Corpus-Based Comparative Study of Google and Youdao Machine Translation Quality

Xinqi Dun[1], Shan Liu[1]

[1]School of Foreign Languages, Wuhan Textile University, Wuhan, Hubei, China

Keywords: Corpus, Machine translation quality, Political publicity text, Technical text

Abstract: In recent years, Machine Translation (MT) has seen great development as well as constant doubts in translation quality. This paper combines corpus research with machine translation text research and then compares translation quality of two machine translation engines, namely, Google and Youdao based on self-built Chinese-to-English translation text corpus mainly from political publicity texts and technical texts, finding the texts that the two translation engines are good at. The research results of this paper are of significant methodological value for the corpus-based comparative study of the translation quality of other translation engines in the future, and practically significant for people to choose a translation engine according to their needs, and shed light on the improvement or professional development of translation engines theoretically.

1. Introduction

With the wide application of machine translation(MT), its quality has gradually attracted scholars' attention. The translation quality assessment is the most critical index to evaluate MT, which plays an important role in improving its quality and the relevant algorithms, also helps to estimate the workload of post-edition. Based on previous studies, the thesis will present an all-around analysis of translation quality and then figure out the differences between Google System and Youdao System when dealing with political publicity text and technical text. It will be of importance for improving the MT quality. Besides, it will help us choose appropriate translation engine according to actual requirements and help other scholars do similar researches.

2. Literature Review

The existing researches have laid a good foundation. Since 1994, Stephen Minnis[1] has studied the quality evaluation of MT based on pre-edition. Subsequently, Jesus Gonzalez-Rubio(2013)[2], Pooja Gupta(2014)[3] and Kashif Shah(2015)[4] evaluated the quality of MT by different methods respectively. In recent years, Ilona Kostikova, 2019[5]; Noureldin Mohamed Abdelaal, 2020[6] began to pay attention to machine translation of different text types and Dragos Stefan Munteanu 2005[7] began to use corpus. Also, few comparative studies on the quality of MT (Haniyeh Sadeghi Azer (2015)[8]).

Chinese researchers studied the quality evaluation of MT from different perspectives. Zou Qili(2014)[9], Hu Wei (2015)[10], Xiong Bing(2016)[11], Tao Ying(2017)[12] et al., take specific text types or specific translation engines as objects to study MT quality. Subsequently, many scholars began to regard translation engines as variables and focus on comparative analysis of the quality of different translation engines (Yang Yamin, 2016[13]; Yin Jiao, 2017[14]; Yan Shenglin, 2018[15]; Sun Jingru, 2018[16]; Ji Yiwei, 2019[17]; Zhang Ziyue, 2019[18]). Meanwhile, few scholars take text type and translation engine as variables at the same time to study the quality of MT. Liao Menglin(2013)[19] studied the quality comparison between Systran and Google translation systems by taking four types of texts as examples. Han Na(2018)[20] selected three types of texts as test corpora and conducted a comparative study on Chinese-English translation by using Google and Youdao. Wu Lei(2019)[21] constructed a self-built comparable corpus to make a comparative analysis of the differences between translations and original English texts.

Overall, there are few studies on the evaluation of MT quality focusing on the comparison of the

  DOI: 10.25236/etmhs.2022.040

translation quality of different text types or translation engines, which are theoretically and practically significant.

3. Methodology

3.1 Evaluation Criterion

Considering the use of corpus tools and the subjective evaluation, this thesis imitates Wu Lei's[21] evaluation criteria. The translated versions of Google and Youdao will be compared with artificial translation from vocabulary richness, vocabulary density, vocabulary difficulty levels and average length of word and sentence respectively.

3.2 Samples of Texts

The 50 bilingual texts of political publicity materials are all from the Official Papers section of the China Network Bilingual Document Library, covering government documents, international speeches and policy documents. The words in each text is not less than 100, not more than 300.

The 50 scientific and technical bilingual texts are all from the bilingual abstracts of scientific and technological papers in the database of China National Knowledge Infrastructure, covering computer, agriculture, aerospace and so on. The words in each text is not less than 100, not more than 300.

3.3 Self-Built Corpora

The first corpus is the “Comparative English-language Corpus of Political Publicity Texts”, which consists of “Artificial English Translation of Political Publicity Texts” (AETPPT), “Google English Translation of Political Publicity Texts” (GETPPT), and “ Youdao English Translation of Political Publicity Texts” (YETPPT), with a total of 17,177 words.

