

Using Computer Assisted Translation tools' Translation Quality Assessment functionalities to assess students' translations

Jun Yang, Dr. Dragoş Ciobanu, Caroline Reiss, Dr. Alina Secară

The Language Scholar Journal

The Language Scholar Journal (LSJ) is an open access and peer-reviewed publication. Its main objective is to provide a platform to promote the teaching and learning scholar activity within the School of Languages, Cultures and Societies at Leeds and beyond.

Alongside the annual printed issue, the LSJ digital's space hosts and showcases contributions, facilitating sharing and exchange of ideas. Submissions can be sent to the editors at any time. However deadlines may be announced for specific issues.

The Language Scholar Journal is published by the Leeds Centre for Excellence in Language Teaching.

ISSN: 2398-8509

The Language Scholar Platform: <http://languagescholar.leeds.ac.uk/>

Using Computer Assisted Translation tools' Translation Quality Assessment functionalities to assess students' translations

Jun Yang, Dr. Dragoş Ciobanu, Caroline Reiss, Dr. Alina Secară

School of Languages, Cultures and Societies, University of Leeds

Abstract

Translation quality assessment (TQA) is essential in translator training. For formative and meaningful feedback, quantitative methods of error-type categories are frequently used to evaluate students' translations. However, because the annotation of the errors made by students still tends to be done in a Word document, which requires redundant manual work and the result often lacks consistency and clarity. We propose using CAT environments for more efficient TQA. On the one hand, it will bridge the gap between the training and industry by familiarising students with current industry translation practices; on the other hand, it will assist with the design and analysis for formative assessment that could guide the learning of not only one student, but of whole cohort of students. The current paper uses SDL Trados Studio as an example to demonstrate how translation evaluation works in a CAT environment. Through the discussion of practical challenges and advantages of implementing a TQA in the translation classroom, we highlight the clarity both of expression and of presentation of feedback and evaluation and the long-term benefits in recording the students' individual and group progress of all times to make data-driven choices regarding the training curriculum.

Keywords: translation quality assessment, computer-assisted translation tools, translator training

INTRODUCTION

The majority of undergraduate and postgraduate translation programmes include a compulsory translation element as part of the training. In addition to this, training in Computer-Assisted Translation (CAT) tools is frequently part of the postgraduate curriculum. However, these two components tend to be taught separately, one being seen as a linguistic task, while the other a technical one. Our current article suggests a framework where formative student translation feedback could be integrated in and supported by a functionality offered by several CAT tools, which allows the semi-automatic identification, annotation and tracking of translation errors. Depending on the CAT tool used, this functionality can be found under the name of Linguistic Quality Assurance (LQA) for users of the memoQ CAT tool, or Translation Quality Assessment (TQA) for users of SDL Trados Studio. Used creatively, this functionality can enable translation trainers not only to automate the detection and tagging of certain translation errors, but also the creation of a translation evaluation corpus – also known as a 'learner corpus' and representing a collection of corrected and annotated translations - which could be analysed per student or per cohort. The use of such corpora is still rare in translation studies research, as well as translation training, mainly because of the difficulty associated with their compilation; their benefits, however, include the possibility of identifying patterns in student translation choices early on during training courses, thus subsequently enabling trainers to implement corrective actions in a more timely and structured manner.

In the words of Doherty (2016:131), 'Translation quality assessment (TQA) is crucial to formative assessment for translators in training, for professional certification and accreditation, and for recruitment and screening of translators.' At the same time, 'trying to apply quality control to intellectual output such as translation is difficult and encounters significant resistance among practitioners' (Samuelsson-Brown, 2006:42). While on the one hand, trainees cannot progress without meaningful feedback; on the other hand, the manner,

consistency and user-friendliness of this feedback are also vital in the training of future translators.

The evolution from general qualitative evaluations of translations to more detailed, quantitative methods was a welcome development which has brought translation training closer to industry translation evaluation methodologies, and nowadays numerous training programs evaluating student translations are using error-type categories for consistency and efficiency purposes. While it is undoubtedly a step in the right direction, more can be done in order to streamline the evaluation process. At the moment, the annotation of the errors made by students still tends to be done in a Word document, which requires assessors first to remember by heart and write redundantly throughout the document the error typology labels, and secondly to add up at the end of the evaluation the numbers of errors made in each category and sub-category. In some training contexts one further evaluation sheet is created for each student with data centralised to various degrees of reliability, which is used outside the specific translation assessment and which invariably requires constant reference back to the assessed translation for context purposes. This practice is prone to human error, as well as more complicated to implement in cases where the error categories are organised on multiple levels and sub-levels, and they are also differentiated according to their severity.

