

Translation Quality Assessment Tools and Processes in Relation to CAT Tools

Viktoriya Petrova

Institute for Bulgarian Language -Bulgarian
Academy of Sciences
v.k.petrova@abv.bg

Abstract

Modern translation QA tools are the latest attempt to overcome the inevitable subjective component of human revisers. This paper analyzes the current situation in the translation industry in respect to those tools and their relationship with CAT tools. The adoption of international standards has set the basic frame that defines “quality”. Because of the clear impossibility to develop a universal QA tool, all of the existing ones have in common a wide variety of settings for the user to choose from. A brief comparison is made between most popular standalone QA tools. In order to verify their results in practice, QA outputs from two of those tools have been compared. Polls that cover a period of 12 years have been collected. Their participants explained what practices they adopted in order to guarantee quality.

1. Introduction

There is no single ideal translation for a given text, but a variety of translations are possible. All of them serve different purposes for different fields. For example, a legal translation will have very distinct requirements in terms of accuracy and adherence to locale-specific norms than that of an advertisement or a user instruction manual. CAT tools are adapted for texts such as contracts, technical texts and others that have in common a standardized and repetitive pattern. In the last 20 years the use of CAT tools increased and overturned human perceptions about the way those texts are processed and worked.

CAT assists human translators during their work by optimizing and managing the translation projects. They include a wide range of features, such as the possibility to work with different types of documents without needing to convert the text to a different format.

Another factor that overturned human perceptions about time and achievability is machine translation. Its improvements (in particular NMT in the last years) and the use of plug-ins allowed its effective use in the CAT environment.

The result was a substantial reduction of delivery times and decrease in budgets, which forced participants in the industry to change their workflows. Consequently, those changes reflected directly on the speed of translation evaluation. Previously an additional difficulty was that translation quality assessment was carried out by humans, thus the subjective component of the “human factor” was even more pronounced (Zehnalová, 2013). QA tools and quality assessment processes are the latest attempt to overcome those limitations. According to their creators, they are able to detect spelling errors, inconsistencies and all sorts of mismatches in an extremely short period. Since there are many such tools, it might be useful to distinguish them as built-in, cloud-based and standalone QA tools. In this paper, the focus will be on the last group because they represent, at least at the time of writing, the most used ones. Probably the advantage of standalone programs is that they can work with different types of files, whereas the others are limited by the format of the program. Section four shows sample output reports from two of those tools and how they behave. This paper analyzes the quality assessment tools that are being adopted by the translation industry. The first section shows the current situation in the industry and use of CAT tools in combination with their help tools (translation memories and terminology bases). The second section traces the adoption of international standards and regulations that every participant in the work chain has to follow. The third section describes the most popular standalone

“QA tools” with their common characteristics, results and reliability. The fourth section shows examples of two different QA reports with real examples of the detected issues. The last chapter presents polls on the practices adopted by the translation industry’s participants that cover a period of 12 years.

2. CAT Tools and Their Help Tools

The main help tools in a CAT tool are translation memory and term bases. The second one is crucial when a translation quality assessment is performed.

Translation memory or TM is “...a database of previous translations, usually on a sentence-by-sentence basis, looking for anything similar enough to the current sentence to be translated” (Somers, 2003). This explains why standardized texts are very much in use and make quite a good combination with CAT tools.

The help tool that deserves more attention for the purpose of this paper is the term base. A term base is a list of specialized terms (related to the fields of engineering, physics, law, medicine etc.)

Practice shows that usually they are prepared in-house and sent to the translation agency or the freelancer to use them during their work. For one thing, this practice saves time for the translator, so they do not have to research the specific term. Moreover, clients may have preferences for one specific term instead of another. Here is also the place to mention the so-called “DNTs” or Do Not Translate lists (mostly brand names that have to remain the same as in the source language). By using terminology tools, translators ensure greater consistency in the use of terminology, which not only makes documents easier to read and understand, but also prevents miscommunication.

This is of great importance at QA stage. A properly-defined term-base will allow the QA tool to identify unfollowed terminology. As this paper later describes, unfollowed terminology is the main reason for sending back a translation and asking to change it.

3. International Standards

There are a number of international standards related to the definition of quality and what may affect it: The final product quality, the skills necessary for the translator to have, the quality of the final revision procedure and the quality of the processes for selecting translators or subcontracting, as well as the management of the whole translation process.

