whether these elements in a given culture will not be misinterpreted and considered inappropriate

In other words, localization gives the product, documentation, or service the look and feel expected by the target audience.

What is the purpose of localization?

Explaining what localization is for.

Localization aims to give the product or service the look and feel of having been created specifically for the target audience, no matter their language, cultural preferences, or location. So, localization gives something the look and feel expected by the target audience. Thanks to this, the content is more appealing, which makes the audience more likely to buy. Consumers are more likely to purchase if the corresponding product information is in their native language. Customers prefer the content created in their native language, especially if their understanding of the content in the source language is low or non-existent.

What is internationalization and how is it different from localization?

Differences between internationalization and localization.

Internationalization starts at the conceptual stage for products and services delivered globally.

According to Wikipedia, internationalization is the process of designing the content to be adapted to various languages and regions without engineering changes. Localization is the process of adapting the internationalized content for a specific region or language by translating text and adding locale-specific components. Localization can be potentially performed multiple times for different locales. It uses the infrastructure provided by internationalization, which is ideally performed only once before localization or as an integral part of ongoing development.
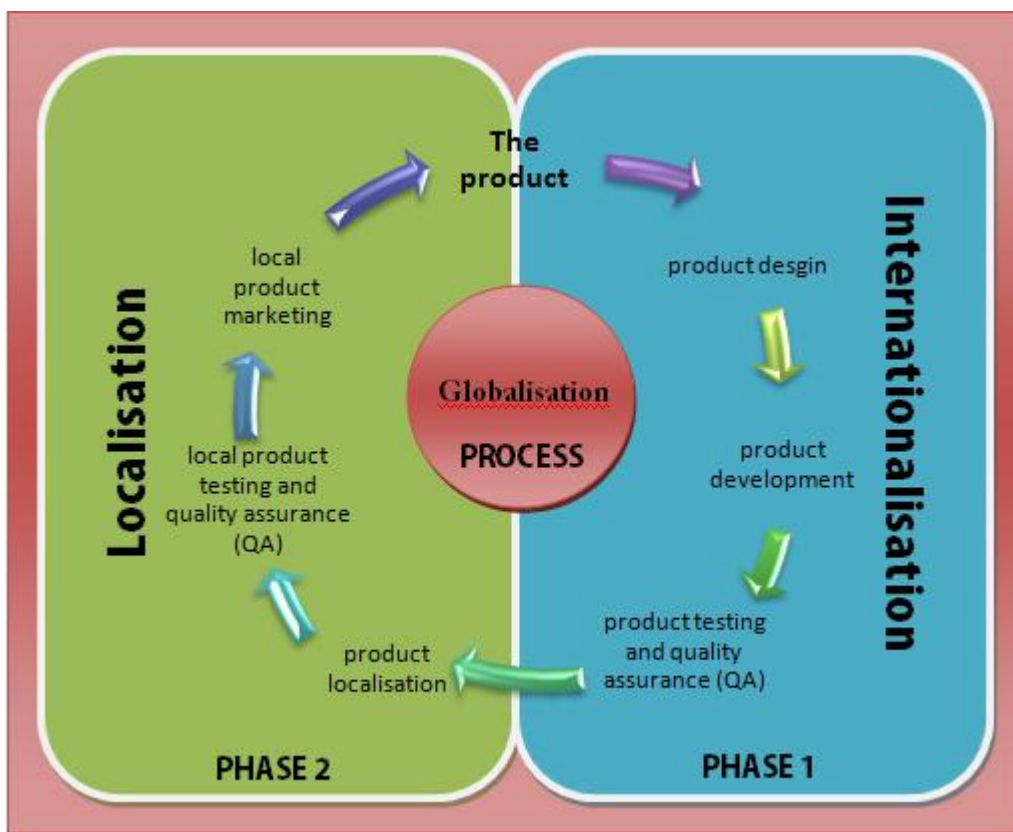
Figure 1: Internationalization and localization process, source: Wikipedia

Related information

https://en.wikipedia.org/wiki/Internationalization_and_localization

What is translation and how is it different from localization?

Differences between translation and localization

Translation is often defined as the literal word-for-word transformation of a source text into the target language. However, this is somewhat of a simplification because translation aims to maintain the original meaning through vocabulary choices and idiomatic choices. A translator converts the content into the target language while respecting syntax and grammar rules. Typically, translation is the most time-consuming component of language localization.

In common usage, the terms "translation" and "localization" are incorrectly interchanged or misused regarding the content's translation. Indeed, translation is a part of localization, but so much more is required for content to become authentic and locally accepted. When localizing the content, you should pay even more attention to the target market's cultural nuances and other specifics.

Figure 2: Differences between translation and localization, source: quicksilvertranslate.com

How should technical documentation be created for localization?

Learn what documentation should be to make it easy for localization.

- Product documentation should be created in such a way that it can be translated into other languages without involving additional effort and time, e.g., for reengineering.
- Product documentation should be planned as early as possible, preferably as early as the product design stage.
- Define the purpose and scope of the documentation and the end-user who will use the documentation.
- Select the style guide, convention, as well as source and target formats.
- Minimize fragmentation of translatable text. Features such as snippet, keyref, conref , etc. allow reuse of text. From the author's point of view, it is beneficial for the consistency of the complete documentation. However, the text fragmented into parts smaller than one sentence should not be intended for translation. The parts will be difficult to translate and localize well.
- Provide information about non-translatable keys or just normalize content before translation. The non-translatable keys are simplifications to make the documentation easier to create, but they can be cumbersome to translate. Normalization means processing the text to make it easier to localize. Some of the keys that can be translated are replaced with plain text.

Related concepts

How does CAT work? on page 9
About CAT concept

What should be translated, and how should it be done?

Decide what and how should be translantable.

Define the text you want to be translated: trademarks, product names, code samples, legal fragments, etc.

Share glossaries with translators: it's the easiest way to provide info what is "Do Not Translate". Also, it's the easiest way to get consistent terminology in translation.

Should the drawing text also be translated and localized?

Learn how to translate and localize the drawing text.

Yes, the text on the drawings also should be translated and localized.

Place the text on a separate layer and avoid text fragmentation.

If possible, use formats or exports that are compatible with CAT solutions (e.g., SVG).

If User Interface is also localized, handle screen captures.

What should be the formats for creating documentation?

Properties of formats for creating documentation.

- Documentation formats should allow for easy localization.
- Documentation text should be easily available for editing or extracting without any additional effort (e.g., reengineering).
- The text should be available for reuse.
- And last but not least: The documentation formats should be translatable.

What are the best formats for creating documentation?

DITA, MadCap, Markdown as formats for creating documentation.

Figure 3: DITA logo

DITA (Darwin Information Typing Architecture)specification defines a set of document types for authoring and organizing topic-oriented information and a set of mechanisms for combining, extending, and constraining document types. It's an open standard that is defined and maintained by the DITA Technical Committee. DITA provides support for translation via the localization attribute group. Set element attributes to indicate whether the content of the part should be translated or not.

Figure 4: MadCap logo

MadCap means the format used in MadCap Flare software. It uses single-source publishing. The method allows content to be used more than once across different media. Flare supports topic-based authoring and content management. Topic-based authoring means the content is written in modular topics that can stand alone and be mixed or reused as needed. Produce online help systems, eLearning modules, knowledge bases, document portals, and various guides and manuals for use online or in print. MadCap Flare provides many ways to reuse content, which saves time maintaining or updating topics.

