# Zen and the Art of Quality Assurance

## Quality Assurance Automation in Translation: Needs, Reality and Expectations

Julia MAKOUSHINA
Palex Languages & Software Ltd.,
ul. Uchebnaya 39/1
634034 Tomsk, RUSSIA

Dr. Hendrik J. KOCKAERT
Lessius/KULeuven
Department of Applied Language Studies
Sint-Andriesstraat 2
2000 Antwerpen, BELGIUM

> Not everyone understands what a completely rational process this is, this maintenance of a motorcycle. <…> A motorcycle functions entirely in accordance with the laws of reason, and a study of the art of motorcycle maintenance is really a miniature study of the art of rationality itself.
>
> Robert M. Pirsig, "Zen and the Art of Motorcycle Maintenance"

### 1. Introduction

Life is constantly getting faster, and we are forced to follow the trend, whether we like it or not. Every year, the translation market pursues improved ROI and is therefore shifting from single-piece production to serial production and further to mass production. To weather successfully this development, the translation business needs appropriate automation. Automated translation management has become unavoidable, hence is booming in today's translation business. Like in manufacturing, automated processes in translation result in high volume increase in less time with less effort. But while automating in manufacturing helps exclude "human errors" and therefore avoid deviations and variance in products being manufactured, which contributes to higher quality, the impact of automation on quality in translation has always been questionable. The most commonly accepted belief is that automation minimizes time and cost but gives also rise to reduced quality. Or, in which manner, if at all, does the maxim "Quality Doesn't Matter" fit in today's translation industry?

However, we need to understand exactly what quality in translation, hence in translation assessment, refers to. It has been debated for ages, and is not defined yet. Turning e.g. to the EN 15038 [2006] standard on Translation services and Service requirements, we notice that the standard considers quality of paramount importance, but does not define quality in this context[1].

> "[EN 15038] encompasses the core translation process and all other related aspects involved in providing the service, including quality assurance and traceability".

That translation quality is a must-have across the entire translation workflow management, including a positive vendor-buyer relationship, has been put forward by Nataly, Beninatto & DePalma [2008]. While translation service providers [TSPs[2]] mainly focus on delivering linguistic quality, buyers of translation services often consider the attitude, the quality of service, the relationship between them and the service provider as essential value-added components in translation quality processes. Probably the most comprehensive [but again uncertain] definition is that translation quality refers to conformance to requirements. Numerous quality standards exist in the industry and many providers try to conform to them, but quality is still very subjective and difficult to measure. There is no existing recipe on how to achieve overall quality, because it is unknown what exactly needs to be achieved.

---

[1] The forthcoming ISO standard on translation services [ISO/TC37/WG6] has not been referred to in this paper.
[2] Person or organisation supplying translation services [EN 15038]

Without knowing what we need to automate exactly, we consider translation as a process which includes automating quality assurance and quality control during this process.

Turning to quality in translation assessment in particular, we follow van Santen [2007:83] and Anckaert, Eyckmans and Segers [2008], and observe that quality has gained momentum recently in that evaluating translations has been crucially linked to legal court actions triggered by non-transparent, and non-objective evaluation practices. Amidst this spirit of the times, [semi-] automated Translation Quality Assurance [TQA] should not forfeit to deliver tools capable of displaying equally objective and transparent evaluation schemas across the board. Together with the increased legal implications of evaluation, TQA has entered into the automation world.

## 2. The needs of quality assurance automation

Together with Nataly, Beninatto & DePalma [2008: 16], and with the evaluation stages in the overall translation workflow recommended by EN 15038 [2006], we can choose to subdivide quality assessment into the following categories: [1] detecting errors in end [delivered] documents known as quality control [QC], [2] locating and improving problems and errors during the translation process, known as quality assurance [QA] and [3] preventing as well as controlling problems, known as quality improvement [QI]. In our exercise of providing a comprehensive translation assessment tool, we aim at integrating translation QC and QA in a continuous QI process, which "focuses on improving overall translation performance". Both QC and QA are considered as being part of the entire quality process, and we recommend a translation assessment software suite that is able to provide assistance during the overall translation workflow, hence the full TQA process.

Translation process are not static, but they change and evolve, and most probably in some years they will be completely different from what we saw a few years ago. It is known that the translate-edit-proof [TEP] schema is the approach most commonly followed by translation service providers and welcomed by buyers. This conventional process is the reality in which most translators work and try to ensure the quality, so we will consider exactly this process. It starts with a set of source files, continues with intermediate files obtained from source files after a certain set of manipulations, and ends with the final files obtained from intermediate ones after another set of manipulations.

