International Journal of Advanced Culture Technology Vol.6 No.4 165-171 (2018)
DOI https://doi.org /10.17703//IJACT2018.6.4.165

IJACT 18-4-21

A Quality Comparison of English Translations of Korean Literature between Human Translation and Post-Editing

IL-JAE LEE

Department of English Language and Literature, Kwangwoon University, Korea
Translation Industry Research Center of Korea, Kwangwoon University, Korea
ijlee@kw.ac.kr

Abstract

As the artificial intelligence (AI) plays a crucial role in machine translation (MT) which has loomed large as a new translation paradigm, concerns have also arisen if MT can produce a quality product as human translation (HT) can. In fact, several MT experimental studies report cases in which the MT product called post-editing (PE) as equally as HT or often superior ([1],[2],[6]). As motivated from those studies on translation quality between HT and PE, this study set up an experimental situation in which Korean literature was translated into English, comparatively, by 3 translators and 3 post-editors. Afterwards, a group of 3 other Koreans checked for accuracy of HT and PE; a group of 3 English native speakers scored for fluency of HT and PE. The findings are (1) HT took the translation time, at least, twice longer than PE. (2) Both HT and PE produced similar error types, and Mistranslation and Omission were the major errors for accuracy and Grammar for fluency. (3) HT turned to be inferior to PE for both accuracy and fluency.

Keywords: *Artificial Intelligence, Machine Translation, Post-Editing, Human Translation, Translation Quality, Translation Memory, EverTran, VisualTran*

1. Introduction

Machine translation (MT) had been an area in which the artificial intelligence (AI) could not chip in much with its technological advancement. But in this era of the Fourth Industrial Revolution with the emergence of a new computational approach to translation, the artificial neural network loomed large to shed light on MT. The multinational technology company *Google* launched a statistical MT program *Google Translate* in 2006, but software and computing power could not sufficiently process even a proper noun or an inverted structure. In November 2016, *Google* introduced a Neural Machine Translation engine, which translates the entire sentences at a time, not piece by piece. It determines a meaningful context implied in the phrasal or clausal level, not word level, to decode the most relevant translation. It then reorganizes the translated text and attunes the meaning to be closer to human translation (HT)—with proper grammar. MT productivity and consistency have been increased [8],[9], while the its quality is comparable to that of genuine HT from scratch [5],[6]. Unlike the previous method of HT, MT engages the translators as post-editors in comprehending the source

Manuscript Received: November 22, 2018 / Revised: November 28, 2018 / Accepted: December 1, 2018
Author: ijlee@kw.ac.kr
Tel: +82-2-940-5429
Department of English Language and Literature, Kwangwoon University, Korea

text and evaluating and revising the MT target text, the process of which is called post-editing (PE) [4]. What became less crucial in the process of translation is once sophisticated syntactic and lexical knowledge of language pairs—the properties of conventional translators. Although commercial and general genres of translation have been replaced with MT among most of the translation services companies in the world [7], the public assumes the superiority of HT, for example, in the translation of literary genres. Under such an academic circumstance related to HT and PE, the research questions that intrigued me in the English translation of Korean novels were the following. Would HT and PE produce similar or distinct kinds of errors? Would native Korean speakers and native English speakers evaluate HT and PE the same or distinctly in terms of accuracy and fluency? And which method between HT and PE produces more or less errors?

