OPEN ACCESS

**Scholars International Journal of Linguistics and Literature**

Abbreviated Key Title: Sch Int J Linguist Lit
ISSN 2616-8677 (Print) | ISSN 2617-3468 (Online)
Scholars Middle East Publishers, Dubai, United Arab Emirates
Journal homepage: https://saudijournals.com

**Review Article**

# Review of Translation Quality Assessment Research: Current Studies and Development

Leyang Wang[1], Qingyun Chen[2*]

[1]Associate Professor, Department of Foreign Studies, North China Electric Power University (Baoding), Baoding, Hebei, China
[2]Department of Foreign Studies, North China Electric Power University (Baoding), Baoding, Hebei, China

**DOI:** 10.36348/sijll.2023.v06i12.003 | **Received:** 09.11.2023 | **Accepted:** 12.12.2023 | **Published:** 15.12.2023

***Corresponding author:** Qingyun Chen
Department of Foreign Studies, North China Electric Power University (Baoding), Baoding, Hebei, China

## Abstract

The development of computer automatic assessment and corpus linguistics has provided a new research idea for translation quality assessment. This paper reviews the current situation of translation quality assessment at home and abroad and its application in translation teaching. Findings are that most of the current translation quality assessments are static, and the empirical research applied in translation teaching is not sufficient. Therefore, how to make the assessment model of translation quality keep pace with translation teachers and learners' needs is an unavoidable issue that cannot be ignored in the future translation quality assessment research.

**Keywords:** Translation quality assessment; quantitative research; static assessment.



## 1. INTRODUCTION

Translation quality assessment can be categorized into qualitative and quantitative assessments. Qualitative assessment evaluates the quality of translations based on established indicators but does not give an exact scoring figure, for example, the model proposed by House (1977) belongs to qualitative assessment, whereas quantitative assessment presents the results of the assessment in the form of figures. At present, the weak point of translation quality assessment research falls on the quantitative research, especially the research on the assessment of human translation and the assessment indicators of machine translation.

There are many ways to quantitatively assess the quality of translations. The most traditional one is the "scoring method", which refers to that, the evaluator gives a certain score according to the reference translation or the ideal translation in the evaluator's mind in the process of assessing the translation, and the final score of the whole translation is the sum of the scores of the different parts constituting the translation scoring model. Another popular quantitative method commonly used in the assessment of machine translation is the "statistical method". In this method, the evaluator tries to match the reference translation with the translator's text according to the similarities such as the structural and formal similarities between the reference translation and the translator's text manually or (semi-)automatically, and counts the results of the mutual matching degrees between them. The more matches there are, the better the quality of the translation is. This method is realized on the premise that there are reference translations and the reference translations are high-quality enough.

Whether it is the "scoring method" or the "statistical method", both of them have one thing in common, that is, they both need reference translations as scoring criteria. The standards to justify the reference translation, or determine whether a certain translation can become a reference translation are often still based on subjective judgment. The definition of "reference translation" brings the research question back to its original starting point, that is, "how to judge whether a translation is good or bad". The purpose of translation quality assessment is precisely to find a method that can objectively and efficiently assess the quality of various types of translations, that is, to evaluate the reliability and validity between the translations and the original text at the macro and micro levels, and to comment on the merits and demerits of translations. The assessment process needs to focus on the comparison between the original text and the translated text, and the assessment results of this comparison requires to be made according to certain standards and indicators. Therefore, the establishment of an objective, accurate and operable system of assessment standards and indicators is the key to building a reasonable translation quality assessment

model.

## 2. Current Status of Research on Indicators for Quantitative Assessment of the Translation Quality

The development of computer technology and corpus linguistics has made it possible to put forward some quantitative norms for the assessment of translation quality, that is to say, the assessment of translation quality is no longer based on a subjective impression, but on the question of whether or not it is possible to put forward some objective standards from the quantitative point of view.

### 2.1. Research on Indicators for Assessing the Translation Quality Based on Automatic Computerized Assessment

With the continuous progress of machine translation, the research on automatic assessment of machine translation has also made some progress. At abroad, N-gram similarity-based assessment, also known as N-tuple similarity-based assessment, follows the principle of "optimal similarity of reference translations", which aims to maximize the N-tuple match between machine translations and human reference translations. One of the most influential assessment method in this category is BLEU (Bilingual Evaluation Understudy) (Papineni *et al.*, 2002), which evaluates the quality of a translation by analyzing the extent to which N-tuples co-occur in the translation to be tested and the reference translation. Later on, a number of scholars have revised and improved the BLEU model, forming assessment indicators such as NIST (Doddington, 2002), METEOR (Banerjee & Lavie, 2005), MBLEU (Agarwal & Lavie, 2008), ChrF (Popović, 2015), and so on. Among them, NIST adds the concept of informativeness, giving different weights to content words and form words based on N-gram; METEOR introduces WordNet synonym database; MBLEU adds morphological restoration; and ChrF calculates text similarities at the lexical level.

