Karina Martínez Ferber
euroscript Deutschland GmbH
Berlin
Karina.Martinez@euroscript.de

DECONTEXTUALIZATION IN A MULTILINGUAL ENVIRONMENT

Abstract

Content reusability and thus decontextualisation is a main requirement within CMS scenarios. Source content creation processes have to change as terminological and stylistic consistency become ever more critical. Optimization of translation processes is often seen rather as a "natural" and automatic side effect than as a pre-requisite for the effectiveness of a multilingual content system. This lecture presents the adaptation of translation processes to the requirements of the CMS as a challenge for the (in-house or external) translation service provider, which can nevertheless only be mastered by close cooperation between the translation unit and technical writers or content authors. What are the "typical" scenarios which the translation service provider encounters within a CM workflow, what aspects of translation services are affected, and how does the service provider interact with content authors, translators and proof-readers in order to ensure the success of the multilingual content management system? Against the background of terminology management as a central issue, typical initial situations, questions and procedures will be dealt with, along with the "classic" pitfalls and fallacies.

INTRODUCTION

Against the background of modularized documentation and content management, translators and translation service providers are faced with changing requirements.

Is this statement really true?

To a certain extent. The fact is that many language service providers are only gradually beginning to see the fundamental changes that are affecting their work. And these changes are even less obvious to companies buying technical translations, that is, to industrial enterprises.

Basically, we will be talking about quality here - but about a *different quality of quality*.

The goal of this presentation is to outline the new quality requirements that result from the modular work process and the typical scenarios that they entail for customers as well as for providers of translation services. At the same time, we will take a closer look at current situations at the beginning of the process and pinpoint conflicting aspects and "gaps" between the quality requirement and the realities of production.

In the concluding considerations, the focus will be on professional terminology work as the linchpin to ensure this new type of quality.

A "DIFFERENT QUALITY OF QUALITY"

An example from the automotive industry: the fact that the parts used to build your automobile are standardized and can be used in several different models and model variants does not in itself increase the quality of your vehicle. Perhaps you would even prefer a prototype specifically manufactured for you that better meets your requirements.

However, your overall advantage (so to speak, the "total benefits of ownership") is undoubtedly greater in the case of a standardized product.

The benefits include lower manufacturing costs and, thus, lower prices; improved availability, better repairability (almost) regardless of the workshop you take your car to, the availability of replacement parts and accessories, etc.

If we look at technical documentation again, we are talking about a kind of text quality that meets the basic functional standard: we are assuming that the contents and the choice of words fulfil the subject requirements and – figuratively speaking – that our "car" has the properties in demand by the market.

However, these texts also share some properties with the standardized components of an automobile. These properties arise when we deconstruct complete documents into functional units in the context of modularized documentation.

Just as in the automotive industry, the aim of the business is not to produce single, customized products, but rather to achieve high-quality mass production.

And, of course, to reduce costs.

When a particular vehicle model is phased out, some of its components can continue to be "re-used" in other models without being changed. The same applies to text components and their translation: they can potentially "survive" their original context of creation (if there ever was one) for an infinite amount of time.

We could be translating for eternity!

TYPICAL SCENARIOS AND CLIENT EXPECTATIONS

There are Content Management projects in which the company takes the aspects of terminology and translation into account well before the project starts.

Unfortunately, however, such projects are the exception.

In the majority of the cases, decisions about which system and process to use in the CM environment are made largely independent of any requirements relating to language structure and processes associated to multilingual document management.

This cost-intensive investment is – of course – made with the expectation of a high degree of process automation, resulting in a quick return on investment.

Even if linguistic aspects are considered, requirements are generally only defined for the language of authorship, that is, the source language. Sometimes style guides and glossaries are created or the technical writers are trained in the creation of structured information.

The properties of the information modules which have been created in this way are then expected to be passed on "automatically" to the translations of these texts.

Whatever is reusable in the source language should be reusable in the target language as well. However, the principles of reusability are often not clearly defined or known with regard to source modules. In these cases, individual decisions of the technical writers are the rule. Their decision whether to rewrite or to reuse is often based on the effort necessary for a retrieval compared to that necessary for a new creation. Unsuitable structuring principles, such as information modules, which are too small, lead to time-consuming searches and to redundant text production.

As the new system and workflow are expected to become operational as quickly as possible, existing documentation material is often transferred without any re-editing. Frequently, the traditional method of document creation continues in parallel with single source publishing for quite some time. So to speak, the engine is replaced "at full speed".