The second corpus is the “ Comparative English-language Corpus of Technical Texts”, which is composed of “ Artificial English Translation of Technical Texts” (AETTT), “Google English Translation of Technical Texts” (GETTT), and “ Youdao English Translation of Technical Texts” (YETTT), with a total of 20,193 words.

4. Results and Analysis

4.1 Analysis of Corpus of Political Publicity Text

4.1.1 Vocabulary Richness

The total number of tokens is the most commonly used unit of measurement for corpus capacity. Type refers to any unique word form in the corpus text. In other words, a token that appears repeatedly in a text can only be counted as one type. The Type-token ratio refers to the ratio of types to tokens in the corpus. The lower the ratio, the more monotonous the vocabulary. For corpora of unequal sizes, it is more appropriate and rational to use Standardized Type-token ratio (STTR), the average Type-token ratio per thousand words. AETPPT, GETPPT, YETPPT are processed by Wordsmith's Wordlist function, and the result is that the tokens of AETPPT is 5,980, the types of it is 1,653, TTR is 27.75 and STTR is 45.86, while the four data of GEPPT is 5,664/1,471/26.09/42.66 and YEPPT is 5,533/1,481/26.89/43.34. Both the TTR and STTR of YETPPT is higher than that of GETPPT, which means that the vocabulary of GETPPT is not as rich as that of YETPPT. The result shows that the translation of Youdao is closer to the human translation.

4.1.2 Vocabulary Density

Lexical density is the ratio of the total number of real words to the total number of words in the corpus multiplied by 100%. A real word is a word with actual meaning. The greater the proportion of real words, the greater the content of textual information, and the greater the difficulty of understanding. CLAWS 5 is used firstly to perform part-of-speech coding on the pure text. Then

AntConc is used to retrieve the statistics of the coding results to obtain the number of different real words. The vocabulary density of AETPPT is 55.30%, while GETPPT is 56.43%, about 1% higher than that of AETPPT, and the vocabulary density of YETPPT is 55.51%, almost the same with AETPPT. There is no doubt that Youdao Translation is closer to human translation.

4.1.3 Vocabulary Difficulty Levels

A word may have many forms of inflectional change. The word itself and its inflection forms consist of a word family. The current legibility test mainly uses the word family as the reference standard. Antwordprofiler is used to get the result that in the first and second levels, AETPPT(511 and 126) and YETPPT(507 and 125) both higher than GETPPT(495 and 119). But when it comes to level three, these two versions(184 and 189) both own a lower rate than GETPPT, which shows that YETPPT is closer to AETPPT in this part.

4.1.4 Average Length of Word and Sentence

Word length refers to the length of the vocabulary in alphabetical units in the corpus, which can affect the efficiency of readers' understanding and information recall ability. Danielson & Bryan proposed (1977)[22] said that the difficulty of English text can be judged based on the average number of words in each unit space and the average number of words in a sentence. Three versions are processed by Wordsmith's Wordlist function and the results is that Youdao translation(5.17 and 24.35) is closer to human translation(5.20 and 23.45) than GETPPT(5.33 and 25.15).

4.2 Analysis of Corpus of Technical Text

4.2.1 Vocabulary Richness

The same method and theory in 4.1.1 are used in this part. In terms of TTR, YETTT(20.40) and AETTT(23.53) differ by about 3.1, while GETTT(22.25) and AETTT differ by about 1.3. In terms of STTR, YETTT(37.88) and AETTT(41.20) differ by around 3.3, while GETTT(39.92) and AETTT differ by around 1.3. It is clearly that Google Translate is closer to artificial translation than Youdao.

4.2.2 Vocabulary Density

Steps in 3.1.3 are repeated to get the following data. The vocabulary density of AETTT is 61.97%, while GETTT is 63.22% and YETTT is 63.34%. Though the difference of between GETTT and YETTT is not so apparent, but GETTT is relatively closer to AETTT.

4.2.3 Vocabulary Difficulty Levels

We also use Antwordprofiler to get the data of vocabulary difficulty. The word family number of AETTT at the first level is 402, the second level is 115 and the third level is 234.The data YETTT is 389, 105 and 219, and GETTT is 389, 112 and 221. It is clear that the vocabulary difficulty of the Google Translation System is closer to the artificial translated version.

4.2.4 Average Length of Word and Sentence

Wordsmith's Wordlist function is used to get the data. The mean word length of Google translation is 5.84 and the mean sentence length is 30.75, which is closer to human translation(5.79 and 25.02) than YETTT(5.92 and 33.30) and the difference is quite apparent in the latter data.

