

HANDS-ON QUALITY ASSURANCE AUTOMATION

Julia Makoushina

Palex Languages and Software (<http://www.palex.ru>)

Tomsk, Russia

julia@palex.ru

Tel.: +7 3822 562725

Fax: +7 3822 562733

Julia Makoushina is the Operations Director of Palex (Tomsk, Russia). She is responsible for adjusting and automating in-house workflows and processes. Her other professional interests include machine translation, localization quality metrics, information and terminology management.

A DECADE IN QA...

When we started our own company in 2002, we were anxious to make sure that our few but valuable customers never had cause for disappointment. I personally reviewed every file to make sure that terms were consistent, all the numbers were correct, the right quotation marks were used and that other language standards were followed. We translated into about ten languages, and I did this kind of quality assurance for all of them. Some experience in several languages and the rules of logic helped me find typos and other formal errors even in Finnish, which I did not understand at all. I would politely ask translators to see if what I found could be an error. It usually was, and the translators were surprised that I picked out errors without any knowledge of their language. That made me realize that proofreading is not always needed for certain types of errors, and moreover, detecting such errors can be automatic and require little human intervention.

For several years before that, I had been doing freelance proof-reading for a client with repetitive projects who used a very small group of translators, and I found myself correcting mostly repetitive and predictable errors. Once I compiled an extensive checklist for those projects, I saved time by running every new project through the checklist and correcting the translation accordingly. I always wished I could write a checklist macro that would only show a list of segments possibly requiring correction. I never got around to it, but the idea was always in the back of my mind. The projects were in Star Transit, and back then I was amazed at its formatting check feature, which was able to adjust spaces in the beginning or at the end of a translated segment according to the source and make sure the translator used correct decimal and thousand separators in localized figures.

In time, as our client list grew longer we selected SDLX as our translation memory solution. In 2005, SDLX began offering a QA Check module that detected many formal errors automatically and made our life easier to some extent. Today we are seeing TM solutions that incorporate some QA functions, as well as standalone QA tools. The main difference between TM features and standalone tools is that the former normally work within the TM environment and are intended for the TM's internal file format, while the latter support numerous file formats, which is very useful for translation agencies that handle different formats simultaneously. While TM tools do allow you to import certain other formats, doing so often compromises the internal structure of the files and you may get unpredictable changes when you export files back from the TM tool.

Despite the fact that QA tools are still young and developing rapidly, they are already mature enough to be researched, benchmarked and classified. Since the purpose of this article is to provide more practical information than theory, I will only briefly summarize our key findings before jumping to more relevant questions: which tools to use for what situations? Readers who need more theory are invited to download my research paper (<http://www.palex.ru/en/downloads.html>).

WHAT ARE PEOPLE USING?

To get a clear picture of the market, we conducted an on-line survey among language service providers which helped us uncover their QA practices, rate the most popular QA tools and reveal the reasons for not using such tools.

We discovered that only about 5% of respondents do not perform QA tasks in-house, instead having their vendors or third-party companies do QA for them. In fact, all the respondents reported that they try to perform QA whenever possible and that QA automation tools often save time.

About 12% of the respondents do not use any QA tools. Interestingly, a quarter of those who do not use QA tools said that they did not even know such tools existed.

