



Reviews

MOORKENS, Joss; CASTILHO, Sheila; GASPARI, Federico; DOHERTY, Stephen (eds.). *Translation Quality Assessment: From Principles to Practice*. Machine Translation Series. Switzerland: Springer International Publishing, 2018, 292p. ISBN 978-3-319-91240-0 / ISBN 978-3-319-91241-7 (eBook) / <https://doi.org/10.1007/978-3-319-91241-7>.

Marileide Dias Esqueda¹

The goal of the book edited by Joss Moorkens, Sheila Castilho, Federico Gaspari, researchers at Dublin City University, located in Dublin, Ireland, and Stephen Doherty, researcher at The University of New South Wales, located in Sydney, in Australia, is to inform about the ways machine translation can be properly evaluated for its successful integration in the language services industry of today. Focusing on the product, with emphasis on Translation Quality Assessment (TQA), rather than on the process of translation, the editors affirm that the lines between human and machine have become blurred, and adaptability to change TQA is a key asset for translators and overall users embedded in a continuing growth of digital content. For Moorkens and colleagues, the landscape of machine translation affects not only all translation

1. Professora Associada do Programa de Pós-Graduação em Estudos Linguísticos e do Curso de Bacharelado em Tradução da Universidade Federal de Uberlândia. Uberlândia, Minas Gerais – Brasil. <https://orcid.org/0000-0002-6941-7926>. E-mail: marileide.esqueda@ufu.br.



This content is licensed under a Creative Commons Attribution License, which permits unrestricted use and distribution, provided the original author and source are credited.

stakeholders, project managers, and language service professionals, but also translation students, educators and researchers alike when it comes to assess quality in translation.

Translation Quality Assessment: From Principles to Practice is written in English, and it is part of the Springer Series titled *Machine Translation: Technologies and Applications*. The book is composed by an introduction, three main parts making a total of 11 chapters, containing an average of 30 pages each, and with a sum of 672 references. The oldest reference used by the authors dates back to 1948, and the latest one dates to 2018, covering almost 70 years of research in machine translation and the problem of translation quality assessment. One of the chapters has 145 references and the one with less reference works presents only four. According to the editors, the chapter which refers to this very small amount of references can be justified by the experience of the author in the field. Accordingly, the book draws upon the authors' experiences with machine translation technology and its applications to both pragmatic texts, which mostly pertain to the localization industry, and audiovisual and literary texts.

Under the title *Scenarios for Translation Quality Assessment*, the first part of the book has four chapters. Chapter 1, titled *Approaches to Human and Machine Translation Quality Assessment*, written by the abovementioned editors Sheila Castilho, Stephen Doherty, Federico Gaspari, and Joss Moorkens, provides an overview of the yet established and developing approaches to the definition and measurement of translation quality in human and machine translation workflows, especially considering TQA as a complex task, for both research and practice, and which involves a range of linguistic and extra-linguistic factors. The chapter reviews a wide range of approaches to TQA in the context of human translation (HT) within Translation Studies, and it then moves to examine MT quality and its assessment, highlighting the strengths and weaknesses of the various approaches and systems. Although TQA is recognized as a key topic in the area of Translation and Localization, the authors affirm that academia and industry differ greatly when it comes to defining and evaluating translation quality. While researchers and academics tend to focus on theoretical and pedagogic concerns related to translation quality, in most sectors of the industry TQA is broadly limited to the application

of somewhat arbitrary “one-size-fits-all”² error typology models that aim to give quantitative indicators of quality. The authors also highlight the fact that the classical strict separation between (professional) HT on the one hand, and MT on the other, seems to become increasingly indistinct today; one need only think about pre-editing, interactive MT within translation memory systems, and the related techniques and tools which are becoming progressively more efficient.