2. Related works

Fiederer and O'Brien reported a judgment test of HT and PE [2]. A section of software user manual in English was extracted to be translated into German. A total of 30 research sentences were selected for 3 human translators and 3 post-editors, and a total of 11 evaluators evaluated the translation outputs. They found that PE output was judged to be higher in the parameters of *clarity* and *accuracy*, while HT was judged to be higher of *style*. Nevertheless, the evaluators selected HT to be their 'favorite' sentences. Plitt and Masselot employed 12 professional localization translators for a two-day test [8]. Each 3 of them performed HT first and then PE from English to French, Italian, German, and Spanish. They found that, first, PE necessitated less keyboard time and pause time than HT; second, HT contained a greater number of errors than PE for all 4 language pairs; lastly, PE, when trained and used on Autodesk data, allows translators to substantially increase their productivity. O'Curran had each 2 translators of Brazilian Portuguese, French, and Spanish to translate a real software User Assistance content in English [6]. They first translated the content from scratch (HT) and later post-edited the content generated by a statistical MT system with translation memories and glossaries. Dedicated third-party reviewers performed the blind quality assessment, meaning that they were not aware if the output was HT or PE. The evaluation was based on accuracy, language, terminology, style, country, and functional criteria. She found that PE had fewer errors than HT, and more errors were found in categories like Punctuation, Tags, and Style. Daems et al. used 10 master's students of translation and 13 professional translators as the participants, who were all Dutch natives [1]. A total of 8 English newspaper articles with a 150/160-word passage at various levels of complexity were selected and translated to Dutch. PE was faster in time per word by a second compared to HT (roughly 5 sec. vs. 6 sec.). Most attention on screen went to the target text for both HT and PE but the difference in attention was greater for PE than HT. Processing of the source text during PE required fewer fixations per word than for HT. Not for HT, but for PE, the students spent less time with the source text but more time with the target text. There is no difference in overall quality of error type between HT an PE, but errors were less common for PE. Lastly, more than professional translators, but it was student translators who seemed to consider PE the least tiring method of translation.

3. Present research

3.1 Research questions

A broad research question intrigued me from the related works if similar translation behaviors can be expected in the English translations of Korean literature between HT and PE: Which translation method between HT and PE can be subject to greater accuracy and fluency? And, are there differences in error types between HT and PE? What follows the next is a brief description of translation contents and participant profiles for the present research.

3.2 Content profiles and adjustments

Table 1. Genre and source texts

	Content	Title
1	Classical literature	'The Rabbit Story'
2	Early modern literature	'The Camellia'
3	Modern literature	'US-China War'
4	Control	'Why do Koreans eat like that?'

After deciding the genre and source texts as in Table 1, adjustments were made as did in related studies [1]. Idiomatic and dialectal expressions were changed to modern-day ones. Each source text consisted of 2 to 3 sentences—each one with no more than 20 words in Korean (*cf.* [3]). In so doing, some functional words, repetitions, connectives that were thought to be unimportant were discarded. Hence, the texts were treated to the level of average high-school students and were read by 3 native Koreans with high academic levels.

3.3 Translator (HT) and post-editor (PE) profiles

Table 2. Translator (HT) profiles

	Age	BA	MA	Specialty	Exp. (yr)
HT1	53	4-year, English	theology, abroad	religion, literature	4
HT2	32	4-year, English	n/a	web-tune	3
HT3	43	4-year, Physics	English, domestic	literature, medical	4