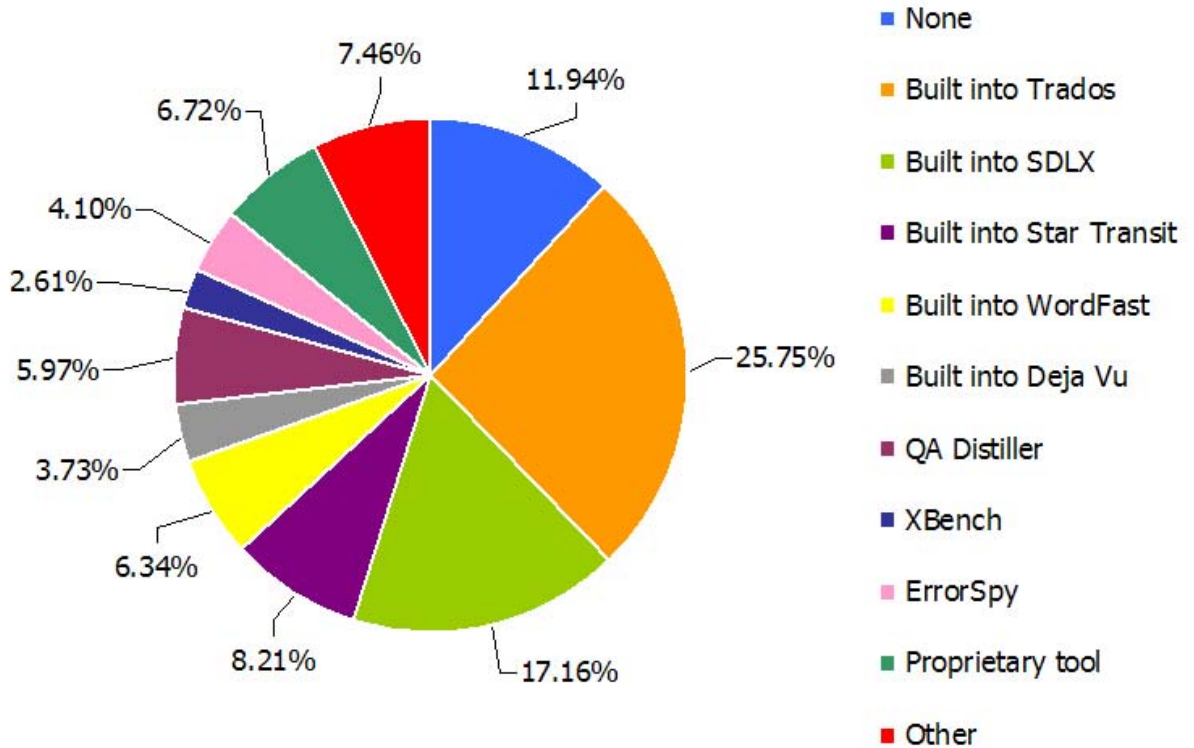


Figure 1. QA Tools by Popularity

Not surprisingly, the most popular QA tools are the ones built into the most popular TM tools. Trados and SDLX provide extensive out-of-the-box QA functionality. While Déjà Vu and Wordfast have limited default QA features, experienced users can extend them using SQL queries and macros. Built-in tools get a boost from the fact that translators who do not have to check files in many different formats definitely prefer to save money by not purchasing a dedicated tool.

Our survey and research focused on the five most popular TM tools: Trados, SDLX, Star Transit, Wordfast and Déjà Vu. These were accompanied by three standalone tools: QA Distiller, ErrorSpy and XBench. We are aware of at least one more standalone tool on the market, Ando Tools. It is a free package of MS Word and Excel add-ons that checks translation quality; however, it was only mentioned once in the survey results and the most recent version of the package is dated December 2005. For those reasons we neither provided it as a survey option nor included it into the benchmark test.

QA Distiller, with approximately 6% market share, is a leader among standalone tools, but the total market share of standalone tools does not exceed 13%. Proprietary QA tools (6.72%) are also quite popular, at least compared to standalone tools.

Another free standalone tool, XBench, is worth describing in more detail because it is not a traditional QA tool. Though it lacks Unicode support and does not provide extensive checks out of the box, XBench's additional features paired with its unbeatable price (it is freeware) make this tool really promising. The tool currently supports the widest variety of input formats and allows users to import numerous reference files for concordance search and consistency checks. It also provides term search capability in user-specified on-line glossaries and dictionaries. Additionally, the tool provides easy and effective support for the personal and client-specific checklists that are a part of every QA manager's life, and no technical knowledge is required to add new types

of detectable errors. While other tools are extensible via regular expressions, SQL queries or macros, all of which are normally out of the realm of a QA manager's experience, extending XBench is almost as easy as using global search and replace in MS Word.

