So, H. J. et al. (Eds.) (2020). Proceedings of the 28th International Conference on Computers in Education. Asia-Pacific Society for Computers in Education

A Comparative Study on the Translation Quality of Specialized and General Machine Translation Outputs

Xiaohan YU[1], Guangming YU[1], Zhiyang, LIU[2] & Chunping ZHENG[1*]
[1]*Beijing University of Posts and Telecommunications, China*
[2]*China National Institute of Standardization, China*
[*]zhengchunping@bupt.edu.cn

Abstract: This study compares the translation quality among two general machine translation systems, Google Translate and Youdao Translation, and one specialized machine translation system, Smart Translation System for Standards (STSS). This pilot study explored the effectiveness of STSS and its translation errors, providing suggestions for language learners, educators, and users of machine translation.

Keywords: Machine Translation, Machine Translation Output, Evaluation Methods, BLEU, Translation Quality

1. Introduction

Apart from general translation platforms like Google Translate, research institutes and enterprises have developed domain-specific MTs to translate texts in a specified technical domain. Unlike general translation systems, however, there are few studies investigated the effectiveness of specified MTs when translating domain-specific documents. The comparative study explores the translation quality and errors of translation output of three machine translation platforms, namely, STSS, Google Translate and Youdao Translation. Based on our research, we provided suggestions for translation quality evaluation and language learning.

It is guided by two research questions:

1. What are the differences in performance of three platforms when translating documents about standards?

2. What are the differences of errors in translation outputs between general and specified machine translation platforms?

2. Research Methodology

2.1.1 The Translation Platforms in the Study and Source Texts

STSS was developed by China National Institute of Standardization (CNIS) as a specified MT for conducting bilingual translation (Chinese and English) of documents about national and international standards. Google Translate is one of the most popular MTs globally while Youdao Translation is a China-based and one of the most used MTs in China. The little difference is that STSS has its own translation memory (TM) which makes it a domain-specific translation engine.

We selected six documents which were included in the STSS platform TM (in-TM), and six not included in the TM (out of-TM). In-TM source texts consist of 6408 Chinese characters and 4532 English words while out of-TM texts consist of 6203 Chinese characters and 5107 English words. That's all the source text used in the study.

2.1.2 Automatic Evaluation (AE) and Human Evaluation (HE)

We applied BLEU (Bilingual Evaluation Understudy) as our automatic evaluation tool. The human references of translation, which is required by BLEU metric, are all well-translated and published documents provided by CNIS. Quality Evaluation Code for Localization Translation and Desktop Publishing is a human evaluation standard issued by the Translators Association of China. Based on the standard, we classified the translation errors. After AE and HE, we are able to compare the performance of the three MT platforms. The comparison was carried out by two professional translators separately and cross-checked to ensure the accuracy.

3. Results and Conclusion

3.1.1 AE-based Comparison

Table 1 and 2 show the BLEU scores in two scenarios. For in-TM texts, STSS platform presents a higher result than Youdao Translation and Google Translator for two-way translation. For out of-TM texts, the three platforms do not show significant difference in scores.

Table 1 *BLEU Scores of In-TM Texts*

Source Text	Language	Yuodao Translation	Google Translate	STSS Platform
In-TM	Chinese-English	0.27	0.24	0.31
	English-Chinese	0.17	0.16	0.29

Table 2 *BLEU Scores of Out of-TM Texts*

Source Text	Language	Yuodao Translation	Google Translate	STSS Platform
Out of-TM	Chinese-English	0.25	0.23	0.27
	English-Chinese	0.16	0.18	0.19

3.1.2 HE-based Comparison

Table 3 and 4 show the error distribution of the two scenarios. Term error is the most frequent type for all three MTs followed by expression error. STSS did much better for in-TM texts while nearly the same as the other.