Table 3. Post-editor (PE) profiles

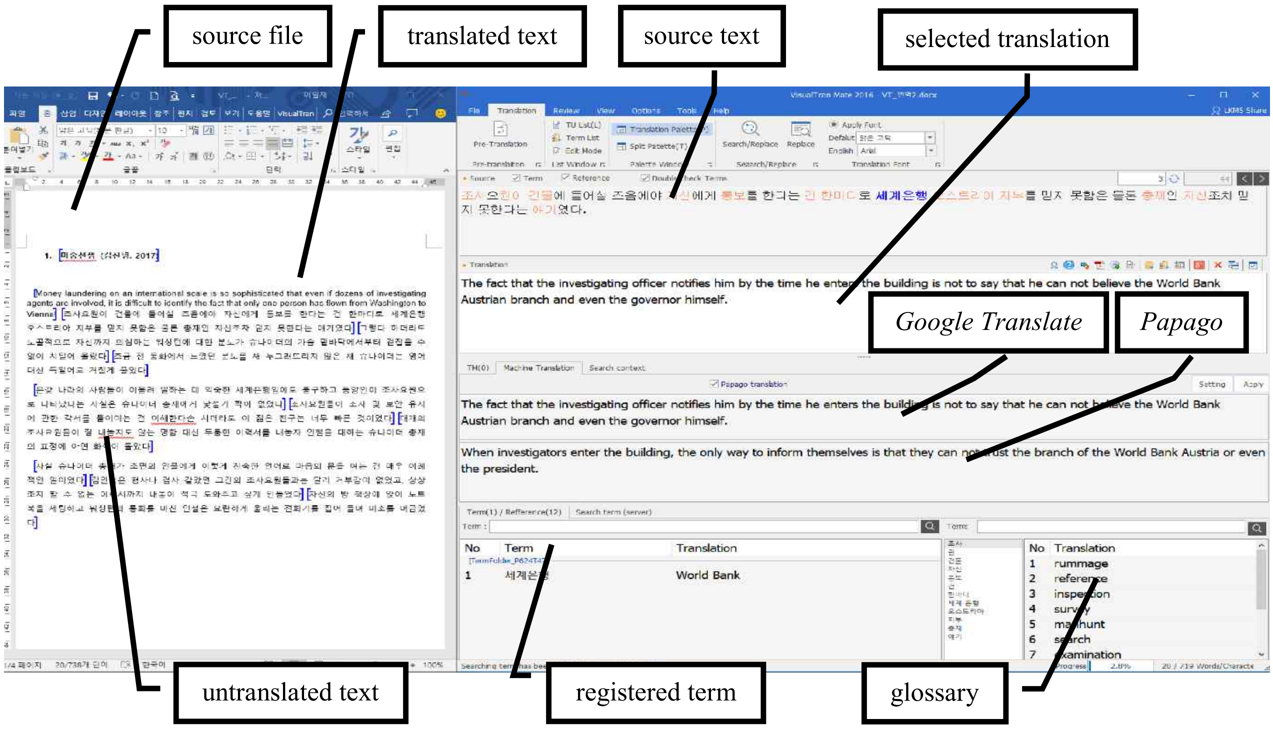
	Age	BA	MA	Specialty	Exp. (yr)
PE1	41	4-year, English	Trans & Inter, AUS	education, contents	4
PE2	40	4-year, USA	n/a	manual	3
PE3	50	4-year, Italian	n/a	finance, contract	4

3.4 MT platform: *VisualTran* of *EverTran Co., Ltd.*[1]

The MT platform used for this research is an interactive computer-aided MT tool called *VisualTran* exclusively programmed by a Korean IT company *EverTran Co. Ltd.* It is interactive in the sense that the post-editor can view not only the source and target texts, but she can interactively decide the better translations between the ones decoded by *Google Translate* and *Papago*. Moreover, the post-editor can refer to the translation memories for the expressions having done by her or others.

[1] *EverTran Co., Ltd.* is a Korean company that provides high-quality translation services including large-volume high quality machine translation, using its self-developed multi-language IT-based translation solution, *VisualTran*. *VisualTran* is, in essence, the first computer-aided translation tool developed in Korea by *EverTran Co. Ltd.*, using differentiated and refined Translation Memory accumulation to ensure consistency of terms and quality across large scale projects.

3.5 Translation processes



1. 미술선생 (김선영, 2017)

Money laundering on an international scale is so sophisticated that even if dozens of investigating agents are involved, it is difficult to identify the fact that only one person has flown from Washington to Vienna

조사요원이 건물에 들어설 즈음에야 자신에게 통보를 한다는 건 한마디로 세계은행 오스트리아 지부를 믿지 못함은 물론 총재인 자신조차 믿지 못한다는 얘기였다.

The fact that the investigating officer notifies him by the time he enters the building is not to say that he can not believe the World Bank Austrian branch and even the governor himself.

☑ Papago translation

The fact that the investigating officer notifies him by the time he enters the building is not to say that he can not believe the World Bank Austrian branch and even the governor himself.

When investigators enter the building, the only way to inform themselves is that they can not trust the branch of the World Bank Austria or even the president.

No	Term	Translation
1	세계은행	World Bank

No	Translation
1	rummage
2	reference
3	inspection
4	survey
5	manhunt
6	search
7	examination

Figure 1. Screenshot of *VisualTran* workbench

Translators were asked with the following guideline:

- ✓ Spend no more than 1 hour per source text.
- ✓ Do not take a long recess during a text translation.
- ✓ Do not use any sort of translation program.
- ✓ Dictionaries and web-searching are allowed.