Doherty (2016) highlights the need to approach TQA in a more objective, empirical manner and move away from a subjective analysis of linguistic features only. Reiss (2000) also supports the view that evidence-based TQA would be a good pedagogical tool for both evaluators and translators themselves. Collecting data on evaluations and then mining this data according to the error categories the trainers and trainees want to focus on at particular times would indeed offer a more methodologically-valid approach to translation training. We share this view, and the current paper will focus on using CAT environments for more efficient translation quality assessment. Adopting CAT tools as part of traditional translation classes will both bridge the gap between training and industry by familiarising students with current industry translation evaluation practices, and will also assist with the design and analysis of

formative assessment that could guide the learning of not only one student, but of whole cohort of students.

While in the Translation Studies literature, it very often looks like evaluation is the final step in a process; in other domains such as Machine Translation (MT), the evaluation is often part of a larger framework and therefore it is built in from the beginning with the idea of supporting a more complex process (Doherty, 2016:133). The framework we are suggesting would imply the trainers either adopting an existing industry translation evaluation framework, or re-creating and using inside a CAT tool their favourite error typology. To make the training process more transparent and meaningful, this framework should also be shared with the students from the beginning of the translation process, progress against all error categories should be regularly tracked at both individual and group level, and students should be trained to use this very same industry-driven evaluation methodology when self- and peer-assessing translation output. Moreover, instead of proceeding with training activities and source text selected intuitively by the trainer, the regular TQA reports should influence the creation of customised training resources based on areas where the data indicates that students need more practice.

Various TQA frameworks have been used for human and machine translation assessment, post-editing effort, estimation of productivity and implementations. For human translation evaluation, the earliest Quality Assurance (QA) models were introduced in the automotive industry - e.g. SAE J2450 - or the localisation industry - e.g. the QA models proposed by the late Localization Industry Standards Association (LISA). These models were based on translation error-types identification and the categories included ranged from accuracy, fluency, terminology and style to formatting and consistency (for a more detailed discussion on the early translation error-category models, see Secară (2015)). Later on, the Translation Automation User Society (TAUS) adopted a dynamic model for evaluation translating quality. In addition to the previous categories, they also evaluated the time spent on the task together with more localisation-specific tasks, such as the post-editing of machine translation. In

parallel, Language Service Providers and organisations have been developing their own translation evaluation schemes, based on their needs. For example, for International Organisations such as the United Nations or the European Commission, a category linked to appropriate research and double-checking of references is always included in the evaluation process. Also, some choose to integrate weighting in their models, deeming some errors more serious (mistranslation) than others (formatting). No matter how many errors one chooses to use in their translation evaluation scheme, their choice should be linked to the purpose of the translation task and attention should be paid to the actual error-types included.

Assuming a translation trainer already has a suitable translation evaluation scheme, in the next section we will discuss how the scheme can be integrated in a CAT tool so that the trainer could semi-automatically tag certain types of error when marking student translations.

USING CAT TOOLS FOR TRANSLATION ERROR ANNOTATION

SDL Trados Studio and memoQ are two popular commercial CAT tools that include the functionality of TQA. SDL Trados Studio followed the example set by memoQ and, starting with its 2015 version, has been supporting the creation of customised TQA models - one could imagine a scenario where tutors would create different models depending on characteristics of individual language combinations. For example, some error typologies in European languages may not be suitable in Asian character-based languages (Chinese, Korean, and Japanese) - for instance, due to the absence of gender and number in Chinese, an error category of the type *Language: Inflection and Agreement* is superfluous and potentially confusing for the learners of this language, while for inflected languages such as any of the Romance languages it is an essential category.

Although there are some differences in implementation, both tools can generate comparable bilingual TQA reports which contain scores, the error typology used and, most importantly, include 'track changes' of with the annotated erroneous translated segments compiled in one document. Since the purpose of using TQA in translator training is to provide systematic and

structured feedback, we believe that the ‘track changes’ functionality is an essential one as it tells students where their errors are and – provided the marker included corrections and/or additional comments – how these errors can be corrected and the student’s performance improved. In this paper, we use SDL Trados Studio 2015 as an example to demonstrate how TQA works in a CAT environment.