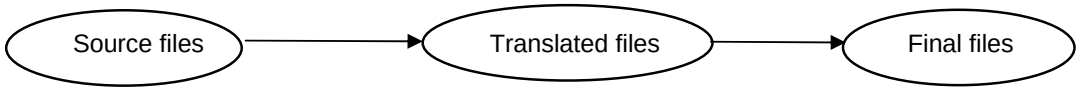


*Figure* 1: Different stages in the translation workflow

It is recommendable to perform translation quality checks at each stage starting from the source files. To aim at a low error rate, source translatable files need to be error-free, clear and unambiguous, especially in the realm of technical translation, as well as formatted and typeset correctly. Translated files need to convey an adequate meaning, use correct and consistent terminology, and ideally sound as if they were originally written by a native speaker of the target language. They need to keep the source formatting [unless required differently by the client]. Final files are identical to translated files, but with the lowest possible translation error rate [ideally zero]. If we take a closer look at what is checked at each stage, we observe that some tasks may be formalised effortlessly [i.e. terminology may be checked against a termbase, formatting requirements are almost always formal, spell-check has been already successfully implemented by many software manufacturers, etc.], but others may not, or only partially. The latter include checks of unambiguity, adequacy of meaning conveyed, native-language approximation, etc. In other words, each stage in the translation process requires both formal and non-formal checks. Table 1 summarises the QA tasks at each translation stage and the manner in which most TSPs currently perform these tasks.

| Source files | | Translated files | | Final files | |
|---|---|---|---|---|---|
| Formal QA | Non-formal QA | Formal QA | Non-formal QA | Formal QA | Non-formal QA |
| Manual check<br>QA software | Editing<br>Controlled language<br>Authoring memory, etc. | Manual check<br>QA software | Proofreading<br>Editing | Manual check<br>QA software | In-country review |

*Table* 1: QA Tasks at different translation stages

The set of files at each stage include not only files to be translated, but also translation memories, project glossaries, reference materials and client instructions. The translator modifies other files directly by feeding translation memories and/or glossaries and indirectly by asking questions about instructions, which may trigger client changes. In this way, the instructions can be made clearer, and contain more or less details.

Translation quality does not depend on high-quality translatables only, but also on other file types of high-quality. Accordingly, translation memories should, from their inception, be fed with high-quality translation pairs. Translation units should be correctly and appropriately segmented. Glossaries should contain correct terms. Terms should be classified and described correctly, allowing the translator to make unambiguous choices. Instructions also need to be clear and intelligible. The same applies to reference materials. Ideally, reference materials need to be converted into a translation-memory or glossary-like database. And in the end, the translation memories and glossaries need to be refined according to any comments or questions, and checked again to make sure they are correct, clear, adequate and appropriate. Table 2 displays a classification of QA tasks performed during translation[3].

| File type/Checks | | Formal | Non-formal |
|---|---|---|---|
| Translatables | | Formatting, typesetting, terminology | Editing, proofreading, revision |
| TMs | | Consistency, text formatting, segmentation etc. | Proofreading, revision |
| Glossaries | | Unambiguity, grammar | Adequacy |
| Reference materials | | Consistency, terminology | Proofreading, revision |
| Instructions | Requirements themselves | N/A | Unambiguity, clearness, understandability |
| | Conformance to | Automatic check of conformance [for those instructions that may be formalized] | Manual check |

*Table* 2: Classification of QA tasks performed during translation

This table is still not complete because it includes not all parts of the process. What is e.g. not included, but directly impacts on the quality of translation is the quality of resources, i.e. translators, editors, proofreaders, and reviewers. The process to control, ensure and improve their quality embraces selecting carefully in advance, evaluating the delivered quality regularly, and accumulating the evaluation database to analyse the progress.

[3] We do not claim to give an exhaustive study here.

Based on the above, Table 3 lists the required tasks in order to ensure translation quality, paired with possible software to automate those tasks.

| Resource Type / Check Type | Formal | Non-formal |
|---|---|---|
| Source files | Single-language QA checker | Controlled language, authoring memory |
| Translated files | Double-language QA checker [including possibility to check conformance to formalized client instructions] | Revision memory, what else? |
| Final files | Double-language QA checker [including possibility to check conformance to formalized client instructions] | Revision memory, client QA checker |
| TMs | QA checker for TMs, TM enhancer | Revision memory |
| Glossaries | Terminology manager | Revision memory, what else? |
| Reference materials | OCR, text aligner, terminology extractor, TM enhancer | Revision memory, what else? |
| Translators, editors etc. | Database of evaluation results | Selection and evaluation utilities |

*Table* 3: Required tasks for ensuring translation quality

That it is difficult to find adequate means to formalise correctly client instructions due to the high diversity rate of projects and project files preparation types, deserves equally attention here.