Post-editors were asked with the following guidelines:

- ✓ Spend no more than 1 hour per source text.
- ✓ Do not take a long recess during a text translation.
- ✓ Use either *Google Translate* or *Papago*[2] as shown on *VisualTran*.
- ✓ Dictionaries and web-searching are allowed.

All guidelines were reportedly respected, except for the time limits. The time of 1 hour given to translate 1 source text for translators was not enough, and they most of the time spent longer than 1 hour. On the other hand, post-editors believed 1 hour was enough, and they spent roughly 30 minutes per text. The translated research texts collected were a total of 72 sentences out of 24 texts: 4 texts by each 3 translators and 3 post-editors. Ordering effect was applied in doing away with related biases, and also for quality evaluation.

[2] It is an AI-based translation and interpretation program developed by a leading IT company, NAVER Labs, in Korea.

3.6 Evaluation criteria

Adopted from the Dynamic Quality Framework (DQF) of 2015 and Quality Translation 21 (QT21) of 2015, the selected issue types for MT diagnostic evaluation were *accuracy* and *fluency* for the Korean-English translation output, while the left-outs were *locale convention* and *terminology*. *Accuracy* evaluates how acceptable and accurate the meaning of PE output is in comparison with the meaning of source text, while *fluency* evaluates how readable and fluent the grammar and vocabulary of PE output are so that the end-user does not discern that she is reading a translated text. Refer to the evaluation criteria used in this research in Table 4 below. A 4-point Likert-scale was used for each sentence: excellent, good, fair, and poor.

Table 4. Evaluation criteria

	Error type	Description
Accuracy	Addition	Addition of new information in the target text
	Mistranslation	Wrong translation
	Omission	Omission of the source text
Fluency	Grammar	Grammatical and structural error
	Register	Stylistic and pragmatic error
	Inconsistency	Semantic error
	Spelling	Wrong spelling

3.7 Evaluator profiles for *accuracy* and *fluency*

Table 5: Accuracy evaluators

	Age	L1	Education	Exp. (yr)
ACC1	45	Korean	- BA in English, USA	12
ACC2	34	Korean	- MA in Trans & Inter, Korea	7
ACC3	37	Korean	- BA, Australia, MA in Trans & Inter, Korea	8

Table 6. Fluency evaluators

	Age	L1	Nationality	Education
FLU1	39	English	Canadian	- BA in engineering, Canada
FLU2	37	English	American	- BA in arts, USA
FLU3	38	English	South African	- BA in education, S. Africa

4. Results

4.1 Accuracy comparison

Three adult Koreans (see Table 5) who had extensive knowledge in English and experiences in translation evaluated the accuracy parameter. Each reviewed 162 sentences (HT and PE, each 81 sentences).

Table 7. Accuracy comparison of HT and PE

	Addition		Mistranslation		Omission	
	HT	PE	HT	PE	HT	PE
ACC1	9	1	39	8	24	20
ACC2	8	1	55	69	22	28
ACC3	23	1	63	41	25	13
Total	40 (14.9%)	3 (1.6%)	157 (58.6%)	118 (64.8%)	71 (26.5%)	61 (33.5%)

According to a descriptive calculation, both HT and PE revealed substantial errors. A greater number of error occurrences for accuracy took place with HT (268=40+157+71) than with PE (182=3+118+61). This finding stands on par with the results reported in [1] and [6]. Nevertheless, there are qualitative similarities and also differences in error types. *Mistranslation* was turned out to be the major type of error for both HT (58.6%) and PE (64.8%), and *Omission* following the next (26.5% and 33.5%, respectively). A peculiar behavior that was detected and undocumented is that HT marked relatively a high occurrence of *Addition* (14.9%), which is literally absent for PE (1.3%).

4.2 Fluency comparison

Three native speakers of English (See Table 6) with different nationalities were recruited to evaluate the fluency parameter. Just like the accuracy evaluation, each reviewed 162 sentences (HT and PE, each 81 sentences) on a 4-point Likert-scale for each error type. The judgment was quantified as follows: excellent (0 point), good (1 point), fair (2 points), and poor (3 points). The higher the score is, the less fluent the sentence would be.

