



Mapping the Industry I: Findings on Translation Technologies and Quality Assessment



Funded by the European Commission – Grant Number: 296347

Copyright: QTLaunchPad and Individual Authors, 2013: Doherty, S. Gaspari, F. Groves, D., van Genabith, J., Specia, L., Burchardt, A., Lommel, A., Uszkoreit, H.



Table of Contents

1. About QTLaunchPad	2
2. Translation Quality Assessment: Models, Tools, and Processes	3
3. TQA and Machine Translation	7
4. Summary	8

1. About QTLaunchPad

QTLaunchPad is a two-year European Commission-funded collaborative research initiative started in July 2012 and dedicated to identifying quality barriers in translation and language technologies and preparing steps for overcoming them. The project's partners (the German Research Centre for Artificial Intelligence, Dublin City University, the University of Sheffield and Athena Research and Innovation Center) are working closely with buyers, vendors, and users from the translation and localisation industries and communities. With the support of GALA we have been consolidating a comprehensive mapping of stakeholders in their vision and usage of translation technologies with foci on translation quality assessment (TQA) and machine translation (MT).

QTLaunchPad focuses on identifying and overcoming the barriers to high-quality translation in a range of applied scenarios. Of particular relevance to this focus are evaluation metrics, quality estimation, and optimization of language resources. In this way, we're preparing the groundwork for a large-scale community-based translation quality initiative. To learn more about these surveys and other QTLaunchPad activities:

- Visit the project homepage for more detailed information, accessible reportage, and listings of upcoming events: www.qt.21.eu/launchpad
- Follow the discussions and highlights on our LinkedIn group: <http://www.linkedin.com/groups?gid=4807518> or connect with us on Twitter @qtlaunchpad and Facebook www.facebook.com/qtlaunchpad

This report's findings are derived from an international survey carried out in May 2013 by the Globalization and Localization Association – GALA. Almost 500 translation and localization buyers and vendors voiced their opinions on translation quality methods and technologies via an online survey.

2. Translation Quality Assessment: Models, Tools, and Processes

Questions of translation quality assessment (TQA) are nothing new, yet with the increasing prevalence of translation and language technologies, especially machine translation, quality is paramount. In order to better understand the current TQA landscape, we focus on the models being used, and related tools and processes.

Which models are currently being used to assess translation quality?

Firstly, of note is that most of the nearly 500 respondents indicated the use of more than one TQA model. In these cases, the models may depend on the area of application, where certain standards already exist, yet other areas remain without an appropriate model. 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%. Tool-specific models, e.g. those 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, while other models accounted for less than 5%. – see Figure 1.

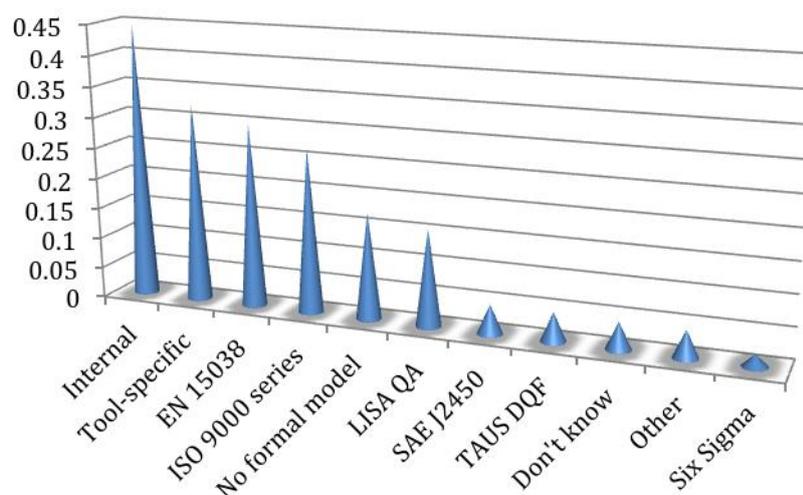


Figure 1: Quality Assessment Models Used %

What are quality scores being used for?