## 3. The reality of quality assurance automation

In this paper, we wanted in the first place to gauge the level at which current QA tools are able to assist [semi-]automated translation processes. We refer to Gerasimov [2007], Zetzsche [2007], and to the survey which was performed by ourselves [Makoushina, 2007], out of which we select the analysis results relevant for this paper.

As we can observe in Table 4, TQA automation tools are currently capable to detect formal errors in translated texts, though not for all languages and encodings, and not for all file formats. Most of the currently available tools are able to reduce the amount of “noise” they report. Trados QA Checker 2.0 is noteworthy: it was released in January 2008 and was probably the best tool able to detect a lot of different errors in many languages while generating the least amount of “noise” possible. The following table summarises which types of software mentioned above are available on the market.

| File type/Checks | Formal | Availability | Non-formal | Availability |
|---|---|---|---|---|
| Source files | Single-language QA checker | Yes | Controlled language, authoring memory | Yes |
| Translated files | Double-language QA checker | Yes | Revision memory | No |
| Final files | Double-language QA checker | Yes | Revision memory, client QA checker | No |
| TMs | QA checker for TMs, TM enhancer | Yes | Revision memory | No |
| Glossaries | Terminology manager | Yes | Revision memory | No |
| Reference materials | OCR, text aligner, terminology extractor, TM enhancer | Yes | Revision memory, what else? | No |
| Translators, editors etc. | Database of evaluation results | Partial[4] | Selection and evaluation utilities | Partial |

*Table* 4: Currently available TQA software types

Table 4 displays various software tools for formal checks, some of which are not commonly known yet. For the non-formal checks, we need to stress the point that the their invasion into TQA is only starting. And as we will illustrate below, it is possible to manage non-formal checks in an efficient way.

## 4. A less known side of quality assurance automation

According to Makoushina [2007; 2008], existing TQA software tools provide significant fine-tuning features to detect more types of formal errors and to generate less “noise”. Eventually, the functionality of the tools evolves into automatically checking the conformance to different client requirements. Although the mechanisms implemented in different applications differ, the most universal and powerful of them is using regular expressions which are like a mini programming language that allows us to describe what exactly we would like to find using text patterns. As it is exampled in Makoushina [2008], regular expressions can extend easily functionality of QA checkers that support them and allow them to detect any forbidden character combinations, repetitive characters and even words. Due to the specific purpose of this paper, we do not intend to explore the analysis of regular expressions. For further reading, we refer to Friedl [2006].

## 5. From the state of the art to future development of TQA

Based on the above findings, QA checker features as well as authoring memories, controlled languages and term extractors are nowadays the most advanced tools to automate QA processes. However, these tools are expected to enhance, extend and customize their functionality. What we have not seen yet, are TQA tools that are able to manage non-formal checks in a systematic and [semi-]automated way.

In this context, a needs analysis was performed in *Lessius* in order to determine which features we would really like to see added in a more comprehensive TQA tool. The outcome was quite straightforward in that the main vacuum seemed to be the deficiency of a memory-based add-in device in an "evaluation memory" [EM] tool [Kockaert, Makoushina and Steurs, 2008]. Such a TQA tool would be able to house all the above formal evaluation features, plus typically non-formal linguistic and stylistic corrections, comments, and even links to external bibliographical resources added by any evaluator[5]. Next, it seemed to be advantageous to complement TQA tools by a statistics chart for the purpose of assisting the

[4] *Partial* here means that those applications are generally built by each translation service provider separately according to their particular needs and evaluation criteria based on any commercially available database management systems.