Table 8. Fluency comparison of HT and PE

	Grammar		Register		Inconsistency		Spelling	
	HT	PE	HT	PE	HT	PE	HT	PE
FLU1	39	24	14	8	2	1	4	3
FLU2	32	38	0	0	15	10	2	2
FLU3	117	54	43	27	0	3	3	2
Total	188 (69.4%)	116 (67.4%)	57 (21.0%)	35 (20.3%)	17 (6.3%)	14 (8.1%)	9 (3.3%)	7 (4.1%)

For fluency, HT (271=188+57+17+9) was evaluated to be less fluent or less readable than PE (172=116+35+14+7). Moreover, just like accuracy, the score variation for error type is the same for HT and PE, and it is *Grammar* that turned out to be the major aspect hindering the reading for HT (69.4%) and PE (67.4%) as well.

5. Summary

Although it can be humbly assumed that HT can produce superior qualities of translation than machine-translated PE, the otherwise experimental cases have been reported in which PE performs as equally as HT or often superior [1],[2],[6],[8]. As motivated from those studies on translation quality between HT and PE, this study set up an experimental situation in which Korean literature was translated into English, comparatively, by 3 translators and 3 post-editors. Afterwards, a group of 3 other Koreans checked for accuracy of HT and PE; a group of 3 English native speakers scored for fluency of HT and PE. The findings are (1) HT took the translation time, at least, twice longer than PE. (2) Both HT and PE produced similar error types, and *Mistranslation* and *Omission* were the major errors for accuracy and *Grammar* for fluency. (3) HT turned to be inferior to PE for both accuracy and fluency.

Acknowledgement

I dearly appreciate *EverTran Co. Ltd.* for helping me out understand and perform *VisualTran* and for recruiting translators and evaluators. Lastly, present findings have been discussed in one of my master's students in her thesis written in Korean.

§ The present research has been conducted by the Research Grant of Kwangwoon University in 2017.

References

[1] J. Daems, S. Vandepitte, R. Hartsuiker, and L. Macken, "Translation Methods and Experience: A Comparative Analysis of Human Translation and Post-editing with Students and Professional Translators," *Meta: Journal des Traducteurs/Meta: Translators' Journal*, Vol. 62, No. 2, pp. 245-270, 2017.

[2] R. Fiederer and S. O'Brien, "Quality and Machine Translation: A Realistic Objective?" *The Journal of Specialised Translation*, Vol. 11, pp. 52-74, 2009.

[3] S.-H. Han, "Feature of CAT Tool-Based Collaborative Translations: Sentence Length and Cohesion," *Interpreting and Translation Studies*, Vol. 20, No. 4, pp. 167-188, 2016. (In Korean)

[4] ISO 18587:2017, Translation services – Post-editing of machine translation. Geneva: International Organization for Standardization, 2017.

[5] I. Lacruz, "Cognitive Effort in Translation, Editing, and Post-Editing," In J. W. Schwieter and A. Ferreira, eds., *The Handbook of Translation and Cognition*, pp. 386-401. New Jersey, USA: John Wiley & Sons, Inc., 2017.

[6] E. O'Curran, "Translation Quality in Post-Edited versus Human-Translated Segments: A Case Study," In S. O'Brien, M. Simard, and L. Specia, eds., *Proceedings of the Third Workshop on Post-Editing Technology and Practice (WPTP-3)*, pp. 113-118, Vancouver, Canada, October 22-26, 2014.

[7] J.-Y. Park, "Language Technologies & the T&I Industry's Future," *Interpreting and Translation Studies*, Vol. 21, No. 1, pp. 137-168, 2017. (In Korean)

[8] M. Plitt and F. Masselot, "A Productivity Test of Statistical Machine Translation Post-Editing in a Typical Localisation Context," *The Prague Bulletin of Mathematical Linguistics*, Vol. 93, pp. 7-16, 2010.

[9] Ö. Temizöz, "Postediting Machine Translation Output: Subject-Matter Experts versus Professional Translators," *Perspectives: Studies in Translatology*, Vol. 24, No. 4, pp. 646-665, 2016.