While tutors using memoQ can immediately select existing TQA templates already loaded in this CAT tool which are already widely used in the industry – e.g. the LISA QA model, MQM, or TAUS – in SDL Trados they need to setup their own TQA by defining the following metrics: categories, severities, scoring, and document types. ‘**Categories**’ allow users to configure the error typology that will be used to annotate the students’ translations (subcategories can also be configured for specific error types). ‘**Severities**’ sets custom metrics that can be used to measure the importance of errors. ‘**Scoring**’ determines the importance of each severity level in the overall assessment and allows users to define the *Pass/Fail Threshold* which is the maximum penalty score that a CAT tool can admit before failing the TQA check for a translation. Finally, ‘**Document Types**’ lists the types of files that the current TQA settings can apply to, which is useful in a translation project which includes files in several formats. The customised TQA settings can be saved as a template for future use.

Figure 1 shows one of the four metrics (‘Scoring’) of the TQA configuration in SDL Trados Studio. When opening the CAT tool, in *File/Options/Translation Quality Assessment*, the user can customise all four fields already mentioned above. All the highlighted areas are editable, including where tutors can define the *Pass/Fail* scores at the word or character level.

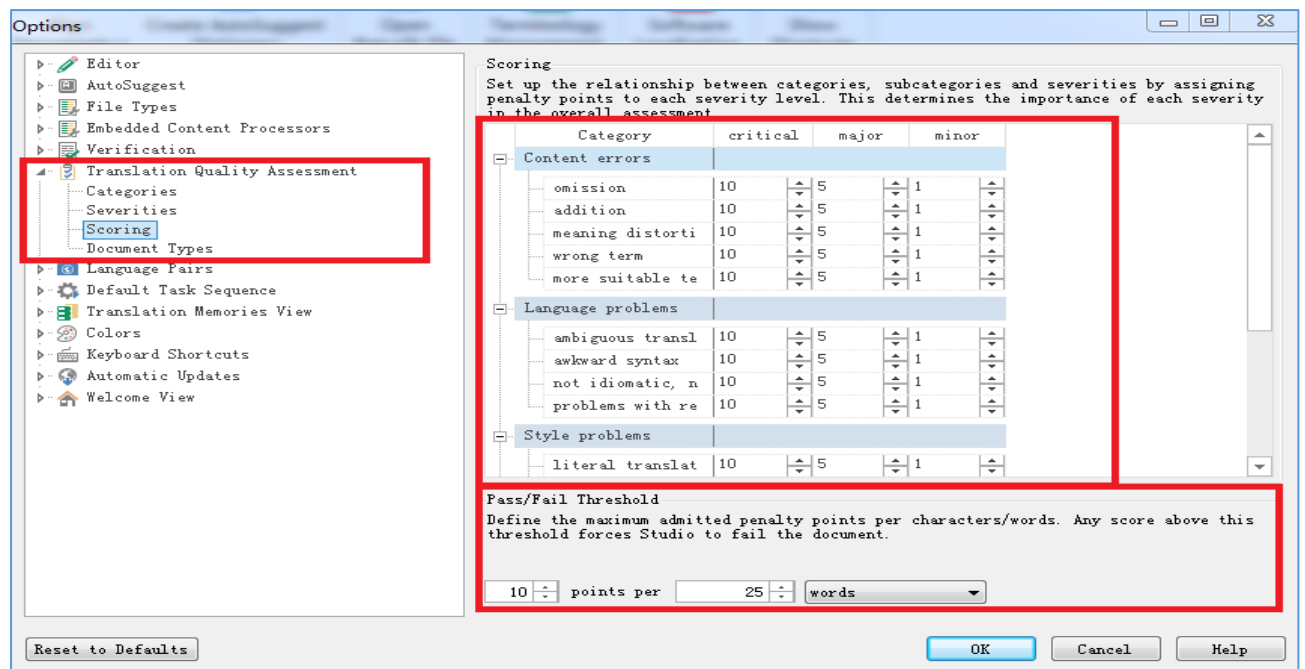


Figure 1. TQA configuration in SDL Trados Studio

Naturally, the easiest way of operating is for the tutor to set up a translation project in a CAT tool and then ask the students to complete it within the same CAT tool before returning it for marking. However, it is also possible to complete the translation outside of a CAT environment, and only use the latter for the error annotation stage. Once the translations have been imported into SDL Studio, the tutor can begin to evaluate the students' translation under the 'Review' pane, in 'Assess quality' mode. When an error is spotted, the tutor can either make revisions similarly to using 'Track changes' in *MS Word*, or highlight the problematic part in one segment and leave a comment. The former option allows tutors to add metadata - e.g. annotations of errors including error type, severity level and revision type (strategies that used to improve the translation: deletion, addition or replacement) - related to each change to give specific feedback on quality. The metadata is captured and displayed in the TQA report. Figure 2 illustrates how to work with the TQA in SDL Trados Studio.