Figure 5: Markdown logo

Markdown is the tool for easy and quick creating and publishing documentation. The tool allows you to write using an easy-to-read, easy-to-write plain text format, then convert it, for example, to structurally valid XHTML or HTML. The format is simple and readable without editing. It's easy to manage a large number of documents and reuse content. Visual issues are separated from the content. The format allows you generate other input and output formats from .md file. Documentation can be expanded to any size, and the file can be modified, or content can be extracted from it.

What other formats can be used to create documentation?

InDesign, HTML, XML as formats for creating documentation.

Figure 6: InDesign logo

Adobe InDesign is a desktop software application for publishing and typesetting. InDesig allows you to create works such as posters, flyers, brochures, magazines, newspapers, presentations, books, and ebooks. InDesign supports XML, style sheets, and other coding markups, making it suitable for exporting tagged text content for use in other digital and online formats.

Figure 7: HTML logo

HTML is a markup language and it's the standard for the Internet and many applications. It's a text format and can be modified. You can extend the documentation to any size. Other input and output formats can be generated, but not all. It's only partially possible to separate visual issues from the content and HTML can look different in different browsers and environments.

Figure 8: XML logo

XML also is a markup language and defines a set of rules for encoding documents in a human-readable and machine-readable format. The design goals of XML emphasize simplicity, generality, and usability across the Internet. It's a textual data format with strong support via Unicode for different human languages. XML design focuses on documents but the language is widely used to represent arbitrary data structures such as those used in web services.

Figure 9: Logo Word and Excel

Word and Excel by Microsoft are sometimes used for documentation purposes because they are widely available and familiar formats. However, they have their limitations and should not be the "formats of choice" for more extensive documentation or for localization. On the other hand, they work well for providing documentation.

Word allows simple formatting and you can generate other input and output formats. You can modify the .docx file or extract the content. However, the format is not suitable for large documents. It's challenging to manage a large number of documents and reuse content. It's hard to separate visual issues from the content. Collaboration between authors and other content consumers is difficult.

Excel is a spreadsheet developed by Microsoft for Windows, macOS, Android, and iOS. Excel has sometimes been used to create documentation (e.g., parts lists, checklists), but it has limitations. The content is difficult to manage and localize.

What formats should NOT be used for documentation?

PDF should not be used for creating documentation.

PDF is suitable for providing documentation (including localized documentation), but it is very unsuitable for creating documentation files intended for localization. For localization purposes, it is always better to take the source files from which the pdf file was created.

PDF file looks the same everywhere and can be read everywhere. Once generated, it's stable for large documents. However, it's a closed document, so you cannot easily modify it or download content from it. Also, It's not possible to generate other input or output formats from .pdf file.

What is Simplified Technical English (STE)?

Learn about Simplified Technical English (STE).

English is the international language of many industries, and the language is the most used for writing technical documentation. However, it often isn't the native language of the readers (or even of the authors) of such documentation.

Simplified Technical English (STE) is a controlled language developed in the early 1980s for commercial aviation to help second-language English speakers understand technical manuals written in English. Many maintenance manuals are written in STEbecause STE makes technical texts easy to understand by all readers. Also, by those who have limited knowledge of English and are confused by complex sentence structures, as well as by the number of meanings and synonyms of English words. STE principles improve text comprehensibility in any industry, and comprehensibility helps with localization.

How does STE work?

About STE concept

The STE Specification provides a set of writing rules and a dictionary of controlled vocabulary. The writing rules cover aspects of grammar and style, and the dictionary specifies the general words that can be used. The words were chosen for their simplicity and ease of recognition. In general, use **one word for one meaning** and **one part of speech for one word**. Writers can use the approved words in the dictionary as a core vocabulary. But they can also use terms that are usual in their companies or industries and applicable to their projects and products.

How to write using STE?

Rules of STE

- BASIC RULE: "one word – one meaning." It means the elimination of synonyms, i.e., using only one allowed word for a particular meaning. E.g., using "to fall" we mean "to move down by the force of gravity" and not "to decrease."
- In the case of English dialects, American is preferred.
- Limit to the most precise forms and eliminate forms difficult to understand.
- Split any long clusters of nouns into smaller logical units (no more than three words).
- Use active voice, not passive.
- Express actions with verbs and not by nouns.
- The sentences should not be too long, 20 to 25 words per sentence. Also, sections should not be too long, up to 6 sentences in a section.
- Warnings and caution notes must be properly marked as such. At the very beginning, indicate the actions to be taken. Descriptions of the danger should be placed later.
- Separately indicate advisory notes and explaining details, and they should contain no commands.

What is CAT software and what is it used for?

Learn about CAT sofware.

Computer-aided translation (CAT) means the use of software to assist a human translator in the translation process. A human creates the translation, and software only facilitates certain aspects of the process.

CAT does not mean machine translation (MT), where a computer creates the translation.

CAT software is an excellent tool for translating and localizing technical documentation. They also make it easy to manage terminology and content simultaneously in many different languages.

CAT programs typically are the tools that specifically facilitate the actual translation process. Most CAT tools provide:

- The ability to translate a variety of source file formats in a single editing environment,
- Translation memory, and
- Integration of various utilities or processes that increase productivity and consistency in translation.

How does CAT work?

About CAT concept

First, CAT tool breaks the source text into translation units or segments (TUs). Usually, TUs correspond to sentences or similar wholes. Then, the tool creates a translation memory (TM) for storing the source and translated segments. The CAT search engine allows the translator to view saved segments, search for segments with similar content, and translate them.

With the help of termbases, glossaries, and translation memories, the CAT tool helps the translator ensure the terminology and writing style of the source file are translated accurately and as intended.

Depending on available options, CAT tools include such features like autosuggest, quality assurance, side-by-side alignment, optionally feature of the Machine Translation service, and much more. Use a spell checker, manage terms, use specific dictionaries, term databases, text indexers, merge the translation and the source text, manage projects, translation memories, and automate translation work in general.

Related concepts

Learn what documentation should be to make it easy for localization.

What CAT software may be used for localization?

Trados, memoQ and Déjà Vu X3 as tools for localization.

The use of CAT software depends, among other things, on the format of the localized documentation. It can be known the best or other formats for localization.

You can use any CAT tool as long as you have the compatible format.

There is a wide variety of CAT tools available on the market. Each CAT program offers both essential functions (e.g., creation of translation memory (TM), concordance search in TM, and more advanced functions (e.g., terminology management, alignment, optionally feature of the Machine Translation service, and many other) or functions that translation agencies use to manage very complex projects. The choice of CAT software will depend on many factors.

See examples of the most popular and commercially available CAT tools. Detailed information about each CAT software is available on the corresponding website.

Trados

Probably **Trados** is the biggest of all the big players in the CAT tool market. It's a complete translation software solution for translating, managing terminology, editing, and running LQA.

memoQ

memoQ is also major CAT software but not as well-known as Trados. It offers practisely the same features and has some useful extras, such as the translation preview pane to see the segment you are translating in context. memoQ can accept Trados and other packages, meaning that translators can usually work together with agencies who use other translation software.