- **EN 15038:** Defines the translation process where quality is guaranteed not only by the translation itself (it is just one phase of the entire process), but by the fact that the translation text is reviewed by a person other than the translator. On a second level, this standard specifies the professional competences of each of the participants in the translation process. This is important especially for new professions such as “Quality Assurance Specialist”. This standard was withdrawn in 2015¹.
- **ISO 17100:** Provides requirements for the processes, resources, and all other aspects that are necessary for the delivery of a quality translation service. It also provides the means by which a translation service provider can demonstrate the capability to deliver a translation service that will meet the client's requirements (as well as those of the TSP itself, and of any relevant industry codes)². Later in the paper examples are given of why it is so important to follow clients' requirements.
- **ISO 9000:** Defines the Quality Management Systems (QMS) and the necessary procedures and practices for organizations to be more efficient and improve customer satisfaction. Later becomes ISO 9001³.

¹ Bulgarian Institute for Standardization (BDS) - БДЦ EN 15038:2006 [http://www.bds-
bg.org/standard/?national_standard_id=51617](http://www.bds-bg.org/standard/?national_standard_id=51617)

² Bulgarian Institute for Standardization (BDS) БДЦ EN ISO 17100:2015

[http://www.bds-
bg.org/standard/?national_standard_id=90404](http://www.bds-
bg.org/standard/?national_standard_id=90404)

³ Bulgarian Institute for Standardization - БДЦ EN ISO 9000:2015
[http://www.bds-
bg.org/bg/standard/?natstandard_document_id=76231](http://www.bds-
bg.org/bg/standard/?natstandard_document_id=76231)

- **SAE J2450:** This standard is used for assessing the quality of automotive service translations⁴.

4. Translation Quality Measurement

Translation Quality Assessment in professional translation is a long-debated issue that is still not settled today partly due to the wide range of possible approaches. Given the elusive nature of the quality concept first it must be defined from a multifaceted and all-embracing viewpoint (Mateo, 2016). Simultaneously and from a textual perspective, the quality notion must be defined as a notion of relative (and not absolute) adequacy with respect to a framework previously agreed on by petitioner and translator. Since the target text (TT) will never be the ideal equivalence of the source text (ST) because of the nature of human languages, and the translation needs to be targeted for a specific situation, purpose and audience, translation quality evaluation needs to be targeted in the same way: For a specific situation, a specific purpose and a specific audience. This is where translation standards set some rules that are to be followed by everyone in the sector.

The question of how the quality of a translation can be measured is a very difficult one. Because of the clear impossibility to develop a universal QA tool, the “7 EAGLES steps”⁵ has been developed. It is a personalized QA that suggests 7 major steps necessary to carry out a successful evaluation of language technology systems or components:

1. Why is the evaluation being done?
2. Elaborate a task model (all relevant role agents).
3. Define top level quality characteristics.
4. Produce detailed requirements for the system under evaluation (on basis of 2. and 3.).
5. Devise the metrics to be applied to the systems for the requirements (produced under 4.).
6. Design the execution of the evaluation (test metrics to support the testing).
7. Execute the evaluation.

4.1 Translation Quality Assurance Tools

The main issue associated with the evaluation of translations is undoubtedly the subjectivity of evaluation. In order to find a solution to this, various software programs for determining

translation quality have been developed and adopted in the last decade.

Quality Assurance (QA) is one of the final steps in the translation workflow. In general, its goal is to finalize the quality of the text by performing a check on consistency and proper use of terminology. The type of errors that a Quality Assurance specialist should track are errors in language register, punctuation, mistakes in numerical values and in Internet links (Debove, 2012). This paragraph is dedicated to the QA tools and the advantages they bring. It should be noted that only standalone tools will be analyzed. It is highly probable that new-generation tools will be cloud-based (one example is the recent lixiQA), however, based on the author’s knowledge, standalone QA tools are currently the most preferred, and in particular those listed here.

The following tools are among the most widespread across the industry: QA Distiller, Xbench, Verifika, ErrorSpy and Ltb. They are the most frequently listed in blogs, companies’ websites and translation forums, and this is why they are included here.