Table 3 *The Error Distribution of In-TM Texts*

Statistics on the Number of Translation Errors				
Error Distribution		Google Translate	Youdao Translation	STSS Platform
Slight Errors	Expression	12	15	8
	Language Style	1	1	4
	Grammar	2	3	3
	Specific Symbol	1	1	5
Minor Errors	Term	28	26	23
Major Errors	Mistranslation	2	3	2

Table 4 *The Error Distribution of Out of-TM Texts*

Statistics on the Number of Translation Errors				
Error Distribution		Google Translate	Youdao Translation	STSS Platform
Slight Errors	Expression	14	18	14
	Language Style	5	3	4
	Grammar	3	4	5
	Specific Symbol	2	2	6

Minor Errors	Term	30	30	29
Major Errors	Mistranslation	3	4	3

STSS performed generally acceptable and significantly better translation output than general machine translation platforms for the in-TM texts. However, its translation quality was nearly the same as that of general MTs for the out of-TM texts. This may due to the translation memory which makes the STSS sharper for translation of standards. We can safely say that the scale and quality of parallel corpus of translation memory is critical to improve the effectiveness of a domain-specific MT. It would be helpful for language learners to choose a domain-specific MT when learning terminologies of the technical field. From the perspective of error distribution, post-editors should pay special attention to term translation when using MT platforms. As we only used two methods for quality evaluation, the assessment needs to be refined by more comprehensive metrics.

Acknowledgements

This research is supported by the Fundamental Research Funds for the Central Universities (2019XD-A04) and the Project of Discipline Innovation and Advancement (PODIA)-Foreign Language Education Studies at Beijing Foreign Studies University (2020SYLZDXM011).

References

Akiko, S. (2019). Unintended consequences of translation technologies: from project managers' perspectives. *Perspectives-Studies in Translation Theory and Practice, 27*(1), 58-73.

Ben, S. (2017). Machine translation and Welsh: Analyzing free Statistical Machine Translation for the professional translation of an under-researched language pair. *Journal of Specialized Translation,* (28), 317-344.

Bentivoglil, L. & Bisazza, A. (2016). Neural versus Phrase-Based Machine Translation Quality: a Case Study. *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, *23*, 257-267.

Castilho, S. & O'Brien, S. (2017). Acceptability of machine-translated content: A multi-language evaluation by translators and end-users. *Linguistica Antverpiensia,* New Series: Themes in Translation Studies*, 16*, 120–136.

Christopher, D. M. (2017). Translators and machine translation: knowledge and skills gaps in translator pedagogy. Interpreter and Translator Trainer, *11*(4), 280-293.

Deng, D. & Xue, N. W. (2017). Translation divergences in Chinese-English machine translation: An empirical investigation. *Computational Linguistics, 43*(3), 521-565.

Doherty, S. (2016). The impact of translation technologies on the process and product of translation. *International Journal of Communication,* 10, 947-969.

Eustace, J., Wang, X.Y., & Cui, Y.Z. (2015). Overlapping community detection using neighborhood ratio matrix. *Physical A: Statistical Mechanics and Its Applications*, *421*:510-521.

Francois, Y. (2019). Quality estimation for machine translation. *Computational Linguistics*, 45(2), 391-394.

Helmut, S. & Alexander, F. (2015). The Operation Sequence Model Combining N-Gram-Based and Phrase-Based Statistical Machine Translation. *Computational Linguistics, 41*(2), 185-214

Kenny, D. & Doherty, S. (2014). Statistical machine translation in the translation curriculum: Overcoming obstacles and empowering translators. *The Interpreter and Translator Trainer*, (2), 276-294.

Koehn, P. & Knowles, R. (2017). Six Challenges for Neural Machine Translation. *Proceedings of the First Workshop on Neural Machine Translation*, *18*, 28-39.

Sanjika, H. & Stephan, V. (2016). Extracting parallel phrases from comparable data for machine translation. *Natural Language Engineering*, *22*(4), 549-573.