Consequently, extensive post-editing is often necessary in the automated process. This procedure may be acceptable for a limited period of time in view of the pressure of circumstances, but it is really a paradox situation: a complex CMS is basically used to continue production in the old way.

TYPICAL SITUATIONS FOR THE SERVICE PROVIDER

For translation service providers, various forms of data transfer are possible for the CMS output.

Sometimes, the published source document (in the target DTP format) is handed in for translation, occasionally complemented with some inserted text blocks that already exist in the target languages. The task of the service provider then is to overwrite the texts that have not yet been translated, usually in a DTP format. If a CAT system is to be used in this process, the texts to be translated must first be extracted from the complete document.

A preferable output method is provided through so-called CAT interfaces. They output the text units to be translated as XML or SGML. A preparation of the data as in the first case does not take place.

However, these interfaces usually do not take into consideration the properties of the software used and the process requirements of the translation. For instance, the sequence of the output XML fragments can deviate from that in the finished document, or the output units exceed the maximum file size that can be processed by the CAT tool editors. In such a case, the service provider usually must split larger units into smaller blocks prior to beginning the translation and then recombine them after the translation. These tasks require a high level of specialized knowledge.

Apart from the text components to be translated, translators also need the surrounding text so that they can understand the overall context. Some CM systems already offer the possibility to output the whole document as a PDF file in which the existing translations are, for example, highlighted in colour. If, additionally, the complete document is available in the source language, then translators can (at least) actively search for and research terminology and ensure that the new sections to be translated are consistent with the surrounding ones.

If translation memory systems were used for the respective customer before the CMS was introduced, a memory might be available as an additional resource. However, due to the tagging differences between the previously used DTP program and XML, the segments' value may be reduced. In other words, the same source sentence is no longer considered a "100% match". Therefore, fuzzy matches cannot be accepted without adapting the tags so that the usable value and any associated cost savings are reduced until a sufficient number of new language pairs has been created again.

In some cases, the customer uses the transfer to a CMS as an opportunity for insourcing more and more of the translation management. Customers might, for instance, want to host translation memories themselves so that they can carry out their own analyses. As a consequence, they will most likely carry out automatic pre-translations before passing the remaining segments that have not yet been translated on to the service provider. These remaining text fragments have, thus, been filtered twice at this point: first on a modular level and then again on a sentence level.

From the standpoint of terminology management, the introduction of modular technical documentation unfortunately does not usually bring about any changes.

Unlike translation memories, which are virtually created as a "by-product", the building-up and maintenance of terminology databases require additional manual processing. They are, therefore, often neither requested nor paid for and left up to the translator's discretion.

Consistency on a terminology level is supposed to be achieved by the interaction of translation memory, supplied reference material and the individual terminology maintenance of the translator.

As mentioned before, the service provider is often unprepared when he learns about the modified production environment, and he is not given the chance to play a constructive role in the planning of new processes and specifications at an early point in time.

Frequently the customer does not see the need to inform the service provider about the basics of the new information structure, the processes, the applicable rules and regulations for the source language (if any) or the objective for reusability.

THE ROLE OF TRANSLATION

Now the translators deserve our more detailed attention, since they are expected to ensure both types of quality we have previously discussed.

To summarize it again briefly, they must

- Create accurate information with correct contents (just as the technical writer for the source language)
- Maintain terminological consistency with regard to any possible context while doing so
- Ensure the reusability of the functional components in multiple contexts

But do they have the necessary information and chances to influence the process?

In the traditional setup, a translation service provider works with a number of internal and external (free-lance) translators. Although it is very important to quality-conscious service providers to maintain a regular working relationship with their free-lancers, they cannot avoid using different translators for the projects of a specific customer, due to personnel availability or fixed deadlines.

Moreover, a service provider frequently supplies more than one language combination.

So the number of translators used for a customer over an extended period of time can be quite large.

As a natural consequence, the amount of time required to transfer information, communicate knowledge and perform checks increases.

Let’s have a quick look at the typical participants within a translation workflow:

The *technical writer* is the author of the source texts. Typically, the technical writer is also the person responsible for the terminology to be used in the source language and thus the "guardian" of in-house knowledge.

Another role that is assigned at the customer's end is that of the *"translations" project manager*; in other words, the customer's interface person who sends jobs to the translation service provider(s).

The translation service provider, in turn, generally provides a project manager who receives jobs and data, clarifies specifications and distributes the tasks to the translators and other resources. This task involves the efficient channelling of data and information streams.

Generally, the translator is the one solely responsible for transporting the text contents into the target language. In most cases, they are also the terminologists for their respective target language because they have to define terms they encounter in the translation process for themselves and decide on an suitable translation.