In the early stages of their development, translation quality assurance tasks were grouped into two categories: Grammar and formatting. Grammar was related to correct spelling, punctuation and target language fluency. Formatting - detecting unnecessary double spaces, redundant full stops at the end of a sentence, glossary inconsistencies, and numerous other tasks, which do not require working knowledge of the target language. If required from the user and properly set, all detected errors are included in a report, which allows convenient correction without the use of external software. A crucial turnover for the industry was that, thanks to those tools, such issues regarding terminology and consistency were immediately detected and marked differently into the reports. Some of them offer the possibility to create checklists that can be used for a specific client, in order to minimize risk of omissions.

⁴ <https://www.sae.org/standardsdev/j2450p1.htm>

⁵ <https://www.issco.unige.ch/en/research/projects/eagles/ewg99/7steps.html>

Xbench⁶

Xbench is a multi-dictionary tool rather than a QA tool in the true sense of the word. It provides the possibility to import different file types simultaneously, which can then be used as glossaries. Another very important feature is the possibility to convert a term base into different formats. Its functionalities are related to the options of checking consistency of the translated text, numbers, omissions, tag verifier, spacing, punctuation and regular expressions. It has a plugin for SDL Trados Studio.

Verifika⁷

Verifika is another tool that locates and resolves formal errors in bilingual translation files and translation memories. As in the previous tool, this one detects formatting, consistency, terminology, grammar and spelling errors in the targeted language. In addition, Verifika features an internal editor for reviewing and amending translations. For many error types, Verifika also offers an auto-correction feature. It has a plugin for SDL Trados Studio.

ErrorSpy⁸

ErrorSpy is the first commercial quality assurance software for translations. As the other two, in this one the reviser receives a list of errors and can either edit the translation or send the error report to the translator. The evaluation is based on metrics from standard SAE J 2450.

Ltb⁹

Ltb (i.e Linguistic ToolBox) provides automated pre-processing and post-processing work documents prior to and after translation, allowing the user to easily perform QA tasks on files. Some of its features include: Batch spell check over multiple files, translation vs. revision comparison, inconsistency and under-translation checks.

QA Distiller¹⁰

QA Distiller detects common errors like double spaces, missing brackets, wrong number formats. It supports omissions, source and target language inconsistencies, language-independent formatting, language-dependent formatting, terminology, search and regular expressions.

As shown in these brief descriptions and the table below, these tools have a lot of common features. For example, all of them can detect URL mismatches, alphanumeric mismatches, unfollowed terminology, tag mismatch, the possibility to create and export a report covering inconsistencies in both the source and target. Advanced settings, such as change report or the option to use profiles, are not common to all the tools.

	Xbench	Verifika	ErrorSpy	Ltb	QA Distiller
Empty segments	X	X	X	X	X
Target text matches the source text*	X	X	X	X	X
Tag mismatch	X	X	X	X	X
Number mismatch	X	X	X	X	X
Grammar		X		X	
URL mismatch	X	X	X	X	X
Spelling	X	X	X	X	X
Alphanumeric mismatch	X	X	X	X	X
Unpaired symbols**	X	X	X	X	X
Partial translation***	X	X		X	X
Double blanks	X	X		X	X
Repeated words	X	X	X	X	
Source consistency	X	X	X	X	X

⁶ <https://www.xbench.net/>

⁷ <http://help.e-verifika.com/>

⁸ <https://www.dog-gmbh.de/en/products/errorspy/>

⁹ <http://autoupdate.lionbridge.com/LTB3/>

¹⁰ <http://www.qa-distiller.com/en>

Target consistency	X	X	X	X	X
Change report				X	
Multiple files	X	X		X	X
CamelCase	X	X	X	X	X
Terminology	X	X	X	X	X
Checklists	X	X			X
PowerSearch* ***	X	X	X		X
Profiles*****		X		X	X
Report	X	X	X	X	X
Command line*****					X
DNT List	X	X	X	X	X

* Potentially untranslated text

** I.e. unpaired parentheses, square brackets, or braces

*** Setting with minimum number of untranslated consecutive words

****Searching modes: Simple, Regular Expressions, and MS Word Wildcards.

***** “Profiles” are custom QA and language settings that are selected for a specific customer

***** It allows to automate the QA tool without processing files via the graphical user interface

Even though those tools have many similar settings, some of them are preferable to others.

