

TRANSLATION AS A PRODUCT

TRANSLATION PROJECT MANAGEMENT IN A MODERN COMPANY

“Each translation project is coordinated by a Project Manager who is responsible for meeting the requirements for all aspects of the production process .“

ISO 17100

ABSTRACT

Managing a technical translation project is a complicated business process comparable to many industrial workflows. This article describes the main stages of the life-cycle of a translation project: from preprocessing of the material to the final proofreading of the complete translation. Its focus is not only on the technical aspects of project management, but also on the role of each stage in the whole process and on the issues that are characteristic of each stage. It is important that all participants of the process, from customers to translators, are aware of the complexity of the translation project management. This awareness also has a significant influence on understanding the issues related to technical language.

CAT Software — a *Computer Aided Translation* tool. It is a program that aids in translation by suggesting phrases from the translation memory and terminology bases and allows to analyze text.

Glossary — a list of terms in the source language and their equivalents in the target language.

OCR Software — an *Optical Character Recognition* tool. It is a program that facilitates the conversion of an image of text (a photograph, fax, or scan) into a computer-editable text.

Proofreading — examination of the revised target language content and applying corrections before printing.

Review — monolingual examination of the target language content for its suitability for the agreed purpose.

Revision — bilingual examination of target language content against source language content for its suitability for the agreed purpose.

Quality Assurance — activities that allow to ensure compliance of the translated text with the requirements of the customer. Quality assurance is a process that precedes working with text and focuses on preparing the glossary and instructions, as well as on team selection and answering any queries.

Quality Control — activities that allow to verify whether the text meets the requirements of the customer. Quality control is performed after working with text and focuses on finding and removing any technical, linguistic and editorial errors.

Text Analysis — searching for previously translated parts of text that are suitable for use in the current translation and making them available in the form of a Translation Memory (TM).

Text Conversion — conversion of text from a non-editable format (e.g. PDF) into an editable format suitable for translation (e.g. DOC).

Text Finalization — conversion of the translated text into a format expected by the end user, for example a PDF file, a publication or a functional software.

Translation Memory — a database containing phrases and their translations. This is a basic component of

1 INTRODUCTION

The traditional definition of translation is: “expressing in the target language of the text content (including oral statements) produced in the source language”. A very high level of generality of this definition and the documents wide scope of the term “translation” in Polish — it refers to written text (from poetry through various to computer software, and even to text displayed by household appliances), oral statements and any other verbal forms of communication — probably makes the impression that translation is considered to be a simple task, like cleaning or cooking which does not require any specific qualifications.

In English, the scope of the Polish term “translation” comprises at least three meanings: *translation* — focusing on translating written documents; *interpreting* — focusing on interpreting oral statements; and *localization* — focusing on localizing software, Internet content, technical documentation and other forms of communication, such as movies, video games and messages displayed by devices.

The narrowing of meaning entails the narrowing of specialization of people performing relevant services which, in turn, increases awareness of the fact that such services are specialized and require proper preparation and expertise, not only knowledge of the source language.

2 MAIN STAGES OF A TRANSLATION PROJECT

Similarly to typical products, translation has its own life-cycle which consists of three main stages:

- Preprocessing
- Translation
- Postprocessing

Each stage comprises activities, the number of which — depending on the complexity of the process — can vary. The preprocessing stage comprises the following activities:

- Conversion of text into a format suitable for translation — this activity is the most difficult when the received text is in a non-editable format (e.g. PDF), or when it contains elements that must not be edited.
- Text analysis — this activity entails searching for previously translated fragments of text that are suitable for use in the present translation and replacing the source text with them or making them available to translators in the form of a translation memory.
- Creating a glossary that includes the terminology found in the text to be translated.
- Preparing translation instructions that include information on the style, conventions, translations of specific phrases or special recommendations.
- Selecting a team of translators or a translator, and providing them with the text, a glossary, instructions, and reference materials.

The stage of working with the text is a cycle including the actual translation and revision. Each translated part of the text is forwarded to the reviewer who provides his or her comments and makes corrections, or sends the text back with corrections to be made by the translator.

