Machine Translation of Korean Statutes Examined from the Perspective of Quality and Productivity

Jieun Lee jieun.lee@ewha.ac.kr
Hyoeun Choi hyoeun.choi@ewha.ac.kr
Graduate School of Translation & Interpretation, Ewha Womans University, Seoul, 03760, ROK

Abstract

Because machine translation (MT) still falls short of human parity, human intervention is needed to ensure quality translation. The existing literature indicates that machine translation post-editing (MTPE) generally enhances translation productivity, but the question of quality remains for domain-specific texts (e.g. Aranberri et al., 2014; Jia et al., 2022; Kim et al., 2019; Lee, 2021a,b). Although legal translation is considered as one of the most complex specialist translation domains, because of the demand surge for legal translation, MT has been utilized to some extent for documents of less importance (Roberts, 2022). Given that little research has examined the productivity and quality of MT and MTPE in Korean-English legal translation, we sought to examine the productivity and quality of MT and MTPE of Korean of statutes, using DeepL, a neural machine translation engine which has recently started the Korean language service. This paper presents the preliminary findings from a research project that investigated DeepL MT quality and the quality and productivity of MTPE outputs and human translations by seven professional translators.

1. Introduction

Human intervention, namely post-editing, is needed to ensure quality translation when machine translation (MT) is used. The existing literature indicates that compared to from-scratch translation, namely human translation (HT), machine translation post-editing (MTPE) enhances translation productivity, but the question of domain-specific MT and MTPE quality still remains to be answered (Aranberri et al., 2014; Jia et al., 2022; Kim et al., 2019; Lee, 2021a). In the case of patent translation, MT quality is still less than adequate (Choi et al., 2023; Lee and Choi, 2022; Tsai 2017) and MTPE may not be efficient to produce HT quality output. Legal translation is also considered as a specialist domain, but facing the demand surge for legal translation, the translation industry and the language service providers have resorted to MT for documents of less importance (Roberts, 2022). However, translation of legal texts, such as statutes and contracts, requires accuracy and generally been reserved for HT. Therefore, MTPE can be effective only when MT quality is good enough, thus not needing heavy post-editing. Otherwise, it would be simply more time-consuming and inefficient to post-edit

Proceedings of Machine Translation Summit XIX, Vol. 2: Users Track, pages 143–151
September 4–8, 2023, Macau SAR, China.

inadequate MT than translate from scratch. Given that little research has examined the productivity and quality of legal translation via MTPE, this paper aims to examine the performance of a general use neural machine translation (NMT) engine, DeepL, and MTPE productivity and quality of Korean statutes in comparison with HT.

The quality of post-editing results may vary depending on the ability, text type, and difficulty level, such as post-editor's experience in translation and post-editing training, native language, and subject knowledge (Kim, 2022b; Lee, 2021a; Seo and Kim, 2020). As such, it may be argued that translation competence is a necessary condition for post-editing competence (Lee, 2021b: 190). Because previous research on MTPE often engaged student translators rather than professional translators or post-editors, who were likely to lack translation and post-editing experiences and skills, we will examine professional translators' HT and MTPE products from the perspective of productivity and quality to find out if legal translation based on MTPE can be a productive alternative without sacrificing quality.

Productivity is not just a matter of time, and the effort required for post-editing can be analyzed in terms of technical, temporal, and cognitive efforts (Krings, 2001; Snover et al., 2006). Technical effort refers to the frequency and amount of correction, whereas temporal effort refers to the time required for task completion, and cognitive effort means the effort required to identify and correct errors in MT (Krings, 2001). Translation Edit Rate (TER) is often used to measure technical effort, such as inserting, deleting, replacing, and moving. However, it can only infer productivity through the modification rate, and thus not an absolute indicator of productivity. Further, because HT cannot calculate TER, it cannot be directly compared with HT (Snover et al., 2006). Although MTPE appears to have similar quality and improved productivity compared to HT, some studies suggest that it requires more cognitive effort than from-scratch translation (Guerberof Arenas, 2020: 347; Krings, 2001; 320; O'Brien, 2017). It is said that the cognitive load is large in correcting syntax problems, word order, mistranslation, and idiomatic mistranslations (Daems et al., 2015, 2017; Teminkova, 2010; Popović et al., 2014). In the following section, we will review the relevant literature, focusing on MT and MTPE involving the Korean language and legal texts.