As has been described in the ISO 17100 standard, client requirements are determined before the start of a translation project. The following files are usually requested at the time of delivery:

“Deliverables:

1. Cleaned files
2. _____ QA report with commented issues”

The blank space usually signifies which type of report if required. "Commented issues” relates to all false positives that are inevitably detected. A few such examples are given below.

5. QA Tools Output Comparison

As already established, those tools have many common characteristics, but also a lot of different ones. Some of them can be connected to a CAT

tool, while others cannot. They all verify terminology, inconsistency, numbers, tags, links, and create an exportable report (mostly in excel format), which can then be verified by a QA specialist, or sent to the translator, who worked on the project. This last step depend on what practices have been adopted by the participants in the project. Although those tools provide an excellent quality when used for the verification of formal characteristics of a translation, they are not perfect. False-positive errors can be a difference in spacing rules from SL to TT, difference in length from source to target, the word forms, instruction regarding numbers. Each specific QA tool is better at detecting something than the rest. For example, in English, a number and its unit measures are written without a space in-between, while for Norwegian it is mandatory to write the number separated by a space. In an Ltb report, this will be indicated as an error. Another false-positive issue is the difference in length from source to target. When the target is 20% longer, Verifika indicates it as a possible error, even though languages have distinct semantic and morphological structures. Xbench is unable to detect linguistic differences as well. In order to achieve the best possible outputs, it is mandatory to set specific settings for every project by installing the proper language and settings.

Below are listed examples from exported reports from Xbench and Ltb. Since they have a lot of common features, it will be interesting to verify how they behave with identical settings.

In addition, it is important to briefly touch upon privacy restrictions. As quality notion is previously agreed upon by petitioner and translator, so are confidentiality agreements. Texts are not to be shared or inserted into machine translation engines under any circumstances. For the needs of this paper, and only with a previously corrected text, that would not contain any sort of references about the client, it was possible to use parts of the hereby-listed examples.

Further down are a few examples of how those tools detect possible errors and visualize them. An identical text has been imported in Xbench and Ltb. Only their general settings are activated. This is due to the fact that each translation project is characterized by specific settings related to the

client's requirement and instructions. The translation is from English to Bulgarian.

EN	Ltb	Xbench
Dear Mr/Mrs [NAME],	Уважаема г-жо/г-не <x id="213" mmq78catalogvalue="<nt s value=""[NAME]" t;/>" mmq78shortcatalogvalue="nts" />,	Уважаема г-жо/г-не [ИМЕ],

Table 1: Link visualization.

Both tools have identified that between the parentheses there is a link, but have visualized it in a different way. In the Ltb report it is far more difficult to see where the issue is.

EN	Ltb	Xbench
xxx@123456group.com	xxx@123456group.com	xxx@123456group.com
<g id="383">2B.</g>		<g id="383">2B.</g>

Table 2. Segment not translated

While both tools have identified that the email addresses have not been translated, only Xbench has identified the other segment as untranslated.

EN	Ltb	Xbench
NA		Неприложимо

Table 3. Uppercase mismatch

This issue has been detected only in the Xbench and not in the Ltb.

EN	Ltb	Xbench
Please answer any incomplete (red) questions before trying to submit.	Отговорете на всички непопълнени (в червено) въпроси преди изпращане.	

Table 4: Difference in error detection.

It frequently occurs that a tool will determine something as a potential error, which another tool will not. An example is the Bulgarian word "непопълнени". The file is less than a 100 words. Xbench has detected no errors, while the Ltb has registered a possible spelling error. Even though here we have only a few examples, it is enough to see that Ltb is better at spelling, while Xbench verifies more possible errors on a segment level.

All of the above are false positives. In a real work situation, those issues will be declared "False" or marked "Ignore" before delivering them to the client. A QA specialist or an experienced translator will immediately understand which of those warnings are real and which are not. Nevertheless these tools help visualize quickly what can be wrong with a text, especially when the settings for the specific project are set correctly.

