

BOOK REVIEW

Translation Quality Assessment: From Principles to Practice
Joss Moorkens, Sheila Castilho, Federico Gaspari and Stephen Doherty (Series Editor: Andy Way)
Springer, 2018 (1st edition). 287 pages.
ISBN: 978-3-319-91240-0.

Reviewed by **Rocío Caro Quintana** 
R.Caro@wlv.ac.uk
University of Wolverhampton, Spain

With the growth of digital content and the consequences of globalization, more content is published every day and it needs to be translated in order to make it accessible to people all over the world. This process is very simple and straightforward thanks to the implementation of Machine Translation (MT), which is the process of translating texts automatically with a computer software in a few seconds. Nevertheless, the quality of texts has to be checked to make them comprehensible, since the quality from MT is still far from perfect. *Translation Quality Assessment: From Principles to Practice*, edited by Joss Moorkens, Sheila Castilho, Federico Gaspari and Stephen Doherty (2018), deals with the different ways (automatic and manual) these translations can be evaluated. The volume covers how the field has changed throughout the decades (from 1978 until 2018), the different methods it can be applied, and some considerations for future Translation Quality Assessment applications.

Translation Quality Assessment (TQA) focuses on the product, not on the process of translation. In one way or another, it affects everyone in the translation process: students, educators, project managers, language service professional and translation scholars and researchers. Therefore, this book is addressed to translation students, lecturers, and researchers who are interested in learning about the industry, research about the topic, or even creating new methods or applications.

The volume consists of 11 chapters that are divided into the following 3 parts:

- Part 1: Scenarios for Translation Quality Assessment (Chapters 1– 4).

- Part 2: Developing Applications of Translation Quality Assessment (5–8).
- Part 3: Translation Quality Assessment in Practice (9–11).

The first chapter, written by the editors, is an introduction to Translation Quality Assessment (TQA) and the different methods it can be applied. As aforementioned, there are two main ways to assess the quality of translated texts: manually and automatically. The manual evaluation can be done in several ways; however, the most known approaches are Dynamic Quality Framework (DQF), Multidimensional Quality Metric (MQM) and the LISA QA (Localization Industry Standard Association Quality Assessment) Model. These approaches evaluate the final quality of a translation (for instance, checking if there are terminology errors or mistranslations). The automatic evaluation also has a variety of approaches, for instance, Bilingual Evaluation Understudy (BLEU, Papineni et al. 2002), Metric for Evaluation of Translation with Explicit Ordering (METEOR, Banerjee and Lavie, 2007), and Translation Error Rate (TER, Snover et al. 2006). These approaches measure the quality of a translated text comparing the final output with one or more reference translations. However, the editors claim that no approach or metric is sufficient to all scenarios and text types (literary translation, audiovisual translation, etc.) and these approaches may be changed by the users accordingly to meet their needs.

The next chapter (Chapter 2) introduces how translation is managed and its quality evaluated in the European Union (EU) institutions. The texts published by the EU are official texts that must be translated into many languages. Therefore, quality must be maintained in all the versions and the consistency must be maintained. There are a lot of quality checks and steps that texts must go through before publishing the official version. As there are many texts published and a lot of languages, the EU outsources a lot of these texts, which have to follow the Directorate General for Translation norms. The EU has created its Translation Memory, MT and a glossary database: IATE. The authors conclude by emphasising that these texts are essential to inform the citizens about the EU projects (especially in a time where the opposition to the EU and populist media with anti-EU agenda is very common) and this is achieved through quality translations.

Chapter 3 explores the new phenomenon of crowdsourcing, in this case, translation crowdsourcing, and how its quality can be measured. Crowdsourcing entails the outsourcing of translation tasks (translation, revision, post-editing) for free or for low rates to large crowds. The problem is evident: as there are a lot of participants it is hard to check the quality of the texts due to stylistic issues. Another problem has to do with the scope of the translation: just for gisting purposes or for dissemination. Moreover, the author posed the following question: “*Who is responsible for quality?*” (p.79). The author argues that, in certain cases, those responsible for the final text may be the Language Service Providers and, in others, the translators and revisers. Although it may be difficult to carry out this process due to the challenges it poses, it has been used in a lot of platforms, such as *Amara*, *Wikipedia* or *Facebook*.

The last chapter of the first part (chapter 4) discusses the lack of education in TQA in degrees and even on postgraduates’ translation courses. The authors advocate that it is crucial to teach translation students the quality evaluation methods to prepare them for the translation marketplace, especially since the use of MT is changing the role of translators into post-editors; thus their primary purpose will be to fix MT outputs.

The second part of the volume focuses on the development of approaches or metrics to assess the quality of translation. The first chapter of this part (chapter 5) analyses three different systems for TQA in depth: DQT, MQM and the harmonisation of the two, called the DQQ/MQM Error Typology. The author remarks that these systems were originally created to support translators with the reviewing process. The history of TQA is summarised, explaining that the first attempts to standardise the reviewing process were two standards: SAE J2450 and LISA QA Model. But as the author states, these approaches had important limitations: the low inter-annotator agreement and that they were not useful to all the possible translation scenarios or text types. As a result, DQF and MQM were created. Since 2015, their integration has become the preferred method.