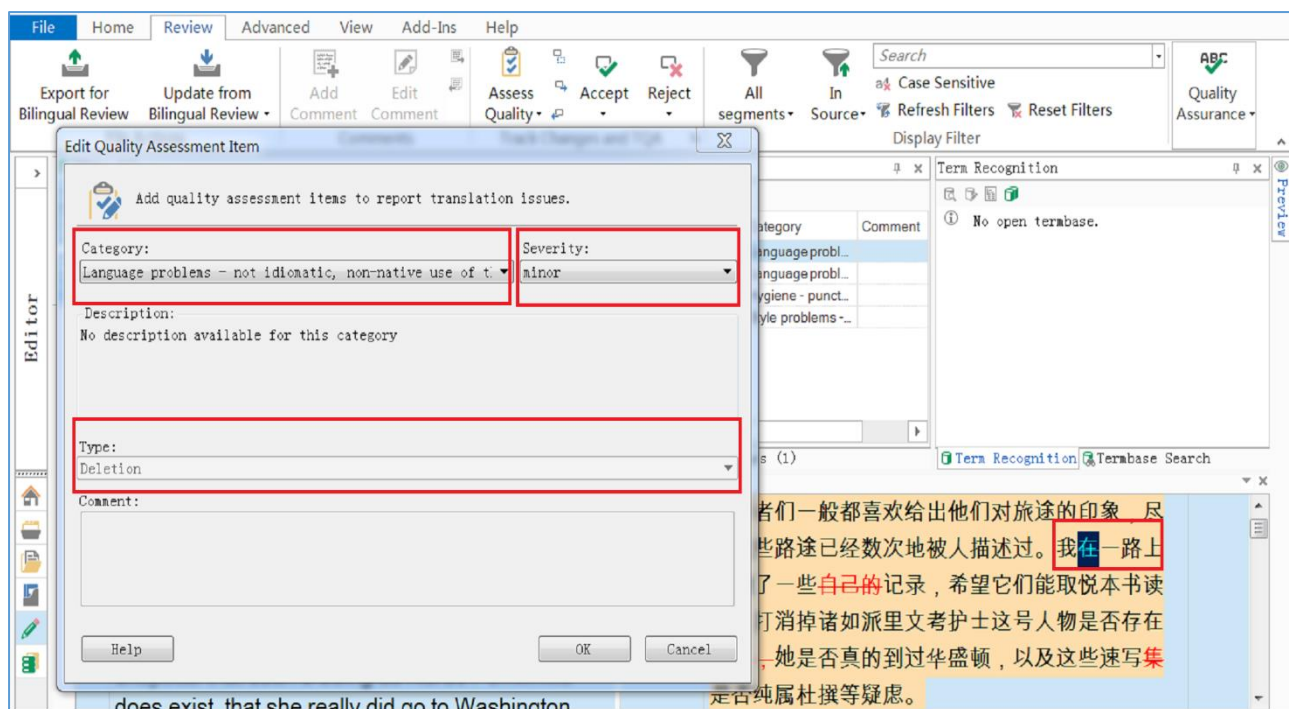


Figure 2. Carrying out a translation evaluation in SDL Trados Studio

In SDL Trados Studio, the TQA annotations that contain the tutor's evaluation and comments are recorded in the 'Editor' pane. As highlighted in Figure 3, TQA annotations are listed together with the segment ID, revision type, document name, author, severity level and category (category-subcategory). Next to 'TQAs', the user can choose to view 'Comments' that were left by the evaluator (circled in red in Figure 3).