6. Polls

Over the years many researchers have attempted to determine what the current state of affairs is within the translation industry. Julia Makoushina describes in her article (2007), among other things, awareness of existing QA automation tools, the distinct approaches to quality assurance, the types of QA checks performed, the readiness to automate QA checks, and the reasons not to. According to her survey, 86.5% of QA tool users represented translation/localization service provider companies, while a few were on the service buyer side, and 2 were software developer representatives. 1/3 reported that they applied quality assurance procedures at the end of each translation. Small companies applied QA before delivery. 30% of respondents applied QA procedures to source files as well as to final ones. Over 5% of respondent companies, mostly large ones, didn't apply any QA procedures in-house and outsourced them. Other QA methods (selected by

4.62% of the respondents) included spot-check of final files and terminology check, while the most popular response in this category was "it depends on a project". The least popular check for that period was word-level consistency, which is often one of the most important checks, but on the other hand is very difficult and time consuming. The most popular QA automation tools were those built into the TM tools - Trados and SDLX. Almost 17% of large companies indicated they used their own QA automation tools. Other tools specified by respondents included Ando tools, Microsoft Word spell-checker and SDL's ToolProof and HTML QA. Also SAE J2450 standard and LISA12 QA model were mentioned which are not in fact QA automation tools, but metrics.

In 2013, QTLaunchPad¹¹ analyzes which models are being used to assess translation quality. Nearly 500 respondents indicated to use more than one TQA model. This happens because in certain cases, the models depend on the area of application. Such shortcomings lead to the use of internal or modified models in addition to the above. Internal models were by far the most dominant at 45%. The QA options included in a CAT tool, were also popular at 32%. The most widely used external standard was EN 15038 followed (30%), followed closely by ISO 9000 series models (27%). Others had no formal model (17%), and 16% employed the LISA QA. To the question which QA tools are being used, most respondents use a built-in QA tool functionality of their existing CAT tools (48%) or their own in-house quality evaluation tools (39%). Here too, in some cases, more than one tool is used. Particularly popular choices were ApSIC Xbench (30%) and Yamagata QA Distiller (12%), yet 22% state they do not use QA tools at all.

The situation has not changed much, as can be seen from a poll from few years ago from SDL Trados¹². The poll is based on the responses from the Translation Technology Insights Research 2016¹³. One of the key findings of the research is the overriding importance of translation quality (it has been pointed as 2.5X more important than speed and 6X more important than cost). At the same time, 64% of the polled have to rework their projects. Terminology is the top challenge. Those

who face rework have to deal with 'Inconsistencies in the use of terminology' - almost 48%. Another fact is that quality assessment is largely subjective. 59% of respondents are not measuring it at all or using ill-defined or purely qualitative criteria. Only 4% are relying entirely on formal, standardized metrics for quality assessment. This result is echoed in a question asking about feedback received: Twice as many receive subjective feedback as getting objective feedback. 59% either don't measure translation quality at all, or use ill-defined or purely qualitative assessment. In details, 35% have no measures or have ill-defined ones. 24% rely on qualitative feedback, 37% have adopted mixed measures and only 4% of respondents have adopted standardized assessment procedures.

According to the same poll, in order to improve translation quality, it is necessary to prioritize terminology management (as terminology inconsistencies are the top cause of rework), participants should familiarize themselves with existing international standards and adopt formal objective approach to measuring quality.

7. Conclusion

Translation quality assurance is a crucial stage of the working process. QA tools are convenient when it comes to both the economical aspect and time-consumption of the work process. Their adoption has helped to create new professions in the industry.

Although the examples that have been shown are mostly false issues, this does not mean that those tools are not able to detect real errors in a text, be it source or target. QA tools are valuable when there is necessity to verify if the right terminology has been followed, and that there are no inconsistencies in the translated text. The last one was previously not considered as important.

¹¹ QTLaunchPad is a two-year European Commission-funded collaborative research initiative dedicated to identifying quality barriers in translation and language technologies and preparing steps for overcoming them. <http://www.qt21.eu/>

¹² <https://www.sdltrados.com/download/the-pursuit-of-perfection-in-translation/99851/>