More than half of QA tool users reported that they use their tools' default configuration, and more than 50% of them do this intentionally after having tried different configurations. Therefore, we decided to verify the efficiency of the default (or sometimes "semi-default") configurations.

THE CONTENDERS

Based on our analysis of QA practices and QA tool capabilities, we selected the 22 most common and serious types of errors and created a test file that contained one occurrence of each error. Then we saved the file in HTML format and translated it into 8 target languages using Trados TagEditor (TTX format). To evaluate terminology check capabilities, we also created a glossary with only one term per language. Target languages were selected based on script and included all common script types:

- Right-to-left (Arabic, Farsi, Hebrew)
- CJK (Chinese Traditional)
- Cyrillic (Russian)
- Eastern European (Polish and Czech)
- Western European (French)

The following tools were benchmarked:

- Déjà Vu X Workgroup version 7.5.302
- ErrorSpy 4.0 (build 001)
- QA Distiller™ 6.0.0 (build 188)
- SDLX 2007 QA Check (build 7014)
- Star Transit XV Professional, version 3.1 SP 21 (build 617)
- Trados QA Checker 2.0, plug-in to SDL Trados 2007
- Wordfast version 5.51t3
- XBench 2.7 (build 0.183)

Each tool was configured as follows: all checks that are not out-of-the-box functionality of the tool but can be enabled by selecting checkboxes were enabled. Any checks that require the use of regular expressions, macros or SQL queries were not enabled.

This simple test provided a lot of interesting findings and statistical information.

BEST IN SHOW

If you are considering using a QA tool, expect some relief in the most routine tasks, but not a miracle.

Almost all QA tools are good for simple tasks and the most common languages. And almost all of them are pretty weak in multilingual batch processing or right-to-left languages. Almost every tool fails to detect some errors of a type it claims to support.

Most of the tools are good at finding empty, forgotten or incomplete translations, inconsistent translations of identical segments, double spaces, matching punctuation at the end of a segment, comparing number values and confirming glossary adherence. However, a terminology check for most of the languages generates numerous false positives ("noise") because the terminology matching technology does not take into account different grammatical forms in different languages.

For all other supported checks, you should expect noise and/or undetected errors.

There is no universal formula for automating QA tasks. Tools that are top performers in some languages are helpless with others, and applications that excel at some checks produce a lot of noise for others. To help readers select an optimal tool, I have classified tools by language groups and typical formal QA tasks.

Table 1 lists the most effective tools for different QA checks according to our benchmark results.

Task	Most effective tools for the task							
	Déjà Vu X	SDLX QA Check	Star Transit	Trados QA Checker	Wordfast	ErrorSpy	QA Distiller	XBench
Find empty segments	✓	✓		✓		✓ ¹	✓	✓
Find untranslated segments		✓		✓			✓	
Find skipped (never opened) segments				✓			✓	
Find partial translations (where some source text was left)							✓	
Find incomplete translations (which are significantly shorter than the source text)						✓ ⁵	✓ ⁵	
Find segments containing corrupt (invalid) characters							✓ ²	
Find identical segments that are translated differently							✓	✓
Find differing segments that are translated the same	✓			✓		✓ ¹	✓	✓
Compare punctuation at the end of a segment		✓ ⁵		✓ ⁵		✓ ⁵	✓ ⁵	
Check the absence (or presence, if they are required) of spaces before punctuation marks						✓ ⁵	✓ ⁵	
Find double spaces		✓		✓	✓	✓ ⁵	✓ ⁵	
Find double full stops		✓ ⁵		✓ ⁵		✓ ¹	✓	
Find double punctuation marks (“.,”)							✓	
Check quotation marks						✓ ⁵	✓	
Check matching parentheses				✓		✓ ¹	✓	
Check number values	✓		✓	✓ ³		✓ ¹	✓ ²	✓
Check number formatting			✓			✓ ⁵	✓ ⁵	
Check digits to text conversion ⁴							✓	
Check adherence to project glossaries		✓ ⁵	✓ ⁵			✓ ⁵	✓ ⁵	✓ ⁵
Check untranslatables (terms that should not be translated)						✓ ⁵	✓ ⁵	
Check identity of tags	✓		✓	✓		✓ ¹		

Table 1. Effective tools for specific checks

¹ ErrorSpy 4.0 build 001 crashed at each attempt to check Russian. Version 3.0 did not crash, but was significantly less efficient.