While some approaches consist of pass/fail results, many models generate scores and threshold values for QA. Respondents indicate several areas of usage for these scores, predominantly, in the selection or evaluation of translators (54%), in assessing the acceptance or rejection of translated content (43%) and in translator training (28%) and risk management (24%) – see Figure 2.

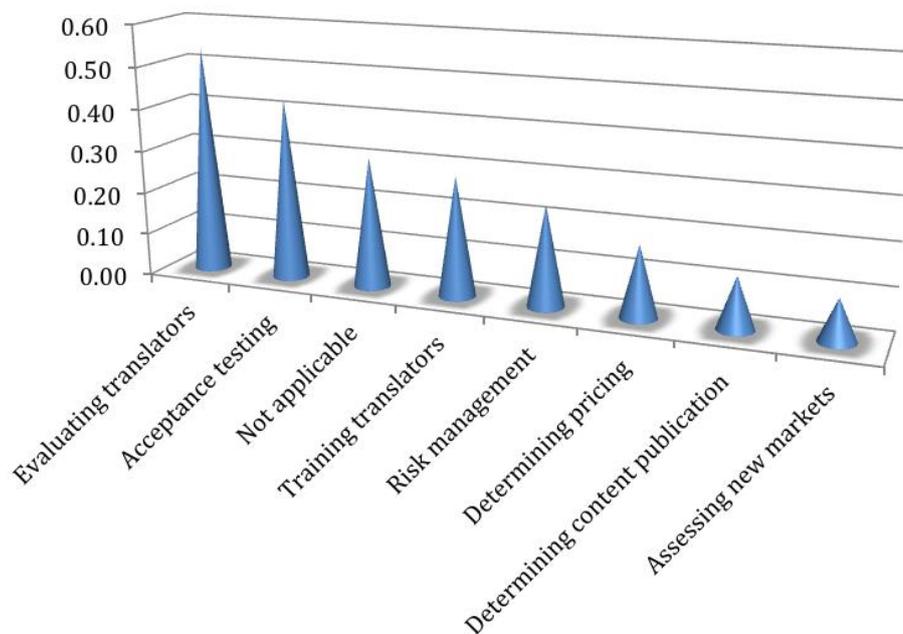


Figure 2: Usage of QA Scores %

Which QA tools are being used?

In terms of the implementation of QA models, most respondents use the built-in QA tool functionality of their existing CAT tools (48%) or their own in-house quality evaluation tools (39%). Once again, 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 – see Figure 3.

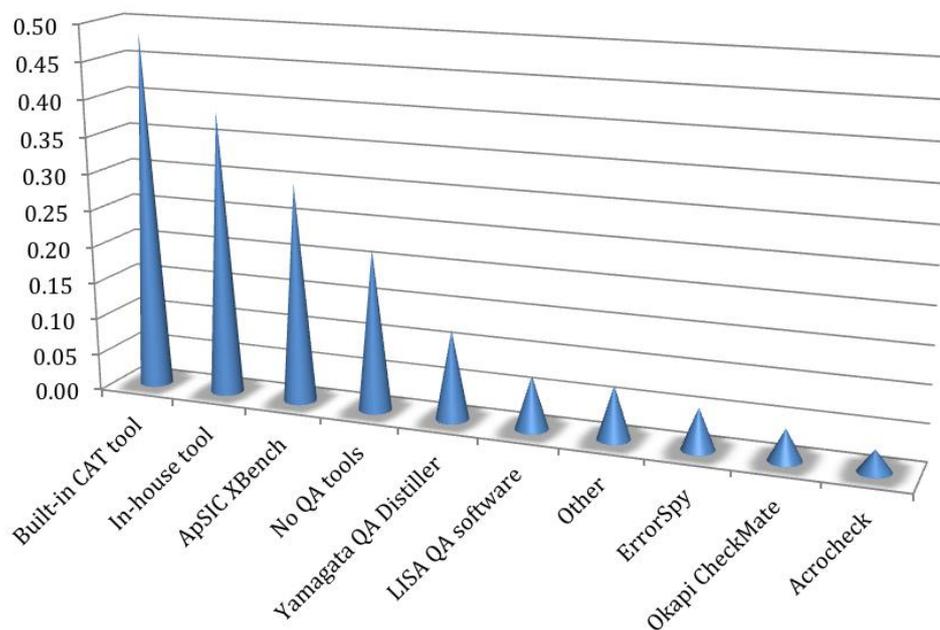


Figure 3: Usage of QA Tools %

How often is QA carried out?