Déjà Vu X3

Déjà Vu is another popular CAT tool that offers a complete host of terminology management and translation memory features. It allows to handle a wide variety of file formats, including Trados and memoQ.

Example: how to localize file using Trados?

Example: documentation files prepared using **Oxygene XML Editor (.dita)** and localized using **TRADOS**

To localize the documentatation files, create a project in Trados.

1. Open Trados.

2. Go to the **Projects View** pane and select **New Project** on the upper ribbon. In the displayed dialog, specify the data of your new project and add source files. Click **Next**.

3. Specify general settings of your new project and click **Next**.

4. Select translation memory TM or create a new one. Click **Next**.
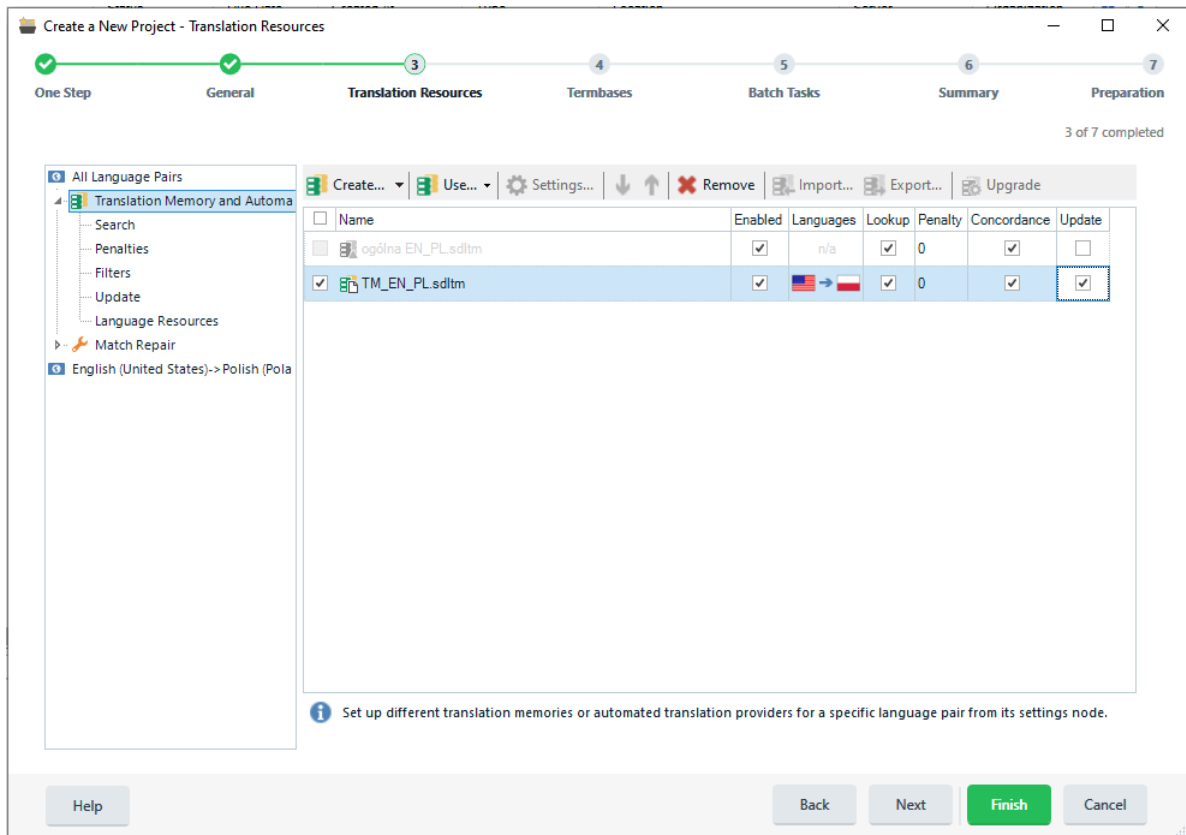
5. Select term base TB or create a new one. Click **Next**.
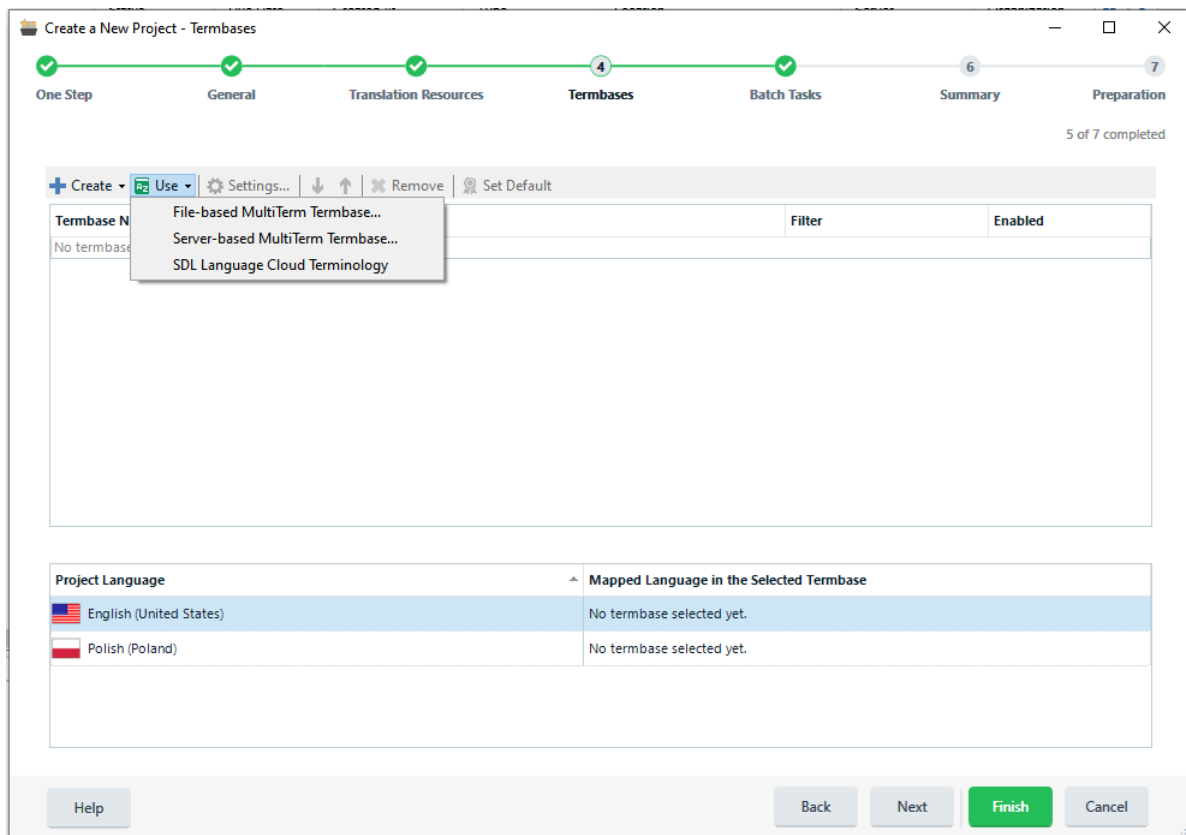
6. Specify batch tasks of your new project and click **Next**.