In the second chapter, titled *Translation Quality, Quality Management and Agency: Principles and Practice in the European Union Institutions*, Joanna Drugan, researcher at University of East Anglia, located in Norwich, United Kingdom, Ingemar Strandvik, and Erkka Vuorinen, researchers at the European Commission in Brussels, Belgium, explain that an important principal for the European Union (EU) is that there are no “original” texts, and that all language versions are equivalent and equally authentic, being consistency in translation strategies and in the approach to quality a critical issue related to ethics, power relations and professional values. Once the translation volumes are massive in the EU, which employs approximately 1,600 in-house translators and 700 other related staff to cope with some 73,000 language service requests and 2.2 million pages in 24 languages, the access to a fitness for purpose³ and reliable translation memory database is a paramount. The authors explain that Euramis, the common central translation memory database for EU translators should fulfil agreed quality requirements, so as to avoid “contamination” of future translation memories retrieved from the database. The authors explain that Euramis also serves as the basis for MT engines, which are now widely and increasingly used as support tools by EU translators, together with IATE (InterActive Terminology for Europe), and ELISE (European Institutions Linguistic Information Storage and Exchange). These tools support rapid exchange of information on individual translations or translation packages amongst translators working on the same file. In

2. The concept of “one-size-fits-all” refers to a degree of standardization in translation quality assessment with restrictive metrics.

3. The concept of “fitness for purpose” refers to a raw or post-edited machine translation that is considered “good enough” for a translation end-user. In Drugan, Strandvik, and Vuorinen (2018, p. 39)’s words, a translation is fit for purpose when it is suitable for its intended communicative use and satisfies the expressed or implied needs and expectations of customers, end-users or any other relevant stakeholders.

Drugan, Strandvik, and Vuorinen's claims, quality management policy empowers and motivates translators by giving them opportunities and responsibilities for taking action to ensure, maintain or improve quality (including through acting in different quality-related roles).

Crowdsourcing and Translation Quality: Novel Approaches in the Language Industry and Translation Studies is the title of the third chapter written by Miguel A. Jiménez-Crespo, researcher at Rutgers University, in New Brunswick, New Jersey, United States of America. For Jiménez-Crespo, crowdsourcing can be defined as outsourcing cognitive tasks and problem-solving activities requested for free, or for low rates, to large crowds of motivated participants. In the author's opinion, crowdsourcing is a technological revolution that allows large groups of people to cooperate at an unimaginable scale. He states that questionings related to crowdsourcing are concerned with possible level of translation quality that arises from this sort of collaborative and distinct translation approach. Nevertheless, as pointed out by Jiménez-Crespo, crowdsourcing can thus no longer be simply associated with non-professional level quality outcomes, as different sectors have extended these practices to include the entire spectrum of possible participants, from lay people to highly skilled professionals, depending on the initiatives. For the author, the problem of assessing the quality amongst these scenarios is the fact that collaborative translations are extremely open, creative, and dynamic, with a wide array of diverging approaches that defy categorization or uniform analysis. According to Jiménez-Crespo, three issues of particular interest in terms of the impact of translation crowdsourcing deserve a more detailed treatment: (i) the blind faith in the process or workflow-based approach to quality, (ii) the consolidation of the "fitness for purpose" approach, (iii) and the sharing of the responsibility for the final quality of the translation. Besides that, the author puts that current scenarios of collaborative translation replicate professional approaches, while others are mainly inspired by machine translation output. At the opposite end of the continuum, quality practices inspired by MT and language automation approaches have emerged. In general, the issue of translation quality in crowdsourcing is one of the most prolific areas of MT research. Since the late 1990s, the intersection of crowdsourcing and MT has been explored as an alternative to both improve and train MT systems. Jiménez-Crespo's chapter also discusses some particular issues related

to TQA measures, such as crowd selection, embedded translator testing, and community-building.

Stephen Doherty, researcher at The University of New South Wales, located in Sydney, Australia, and his already mentioned colleagues from Dublin City University, located in Dublin, Ireland, Joss Moorkens, Federico Gaspari, and Sheila Castilho, are the authors of the last chapter of the first part of the book. Titled *On Education and Training in Translation Quality Assessment*, Doherty and colleagues highlight at the very beginning of the chapter that TQA has been neglected for most stakeholders, translators, post-editors, reviewers, and academia, although some initiatives can recently be seen in the Brazilian context, with the issue on this topic published by the Journal *Letras & Letras* of the Federal University of Uberlandia, located in Uberlandia, Minas Gerais, Brazil, and edited by me and colleagues (ESQUEDA; ECHEVERRI, 2019). Doherty and colleagues revisit some of the key issues focusing on academic applications of TQA, affirming that the teaching of contemporary evaluation methodologies provides translation graduates with skills that we can already see prove valuable, who can move on to advisory roles in the language industry, and use their expertise to take on tasks such as workflow design, project preparation, and MT training data selection. In the authors' opinions, familiarity with TQA measures prepares translation graduates for the standards and quality expectations applied in the translation industry. For these reasons, they advocate adding TQA to translation curricula. It is worth mentioning that words such as education, students, and training, used more than 100 times in the book, reveal its pedagogical purposes.