For the vast majority of respondents, QA is a regular feature in their translation workflows: 38% always carry out QA, in a systematic way, and 37% perform regular checks according to pre-defined criteria and/or schedule. Random checks, when time permits, represent 15%, where no fixed criteria or schedules are used. Lastly, 15% perform irregular, one-off, checks when circumstances change, e.g. for new translators, new clients, or new domains – see Figure 4.

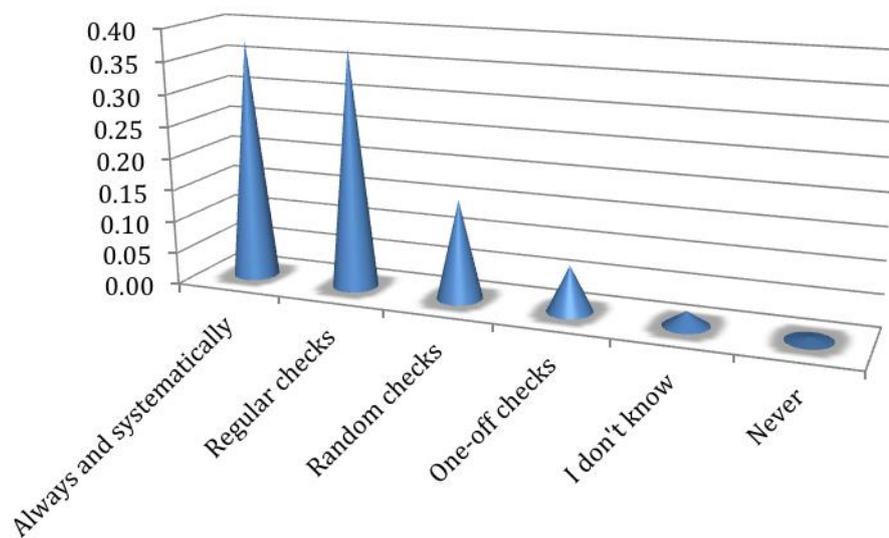


Figure 4: Frequency of QA Checking %

Is there a need to change current QA processes?

Most respondents feel a need to make incremental, yet meaningful, improvements to their QA processes (70%), such as the use of better software, and standardisation of metrics. An additional 10% think that changes are urgent. On the other hand, 15% feel satisfied with their QA processes and believe that their needs are currently being met – see Figure 5.

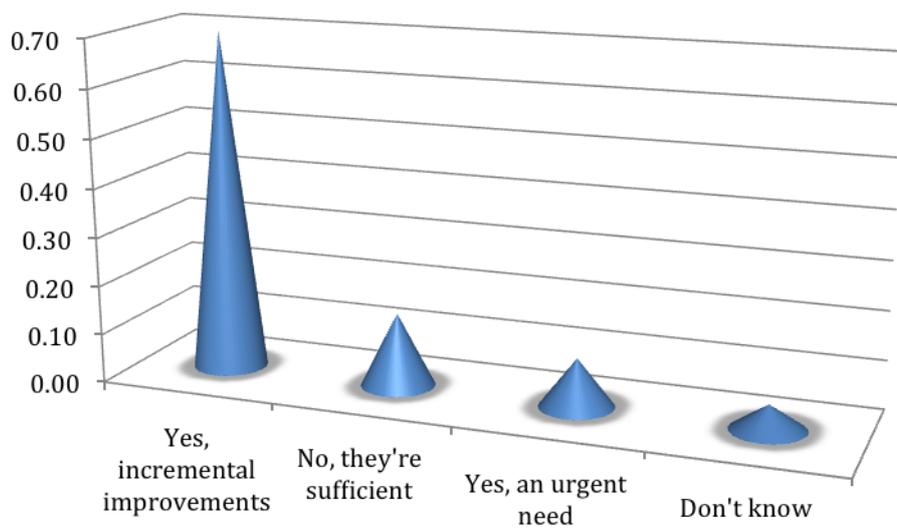


Figure 5: *Is there a need to change current QA processes? %*

3. TQA and Machine Translation

How widespread is machine translation use in industry applications?