Some other scholars have proposed the Translation Edit Rate (TER) as an index for assessing the quality of translations (Snover *et al.*, 2006). TER can be simply understood as the minimum edit distance required to modify a machine translation into a reference translation. The quality of translation is measured by calculating the minimum number of editing operations (including four types: word insertion, deletion, substitution and word block shifting), the lower the minimum number of editing operations, the more similar the machine translation is to the reference translation, and the higher the quality of translation is. In recent years, some improved models based on TER have also appeared, such as character (Wang *et al.*, 2016) and (Panja & Naskar, 2018). The former calculates the minimum edit distance between the machine translation and the reference translation at the lexical level, and the latter adds morphological restoration to the model.

### 2.2. Research on Corpus-Based Indicators for Assessing the Translation Quality

Assessment of translations, including judgments about the appropriate use of language, cannot be based on personal intuition or on certain specific examples. This kind of research requires empirical analysis of a large number of real texts, which can be realized by means of corpus. In recent years, the research on translation quality assessment based on corpus research is broadly divided into two categories: one is to use the corpus as an auxiliary tool to supplement translation quality assessment by human; the other is to use the corpus as the main basis for assessing the quality of translations, and to carry out data-driven assessment.

The corpus-based translation quality assessment method takes the linguistic feature data extracted from the corpus as the main basis. Initially, corpus-based assessment mainly assessed the quality of translations from a single linguistic feature, such as passive voice (Xiao *et al*., 2006), verb phrases (Loock, 2017), and so on. It later developed into the comparison of multilingual feature data to explore the methods of translation quality assessment from a more comprehensive perspective, such as through the comprehensive consideration of quantifiers, modifiers and past tense expressions. It was found that similarities between grammatical structures of the translation and that of the target language could reflect the overall quality situation of the translation (Rabadan *et al*., 2009); Leiva Rojo (2018) examined the subjects' phraseological competence based on corpus, and the results proved that there is a correlation between phraseological competence and translation quality. Most of the corpus-based translation quality assessment studies focus on the lexical level, with fewer studies on the sentence level, and only a few studies have added sentence-level measures, and in practice, the analysis indicators of the sentences are still ultimately returned to the words or phrases inside the sentences (Zhao Y *et al.*, 2015). However, there are still scholars who keep trying to improve the assessment system at the sentence level, De Sutter *et al*., (2017) increased the number of linguistic features to 20, which include type-token ratio, lexical density, average word length, and added average sentence length, etc., and compared the students' translation corpus with that of professional translators. Liu Yanmeng (2021) assessed the quality of English majors' translations from the point of view of acceptability, and added the assessment index of the proportion of complex sentences at the sentence level.

## 3. The Application of Quantitative Research on the Assessment of Translation Quality to the Teaching of Translation

From an interdisciplinary perspective, the development of automatic assessment of machine translations and corpus linguistics provides new ideas for the quality assessment of human translations to a certain extent. Some translation scholars develop computer-

assisted assessment models for human translations from the perspective of translation teaching, based on automatic computer scoring technology and corpus studies. Jiang Jinlin *et al.,* (2012) attempted to use the number of N-tuples matches and the number of manually selected word alignments between the translation and the reference translation as features of translation quality, and constructed several scoring regression models based on different sizes of training corpus. Jiang Jinlin (2013) established a multivariate linear regression model of translation quality based on more than twenty language form feature variables. Wang Lei *et al.,* (2009) proposed an assessment method based on the number of assessment point matches and similarities, considering the features of translation sentence length and the similarities of the translation. They used machine translation word alignment technology to align student translations with reference translations, and determined the rating level of student translations based on the alignment results. Similar to this method, Tian Yan (2011, 2015) developed an online translation scoring system, which calculates the matching degree of semantic similarity and sentence templates between the translation and the reference translation through using keyword matching and latent semantic methods from the perspectives of part-of-speech classification and sentence patterns. Wang Jinquan *et al.,* (2017) constructed a 'Chinese-to-English Automatic Scoring System' by utilizing natural language processing and other related techniques to extract textual variables, formal variables, and semantic variables. Wang Jinquan *et al.,* (2021) constructed a translation quality assessment framework based on lexical measurement features, and explored its predictability from six aspects: fluency, lexical diversity, lexical frequency breadth, lexical difficulty, lexical density, and lexical semantics. The specific quantitative indicators at these six levels are still evaluated based on the number of types and tokens, average word length, type/token ratio, and part-of-speech analysis. Compared with previous automatic assessment indicators based on the lexical level, Wang Jinquan *et al*., (2021) have a more comprehensive consideration of the lexical level, but evaluating translation quality only from the lexical level is not enough to represent the overall quality level of the translation. Obviously, other assessment indicators should be added. Based on the adequacy of the translation content and the fluency of the language, Yuan Yu (2016) introduced a set of 167 vocabulary, syntactic, semantic, and discourse level feature variables, including single-language features, bilingual features, and language model features, and added sentence-level assessment indicators such as average sentence length and number of sentences.