Following this research, the following chapter (chapter 6) focuses on the analysis of the errors found in MT. While the previous approaches described in chapter 5 could be used for human or machine translation, the main focus in this chapter is on the error analysis of MT outputs. The evaluation of MT is usually carried out during the post-editing process; therefore, the author states that the classification of MT errors or post-editing

operations is performed to analyse the process, not translation errors. This error classification can be done manually, automatically or with a combination of the two. There is not, however, a standard system to evaluate MT output.

Similarly, Chapter 7 discusses how MT output is evaluated. The author describes different human and automatic evaluations and their problems. There are three main different human evaluation types: Typological evaluation, declarative evaluation and operational evaluation. Regarding automatic evaluation, the following problems challenge the translation assessment task: 1) they do not compare the translation with the source segment; 2) they usually work with only one reference translation; 3) there is not a “perfect translation”; and 4) the human translation (used as reference translation) could be incorrect. To conclude, the author affirms that novel metrics are needed to improve the outputs of MT engines.

The second part of the volume concludes with chapter 8, which briefly describes audiovisual translation (AVT). It delves into the main features of this field, particularly into spatial and temporal restrictions, which produces a different set of norms and standards than differ from other text types. The authors describe how the Computer-Assisted Tools and MT are also being implemented in AVT, especially to improve the productivity of translators and preserve the consistency of the texts (for instance, on TV shows). Quality is still difficult to assess on these texts as metrics such as NER (Net Error Rate, Romero-Fresco & Pérez, 2015) or WER (Word Error Rate, Nießen et al., 2000) are not useful due to the inherent characteristics of AVT mentioned above.

The third and last part of the book includes chapters which analyse TQA in practice in different fields. Chapter 9 delves into Translation Quality Estimation (TQE) which differs slightly from TQA since TQE does not require a reference translation to estimate how good a translation provided by an MT engine is. The goal of the authors in this paper is to successfully implement TQE methods that can distinguish between “good” and “bad” translations. If the translation is “good”, the MT output is post-edited; and if the output is deemed “bad”, it will be translated from scratch. While this chapter is of interest, it may not be accessible to everyone as it has a lot of terms and mathematical formulas that only people that are familiar with Computational Linguistics may understand.

Chapter 10 explores the use of MT in Academic Texts. English has become a *lingua franca* worldwide and many scholars have to use it in order to publish their work. However, in many cases, English is not their first language, and this could produce some problems with the quality of the texts. The authors posed the following questions: “*is [MT] actually a useful aid for academic writing and what impact it might have on the quality of the written product?*” (p. 238). To this end, the authors conducted some experiments where 10 participants were asked to write half a text in English, and the other half in their native language, and this was later translated to English with an MT engine. Then, the texts were revised. The results of these experiments showed that the revision of the texts written in English was shorter and the opinions of the translators were mixed in terms of efforts and whether they would use MT again for this purpose. The texts were also checked with an automatic grammar and style checker, but there were no major differences in terms of quality.

Finally, the last chapter of this part and this volume (chapter 11) goes into research the use of Neural Machine Translation (NMT) into Literary texts. The authors’ objective is to check whether literary texts can be translated correctly through NMT, namely novels from English into Catalan. To do this, they built a literary-adapted NMT system and compared the results with a Phrase-Based Statistical Machine Translation engine. The quality was checked with automatic metrics (BLEU) and manual evaluation and, as the authors expected, the results proved favourable to NMT.

All things considered, this volume is an excellent reference to learn and understand the different approaches and methods of TQA. It provides a very insightful look at the basics of TQA. The editors do not only present useful chapters about the basics of the theory, but they also present examples where these methods have been and could be applied. Hence, it will be very useful to scholars and translation students, whether they want to focus on research or the industry.

REFERENCES

- Banerjee, S., & Lavie, A. (2005). METEOR: An automatic metric for MT evaluation with improved correlation with human judgments. In *Proceedings of the ACL workshop on intrinsic and extrinsic evaluation measures for machine translation and/or summarization* (pp. 65-72). Michigan: Association for Computational Linguistics.
- Nießen S., Och, F.J., Leusch, G., & Ney, H. (2000). An evaluation tool for machine translation: fast evaluation for MT research. In *Proceedings of the second international conference on language resources and evaluation* (pp.39-45). Athens: European Language Resources Association (ELRA).
- Papineni, K., Salim, R., Todd, W. & Wei-Jing, Z. (2002). BLEU: a method for automatic evaluation of machine translation. In *Proceedings of the 40th annual meeting on association for computational linguistics* (pp.311-318). Philadelphia: Association for Computational Linguistics.
- Romero-Fresco P., & Pérez, J.M. (2015). Accuracy rate in live subtitling: the NER model. In J. Díaz Cintas & R. Baños Piñero (Eds.), *Audiovisual translation in a global context* (pp.28-50). London: Palgrave Macmillan.
- Snover, M., Dorr, B., Richard, S., Micciulla, L., & Makhoul, J. (2006). A study of translation edit rate with targeted human annotation. In *Proceedings of the 7th Conference of the Association for Machine Translation in the Americas* (pp.223-231). Cambridge: The Association for Machine Translation in the Americas.

Received: 18 November 2020

Accepted: 24 November 2020