2. Literature Review

2.1. MTPE studies

Recent MTPE studies generally indicate that MTPE is superior to HT in terms of speed while maintaining similar translation quality (Cadwell et al., 2016; Guerberof Arenas, 2009; Kim et al., 2019; Kim, 2022a,b; Lee and Kim, 2022; Seo and Kim, 2020). Jia et al. (2019) compared and analyzed the results of HT and Google NMT post-editing for two types of text in the English-Chinese direction. They investigated 30 postgraduate translation students' translation and MTPE processes and output quality, using two types of texts— two in specific fields such as patient description materials and dishwasher manuals, and the other two general texts (beverage brand promotion brochures). They noted that for the domain-specific texts, the participants completed MTPE a little faster than HT, and that cognitive efforts decreased in MTPE of both domain-specific and general text types. As for quality, MTPE output quality showed an equal level of accuracy and fluency as HT. In Jia et al. (2019), four evaluators—two professional translators and two Ph.D. students majoring in translation—evaluated a total of 154 sentences. The quality in terms of accuracy was as follows: The average score of the domain-specific text MT was 2.76, 3.2 for MTPE results, and 3.29 for HT, revealing statistically significant differences between the different modes of translation (Jia et al., 2019: 74). In the case of general texts, the difference in evaluation scores was narrower, with post-editing averaging at 3.19 and HT at 3.16, slightly lower than MTPE. Regarding fluency, the domain-specific text MT result scored an average of 2.88, MTPE result was significantly higher at 3.25, and HT was the highest at 3.31. On the other hand, in the case of general texts, the results of HT and PE were 3.19 and 3.33, respectively, showing no statistically significant difference.

MTPE studies involving Korean also demonstrated that MTPE was productive compared to HT. Kim et al. (2019) looked at the time effort required for correction along with the correction rate as an index of productivity. They examined the productivity of light post-editing of the English-Korean MT generated by three general-use NMT engines, namely Google Translate, Papago and Kakao*i*. The participants translated and post-edited without a time limit, and worked on IT manuals, which apparently has contributed to the enhancement of productivity. Based on the number of processed words per minute in HT and MTPE, MTPE productivity increased at least 78% higher than HT, and the Translation Edit Rate (TER) was 3% for Google NMT, 5.9% for Papago, and 5.4% for Kakao*i* (Kim et al., 2019: 65). Except for Kim et al. (2019), the other Korean MTPE studies investigated full post-editing.

Lee and Lee (2021) compared the quality of Korean-English news text MTPE and HT by undergraduate translation students, and found that the productivity as well as the quality of MTPE was better than those of HT. Lee (2021a) examined the difference between HT and MTPE productivity and cognitive processes by having five professional translators translate and post-edit technical texts (IT manuals) in the Korean-English direction. The participants translated around 100 word-long texts in two ways, HT and MTPE, in 10-minute-time frames, respectively, which were subject to evaluation by two experts. He noted that MTPE productivity was higher than HT productivity in terms of task completion time, and that despite individual variations, MTPE quality was not inferior to that of HT. Another study by Lee (2021b), which was based on nine translators' HT and MTPE, confirmed approximately 34% increase of MTPE productivity measured in terms of time, compared to HT. Both Lee (2021a,b) also demonstrated that the MTPE output quality was not inferior to HT in the case of technical texts.

Kim (2022a) also confirmed MTPE's productivity by analyzing Korean-English HT and MTPE outputs provided by 13 postgraduate translation students, comparing technical effort and search effort. She analyzed the English translation of Korean economic text of 76 words for HT and MTPE of 360 word-long economic text generated by Google NMT.

Lee and Kim (2022) analyzed the quality and productivity of English-Korean MTPE based on TER, word throughput, and output quality evaluation. Fourteen undergraduate and graduate students, who had received PE training through regular university courses or special lectures, participated in their research. The task completion time for HT or post-editing was set to 20 minutes (Lee and Kim, 2022: 128-129). Similarly, MTPE demonstrated a productivity advantage of nearly 70% compared to HT, comparable to the productivity of light post-editing in their earlier work (Kim et al., 2019). In addition, the quality of the post-editing results was not inferior to that of HT (Lee and Kim, 2022: 134, 140).