² As long as the language is preset; for others (such as Farsi) it generates a huge number of false positives.

³ Except for Hebrew.

⁴ First you have to create something like a mini-glossary; it is not preset for all languages.

⁵ There are some false positives or non-detected errors.

Most effective tools for each language group are sorted in Table 2 (in order of efficiency).

Western European	Eastern European	Cyrillic	CJK	Right to left
QA Distiller	QA Distiller	QA Distiller	QA Distiller	ErrorSpy
ErrorSpy	ErrorSpy	Trados QA Checker	Trados QA Checker	Trados QA Checker
Trados QA Checker	Trados QA Checker	XBench	ErrorSpy	QA Distiller
		SDLX QA Check		

Table 2. Most effective tools by language group

Table 3 shows the number of false positives (FP) and non-detected errors (ND) each tool generated on the test run. It only shows total values for all eight languages that were tested. Space constraints for this article make it impossible to provide detailed information for each particular language, but anyone who is interested is welcome to contact me privately.

Translation quality assurance tools nowadays are a great help in detecting formal errors. Although they often generate too much noise and can never be used as the only means of assuring translation quality, they are a practical way to cut down the number of mistakes made through lack of attention. Most translators do not have to invest additional money in dedicated QA tools as all popular CAT tools feature at least several basic checks which are typically enough for everyday needs. More complex and sophisticated checks can be expected in the near future, which, paired with reduced noise level, will lead to ensuring higher quality with less efforts and resources.

Though this article discusses which tools best fit various QA tasks and circumstances, the results of our study are by no means definitive and final. The situation is changing quickly, and you are welcome to share your opinions and experiences. I would also like to hear your comments, ideas and news on QA tools and their use: julia@palex.ru.

Check type	Déjà Vu X		SDLX QA Check		Star Transit		Trados QA Checker		Wordfast		ErrorSpy		QA Distiller		XBench	
	FP	ND	FP	ND	FP	ND	FP	ND	FP	ND	FP	ND	FP	ND	FP	ND
Empty Translations	0	0	0	0			0	0			0	1	0	0	0	0
Forgotten Translations			0	0			0	0			7	1	0	0		
Skipped Translations							0	0			0	8	0	0		
Partial Translations			34	0							1	7	0	0		
Incomplete Translations			0	8			19	0			1	1	1	1		
Corrupt Characters			0	8			4	5					479 ¹	0		
Inconsistent Sentence Count			0	8												
Source Inconsistency											0	8	0	0	0	0
Target Inconsistency	0	0	28	0			0	0			0	1	0	0	0	0
Punctuation at the End of Segments			21	0			18	0			20	1	3	0		
Spaces Before Punctuation			0	7							0	2	0	5		
Double Spacing			0	0			0	0	0	0	3	1	2	0		
Double Dots			0	1			0	1			0	1	0	0		
Double Punctuation			0	7							0	8	0	0		
Quotation Marks											6	1	0	0		
Brackets and Parentheses							0	0			0	1	0	0		
Number Values	0	0			0	0	19	0	0	0	0	1	1	0	0	0
Number Formatting					0	0					1	5	0	2		
Measurement Unit Conversion													0	8		
Digit to Text Conversion											0	8	0	0		
Project Glossaries Adherence	42	0	16	0	23	0			159 ²	0	11	2	49	0	97	1
Identical Untranslatables											0	8	0	8		
Identical Tags	0	0			0	0	0	0			0	1	0	8		

Table 3. Benchmarked checks support by each tool.

¹ QA Distiller did not support Farsi and therefore counted all Farsi characters as corrupt

² We could not make it check terminology against the correct glossary. That may be related to this particular installation of Wordfast