5. Summary

First of all, with manual translation as the standard,this article compares the Google translation of political publicity text and technical text with Youdao translation from four aspects. Data shows that Youdao is superior to Google when translating political publicity texts, which may relate to the Chinese-centric language of Youdao’s system. But, for the translation of technical texts, the Google system does better than Youdao. Secondly, we can see that although it is still inferior to manual translation, the quality of MT has been improved a lot because of the unremitting efforts.

There are still some limitations: Firstly, the results are only based on the corpus constructed by randomly selected texts, which is limited. Secondly, only Google and Youdao, two most widely used translation engines currently, are selected for the study. Finally, the quality of MT is improving with time, so the results may change accordingly. After a period of time, the conclusions of this study may no longer be applicable. Future research should be based on a larger and richer corpus and take other translation engines as objects to conduct comparative analysis from different perspectives. And More factors should be considered and more advanced research methods should be adopted to achieve more accurate, more objective and more comprehensive results.

Acknowledgements

This research was financially supported by the Postgraduate Innovation Fund Project of Wuhan Textile University (Grant NO. 52300101).

References

[1] Minnis, S. A simple and practical method for evaluating machine translation quality[J]. Mach Translat, 1994.

[2] Jesús González-Rubio,J.Ramón Navarro-Cerdán,Francisco Casacuberta. Dimensionality reduction methods for machine translation quality estimation[J]. Machine Translation,2013,27(3-4).

[3] Pooja Gupta, Nisheeth Joshi, Iti Mathur. Quality Estimation of Machine Translation Outputs Through Stemming[J]. International Journal on Computational Science & Applications,2014,4(3).

[4] Kashif Shah,Trevor Cohn,Lucia Specia. A Bayesian non-linear method for feature selection in machine translation quality estimation[J]. Machine Translation,2015,29(2).

[5] Ilona Kostikova; Alona Shevchenko; Liudmyla Holubnycha.Use of Machine Translation Technology For Understanding Scientific and Technical Texts[J].Journal of Theoretical and Applied Information Technology,2019,97(4).

[6] Noureldin Mohamed Abdelaal, Abdulkhaliq Alazzawie. Machine Translation: The Case of Arabic- English Translation of News Texts. 2020, 10(4).

[7] Dragos S. M.; Daniel M. Improving Machine Translation Performance by Exploiting Non-Parallel Corpora[J]. Computational Linguistics,2005.

[8] AZER, H. S.; AGHAYI, M. B. An Evaluation of Output Quality of Machine Translation (Padideh Software vs. Google Translate) [J]. Advances in Language and Literary Studies, [s.l.],2015.

[9] Zou Qili. Machine Translation Error Analysis for Management Text [D]. Lanzhou University,2014.

[10] Hu Wei. Quality Improvement Strategy of Baidu Translation [J]. China Science and Technology Information,2015(Z4).

[11] Xiong Bing, et al. A study on the Lexical Features of Tourist Texts in English Translation [J]. Journal of central China normal university (humanities and social sciences edition),2016,55(05).

[12] Tao Ying. Google Translation Quality Analysis based on Corpus [J]. Journal of Liaoning Teachers college (Social Science Edition), 2017(06).

[13] Yang Yamin. A Comparative Study on Quality of Online Machine Translation System [D]. Yangzhou University,2016.

[14] Yin Jiao. Machine Translation in Chinese language [C] / / Proceedings of the Ieee International Conference on Computational Linguistics,2017.

[15] Yan Shenglin. Comparison of Machine Translation Software -- Taking Legal Text as an Example [J]. Modern commerce & trade industry,2019,40(01).

[16] Sun Jingru, Wan Li. A Comparison of Inverted Sentences in Machine Translation Software [J]. Electronic Technology and Software Engineering,2018(11).

[17] Ji Yiwei. Quality Evaluation of Chinese Passive Voice in Machine Translation: A Case Study of Google Translate and Youdao Translation [J]. English Square, 2019 (03).

[18] Zhang Ziyue. A Comparative Study on Translation Quality of Youdao Dictionary, Baidu Translation and Kingsoft Iciba [J]. Comparative studies in cultural innovation, 2019,3 (30).

[19] Liao Menglin. Quality Evaluation of Systran and Google Translation system: A Case study of Four types of Text translation [J]. Journal of hubei normal university (philosophy and social sciences edition),2013,33(03).

[20] Han Na. A new approach to translation based on neural machine translation [J]. Journal of shanxi university of energy,2018,31(05).

[21] Wu Lei. Analysis of the English Translation of Garment Exhibition Website Publicity Text based on Comparable Corpus. Journal of Donghua University (social science edition), 2019,19(03).

[22] Fry,E.B.Elementary Reading Instruction.New York:McGraw-Hill,1977.