[5] In this paper, the evaluator's job refers, accordingly to EN 15038 [2006], to reviews, revisions, and proofreading.

evaluator and/or the translator in the evaluation and translation processes respectively by showing error frequency and error types. Introducing statistics in a TQA tool will allow the end-user to benefit from a large scope translation flow including error data for any subsequent translation or authoring task. Moreover, error statistics will prove useful for training reasons as well in translation studies colleges/universities, EU, NATO and UN departments of translation, and any translation department which is concerned about improving QA. The idea of introducing customisable assessment statistics in a TQA suite, has been particularly welcomed for the purpose of compatibility with the recently published EN 15038 standard on translation services. The development of a TQA suite which is able to impact on each stage involved in the evaluation of the entire translation process will undoubtedly benefit the entire workflow process management. Across the revising, reviewing and proofreading processes, it is recommended to cater for a TQA which can play a role in each of the workflow stages recommended by EN 15038 [2006].

From the technical side, a stand-alone tool that can be used independently from any other translation tool, in e.g. Microsoft Word documents, was put high on the wish list.

In the next lines, we present a TQA tool, which is hoped to face the challenges that arise from the desiderata put forward by the surveyed translators and evaluators. What needs to be added to the above discussed formal checks, seems to be a computer assisted process able to interact with, hence to partially supplement, intrinsic human activities of evaluation, not the least the chance of adding linguistic and stylistic comments in both source and target documentation, which are not automatically triggered by errors that are retrieved and handled in formal automation processes.

When we compare with the above tested TQAs, the most stringent dimension we need to add, seems to be the opportunity of having a manual building up of corrections/comments by a human translator/evaluator. A device able to complement the above described automated evaluation features with manually added comments, which can be stocked in a database, or an EM, is hoped to turn into a TQA tool we need. The Department of Applied Language Studies at *Lessius* and *Palex Languages and Software* have developed a design document of such an innovative device, and we have code named it as RQ.

### *5.1. Wishful thinking?*

In the next lines, we portray a virtual TQA tool that is hoped to near this goal. The evaluator starts correcting one project by adding the necessary comments, suggestions and corrections to be dealt with by the translator. In this way, he builds up a revision/correction memory; for each next translation [within a similar domain and/or text type], the memory can be updated, which is comparable to a Translation Memory [TM]. The memory can be saved and re-used later by the same evaluator, another evaluator, and even the same and/or another translator. Incorporating the main existing automated TQA features in RQ can be a plus, which will lead to a fusion of the above mentioned automated TQA features, and the manually built up database of earlier comments/suggestions. Based on the EM of an earlier translation, a window will pop up in the margin of each new translation assignment while moving the cursor over those segments that have already been evaluated earlier. The automated detection of the items where the evaluator has already made some comments is crucial: therefore, the EM is linked to the source text and the comments/suggestions already been entered. When RQ detects a segment for which a comment has been entered, it will display automatically the comments/suggestions the evaluator made during the evaluation of a previous translation of the text. In this way, RQ will guarantee an overall consistency in each revision, at all levels, ranging from punctuation to language register. It all depends on what the evaluator enters.

The evaluator will always be able to edit anew each comment/suggestion thanks to an Edit mode. Each new editing operation in an earlier added comment will generate an automatic update in earlier evaluated translations, when opened. This feature will offer maximised flexibility and back/forward change management to each evaluator. What will be particularly useful in the context of objective evaluation criteria across the translators [including examinees], is the fact that each next revision of the same [similar] source text will generate exactly the same revision comments, hence a huge gain in evaluation consistency at all levels, technical, linguistic and stylistic. Thanks to the constantly updated EM, each translator will receive an identical feedback, as broad as possible. In this way, RQ embraces present-day assessment policies of translations, which are more and more integrated into the University/College examination guidelines. The EN 15038 standard is the first concrete embodiment of this new trend.

Assessing a number of similar texts will reduce enormously the very time-consuming task of evaluating translations, because the EM will always display each previously added and updated comment, allowing to re-enter this comment at any time when needed. Regarding the evaluation itself, RQ will integrate sufficient statistics features to allow each translation project to be quoted in exactly the same way, which guarantees an objective and fully exchangeable evaluation scheme, and which harmonizes each evaluation which in turn leads to fairly quoted translations.

Technically, RQ will be able to count and display the errors automatically, and to display the marks in a separate window in the statistics chart area. Marking will be done based on an evaluation matrix set by each individual evaluator [Segers, 2007][6], or a default matrix. Each comment can be referenced exhaustively on the account that the evaluator needs to add all the necessary comments and references only once. Severity bench-marks based on international standards [EN, ISO, JA2450, ...] could also be applied to RQ. Thanks to the expertise developed by *Palex Languages and Software*, it will be possible to integrate RQ into their already advanced automated TQA. The EM will hopefully be linked to existing or new databases/TMs in order to ease further the workload. This will alleviate at the same time the searching job for the exact terminology and phrases/context. In other words, external databases will help the evaluator and the translator choose the right terms/phrases before adding these to the RQ memory.