The second part of the book, titled *Developing Applications of Translation Quality Assessment*, equally contains four chapters. The first one, titled *Metrics for Translation Quality Assessment: A Case for Standardising Error Typologies*, is written by Arle Lommel, researcher at Indiana University, located in Bloomington, Indiana, United States of America. With his long history of studies in the translation and localization industries, the author provides an overview of three systems designed for TQA: i) Multidimensional Quality Methods, developed by the German Research Center for Artificial Intelligence in Berlin, Germany (*Deutsches Forschungszentrum für Künstliche Intelligenz*, DFKI), ii) TAUS Dynamic Quality Framework Error

Typology, developed by the Amsterdam-based Translation Automation User Society (TAUS), and iii) the harmonization of the two, carried out as a collaborative effort by DFKI and TAUS within the EU-funded QT21 project. According to Lommel, these projects have been created during the years of 2012 to 2015, and even though they were perceived as competing projects, they were harmonized to create emerging TQA approaches. The author starts explaining the early difficulties in establishing systematic quality evaluation, mainly due to the fact that the scores were ultimately unverifiable, once the only link to the text was in the mind of the reviewer; it was unclear if the scores they generated correlated with audience or customer requirements. In Lommel's claims, it was in the 1990s that we witnessed two systematic efforts to address the *ad hoc* nature of translation quality assessment, the SAE J2450 (Society of Automotive Engineers), developed by General Motors to improve translation quality related to automotive documentation, with six error types, and LISA (Localization Industry Standards Association) Quality Assessment Model, with almost 20 categories of errors. After these first attempts, other groups began active work on translation quality assessment and developed other extensive translation error typologies for use in detailed analysis of human and machine translation. These error typology models have been constantly reorganized, mainly with the purpose of not defining a single metric, but rather a common vocabulary for declaring metrics.

Maja Popović, researcher at Humboldt University of Berlin, located in Berlin, Germany, is the writer of the following chapter titled *Error Classification and Analysis for Machine Translation Quality Assessment*. Popović describes the state-of-the-art of automatic, human, and computer-aided annotation of MT errors according to various error typologies as a way to compare MT systems or as a diagnostic tool for MT developers. Popović explains that human-annotated translations can give deep insight, but tend to suffer from low inter-annotator agreement, especially when error classes are not clearly defined. Popović explains why automatic tools struggle to accurately identify very specific error types and tend to confuse mistranslations, omissions, and additions. The author also discusses the evolution of MT error typologies, and describes experiments with different analysis methods (including the MQM (Multidimensional Quality Metrics), also described in detail in Lommel's chapter), such as attempts to employ linguistic check-points

to identify specific linguistic phenomena that cause particular problems. Popović's chapter also brings up the need for consolidating disparate MT evaluation typologies, in order to improve consistency. Some important conclusions of Popović are that widespread use of MQM for MT evaluation would allow subsets of a single unified metric to be used for both human and MT evaluation, and that errors involving different types of multi-word expressions are associated with high cognitive and temporal effort.

Quality Expectations of Machine Translation is the title of the following chapter written by Andy Way, researcher at Dublin City University, located in Dublin, Ireland. Way starts by affirming that machine translation needs to be measured rather than rejected out-of-hand merely as a knee-jerk reaction to the onset of this new technology. After posing some question related to how translations should be used in the future, and for how long will we need to consult them, the author describes appropriate use for MT based on the perishability of texts. Cognizant of this, Way updates the assessment of MT today in his contribution, explaining the "proper place" of MT, human and automatic evaluation metrics, and task-based MT evaluation. He addresses the weaknesses of automatic evaluation, and describes the changing nature of MT systems. Finally, he examines how MT is currently deployed, and considers associated questions of MT quality expectations and perception. Way also predicts machine translation continued use as a production tool alongside with translation memory systems.