Over one third of the respondents reported that they are currently using machine translation (MT), while a slightly higher percentage stated that their businesses are currently not using MT, but are planning to do so, either within one year (13%) or in the longer term (22%). Conversely, 28% of the respondents said they do not use MT and have no plans to start doing so – see Figure 6.

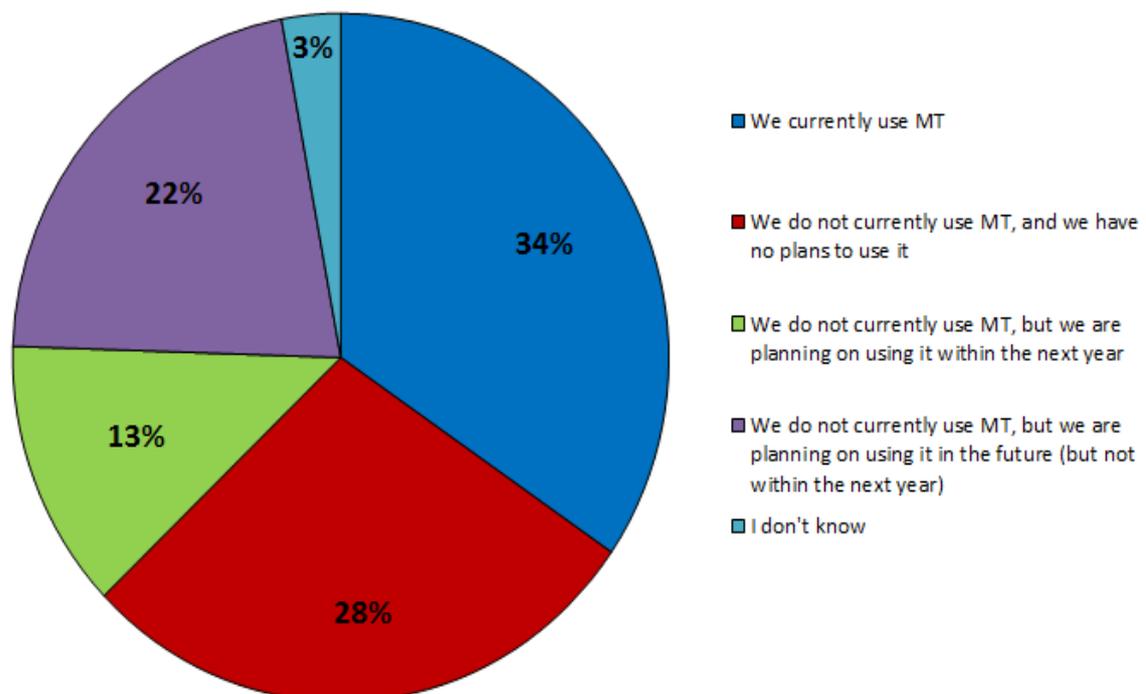


Figure 6: Current MT Usage

What are the leading approaches to MT?

The survey also investigated the type of MT systems used by the above participants. Just over half of the MT adopters use statistical systems (SMT); hybrid MT is also a popular choice (36%), followed by rule-based systems (22%). Of interest too is the finding that a third of all MT adopters use free external online MT services such as Google Translate, BabelFish or Bing in their work.

In terms of using off-the-shelf systems, the majority, 84%, stated that they had implemented some sort of customization of their MT systems. Popular modifications lie in areas of terminology (61%), in the use of additional domain-specific corpora (32%), and by providing tailor-made linguistic rules (21%). Interestingly, 16% indicated the adoption of controlled authoring or linguistic pre-processing to enhance MT output quality, while another 16% had not carried out any customization of their MT system in any way.