7. Check the summary and click **Finish**.

8. Trados will prepare your new project. Click **Close**, go to the new project, click file or files and translate.
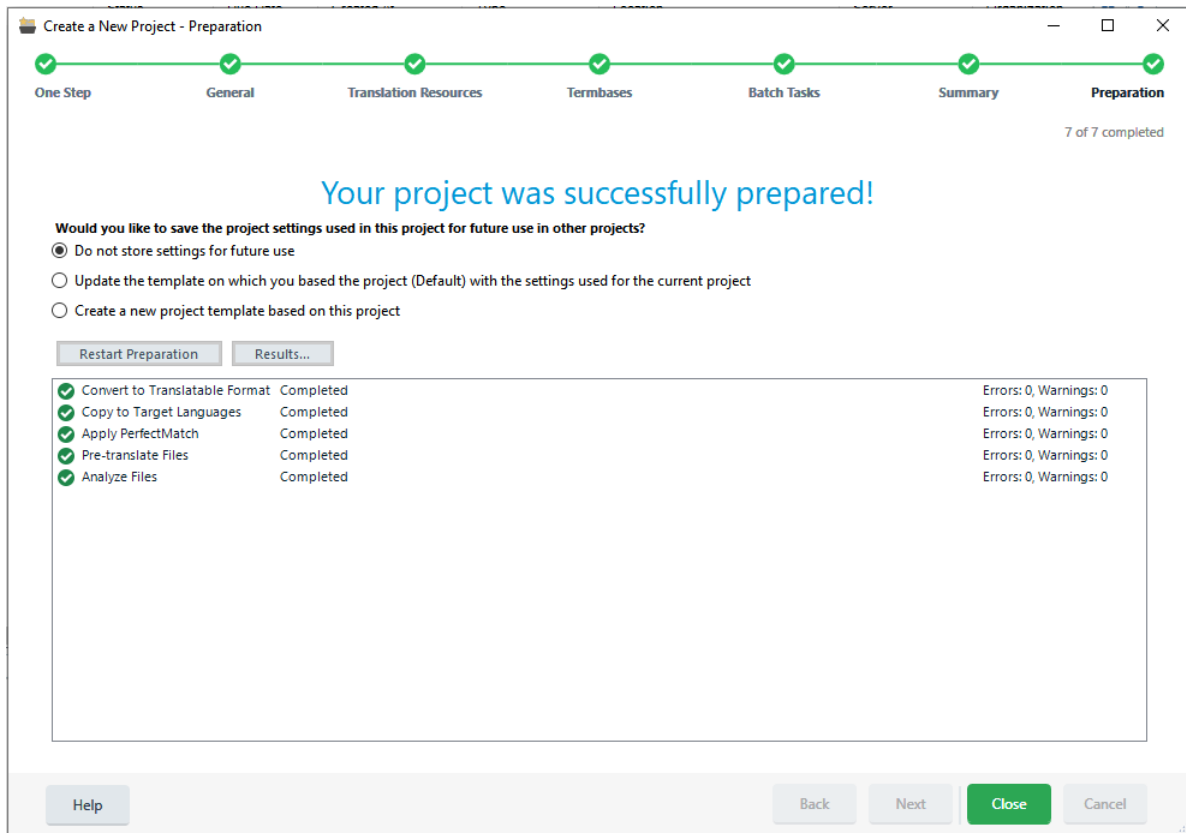
9. Translate project files, consulting source files as necessary. To start translating, click in the first segment in the target column and type in your translation. Click **Ctrl+Enter**. This will confirm the first segment, send it to the TM and move the cursor to the next segment. Translate the other segments.

10. When file is translated, go to **File** tab in upper pane and select **Save Target As**. Specify, where you want your destination file to be saved.
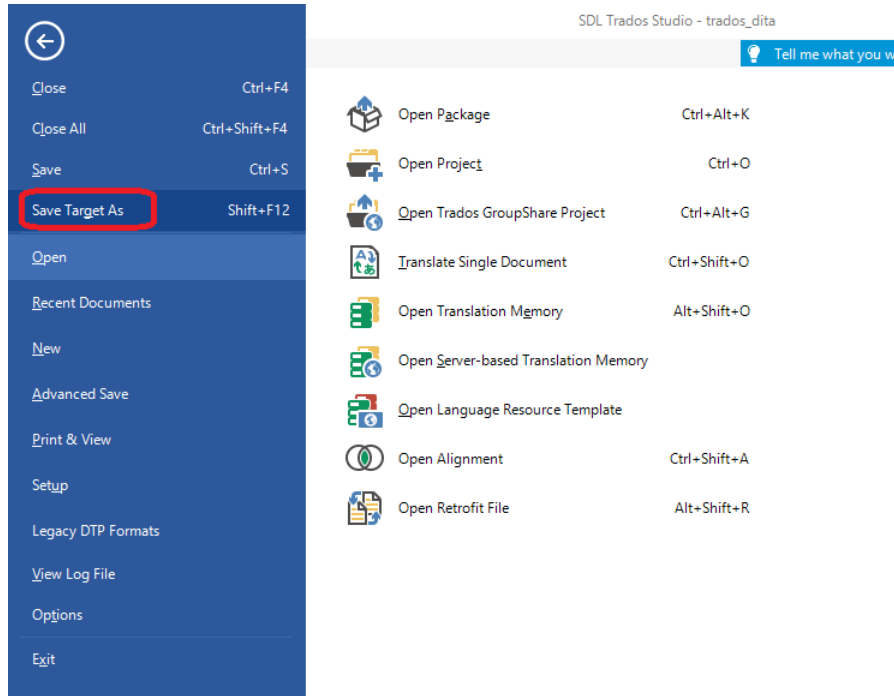
11. The exported file is saved in the source format.

Example: how to localize file using memoQ?

Example: documentation files prepared using **Markdown (.md)** and localized using **memoQ**