In addition to an evaluator's tool, the TQA tool is expected to use RQ as an assistant tool for translators while in the process of translating. Depending on default or customized access rights, RQ displays the previously checked comments, which are known to be valuable for the translators on the account of the evaluators' expertise, their previous search results, and advice from externally contacted domain experts.

### *5.2. Descent into reality*

Let us now descend to reality and distil the performable features out of the above desiderata. The objective of this TQA project is to deliver a system that can be used to “capture” changes and corrections the evaluator makes in source and target documentation, to add comments for new users and to accumulate such information in a database for re-use. With each new segment in a text, the EM database is scanned for a similar record, and if one or more are detected, the information is retrieved and made available to each user involved in the translation process.

### *5.3. RQ: Workbench and MS Word*

With RQ coupled to SDL Trados Workbench, the engine of the system is designed to manage with adequate success the EM database. The functions of the application are the following:

Open/close correction and authoring memories; Import/export correction and authoring memories; Populate correction memory by adding [or changing] records of source, target and corrected segments into the database [as well as any comments made by the evaluator] after the segment closure; Populate authoring memory by adding [or changing] records of source and corrected segments into the database after confirmation; Look up segments in the correction/authoring memory based on user selectable criteria and visualize the options to the user. This means the user will see the previously entered corrections/comments on similar segments when translating a new [similar] text; Export partial data from correction memory to authoring memory; Search correction and authoring memories for user-specified text [concordance search].

Its main window might look as follows: the commented segments may be highlighted in *Source text*, *Original translation*, and in *Corrected translation* boxes. It is in this window that the "scanning or detecting" [by cursor gliding over segments] of translatable [or pre-translated by SDL Trados] segments in the *Source text*, while the translator is performing the translation of these segments in the *Original translation* window, will trigger the pop-up in the *Comments* box of comments/corrections that were already added by the evaluator in previous translations.

[6] The assessment policies and technical feasibility to convert a gross score into an individual mark deserves further attention in another context.

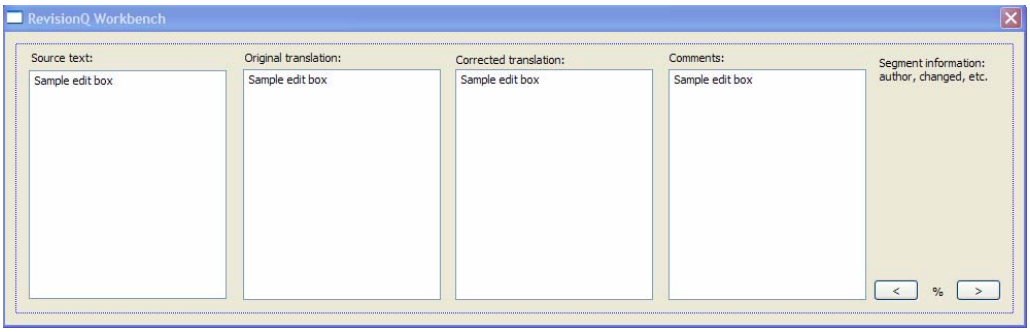


*Figure* 2: RQ Workbench User Interface

The user interface will show four panes in a window. The buttons [< and >] on the right allow to switch between different matches with higher or lower percentage of similarity. The user will be able to switch different views in future versions [this is not the high priority feature]. The information about the database record is shown above the buttons: author and time of segment translation, author and time of modifications, and file names.

The proposed TQA tool is not necessarily linked to an entire translation memory tool, such as among others SDL Trados Workbench or the recently released MemoQ 3.0 tool. Compatibility with other translation environment tools will be complemented by a Microsoft compatible tool. Let us take a glance at the MS Word format. In this respect, RQ differs, in that it offers a stand-alone TQA tool, directly applicable on MS Word and Excel documents, without being linked exclusively to a particular translation memory tool. In MS Word, the evaluation macro set allows users to open source and target texts, to correct translation and to add comments to the file before finally sending the evaluation data to the database management program. In MS Word, RQ works in a Track Changes mode where the evaluator inserts comments and corrections, and references to terminological databases.