Is MT able to meet the demands of high-quality translated content?

The quality of MT is always of great concern. For most, the successful implementation of MT represents a huge leap and long-term investment in resources and expertise before return on investment becomes apparent. Larger buyers and providers with more resources (and typically more content) still struggle with quality issues and delivering quantifiable results to their customers in a meaningful way. As technologies improve, the case for MT only being good for *gisting* is being challenged, but what about real-world usable and high-quality outbound translation?

Concerning how much MT output is intended for dissemination purposes, 69% of users felt that less than half of their outbound translation requirements were satisfied with MT, where 12% used MT for more than half of their disseminated translated content, i.e. high-quality published content. Interestingly, 4% of respondents stated that they use MT for all their translation content. Additionally, the majority of those already using MT (77%) witness increasing requests for high-quality MT as part of their work, as opposed to *gist* or *information only* quality MT.

A related question elicited opinions on the quality of the MT output provided by the systems used by the respondents. A predominantly positive assessment was given where most responses rated MT output as *fair* (43%), *good* (41%), and even *excellent* (2%). Only 7% of the answers indicated that MT output was *poor*.

How is MT quality being assessed?

The evaluation of translation quality represents an area where the absence of reliable and meaningful standardisation and evaluation methods for buyers, vendors, MT adopters, etc. is particularly serious. A crucial part of the survey focussed on issues concerning translation quality from the point of view of the user, starting with their preferred assessment methodologies.

In general, human translation is quality checked always (38%) or regularly (37%) using internal (44%) or tool-specific (32%) assessment models, eventually in combination with standards such as EN 15038 (30%), ISO 9000 (27%) or LISA QA (15%).

The most popular choice in evaluating MT quality is human evaluation (69%), e.g. having internal staff or external experts assess the fluency and adequacy of MT output. 22% of the respondents indicated the use of state-of-the-art automatic evaluation metrics (such as BLEU, METEOR and TER), and 13% reported the adoption of in-house or internally developed automatic evaluation methods. Finally, 35% of the respondents opt for a combination of human and automatic evaluation methods, while 7% did not actually perform any formal quality assessment on their MT output.

4. Summary

Moving beyond the current practice and state-of-the-art

Overall, we have seen how internal and tool-specific models are the most popular, but many of the respondents to this survey are not using a formal model at all. While scores from TQAs are useful for a wide range of applications outside of quality assessment, e.g. training, screening, publication and dissemination level decisions, there remain many shortcomings of existing approaches. This is reflected in the popularity of built-in CAT tools and in-house software for TQA. Notwithstanding, quality assessment is typically being carried out in each and every use case and in systematic ways. Even when resources are tighter, regular checks are still seen as essential as expectations for high-quality translated content still need to be met. All things considered, there is indeed a strong voice for the need to make meaningful incremental improvements to current TQA models, tools, and processes.

For MT, a clear trend is emerging in its uptake becoming more commonplace. Interesting steps are being taken to customize state-of-the-art MT systems to improve their output, yet many users may still face a knowledge and resource gap in the face of this advancement. TQA is an area where there is a need to develop reliable and meaningful standardised methods of evaluation, and MT is no exception. Meaningful results in terms of high-quality MT output can then be substantiated with evidence of productivity gains (e.g. in post-editing scenarios), improvements in resource management (e.g. terminology management), and savings in resources in the face of a growing diversity and volume of content.

Overall, the urgent need for a flexible and open quality metric that can be used for all types of project as well as the urgent need for high quality machine translation that the project assumed have been confirmed by this survey. In fact, the project already launched the second version of its Multidimensional Quality Metrics and integrated Quality Estimation technologies needed for the improvement of MT and TQA. To learn more about these QTLaunchPad activities, visit www.qt21.eu

You can help making quality translation tools and technologies better by testing and providing feedback. If you are interested, contact us at

info@qt21.eu