Postprocessing stage mostly consists of finalizing the project which means editing the text to a form in which it has been ordered. It involves removing all traces of the actions of CAT and quality control tools, as well as the conversion of the text to the target format. Another important step is also revision — checking whether the translated text really meets customer's requirements, and, finally, cost settlement with the customer and team members who have performed the work.

3 TRANSLATION PROCESS IN DETAIL

3.1 Conversion

Materials to be translated can be prepared in many different formats. Unless the customer ordering the translation uses CAT tools, which is the case with big IT companies, the materials must be prepared properly, so that the translation with the aid of CAT software was possible. Most CAT tools directly support the most common file formats (such as MS Word, HTML, text, or DTP system files), but there are formats that require advanced conversion, such as PDF, MS Excel files or files with software source code. Interestingly, the files that are sent over most frequently belong to this second group. That is why the text conversion is one of the most important stages in the translation cycle as it determines all following stages.

In the case of PDF files the conversion stage is often a key stage in the translation process, as this format is designed for viewing and printing, and not for editing. PDF files may include scanned images or text available for direct conversion. Each case requires a different approach. Scanned text needs to be processed with OCR software (*Optical Character Recognition*), and then manually checked, as character recognition is not perfect and it is impossible to avoid errors. On the other hand, the text saved in the PDF format may be directly converted to editable format, which still requires special software.

Files with software interface elements are the next problematic format. These are usually text files, so they are technically quite simple. However, the text often contains various control and wildcard characters which need protection, so that the translator does not delete them accidentally. For example the sentence “*User %s created %d files\nand saved only %d of them.*” will be displayed in the following way:

*User John Smith created 12 files
and saved only 5 of them.*

It contains 4 special characters (%s, %d, \n, and %d) that must remain unchanged in the translation. Otherwise, the program may not work properly. In order to prevent this sentence from being accidentally damaged, it is necessary to perform the operation of tagging — to tag these characters, which means to change them into tags recognized by CAT tools.

Once this has been done, the sentence will look as shown below: “*User [tag] created [tag] files[tag]and saved only [tag] of them.*” The only CAT tool that allows for tagging with the use of regular expressions is memoQ. With other tools this operation must be performed individually, for example by creating a proper script, or using editor supporting regular expressions.

3.2 Analysis

The analysis of translation is a key stage of the localization project, and is usually performed by the project manager. It entails the content-related assessment of the text, analysis of the volume of translation, selection of a properly qualified team of translators and proofreaders, as well as collection of all available information and reference materials that may be helpful during works connected to the project.

First, the volume of translation is taken into account. In order to check it, a CAT tool used in the translation will count the number of words in the source text and determine to what extent it matches the units in available translation memories. Based on that information the time needed to complete individual stages of localization procedure: translation, proofreading, finalization, etc. are estimated. CAT tools allow to check if there are any similar matches saved in the translation memory for the untranslated text, and to use such existing entries saved in the memory.

After checking (confirming) the volume of translation, the difficulty level is estimated and the content category is determined in order to select the team of translators and proofreaders with proper qualifications and experience. If the translated project is an update of a document or software localized in the past, the materials (e.g. bilingual files of the earlier version or localized documents) are collected, allowing to keep consistency with the software/document being updated. It is especially important in the case of localization e.g. of the user interface in the software being updated.

CAT tools used to analyze the volume of the project (to generate calculations report) offer a very helpful function of calculating the number of repetitions. Localization projects, depending on their size, are completed by one or several translators. If the project includes a significant amount of repeated phrases, CAT software allows to extract them from translation and translate such segments first, and then to retrieve them from TM to the target text (already translated parts that are present in translation memory can also be retrieved — this way there is no need to repeat the task and the translation is consistent with the previous translation, and the phrases retrieved from the memory are checked during proofreading). The goal of this procedure is to maintain consistency between translators, when there are many repetitions and the translation is divided between several people. The translators will receive files which contain partial, consistent translations of repeated phrases.