To localize the documentation files, create a project in memoQ

1. Open memoQ.

2. On the memoQ start screen, choose **New Project** and **New Project From Template.**
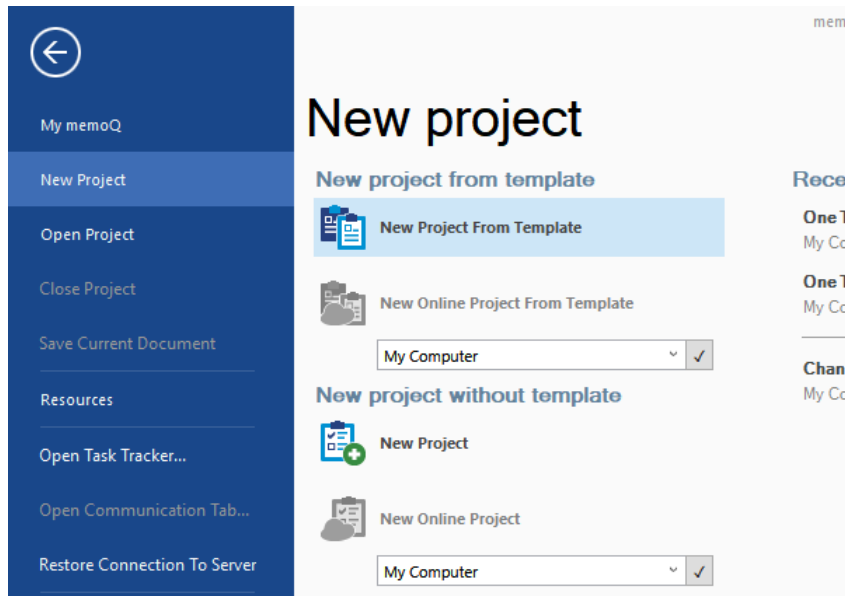
3. In displayed dialog choose **From local computer**, click **Add documents** and **Next.**
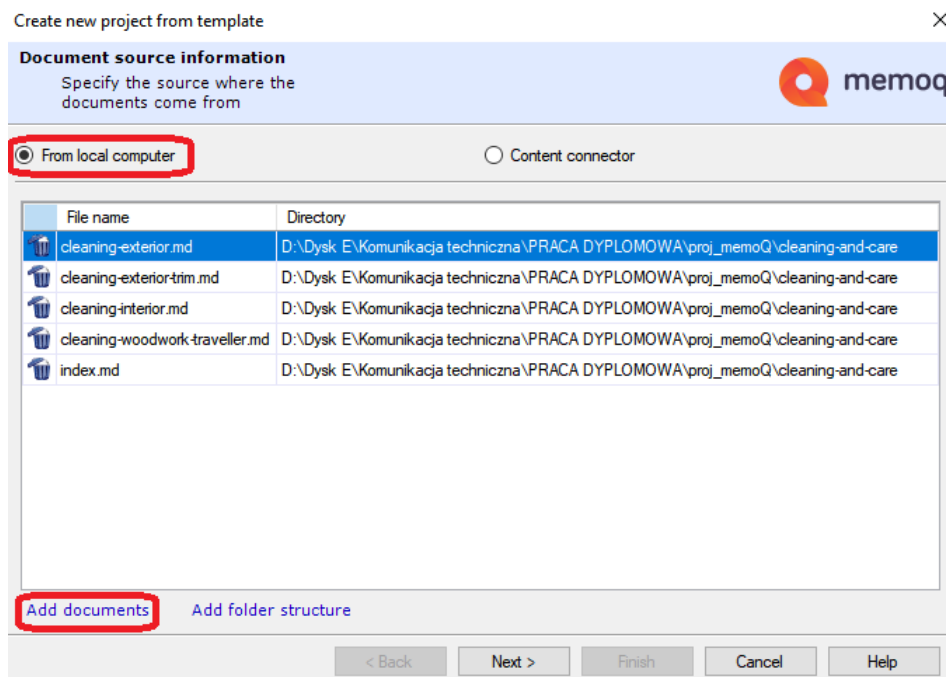
4. On the next page, choose a template, language pair, and project details. Click **Finnish**.

5. When importing source files, errors can occur. Click **Change filter & configuration** in the displaed **Document import options** window.

6. From displayed dialog choose **Plain text filter**.
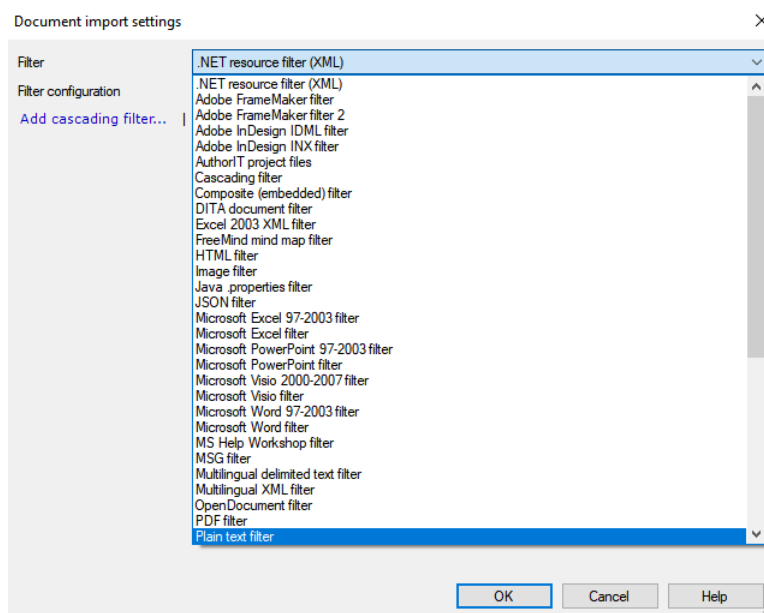
7. ✏️ **Note:** Some formats (DITA, HTM) are fully supported by CAT tools, while others (Markdown) are simply loaded as text. In the case of a more advanced Markdown, it is beneficial to work on creating a filter. More information can be found in the appropriate CAT software documentation.