In summary, the existing Korean MTPE productivity-related research pointed to an average of more than 30% productivity enhancement compared to HT. However, the texts used in the previous research were mainly manuals, news and economic texts. Few Korean researchers investigated MT or MTPE of legal texts. In this paper, we will present the preliminary findings from our research on the Korean-English MT and MTPE, focusing on the temporal and technical efforts as productivity indicators, and the output quality relative to HT.

2.2. Legal Text MTPE

There is a lacuna in the literature on MTPE of legal texts, particularly in Korean and English language combinations. When it comes to English and Spanish, Killman and Rodríguez-Castro

(2022) analyzed 26 translators' from-scratch translations of English-Spanish legal texts and their Google Translate (SMT) post-editing. They reported that post-editing was superior in quality and productivity. The found that MTPE reduced time by 16%, with an average of 56.6 minutes for MTPE and 67.1 minutes for HT. Human evaluations revealed that MTPE contained fewer translation errors, with an average of 14.5 errors, whereas HT contained an average of 22.9 errors (Killman and Rodríguez-Castro, 2022: 63). The results did not indicate that the participants' translation experience and translation training made any significant difference in quality and time.

There are few studies that investigated MT or MTPE of Korean legal texts. For instance, Lee (2022) undertook an evaluation of Korean-English contract MTs produced by Google NMT and a legal domain-specific NMT. He found that the domain-specific NMT produced a better quality output than the generic NMT. While Lee (2022) examined MT quality of Korean-English contract translation, Lee and Choi (2023) examined the quality of English translations of Korean statutes generated by three NMT engines. They analyzed the output quality of two general-use NMT engines, Papago and Google Translate, and a legal domain-specific NTM engine, Otran, drawing on human and automatic evaluations. Four experienced legal translators evaluated the output quality, using a five-point rating scale—0 to 4—based the criteria of accuracy, fluency, and terminology. The human evaluation resulted in an average of 2.8 for Google NMT and Papago, and 3.5 for Otran (Lee and Choi, 2023:83). BLEU score for each NMT recorded 0.421, 0.395 and 0.585 (Lee and Choi, 2023: 82). As the figures suggest, the legal domain-specific NMT outperformed the other two generic NMT engines in both human and automatic evaluations. However, there was still a substantial gap between HT and MT, requiring human intervention in order to produce legal translation of publishable quality.

Considering that the largest Korean government-funded legal translation service provider, Korean Legal Research Institute's Translation Center, has sought to improve the efficiency and quality of its legal translation services by introducing computer-assisted-translation and translation automation (Lee, 2021), the current research is expected to throw some light on the prospect of MTPE in the legal translation domain from the perspective of productivity and quality. Further, DeepL recently launched its Korean language services in 2023, and merits scholarly investigation of its performance in the legal domain. Against this backdrop, this paper aims to investigate the DeepL Korean-English legal MT quality and the MTPE productivity and quality in comparison with HT.

3. The Study

The current research was designed to answer the folllowing research questions:

1. What is the quality of DeepL Korean-English legal translation according to human and automatic evaluations?
2. What is the quality of DeepL MTPE and HT according to human evaluation?
3. What is the productivity of MTPE in comparison with HT in terms of temporal and technical efforts?

3.1. Research Methods

For this study, we used extracts from two Korean statutes for source texts. Text 1 (241 Korean words/21 segments) was extracted from the Act on the Punishment of Stalking Crimes and was used for HT. Text 2 (242 Korean words/24 segnents) from the Act on Registration and Inspection of Water Leisure Devices was translated by DeepL and the MT output was post-edited by seven professional translators. The text difficulty was about the same in terms of readability (Flesch Kincaid Score was 31.04 and 24.42 respectively) and lexical density (52.33 and 57.93 respectively).

We recruited seven professional translators who had at least three years' experience in legal translation after a MA in Translation. The participants were requested to translate one text from scratch and post-edit the other to produce a HT quality output referring to the translation and post-editing guidelines we had provided. They were given 90 minutes each for HT and MTPE with 10 minutes' break in between. We measured their translation and post-editing time and calculated the words per minute for productivity analysis, and analysed the evaluation results and errors analyses provided by evaluators, who had assessed the MTPE and HT outputs according to the evaluation criteria of accuracy, fluency, and terminology.

We engaged three evaluators, two veteran professional legal translators and a translator trainer who are familiar with MTPE. They were requested to evaluate not only the raw MT output but also the seven HT outputs of Text 1 and seven MTPE outputs of Text 2, using a five-point rating scale (zero to four points), and also annotate errors.