The MS Word macro set will provide the user interface to the memory. It should work approximately as the SDL Trados Word macro set. The macro switches consequently segments of previously translated bilingual text and: Checks the open EM to find segments with the similarity percentage specified in the Options; Checks if any “terms” saved in the memory are found in the open segment; Provides both perfect/fuzzy matches and “term” matches to the user; Shows Track Changes the user adds in the translation. As soon as the user gets to the next segment, the macro set saves the changes and/or comments in the EM. If the same segment with other corrections already exists in the EM, it should behave according to the settings [overwrite after allowing sufficient access rights, keep the memory version, depending on the users' access rights].

## 6. Any other business?

Based on the latest developments in broader translation environments, such as "collaborative translation", we refer to Nataly et al. [2008]. In this new development, TQA needs to be paced in a broader context, i.e. including the LSPs buyers' perspective, and in a dynamic environment, i. e. in a translation environment which adopts, or if not, at least may experiment working in a "collaborative translation model" [Beninatto & DePalma, 2007]. Our TQA suite seems to be all-round enough to step into this new translation workflow mode.

## 7. Conclusion

Translation has always been a single-piece production process. Just a few years ago, no one could really believe machine translation would become part of our life. It was hard to imagine that machines would be able to perform tasks that are still considered as Art: to join words into readable and correct phrases. And although machines cannot perform these tasks on their own, machine translation has become the reality of our days. Although mass production via machine translation does not seem to generate high-quality products, it may result into a powerful tool allowing to translate hundreds of words in minutes, when paired with automatic QA and QC process.

Today, we wanted to show a new dimension in TQA automation. Humanly entered add-ins are obviously not fully automated, but are a substantial part of the entire evaluation process, and are indeed valuable material that will undoubtedly benefit future translations. The fusion of human and computer-aided evaluation processes in an evaluation memory manager is expected to deliver a more comprehensive TQA tool for translators, evaluators and proofreaders than experienced at the present time.

The RQ functionality differs from currently available software products in that it produces not only the history of earlier translations, but the history of evaluators' comments, suggestions, resources, and corrections on each translatable item. What has not been seen yet, is the evaluator's motivation/justification of correction, which is important for feedback to the translators or examinees in translation exams or entrance exams.

Last but not least, when we refer to the rise of collaborative translation communities, RQ adds to today's state of the art, the functionality of a Review/Revision/Proofreading Queue or History, and not only a set of formal checks, or a history of translated segments.

## 8. References


Anckaert, Philippe , June Eyckmans & Winibert Segers. 2008. "Pour une évaluation normative de la compétence de traduction" In *ITL - International Journal of Applied Linguistics,* 155: 53-76.

Beninatto, Renato S. & Donald A. DePalma. 2007. "Collaborative Translation". In *Common Sense Advisory*, Inc., Lowell, Massachusetts [www.commonsenseadvisory.com, last consulted: 5 Sep 2008].

EN 15038. 2006. *Translation services - Service requirements. Technical Committee*. CEN/BT/TF 138, European Committee for Standardization.

Friedl, Jeffrey. 2006[3]. *Mastering regular expressions.* O'Reilly Media Inc.

Gerasimov, Andre. 2007. "A comparison of translation QA products". *MultiLingual*, 85 [18]: 22-25.

Kockaert, Hendrik, Julia Makoushina & Frieda Steurs. 2008. "Translation quality assurance: what is missing? And what can be done?". XVIII[th] FIT World Congress in Shanghai.

Makoushina, Julia. 2007. "Translation Quality Assurance Tools: Current State and Future Approaches". *Proceedings of the 29th International Conference on Translating and the Computer..*

Makoushina, Julia, 2008. "Automated quality assurance and quality control in translation: getting more from existing tools". *Proceedings of the tcworld 2008 conference*.

Nataly, Kelly, Renato S. Beninatto & Donald A. DePalma. 2008. "Buyer-Defined Translation Quality" In *Common Sense Advisory*, Inc., Lowell, Massachusetts [www.commonsenseadvisory.com, last consulted: 5 Sep 2008].

Segers, Winibert "Ijkpuntenmethoden". In Van de Poel, Chris. & Winibert Segers [eds]. 2007. *Vertalingen objectief evalueren*. Acco: Louvain, 21-25.

Van Santen, Fedde. 2007. "Evaluatie in context, de tijdgeest verklaard". In Van de Poel, Chris. & Winibert Segers [eds.]. *Vertalingen objectief evalueren*. Acco: Louvain, 83-90.

Zetzsche, Jost. 2007. *The Translator's Tool Box. A Computer Primer for Translators* [5.5]. International Writers' Group, LLC.