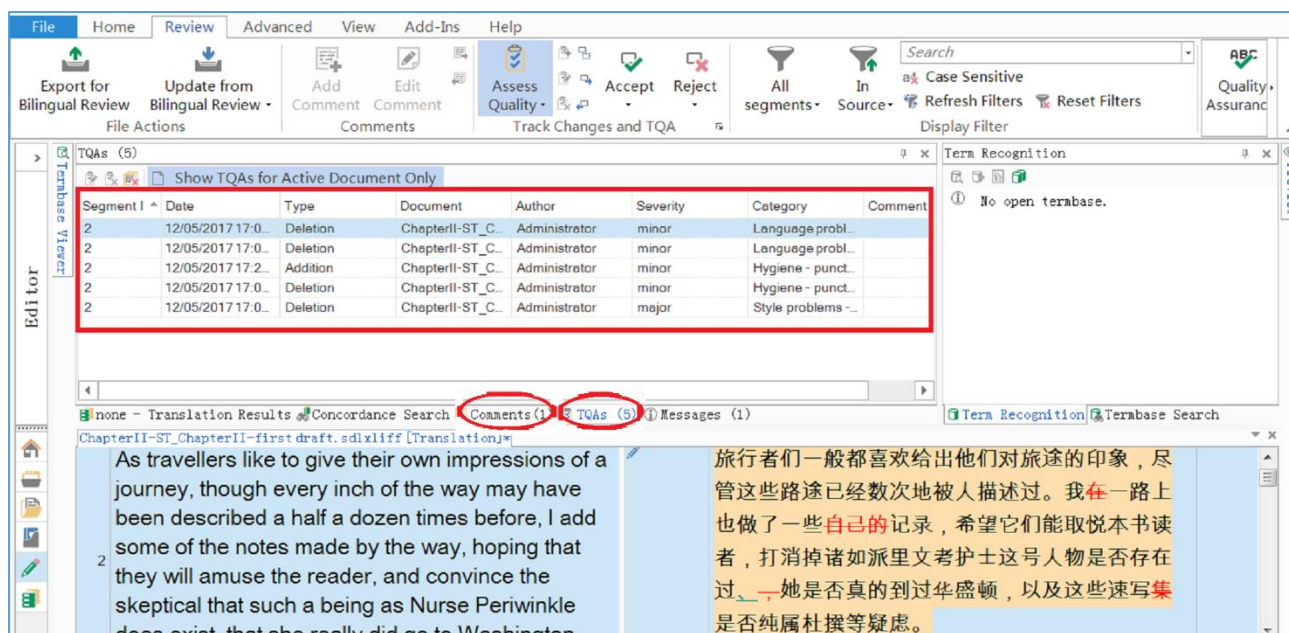


Figure 3. Example of TQA items

When the tutor finishes evaluating the students' translation, a TQA report can be generated through 'Batch tasks' under the 'Home' window. The new TQA report can then be viewed in the 'Reports' pane, from where it can also be saved in other file formats (.html, .xlsx, .mht and .xml). For the common purposes of translator trainers, the TQA report in Excel format offers the most flexibility, as it is clearly-structured and easily-editable. Figures 4-6 illustrate the format of the TQA report in Excel. There are four essential parts in the TQA report: the first part gives a summary of TQA together with the final evaluation score: *Pass* or *Fail*; the second part details the scoring model that has been used; the third part indicates the occurrence of each error type with its specific penalty score; lastly and most importantly for formative purposes, the fourth part of the TQA report provides a comparable table with columns containing the source content and the student's translations, including the original translation, the tutor's feedback (revised translation), the revision strategy (TQA type), the error category, the severity level and any formative comments inserted by the marker.

	A	B	C	D	E
1	Translation Quality Assessment Report				
2	Summary				
3	Task:	Translation Quality Assessment			
4	Project:	Chapter 2			
5	Files:	1			
6	Created At:	01/11/2017 17:10			
7	Pass/Fail Threshold (units):	10/25 (Words)			
8	Project Pass Result:	Pass			
9	Totals				
10	Total	Segments	Words	Characters	
11	Files: 1		96	3070	14132
12	Scoring				
13	Categories	Description	Severities		
14			critical	major	minor
15	content errors				
16	omission		10	5	1
17	addition		10	5	1
18	meaning distortion		10	5	1
19	wrong term		10	5	1
20	more suitable term available		10	5	1
21	language problems				
22	ambiguous translation		10	5	1

Figure 4. Example of a TQA report in Excel (Part 1 and 2)