In the following dialog, **Plain text filter** and **Unicode (UTF-8)** should be selected. Click **OK**.
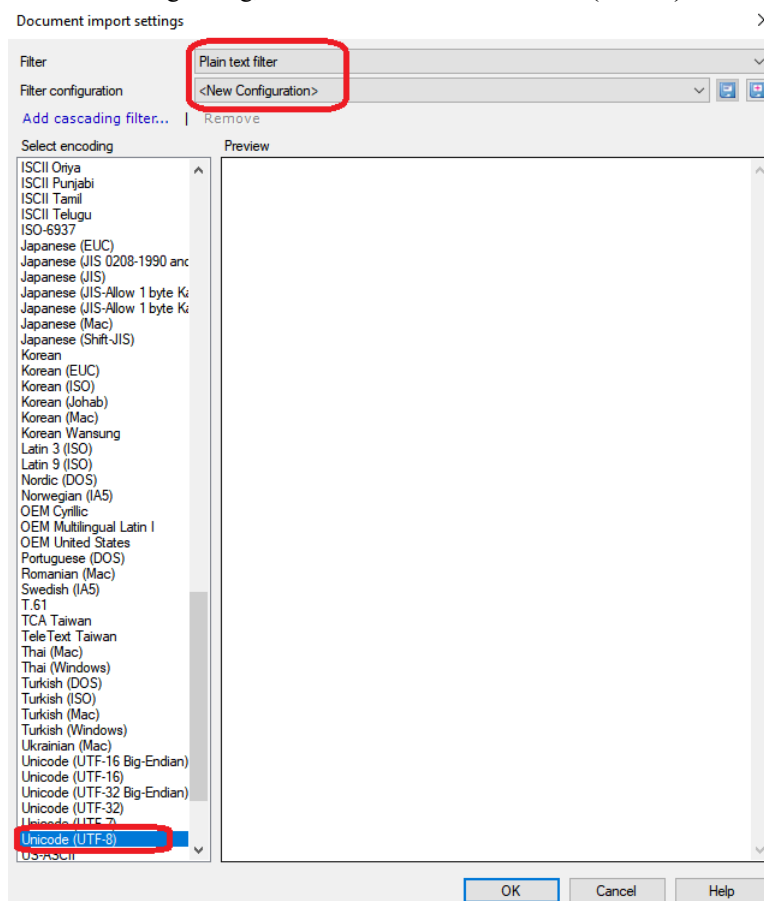
8. The **Document import options** window appears again, and click **OK**.
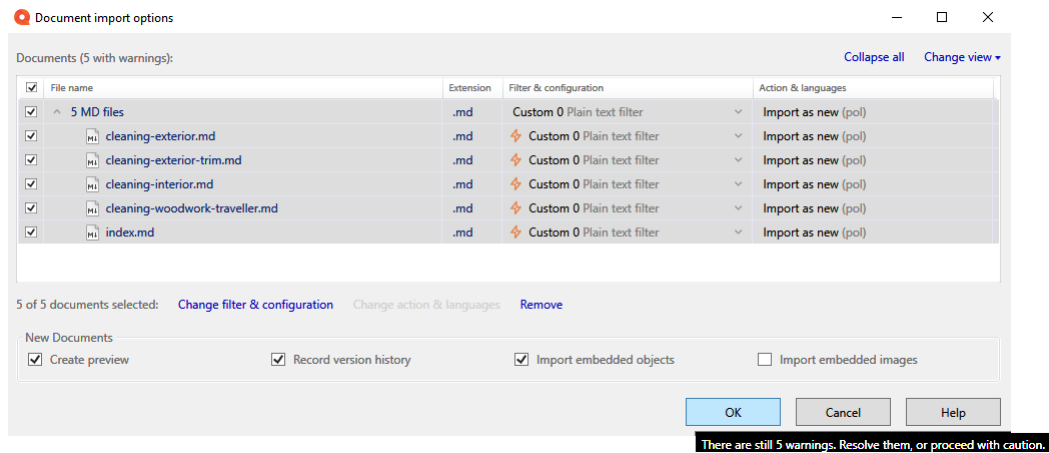
9. Then memoQ will prepare the project and display a window with the project files. Select a file, open it and translate.
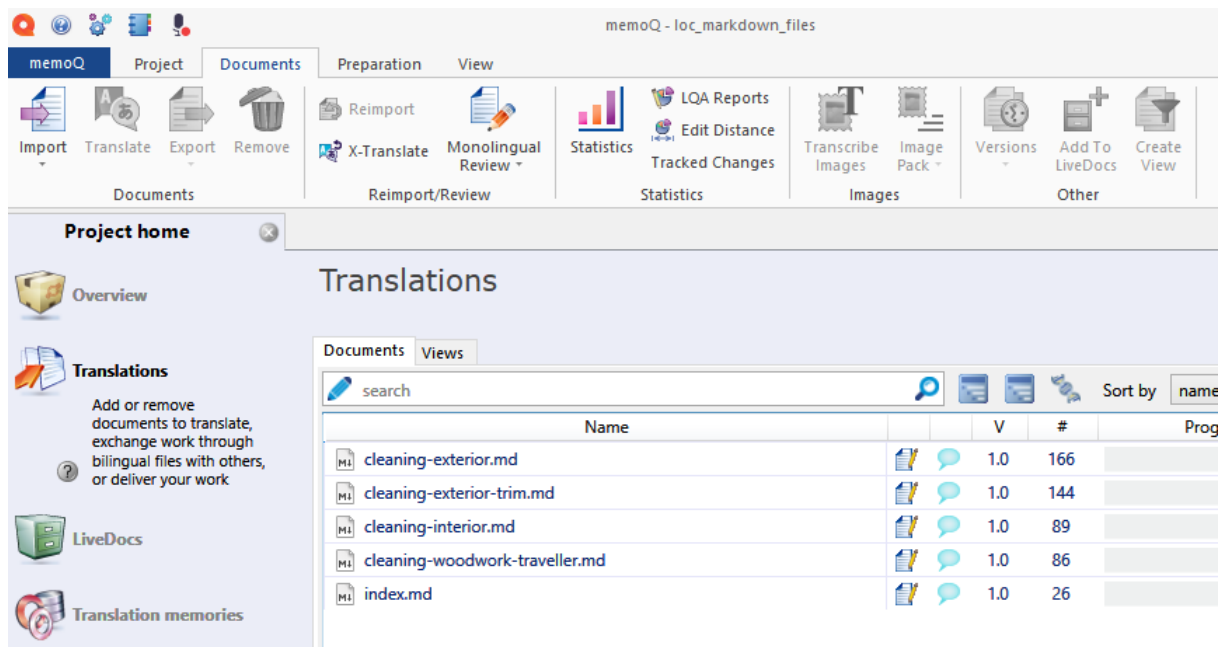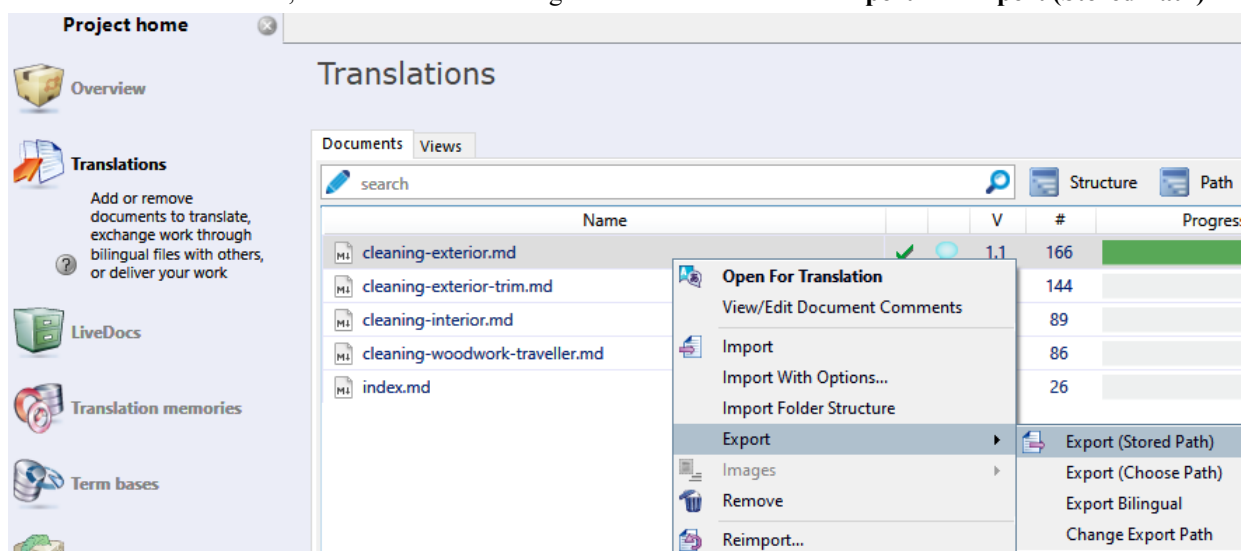
10. Specify translation memory for use.
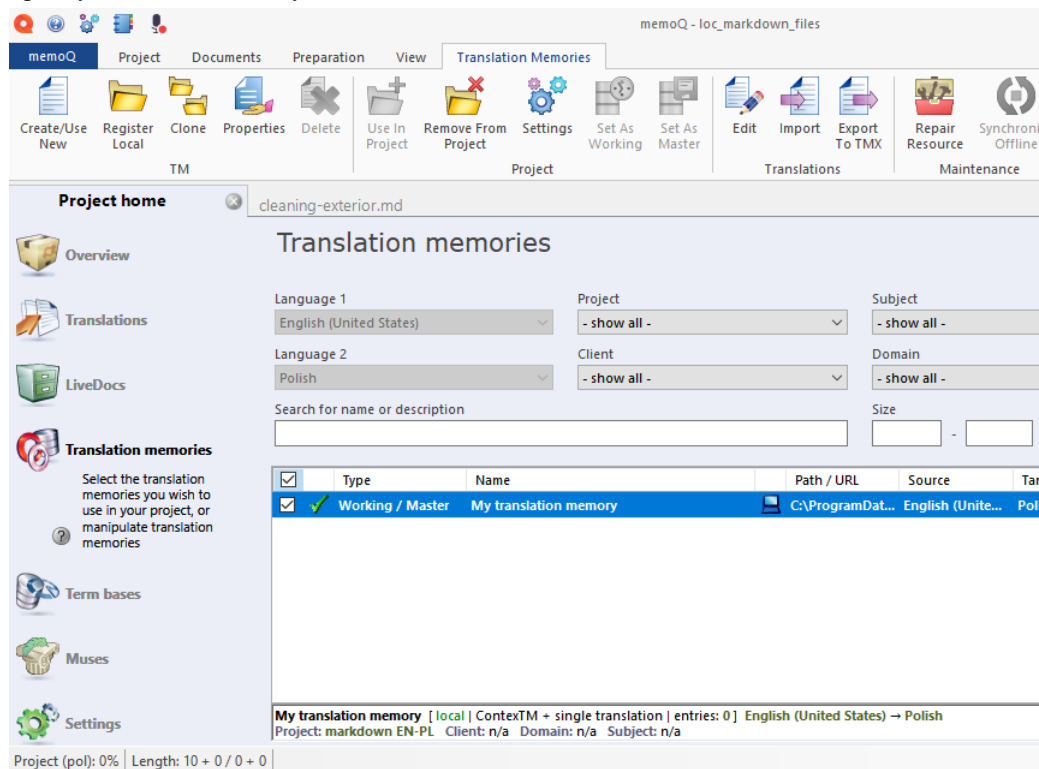
11. Translate project files, consulting source files as necessary.
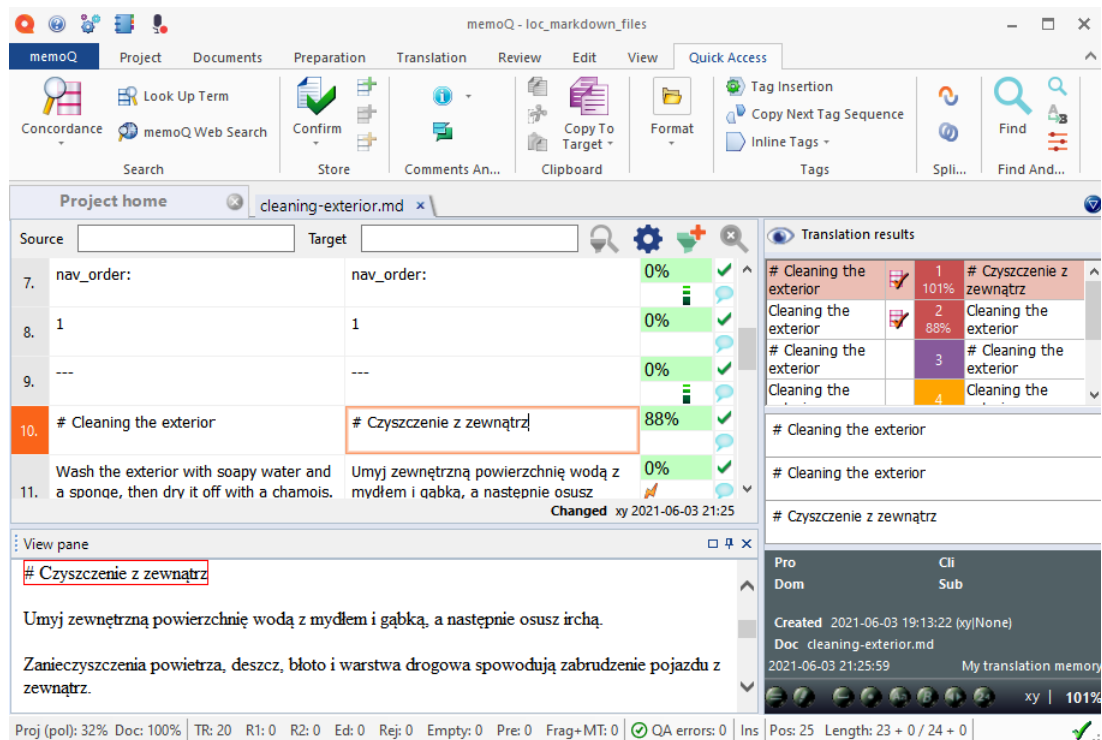
12. When the file is translated, select it and click the right mouse button. Choose **Export** and **Export (Stored Path)**.