Assessing Quality in Human- and Machine-Generated Subtitles and Captions is the title of the chapter written by Jan-Louis Kruger, researcher at Macquarie University, located in Sydney, Australia, and his colleague Stephen Doherty, researcher at The University of New South Wales, also located in Sydney, Australia. The authors affirm that the area of audiovisual translation (AVT) is becoming increasingly merged with language technologies, including computer-assisted translation tools, machine translation, automated subtitling and captioning software, and automatic speech recognition systems. In their claims, AVT has not been exempted from these technological developments with a wave of new tools becoming available, including manual, semi- and fully-automated subtitling and captioning software, speech-to-text systems, and machine

translation. In Kruger and Doherty's opinion, the interest in and applications of AVT have experienced a boom where traditional usage of subtitles for foreign movies and for the deaf and hard-of-hearing has been supplemented by new usage scenarios for language education, literacy, language learning, accessibility, clinical applications, and specialized and general education. According to the authors, unique to AVT are the spatial and temporal restrictions inherent in subtitling and captioning which often force the usage of indirect translation techniques (especially condensation, reformulation, and omission of linguistic elements), in order to achieve functional purposes, e.g. comprehension, education, and entertainment. These restrictions therefore severely limit the usage of translation choices and result in a general preference for approaches to translation quality that champion functionalism and pragmatic equivalence. As a result, they affirm that TQA in AVT is carried out in a diverse range of contexts, including in-house at the broadcaster, within LSPs, and by freelancers, leading to a variety of requirements for assessing quality in individual and ongoing projects (e.g. for a TV series) as well as once-off assessments (e.g. for a feature film or video game). As projects, client requirements, and genres vary substantially, these parameters are typically taken into account as their impact on expectations is significant.

In third part of book, titled *Translation Quality Assessment in Practice*, Lucia Specia, researcher at University of Sheffield, located in Sheffield, United Kingdom, and Kashif Shah, researcher at eBay Research, located in Sao Jose, California, United States of America, are the authors of the chapter *Machine Translation Quality Estimation: Applications and Future Perspectives*. In their opinion, predicting the quality of machine translation output is a topic that has been attracting significant attention. By automatically distinguishing bad from good quality translations, quality estimation has the potential to make MT more useful in a number of applications. In this chapter, Specia and Shah review various practical applications where quality estimation (QE) at sentence level has shown positive results: filtering low quality cases from post-editing, selecting the best MT system when multiple options are available, improving MT performance by selecting additional parallel data, and sampling for quality assurance by humans. They also discuss QE at other levels (word and document)

and general challenges in the field, as well as perspectives for novel directions and applications.

Machine Translation and Self-post-editing for Academic Writing Support: Quality Explorations is the title of the chapter written by Sharon O'Brien, researcher at Dublin City University, located in Dublin, Ireland, Michel Simard, researcher at the National Research Council, in Ottawa, Ontario, Canada, and Marie-Josée Goulet, researcher at Université du Québec en Outaouais, located in Gatineau, Québec, Canada. The authors affirm that non-native speakers of English are in disadvantage when it comes to internationally publish their studies in this language, and that machine translation is the tool that can make it possible, a similar discussion also conducted by Bowker and Buitrago-Ciro (2019). They explore the potential of using MT and self-post-editing as a second-language academic writing aid. The authors choose an interesting range of quality assessment measures, comparing participant perceptions, temporal effort (time spent), and revisions required when participants write an academic abstract in their first language, then machine translate and self-post-edit it, and when they write the abstract in English (their L2). O'Brien and her colleagues have compared these results using an automatic grammar- and style-checking tool, and found out that participants were generally impressed with the quality of MT output, but some had difficulty in finding the appropriate terminology in their native language, as they were habituated to using English language terms. The authors also demonstrate the potential for reducing the cognitive burden of authors when accessing international academic publishing via the current *lingua franca* of English.