After the translation is completed, the data records are saved in the system and the document is automatically published or formatted manually. In the context of content management systems, this task is more and more often performed by the customer himself or by another one of his service providers. Ultimately, the goal is to automate the publishing process to a large extent.

After publishing the document – sometimes even immediately following the translation – the next step in the workflow may consist in a translation quality check. Frequently, this is where the customer's foreign subsidiaries are involved in order to confirm the correctness of the translation with regard to content or terminology and to make specific modifications. Their special task within modular documentation processes will be defined later on.

Translators who are familiar with CAT systems have the know-how required for editing tagged formats such as HTML or even special formats created through filtering processes (e.g. Frame Maker or Quicksilver to Trados ttx-format).

However, unlike these tagged formats, XML offers much more comprehensive options for structuring and thus also more options for restrictions. For example, while translators can delete or duplicate internal formatting tags created out of DTP formats without any major consequences, such actions may result in a loss of validity when dealing with XML. Complex structures that define, for example, the sequence of specific elements of a section of text can hinder the natural flow of text in a foreign language.

Consider the following example of a XML structure:

> <source>Improper operation can lead </source>
> <consequence> to dangerous situations. </consequence>

Syntactical or simply stylistic rules in certain languages may require an inverse sequence of the logical elements of the phrase. However, simply switching the sequence of the structuring tags was impossible, because the rule "source precedes consequence" had been defined as a criterion for validity.

Therefore, at minimum, translators need to have advanced training and their work should be checked in the beginning. In cases like this, the project manager of the service provider must assume the function of a multiplier.

Besides transferring skills and information to the translator, there must also be a feedback loop to the technical writer. XML structures that should be avoided (as shown in the above example) due to the characteristics of some languages must be identified and solutions must be worked out together. This is the only way to ensure that the system will work for all languages that are concerned.

Apart from the technical restrictions, however, the translator also must observe content-related rules that go far beyond what is usually taught in translation curricula. Merely using CAT systems limits possibilities for translators.

In translation theory, "text" is generally defined as the "original linguistic sign" (Hartmann 1981: 9). This implies that the "correctness" of a translation is measured by the congruence of the content of the overall text with the original document. The number, structure and content of individual elements of the text, like sentences, can differ from those in the original document without adversely affecting the overall content.

As a matter of fact, my personal experience at university was that translations were considered particularly good whenever their structure deviated from the original so far that the characteristics of the target language were optimally represented while the meaning remained unchanged. And depending on the target language, these characteristics can differ considerably from those of the source language.

The presence of different grammatical and stylistic principles in different languages is a fact. But even without the presence of such particularities in the target language, a translator working on individual sentences may not always see the need to translate them one-to-one to correspond with the source language sentences.

Let's take a look at the following example:
Source text:

> The lid is held by four screws. Remove the screws beginning with screw no. 1.

A correct translation of this short text into German could read as follows:

> „Der Deckel ist angeschraubt. Lösen Sie die vier Schrauben und beginnen Sie dabei mit Schraube Nr. 1."

If you look at the individual sentences by themselves, it becomes clear that, indeed, contents can be moved between sentences without changing the overall meaning of the text. Regardless of whether such a transfer of content is necessary or purely a matter of taste – from the translator's point of view, there's basically nothing wrong with taking this liberty.

However, as the work in a CAT system is segment-based, translators are getting more and more used to decontextualized translating. In the "traditional" CAT process, however, translators generally continue to have the whole text in front of them.

They are used to modifying full or fuzzy matches and re-reading the text completely at the end so that they can compensate for possible deficiencies in consistency and other weaknesses.

Working with text modules takes away these possibilities. In the idealized case of a fully automated process, the translator will – so to speak - never see the final product.

It is amazing how the qualification of the translator for these tasks is taken for granted, while technical writers in many cases are provided, at minimum, with training and with a learning phase.

THE ROLE OF REVISERS

It is not only common to overlook the translators, but also those persons who might be doing proof-reading, e.g. in foreign subsidiaries.

And here is yet another danger in the system that we cannot afford to ignore:

- It takes a considerable amount of effort and time to build a functioning chain from technical writing to translation to publishing.
- It takes system support through terminology databases.
- It takes training and qualification of all people involved.
- It takes a constant, well-organized flow of information.

Texts that have been created within such a strictly governed system conform to many more rules and regulations than defined by the natural languages.

The reviser constitutes just an additional element of this system.

It is obvious that these participants must follow the same rules and specifications the translator used before. They must be provided with the same information, training measures and systems.