Most of the above computer-aided translation assessment models from the perspective of translation teaching utilize static translation quality assessment methods. In translation teaching, translation learners need to experience a long-term translation practice, and the quality of their translations will be affected by different factors, which are often neglected by static translation quality assessment. In order to create a more dynamic and multifunctional assessment model, researchers at Université Rennes2 (France) proposed the TRASILT three-dimensional translation quality assessment model, which starts from the translation industry-preferred perspective of error types and adds two dimensions: the impact of errors on translation quality and the severity of the impact. The TRASILT model is a conceptually clear, comprehensive, and dynamic three-dimensional quality assessment model. The three dimensions are (1) the type of error (error type); (2) the effect on quality; and (3) the degree of severity (degree of criticality) (Katell *et al*., 2017). The model was first applied in a modified version of the experiment "Translation Technology and Translation Students' Performance" to assess the quality of translations produced by master's degree students majoring in translation by using three translation technologies: translation memory, machine translation and speech recognition. This experiment is a preliminary exploration of the application of the TRASILT model to the teaching of translation majors, and it can be seen that its main advantage lies in its ability to help teachers and other evaluators to analyze the quality of students' translations in a more objective way, from the surface to the deep inside, which reflects its practicality in translation teaching: as the model takes into account the impact of translation methods on translation quality, the evaluators are able to analyze the differences in the quality of students' translations when they use different tools with the model results. The model helps the evaluator to analyze the differences in the quality of students' translations when using different tools; moreover, the model helps the evaluator to dig deeper into the causes of the students' translations while evaluating the quality of the students' translations in the light of the professional translation assessment criteria.

However, it is worth noticing that the number of translations evaluated in this experiment is limited, only 19 students' translations were evaluated in five weeks, and only 12 students' translations were finally analyzed. Besides that, the focus on the impact of translation methods on the translation quality considered by TRASILT can also indirectly suggest that future translation quality assessments can use more diverse assessment methods, including but not limited to, corpora, computer-automated assessment, AI, and so on.

## 4. CONCLUSION

The models and indicators developed for machine and human translation assessment are closely integrated with the development of computer technology. It is precisely because of this deep integration with the field of computer science that research on translation quality assessment has produced more results with the support of computer knowledge and technology. Corpus-based assessment methods

further explore the potential of corpora, improving the objectivity of translation quality assessment. However, it poses a significant challenge to obtain a large amount of data to be analyzed and evaluated. In addition, having too many assessment indicators requires researchers to select reasonable assessment indicators, which is also one of the difficulties of corpus-based translation quality assessment. Therefore, for quantitative research on translation quality assessment, it is necessary to explore more comprehensive assessment dimensions and select appropriate assessment indicators based on different corpora.

On the other hand, as far as the research and practice related to translation teaching and quality assessment are concerned, there are still the following shortcomings: Firstly, even though most of the current studies on translation quality assessment based on automatic computer assessment technology and corpus linguistics stems from translation teaching, they still return to theoretical research, and there is still a lack of enough empirical research on how to apply the theoretical progress to translation teaching in a maximized way. Secondly, most of the quantitative research on translation quality is still a kind of static assessment of translation, which is difficult for learners to realize their own shortcomings in long-term translation practice. Therefore, how to make the translation quality assessment model keep pace with the assessment needs of translation teachers and learners is a major issue that cannot be ignored in the future research on translation quality assessment.

## ACKNOWLEDGEMENTS

This research was supported by Hebei Education Department's Teaching Case Database Program for Professional Master Degree Courses (KCJSZ2022107).