Severities Occurrence			
Category: language problems - not idiomatic		Total: 54	
Severity	Penalty	Occurrences	
minor		1	54
Category: Hygiene - incorrect punctuation		Total: 21	
Severity	Penalty	Occurrences	
minor		1	21
Category: content errors - more suitable term		Total: 30	
Severity	Penalty	Occurrences	
minor		1	30
Category: style problems - inconsistency with		Total: 1	
Severity	Penalty	Occurrences	
major		5	1
Category: language problems - ambiguous t		Total: 13	
Severity	Penalty	Occurrences	
major		5	13
Category: content errors - term not correct		Total: 11	
Severity	Penalty	Occurrences	
major		5	8
minor		1	1

Figure 5. Example of a TQA report in Excel (Part 3)

Chinese (Simplified, PRC)							
File	Segments	Words	Characters				
Chapter II.docx.sdlxliff		96	3070	14132			
Penalty/Point: Allowed (units): 419/92100 (Words)							
TQA configuration result: Pass							
Document type:							
Evaluation comment:							
Reviewer assessment: Not Set							
TQA							
Segment ID	Source Content	Original Translation	Revised Translation	TQA Type	Category	Severity	Comment
2	As travellers like to give their own impressions of a journey, though every inch of the way may have been described a half a dozen times before, I add some of the notes made by the way, hoping that they will amuse the reader, and convince the skeptical that such a being as Nurse Periwinckle does exist, that she really did go to Washington, and that these Sketches are not romance.	旅行者们一般都喜欢给出他们对旅途的印象，尽管这些旅途已被人描述过数遍已经数次地被描述过。我在一路上也做了一些自己的记录，希望它们能取悦本书读者，打消诸如派里文考护士这号人物是否存在过，她是否真的到过华盛顿，以及这些描写是否纯属杜撰等疑虑。	旅行者们一般都喜欢给出他们对旅途的印象，尽管这些旅途已被人描述过数遍已经数次地被描述过。我在一路上也做了一些自己的记录，希望它们能取悦本书读者，打消诸如派里文考护士这号人物是否存在过，她是否真的到过华盛顿，以及这些描写是否纯属杜撰等疑虑。	Deletion	language problems - not idiomatic, non-native use of TT language	minor	
				Deletion	language problems - not idiomatic, non-native use of TT language	minor	
				Replacement	Hygiene - incorrect punctuation	minor	
				Deletion	content errors - more suitable term available	minor	

Figure 6. Example of a TQA report in Excel (Part 4)

The TQA report presents a type of structured feedback for translator training which is beneficial for both tutor and students. On the one hand, the tutor has a quantitative record of the students' performance that enables the diachronic tracking of both individual and group performance, as well as the clear identification of error categories for which additional training is necessary. On the other hand, the students are able to compare their translation with a reference translation that highlights the errors and explains why they are problematic and, if a learner corpus – as defined in the beginning of the article – is updated with the students' work at the same time, investigate the effectiveness of alternative translation strategies or the way in which their colleagues have dealt with particularly challenging source text passages.

Implementing a TQA in the translation classroom: practical challenges and advantages

The use of a translation quality assessment model to mark and provide feedback on student translations offers opportunities to enhance the students' awareness and understanding of theoretical concepts and to enable tutors to further guide students' learning and development. When the assessment of student translations focuses on identifying language and translation errors (Waddington, 2001), tutors should use either the standards for quality assessment currently employed within the translation and localisation industry - such as the TAUS

framework - or, alternatively, adapt the TQA model according to the characteristics of the language pair being taught. In either case, students will need to develop a clear understanding of the role of translation quality assessment, the TQA model's error categories and, in particular, the implications of their differing severity. Broadly speaking, there are four commonly used error severity levels, irrespective of the technology used for evaluation purposes: Neutral/Preferential errors, Minor, Major, and Critical. In order to ensure that students understand fully the meaning and differences between these error severities and categories, the tutor should provide in advance a text-based version of the TQA model, containing practical examples to contextualise the error categories.

The use of a TQA model to categorise errors as pragmatic, semantic, idiomatic, orthographic, linguistic or stylistic (ibid.) could enhance group discussions with the students by highlighting practical, comparable examples of translation theory in action and by introducing the use of industry-wide standards and the importance of complying with a translation brief. The TQA reports could further clarify and consolidate students' theoretical understanding by allowing them to visualise for each text-type translated what kinds of translation errors they have made, and how many. Furthermore, by using a learner corpus, they could compare their translation choices with those of their peers, as well as reference translations.