3.3 Creation of a Glossary and Instructions

This stage of translation project management is a typical example of operations which cannot be omitted or performed without due attention, as the failure to perform it in a proper way may increase the amount of work which has to be done in the final stage of verification of completed translation. As an example let us analyze a project of translating an operator's manual consisting of 100 pages and containing 20 specialist terms. Because of the short deadline for translation, the manual was divided sent to 3 translators, but the person managing the project did not prepare any glossary or translation instructions. Upon delivering of the completed translation, it turned out that each translator had used different terminology. Furthermore, one of the translators used the imperative (“otwórz”, “włącz”, “wyłącz”), the other one used the impersonal form (“otworzyć”, “włączyć”, “wyłączyć”), and another one used courtesy expressions (“proszę otworzyć”, “niech Państwo włączą”, “prosimy o wyłączenie”).

Moreover, a frequently used phrase “*See picture...*” was translated by one of the translators as “*Patrz rysunek...*,” whereas another one translated it as “*Porównaj rysunek...*,” and the other one used “*Zobacz rysunek...*” Even though none of the translators made any errors and all of them followed certain rules accepted in Polish as the correct ones, the text would require many hours of work to make it internally consistent and, as a result, easy to read. It is possible to avoid such situations by devoting some time prior to the start of translation and create a glossary of basic terms and prepare instructions that include information about style, the way of translating frequently repeated phrases, which sentences should not be translated, etc.

Creating glossaries is an operation that may be automated to a large extent when using proper software. Programs performing *terminology extraction* use statistics rules and suggest the terms from the words or phrases that are used most frequently. Glossaries may be created in two languages on the basis of earlier translated materials with a similar thematic content, or in one language on the basis of new materials.

Preparing translation instructions is a fully manual operation and cannot be automated. It consists on skimming through the text, determining the style to be applied during translation (e.g. descriptive texts to be translated in impersonal style, instructions with the use of imperative, marketing materials in the publicist style, etc.), finding references, fragments that need to stay in the source language, and many other significant characteristics of the text. Ready instructions are sent to the translators together with the glossary and the text to be translated. All these files belong to the set of materials forming a well-prepared translation project.

3.4 Translation with CAT Tools

Translation of technical documentation or software is quite complicated. It is usually the longest stage of localization project and it requires not only localization team's experience and proper qualifications, but also a dedicated software supporting the actual process. This software not only supports the process of translation, but also enables to keep consistency of translated documents, manage terminology, and to assure and control quality of performed work. Translation aid software is called CAT programs (*Computer Aided Translation*). Work with CAT software should not be mistaken with so called *Machine Translation, MT*, which means using advanced statistic programs based on language corpora and providing translations without any human intervention (e.g. Google Translate). With CAT tools, it is the translator who decides on the final version of the translated text, and the computer only makes the work faster, suggesting terms and previously translated phrases.

Currently, there are many applications available in the market, both commercial and available for free, or more or less advanced. Though all of them support many file formats and offer unique solutions, translation aid software is generally based on the usage of *translation memories*, where the previously performed translations are stored.

During the translation using CAT software the text is divided into segments — sentences, phrases, or individual words, depending on the configuration of the program. Each translated segment (e.g. a sentence in the source language and its translation) is stored into translation memory, forming a translation unit. Each subsequent segment is then compared to the translation units saved in translation memory and the percentage level of how much the segment matches existing units is determined. Depending on how much the segment matches translation memory, the suggestion is retrieved, allowing the translator to easily use previous translations.

CAT software also allows to use glossaries and (in the case of certain applications) to retrieve terminology from the source text. It is designed to facilitate key terminology management and to keep terminological consistency even when dividing the text between two or more translators.

Clear advantages of CAT tools are: significantly faster translation, support of a collaborative effort of the translation team, and precise allocation of resources needed to complete translation which meets all quality requirements.

3.5 Quality Assurance and Control

According to the requirements of the quality standards, such as ISO 17100, the standard applying exclusively to the translation industry, it is mandatory to use the quality control tools. On the one hand, these regulations assure high quality of completed orders. On the other hand, they drive companies to look for solutions which facilitate quality assurance and work optimization,

and automation. As with translation in the CAT tools, human continues to be the one who makes the decisions.