## REFERENCES


- Agarwal, A., & Lavie, A. (2008, June). Meteor, m-bleu and m-ter: Evaluation metrics for high-correlation with human rankings of machine translation output. In Proceedings of the Third Workshop on Statistical Machine Translation (pp. 115-118).
- Banerjee, S., & Lavie, A. (2005, June). 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).
- Doddington, G. (2002, March). Automatic evaluation of machine translation quality using n-gram co-occurrence statistics. In Proceedings of the second international conference on Human Language Technology Research (pp. 138-145).
- Du, L., Wu, L. (2021). A New Translation Quality Assessment Model for Pedagogical Purposes: The Three-dimensional TRASILT Grid. Foreign Language Education & Research, 009(002), P.47-55.
- Jiang, J. L. (2013). An automatic approach to evaluating the linguistic quality of English-Chinese translations. *Modern Foreign Languages (Quarterly)*, (01), 85-91+110.
- Jiang, J. L., & Wen, Q. F. (2012). Computer Scoring Models for EFL Learners' English-Chinese Translation in Large-Scale Tests. *Technology Enhanced Foreign Language Education*, (02), 3-8.
- Juliane, H. (1977). A model for translation quality assessment. Gunter Na v.
- Lauscher, S. (2014). Translation quality assessment: Where can theory and practice meet?. In Evaluation and Translation (pp. 149-168). Routledge.
- Liu, Y. M. (2021). A corpus-based translation quality assessment method to quantitatively evaluating translation acceptability. *Corpus Linguistics*, (1), 11.
- Modarresi, G., & Ghoreyshi, S. V. (2018). Student-centred corrections of translations and translation accuracy: A case of BA translation students. *Iranian Journal of Translation Studies*, *15*(60).
- Panja, J., & Naskar, S. K. (2018, October). Iter: Improving translation edit rate through optimizable edit costs. In Proceedings of the Third Conference on Machine Translation: Shared Task Papers (pp. 746-750).
- Papineni, K., Roukos, S., Ward, T., & Zhu, W. J. (2002, July). Bleu: a method for automatic evaluation of machine translation. In Proceedings of the 40th annual meeting of the Association for Computational Linguistics (pp. 311-318).
- Popović, M. (2015, September). chrF: character n-gram F-score for automatic MT evaluation. In Proceedings of the tenth workshop on statistical machine translation (pp. 392-395).
- Snover, M., Dorr, B., Schwartz, R., 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: Technical Papers (pp. 223-231).
- Sutter, G. D., Cappelle, B., Orphée De Clercq, Loock, R., & Plevoets, K. (2018). Towards a corpus-based, statistical approach to translation quality: measuring and visualizing linguistic deviance in student translations. Department of Applied Linguistics, Translators and Interpreters, University of Antwerp.
- Tian, Y. (2011). Exploring the Practice of Online English-to-Chinese Automatic Scoring. *Chinese Translators Journal*, (02), 38-41.
- Waddington, C. (2001). Different methods of evaluating student translations: The question of validity. *Meta*, *46*(2), 311-325.
- Wang, J. S. (2019). A new direction for translation quality assessment: DQF dynamic quality assessment framework. *Chinese Science and*

*Technology Translators Journal*, (03), 27-29. doi: 10.16024/j.cnki.issn1002-0489.2019.03.009.

- Wang, J. Q., & Zhu, Z. Y. (2017). A Study on the Automated Assessment of Chinese-English Translation Competence. *Foreign Languages in China*, (02), 66-71. doi: 10.13564/j.cnki.issn.1672-9382.2017.02.009.
- Wang, J., Wan, X., & Dong, Z. (2018). Translation quality evaluation methods and their application in computer translation evaluation systems. *Chinese Translators Journal*, *39*(4), 6.
- Wang, J. Q., Yu, X., & Wu, W. (2021). A Study on Translation Quality Evaluation Based on Lexical Metric Features. *Chinese Translators Journal*, 42(5), 8.
- Wang, W., Peter, J. T., Rosendahl, H., & Ney, H. (2016, August). Character: Translation edit rate on character level. In Proceedings of the First Conference on Machine Translation: Volume 2, Shared Task Papers (pp. 505-510).
- Wu, G. J. (2007). Progress, meta-evaluation, and development direction of contemporary Chinese-English translation quality assessment models. *Foreign Languages Research*, (04), 73-79.
- Yang, Z. H. (2012). Quantitative Evaluation of Translation Quality: Patterns, Trends, and Implications. *Foreign Languages Research*, (06), 65-69+112. doi: 10.13978/j.cnki.wyyj.2012.06.013.
- Yuan, Y. (2016). A feature set for automatic human translation quality. *Foreign Language Teaching and Research (bimonthly)*, (05), 776-787+801.
- Zhang, X. J. (2007). A General Review of Studies on Quantized Translation Quality Assessment. *Foreign Languages Research*, (04), 80-84+112.
- Zhao, Y., & Liu, L. T. (2015). A preliminary exploration of the quantitative model for translation quality assessment: An empirical study based on learner corpora. *Jiangsu Foreign Language Teaching and Research*.