Penalty points can also be applied to errors of differing severity levels. Typically, a neutral error would receive a score of 0, a minor error receives 1, a major receives 5 and a critical error would receive 10, although the CAT tools allow the user to define the values of penalty points. This approach offers students the opportunity to familiarise themselves further with the TQA process as it occurs within the translation and localisation industry and to gauge the professional quality of their translation. If necessary, it is also possible to deselect the use of penalty points without negatively impacting on the usefulness of any TQA output reports produced by the CAT tool. The tutor may choose not to give penalty points, and instead just annotate the error types and TQA types when assessing the students' work. The final TQA report will still illustrate the tutor's specific feedback as shown in Figure 6 for formative

purpose. However, such an implementation loses the advantage of quantifying the students' performance in a longitudinal study.

Assessing student translations in this manner over the course of a semester or academic year may provide students with a meaningful record of their progress, as well as highlight areas for further improvement. On completion of the TQA process, SDL Trados Studio produces the chart shown below (Figure 7) where error types are colour-coded, providing a clear representation of the range of errors identified.

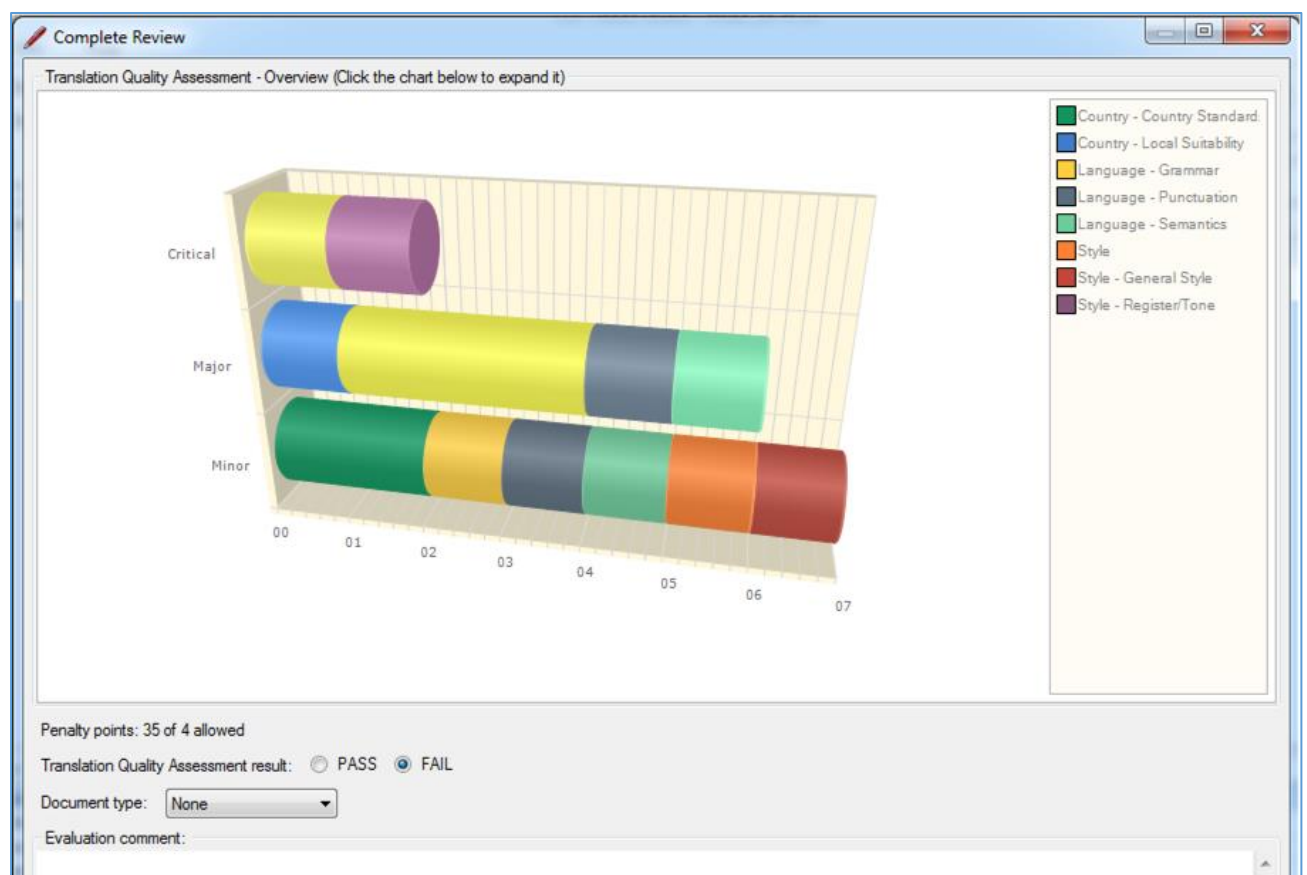


Figure 7. Example of a TQA chart

A comparison of students' output reports in this format could serve to highlight common errors within the group and provide a basis on which to plan further teaching or discussion. The

charts and reports may enable students to focus their attention on particular areas when it comes to revising their own translations.

However, one should acknowledge that, at least initially, integrating the use of a CAT tool with the tutor's regular assessment and feedback processes will require some time and effort. CAT tool integration would expect the tutor to dedicate approximately one day to learning to use the TQA function of the CAT tool and plan the structure of translation projects. During this stage, the tutor would also create translation memories to attach to each student project as a repository for the student's translations. Project and language resource settings, such as adjustments to the translation memory's segmentation rules, can then be saved for future re-use. This may appear slower in comparison to the traditional way of assessing through paper work or the easily-managed tool *MS Word*, but it also brings the advantages of consistency, accurate individual and group progress-tracking, and error-pattern identification already discussed.

CONCLUSION