Experience shows, however, that in most cases the proof-reading of translations is only a marginal task of these individuals. Most of them do neither have an appropriate basic qualification nor do they have the necessary technical aids and time at their disposal.

Therefore, the most important question is: what goal should be accomplished by integrating a proof-reading step into the process?

Just imagine the potential costs of changing a single technical term in hundreds of Translation Memory segments and a large number of modules within a CMS. Should a decision to change a specific term – entailing substantial costs – be left to the reviser or the translator alone?

A clear decision structure is an absolute necessity in this respect, and the responsibility cannot be passed to the foreign subsidiary or to the service provider alone.

A modular system won‘t work without every participant observing a quite sophisticated set of rules. One of the crucial questions is therefore: Will every participant have the necessary information, time and technical support to deal with these requirements? Whatever the answer and the corresponding workflow decision may be, the point is to ask these questions well before implementing the system.

CONSISTENCY AT THE SOURCE AND THROUGHOUT THE PROCESS

The changed requirements regarding the quality of source and target texts and the new processes will be reflected in an expanded and newly defined range of services offered by translation companies. Few

client companies have in-house staff who are able to provide the necessary expertise and take care of all the required coordination tasks.

Translation service providers must meet the challenge of an expanded concept of quality and convey it to their in-house and free-lance employees. At the same time, the service provider has to define processes that allow him to ensure quality in rendering the new services.

Some of the crucial questions are:

- How can we ensure consistent, equivalent terminology work for a large number of languages at the same time?
- How can we ensure the maintenance and updating of translation memories and who should be responsible for that task in the process?
- What can we do to restrain the translator from "unpermitted creativity"? How do we communicate to translators and revisers the possible quality problems and costs that a simple terminology change entails?
- How should feedback loops to the customer's technical writers be set up?
- How can feedback be ensured and organized for a large number of language directions at all?
- How do we choose and train our translators?
- How do we train the employees who interface between customers, translators and engineers?

The changed requirements should be reflected in an expanded and newly defined range of services offered by translation companies. Consequently, "Service level" or framework agreements could become more important in the cooperation with the customer.

In these agreements, the processes and services should be described in detail. Among other things, such an agreement should include:

- Who delivers what at which stage in the process and how?
- What services are included in the "translation"?
    - Terminology work (to what degree and in which form)?
    - Terminology management (maintenance, provision...)?
    - Context check when files have been pretranslated, or no checking of pretranslated segments whatsoever?
    - Feedback for system optimization?
    - Observance of complex editorial rules; possibly (automated) conformity checks and validations?
    - Checking documents in the layout after an automatic layout creation?
    - Checking, evaluating and incorporating corrections from foreign subsidiaries?
    - Special technical tasks (e.g. programming interfaces, performing country-specific adaptations in the XML structure, etc.)?
    - Supervision and control of all steps and information by project management?
    - Consultative services?
- Which properties of the translation must the service provider ensure?
- Which assurances should be given with regard to the qualification of the employees involved?
- Who is, in fact, liable for a complete document, generated from modules? Is there a warranty that text modules created at different stages of the product life cycle will fit together and make up a usable documentation?
- Decision-making structures: Which employees have the right to make decisions with regard to terminology, style and structure? What do the decision-making processes look like - in other words, what has to happen for a terminology change with potentially far-reaching consequences and costs to be implemented?

CONCLUSION

The meaning of quality in connection with the service "translation" is shifting in the environment of modularized content management: quality characteristics with respect to language content remain an indispensable requirement, but are being expanded considerably by aspects relating to form and form of contents.

As a side-effect, translation quality should become a clearly defined and thus measurable feature. Translators and translation clients all over the world are going to hear this with relief.

Under present conditions, however, knowledge concerning the linguistic part of the system requirements is building up very slowly: system integration and language service are usually completely independent of each other within the overall process. But consistency, which is one of the most important requirements, can only be ensured if, for example, terminology work is performed consistently and professionally as a link in the process chain.

An entirely new spectrum of services is being created around the linguistic conformity of texts which go well beyond the traditional meaning of the term "translation".

Translators and translation service providers should embrace this challenge as a chance to enlarge their ranges of activities and to gain efficiency and performance in their work.

And finally: consulting as well as accompanying the process and providing knowledge transfer between the parties involved in the process are areas of tasks for which the language service provider, the specialist in "building bridges", is predestined.

Face the challenge!

REFERENCES

HARTMANN, P. (1981): The Text as a Linguistic Object. In Stempel, W. D. (ed.): Contributions to the Linguistics of the Text. München.