13. The exported file is displayed in the source format.
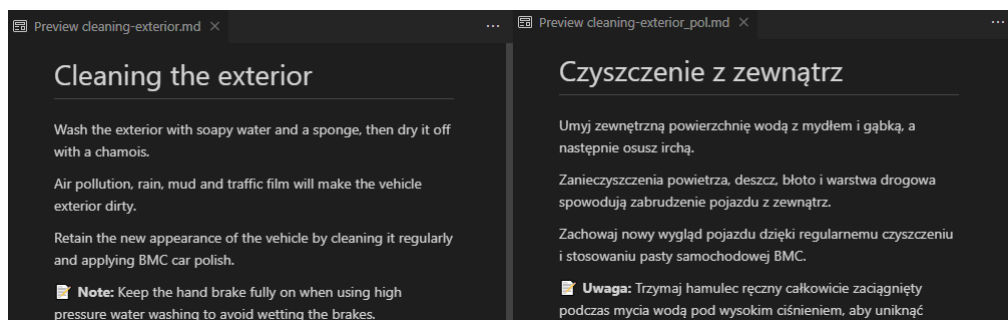
Figure 10: Source and target files .md when localized using memoQ

Example: how to localize file using Déjà Vu X3?

Example: documentation files prepared using **MadCap Flare (.htm)** and localized using **Déjà Vu X3**.

To localize the documentation files, create a project in Déjà Vu X3.

1. Open Déjà Vu X3.

2. On the Déjà Vu X3 start screen, in the section labelled *New*, click on**Translation Project** and next **Project:**
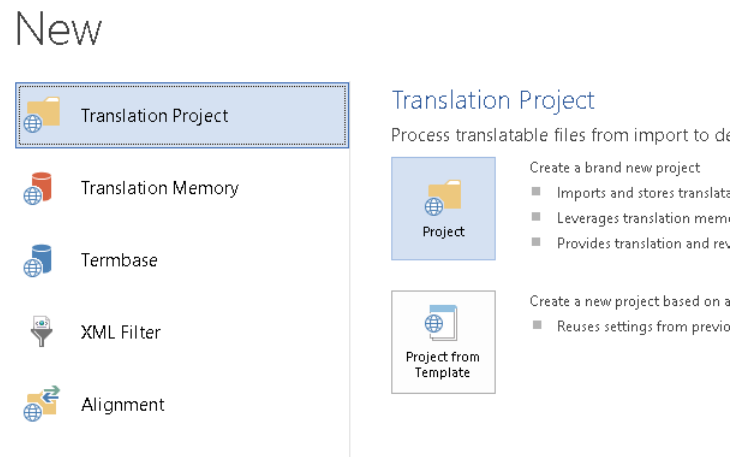
3. **New Project Wizard** appears. Specify where do you want to store your new project.