Overall, using the TQA approach will initially require a greater time investment from the tutor compared to working with *MS Word's* Track Changes features or using another online platform with which markers may already be familiar with. However, it will bring clarity both of expression and of presentation of feedback and evaluation. The most popular current electronic marking practice already poses challenges where adding edits or comments can render both the translation and its annotations difficult to read: in *MS Word*, comments can mount up in the margins and become misaligned with the relevant portion of text. While other electronic systems may include comments in such a way as to overlay the student's translation, the Language Services Industry is fully committed to the use of CAT tools, and students and tutors alike should become familiar with industry standards and approaches in this area as early as possible in order to maximise their employability prospects in this sector.

Within the CAT tool environment, errors and/or their annotations can be clearly presented in a graphical or tabular form to be viewed or printed separately and returned to the student, which makes feedback more visually accessible and, ultimately, more meaningful for learner development and discussion. The short investment of time dedicated to mastering the CAT tool and to embedding them in updated translation training sessions will soon be outweighed by the benefit of knowing the students' individual and group progress at all times, as well as of being able to make data-driven choices regarding the training curriculum.

Addresses for correspondence:

Jun Yang (ml11j3y@leeds.ac.uk), Caroline Reiss (c.l.reiss@leeds.ac.uk), Dr. Alina Secară (a.secara@leeds.ac.uk), Dr. Dragoş Ciobanu (d.i.ciobanu@leeds.ac.uk)

BIBLIOGRAPHY

Depraetere, Ilse (ed.), *Perspectives on Translation Quality* (Berlin: De Gruyter, 2011)

Doherty, Stephen, 'Issues in human and automatic translation quality assessment', in *Human Issues in Translation Technology: The IATIS Yearbook*, ed. by Dorothy Kenny (London and New York: Routledge, 2016)

Reiss, Katharina, *Translation Criticism, the Potentials and Limitations: Categories and Criteria for Translation Quality Assessment* (London and New York: Routledge, 2000)

SAE J2450 Translation Quality Metric Task Force, *Quality Metric for Language Translation of Service Information* <<http://www.sae.org/standardsdev/j2450p1.htm>> [accessed 1 May 2017]

Samuelsson-Brown, Geoffrey, *Managing Translation Services* (Clevedon, UK ; Buffalo, NY: Multilingual Matters Ltd., 2006)

Secară, Alina, 'Translation Evaluation – a State of the Art Survey', in *Proceeding of the eCoLoRe/MeLLANGE Workshop, Leeds. Translation Studies Abstracts* (St. Jerome Publishing, 2005), pp. 39-44

TAUS, *TAUS DQF* <<https://www.taus.net/evaluate/dqf-tools#error-typology>> [accessed 1 May 2017]

Waddington, Christopher, 'Different Methods of Evaluating Student Translations: The Question of Validity', *Meta*, 462 (2001), 311–325