In the last chapter of the book, titled *What level of quality can neural machine translation attain on literary text*, Antonio Toral, researcher at University of Groningen, located in Groningen, Netherlands, and Andy Way, researcher at Dublin City University, located in Dublin, Ireland, affirm that due to the rise of the new neural approach to machine translation and its promising performance on different text types, there is room to assess the quality it can attain on the greatest challenge for MT: literary texts. Toral and Way have built a literary-adapted NMT system for the English-to-Catalan translation direction and have evaluated it against a system pertaining to the previous

dominant paradigm in MT: statistical phrase-based MT (PBSMT). The researchers have trained MT systems, both in NMT and PBSMT approaches, on large amounts of literary texts (over 100 million words) and evaluated them on a set of 12 widely known novels spanning from the 1920s to the present day (2018, the year of the publication of their chapter). Their conclusions rely on the fact that NMT has resulted in an 11% relative improvement over PBSMT. They also concluded that a complementary human evaluation on three of the texts shows that between 17% and 34% of the translations produced by NMT (versus 8% and 20% with PBSMT) are perceived by native speakers of the target language to be of equivalent quality to translations produced by a professional human translator. Although Way advises the use of MT for highly perishable texts in his other contribution (the third chapter of the second part of the book), with Toral he investigates the results when that advice is completely disregarded, translating a non-perishable and difficult content type.

After finishing reading and studying all the 11 chapters of the book, its introduction and references, and considering that translation work processes are technologically and economically under accelerated change, one can affirm that this and other similar proposals are essential to continually update knowledge of translation technologies and translation quality assessment, not only to improve the work of professional translators and the services they provide to society, but also to situate translation teachers, students, and researchers on current developments and future trends involving the growing fusion between human and machine translation. This is why, in my view, including training reflections on how human and machine translation are currently united and how to assess their quality outputs is the most important feature of the book.

Although it has not been specially written to teachers, graduate or undergraduate students of translation, Moorkens and colleagues' book inspire translation teaching programs to optimize the training on machine translation in the translator's education, especially when the authors discuss the results of empirical surveys showing how human and machine translation converge and differ at the same time. Explored by the authors in most of the chapters, these surveys also show that evaluators of all sorts sometimes may not distinguish human from

machine outputs, and vice versa. Accordingly, a number of common error is important to distinguish, besides the fact that the characteristics of the source language text interfere in the output, such as word order problems when translating from Germanic to Romance languages.

The entire volume also attempts to disentangle and shed further light on crucial issues on translation quality assessment from multiple perspectives, including studies involving pragmatic texts from the localization industry and from the EU, to audiovisual and literary texts with pedagogical reflections.

To sum up, I will certainly use Moorkens and colleagues' book to (re)design translation technology courses I am responsible for at the Translation Program of Federal University of Uberlandia (Minas Gerais – Brazil). Alike any other translation teacher, I am searching for inspirations to enhance my students' instrumental competence (HURTADO ALBIR, 2017; ESQUEDA, 2020) which nowadays is pretty much steered by technological impacts like the ones promoted by machine translation engines and the quality of their outputs.

References

- BOWKER, Lynne; BUITRAGO CIRO, Jairo. *Machine Translation and Global Research: Towards Improved Machine Translation Literacy in the Scholarly Community*. United Kingdom: Emerald Publishing Limited, 2019.
- DRUGAN, Joanna; STRANDVIK, Ingemar; VUORINEN, Erkka. Translation Quality, Quality Management and Agency: Principles and Practice in the European Union Institutions. *In*: MOORKENS, Joss; CASTILHO, Sheila; GASPARI, Federico; DOHERTY, Stephen. (eds.). *Translation Quality Assessment: From Principles to Practice*. Machine Translation Series. Switzerland: Springer International Publishing, 2018.
- ESQUEDA, Marileide D. (ed.). *Ensino de Tradução: proposições didáticas à luz da competência tradutória*. Uberlândia: EDUFU, 2020.
- ESQUEDA, Marileide D.; ECHEVERRI, Álvaro. Avaliação de Traduções. *Letras & Letras*, v. 35, n. 2, p. i-xiii, 2019. Retrieved from: <http://www.seer.ufu.br/index.php/letraseletras/article/view/52967>. Latest access: Mar 27 2020. DOI: <https://doi.org/10.14393/LL63-v35n2-2019-0>.

HURTADO ALBIR, Amparo. (ed.). *Researching translation competence by PACTE group*. Amsterdam / Philadelphia: John Benjamins Publishing Company, 2017.

Recebido em: 30/03/2020

Aprovado em: 27/05/2020