4. Click **Next**. On the next page, select the source language and the target language(s) for your project and click **Next**.

5. Choose to use an existing translation memory (or create a new one) and click **Next**.

6. Select or create the Termbase for this project and click **Next**.

Specify a Termbase

In order to store and retrieve terminology entries while working on the project, Déjà Vu X3 needs a termbase. Use the Add and Remove buttons to select the termbases you want to use with this project.

Termbases, in order of use:

Read	Write	% Penalty	Path
☑	☑	0	D:\Content_translation_project\TB\Content_PL_EN.d...

7. **Note:** Like other CAT tools, Déjà Vu X3 also allows you to connect to the Machine Translation service. More information can be found in the appropriate CAT software documentation.

The next page optionally allows you to specify what Machine Translation provider you want to use for the project. Specify the appropriate data, or not, click **Next** and go to the next page. Select a client and a subject for your project or use the default ones.And click **Next**.

8. Add source files to your project.

Specify location of source files

In order to translate a set of files with Déjà Vu X, you need to add them to the project. Use the Add and Remove buttons to select the files you want to add to the project.

D:\Content_translation_project\project\Content\A_Opisy\1_Dla kogo jest ten przewodnik.htm
D:\Content_translation_project\project\Content\A_Opisy\2_Co to jest wino.htm
D:\Content_translation_project\project\Content\A_Opisy\3_Historia wina.htm
D:\Content_translation_project\project\Content\A_Opisy\4_Rola wina w kulturze.htm
D:\Content_translation_project\project\Content\A_Opisy\5_Co to jest wino domowe.htm

9. Click **Next** and your project will be created. A displayed wizard will show the progress of the import process.


Your project has been created

10. When the process is finished, click **Close**. The main interface of Déjà Vu X3 is displayed, and then click the files and translate them.
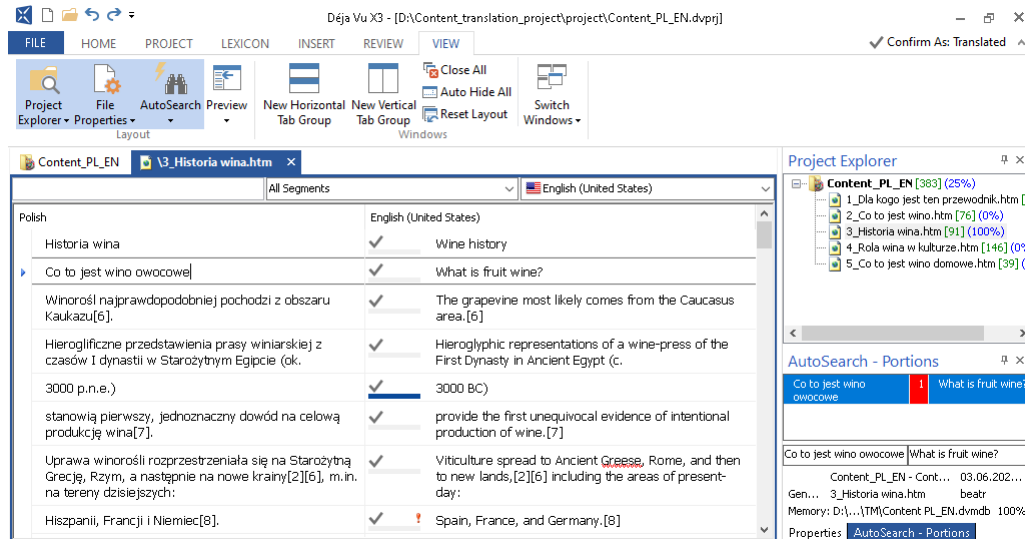
11. When the translation process is finished, the final step is to export the project, i.e., to generate the translated **.htm** files. Select destination folder and click **Export**. Then, it will validate all the translated sentences, checking for problems with tags. If Déjà Vu X3 finds a problem, it will prompt you to fix it. Fix the problem and click **OK**. Click **Yes**, if Déjà Vu X3 asks you to add spaces at the end of each segment (each sentence in the exported file will be separated by space, as they should be.)
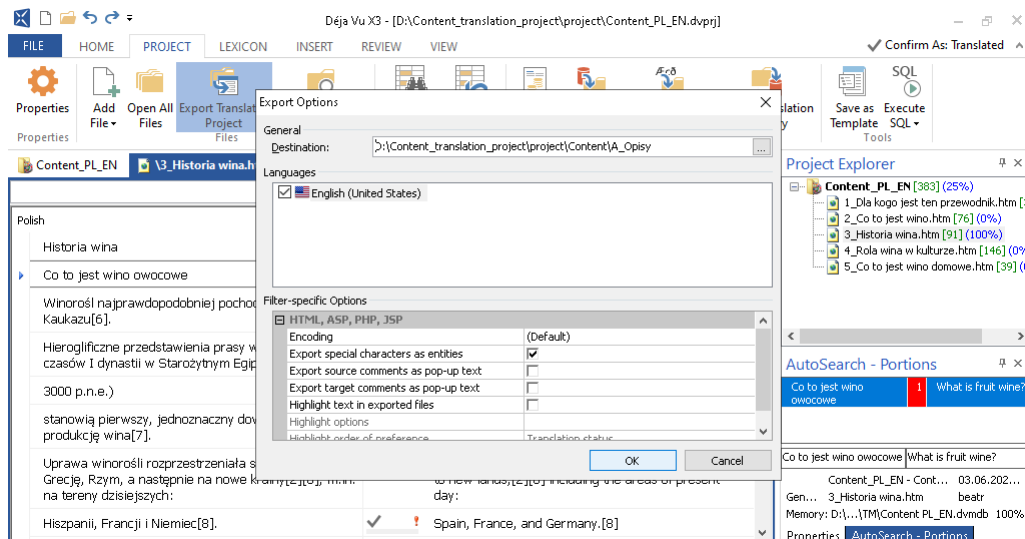
12. A subfolder will be created inside the export directory and named as the code for the specific locale of the target language. Open your source and target files and check them according to your established criteria.
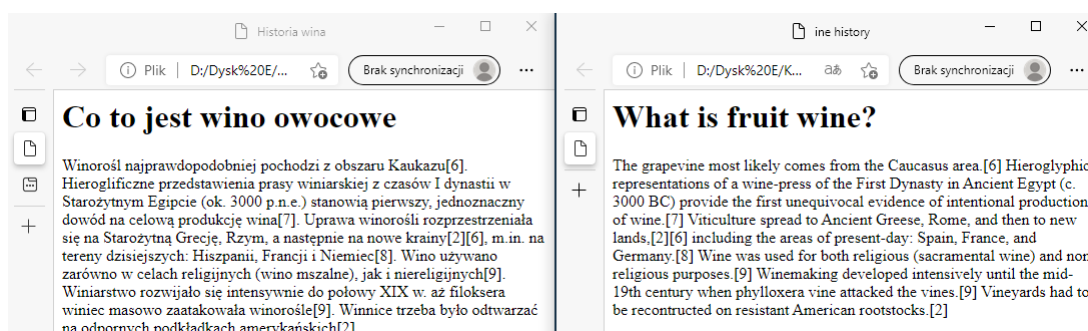

Figure 11: Example source and target files