¹³ <https://www.sdl.com/software-and-services/translation-software/research/>

References

- Antonia Debove, Sabrina Furlan and Ilse Depraetere. 2012. “Five automated QA tools (QA Distiller 6.5.8, Xbench 2.8, ErrorSpy 5.0, SDL Trados 2007 QA Checker 2.0 and SDLX 2007 SP2 QA Check)”- Perspectives on translation quality
https://books.google.bg/books?hl=en&lr=&id=KAmI_xMyFfeAC&oi=fnd&pg=PA161&dq=translation+QA+tools&ots=xpMe0j_cer&sig=bS0WEjNWI6p_XI068NpS5e6-yhc&redir_esc=y#v=onepage&q&f=false
- “Benjamins Translation Library”, Computers and Translation, A translator’s guide, Chapter 4, “Terminology tools for translators” Lynne Brwker, University of Ottawa, Canada, 2003
https://books.google.bg/books?hl=en&lr=&id=-WU9AAAAQBAJ&oi=fnd&pg=PA31&dq=translation+memory&ots=7sq2g4unoS&sig=vveWdiX5SZHOFCoA6wnkjMyCzLM&redir_esc=y#v=onepage&q=translation%20memory&f=false
- Bulgarian Institute for Standardization (BDS) - БДЦ EN 15038:2006
http://www.bds-bg.org/standard/?national_standard_id=51617
- Bulgarian Institute for Standardization - БДЦ EN ISO 9000:2015
http://www.bds-bg.org/bg/standard/?natstandard_document_id=76231
- Bulgarian Institute for Standardization - БДЦ EN ISO 17100:2015
http://www.bds-bg.org/standard/?national_standard_id=90404
- Doherty, S. Gaspari, F. Groves, D., van Genabith, J., Specia, L., Burchardt, A., Lommel, A., Uszkoreit, H. 2013. “Mapping the Industry I: Findings on Translation Technologies and Quality Assessment”- Funded by the European Commission – Grant Number: 296347
http://doras.dcu.ie/19474/1/Version_Participants_Final.pdf
- ErrorSpy
<https://www.dog-gmbh.de/en/products/errorspy/>
- Harold Somers. 2003. “Benjamins Translation Library”, Computers and Translation, A translator’s guide, Chapter 3, “Translation memory systems”, UMIST, Manchester, England
https://books.google.bg/books?hl=en&lr=&id=-WU9AAAAQBAJ&oi=fnd&pg=PA31&dq=translation+memory&ots=7sq1f2vnhP&sig=C3fzF4zADUOs4jGyJoxfJZBBP8&redir_esc=y#v=onepage&q=translation%20memory&f=false
- Ilse Depraetere (Ed.). 2007. “Perspectives on Translation Quality” - A contrastive analysis of Antonia Debove, Sabrina Furlan and Ilse Depraetere
- Jitka Zehnalová. 2013. „Tradition and Trends in Translation Quality Assessment“
Palacký University, Philosophical Faculty, Department of English and American Studies, Křížkovského 10, 771 80 Olomouc, Czech Republic
https://www.researchgate.net/publication/294260655_Tradition_and_Trends_in_Translation_Quality_assessment
- Julia Makoushina. 2007. “Translation Quality Assurance Tools: Current State and Future Approaches”
[Translating and the Computer 29, November 2007]
Palex Languages and Software
Tomsk, Russia
julia@palex.ru
<http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.464.453&rep=rep1&type=pdf>
- Lynne Brwker. 2003. “Benjamins Translation Library”, Computers and Translation, A translator’s guide, Chapter 4, “Terminology tools for translators”, University of Ottawa, Canada
https://books.google.bg/books?hl=en&lr=&id=-WU9AAAAQBAJ&oi=fnd&pg=PA31&dq=translation+memory&ots=7sq2g4unoS&sig=vveWdiX5SZHOFCoA6wnkjMyCzLM&redir_esc=y#v=onepage&q=translation%20memory&f=false
- QTLaunchPad
<http://www.qt21.eu/>
- Roberto Martínez Mateo. 2016. *Aligning qualitative and quantitative approaches in professional translation quality assessment*, University of Castilla La Mancha
<https://ebuah.uah.es/dspace/handle/10017/30326>
- SAE J2450 Translation Quality Metric Task Force
<https://www.sae.org/standardsdev/j2450p1.htm>
- SDL Research Study 2016: Translation Technology Insights
<https://www.sdl.com/software-and-services/translation-software/research/>
- “SDL Translation Technology Insights. Quality”
<https://www.sdltrados.com/download/the-pursuit-of-perfection-in-translation/99851/>

Xbench
<https://www.xbench.net/>

Verifika
<http://help.e-verifika.com>