The quality of translation may be measured as the degree of conformity with the customer's requirements. Quality assurance involves activities performed during the work on the project: selection of a team with proper qualifications, preparation and usage of dictionaries and guides, instructions, control procedures performance, etc. The last stage, at least in theory, is the quality control, which means the actions performed once the translation has been finished. The word “theory” is not accidental here, as in practice some of the quality control activities are performed already by the translator (e.g. quality checks done in quality assurance programs). Except for the quality checks done in dedicated tools, quality control procedures include, among others, translation revision by an in-house reviewer (employed by the company processing the order) and translation review by an external reviewer — the person from outside the company processing the order, who checks the quality of a completed project for the customer. Prior to the external proofreading, localization companies assure quality of their work using available software.

3.6 Finalization

Finalization can be explained in the most general way as a conversion reverse action. All changes made to adapt the text format to the used CAT tool need to be removed to make the text look the same as the original, only written in different language. Consequently, the more work was needed in the conversion stage, the more work it will need during finalization.

The first part of finalization is always the so-called files *clean-up*. All CAT tools create bilingual files which contain both the original text and its translation. It is very convenient during translation and for quality assurance, as it allows for the easy comparison of translation to the source text. However, finished file must contain only the target text. This task is performed in different ways by various CAT tools. It may consist in the typical “clean-up”, which means removing the original version from external files or exporting the files to the target format.

Regardless of the method used, cleaned-up file will look like the original one before it was loaded to the CAT tool, which may not be the form expected by the customer.

Next part of finalization is to modify the file, so that its format meets the customer's expectations, so e.g. PDF, XLS, etc. If during conversion stage some tags were introduced, e.g. if the sign “\n” was replaced with the tag <newline>, the tags must be removed by performing the reverse operation.

If MS Excel files were translated as XML, they must be again entered into MS Excel. Finalization process is the most difficult, if the source format was PDF, as in this case it is necessary to perform desktop publishing (*DTP*) procedures.

DTP is a complicated process. In a nutshell, it involves adjustment of the target text length to the source document structure. The text cannot overlap with the figures, it must fit its designated fields, and allow for proper page division. What is more, it is also necessary to find suitable fonts that contain national characters, to check the formatting, and often to change the figures into those containing translated content.

Certification may also be a part of finalization, in the case of translation that requires it.

3.7 Proofreading

Proofreading is the final stage of translation project processing. Its scope depends on the nature of the project, files format, etc. During verification the following elements are checked:

- The number of files to translate is the same as the number of translated files.
- The names of files after translation are the same as before translation (or they are as required by the customer, e.g. they include a language code added).
- Files are not damaged, open properly, and contain the text in the right language with correct special characters.
- The format of target files is the same as that of the source files.

The last point requires a bit more detailed explanation. It often happens that customers perform DTP of their materials themselves. Documentation part of translated into many languages, and the people performing DTP do not know these foreign languages. In a case like that a finished PDF needs to be thoroughly proofread, in order to check if during DTP there were no format changes that do not conform to the target language conventions, or no errors were introduced. Such proofreading means that the finished document in the target language must be carefully read, its layout must be compared to the source document, and the comments need to be entered. They will be forwarded to the DTP department, so that the corrections are made.

In the case of software localization proofreading is replaced by testing. The tester runs the localized software in the target version and performs operations allowing to test all functions and display all screens containing localized text. It is usually performed with the use of test scenarios prepared by the customer. During testing it is possible to find translation errors (e.g. too long phrases that do not fit the field size — the translation has to be changed), and also other errors that require changes in the program code.

4 SUMMARY

Individual stages of project management are interrelated, like communicating vessels. Omitting or performing one of the stages without due attention will automatically increase the amount of work necessary on the remaining ones. Some stages depend on the type of project, file size or format, in some situations they require more work, in other ones — much less of it. The actual translation is just one of the stages. It is obviously very important, but quite often it is not the most labor-intensive one. Project management quality affects not only the quality of the finished translation, but also how well the agency and translators cooperate.

Awareness of the complexity of the process of translation (localization) project management is especially significant both for the customer, and for all people working on the project, especially the translators. The customer must be aware that he or she cannot expect high quality if the time given for order processing is too short to correctly perform all stages of the process. On the other hand, the translator receiving a specific material needs to remember that he or she is a member of a bigger team, and his or her quality and speed of work determines the organization of other people's work.

Thus the final quality of the “end product”, e.g. localized technical documentation, depends not only on translators' language skills and experience, but also, if not mostly, on the ability to seamlessly perform the process of project management in the localization company.