To assess the MT quality through automatic evaluation, we checked the BLEU score of DeepL MT output.

3.2. Results

Both human and automatic evaluations revealed that DeepL MT quality was not bad. The raw MT output received an average of 3.15 points (segment average scores) in human evaluation, and the BLEU score recorded 29.92.[1] As requested, the evaluators identified errors in MT, MTPE and HT outputs. Twenty six errors were identified in the raw MT output by at least two of the three evaluators (68 out of 450 English words), giving a TER score of 15.1%.

Regarding the text difficulty level, the translator participants considered both texts not too easy nor too difficult for legal translation. There was a consensus among the participants. However, there was some disagreement among the evaluators because they agreed on the medium difficulty level of Text 1, but they were divided in their opinion on Text 2, each selecting high, medium, and low difficulty.

Participant	HT (min.)	MTPE (min.)	HT word/min.	MTPE word/min.	Productivity growth (%)
P1	90	90	2.7	2.7	0
P2	80	74	3.0	3.3	9
P3	80	53	3.0	4.6	52
P4	59	47	4.1	5.2	26
P5	65	65	3.7	3.7	0
P6	83	52	2.9	4.7	60
P7	90	90	2.7	2.7	0
Average	**78.1**	**57.3**	**3.2**	**3.8**	**21**

Table 1. Comparison of HT and MTPE productivity

In terms of temporal productivity, as Table 1 shows, the seven participants (P1 to P7) tended to spend less time on post-editing than on from scratch translation, HT. The average time they spent on MTPE recorded 57.3 minutes whereas the average time spent on HT was 78.1 minutes. A t-test revealed a statistically significant differences at the 90% confidence level ($p<0.1$). As shown in Table 1, the number of words processed during MTPE was higher

[1] The results indicated that DeepL outperformed other general-use NMTs, such as Google Translate and Papago, but its performance was inferior to that of a domain-specific NMT, Otran in Korean-English legal translations (see Lee and Choi, 2023).

than that of HT, confirming enhanced productivity observed in the MTPE research discussed in this paper (e.g. Kim, 2022a,b; Kim et al., 2019; Lee and Kim, 2022). The average productivity growth of 21% in legal MTPE is smaller than the average productivity of MTPE of non-legal texts in these previous studies involving non-legal texts, which hovered above 30% on average. Therefore, it may be argued that MTPE productivity is better than HT in general, but legal MTPE productivity may not be as good as other text types'. Further, individual differences were quite large as shown in Table 1. It may be partly due to the fact that the participants were allowed to spend up to 90 minutes for each task and were encouraged to work at a normal speed to avoid affecting participants' translating and post-editing behaviour. Therefore, individual differences might have affected the temporal aspect of task completion. Some participants were observed to have completed the task early and spend the rest of the time reviewing their work, spending 90 minutes for each task to the full extent.

Participant	Number of edited words	TER
P1	128	0.28
P2	100	0.22
P3	132	0.29
P4	105	0.23
P5	84	0.19
P6	76	0.17
P7	125	0.28
Average	**107.14**	**0.24**

Table 2. TER of MTPE Results

As for MTPE technical productivity, TER of each MTPE output was calculated. Based on Snover et al. (2006), we counted insertion, deletion, substitution, and shift in each of the seven MTPE results by using JavaScript. As shown in Table 2, the average TER was 0.237, which means that 23.7% of MT were edited on average. According to Kim et al. (2019: 65), TER of light post-editing was recorded 3% whereas the average TER of full post-editing involing Korean was 0.230 (Lee and Kim, 2002 : 135). As such, our TER results suggest that the seven participants' MTPE was carried out on a full post-editing scale. However, TER and MTPE time did not tend to correlate. For example, P3 edited the largest number of words, while spending only 53 minutes in MTPE. In other words, more editing does not always mean more post-editing time, which was also observed in Lee and Kim (2022 : 135). Meanwhile, P3's correction rate was lower than the other participants' as shown in Table 5, so processed words may not always be considered as an indicator of MTPE quality either.

In addition to productivity, we compared the quality of HT and MTPE products through human evaluation by three evaluators (E1 to E3). The evaluation results and the average scores for each participant are presented below (see Table 3 and Table 4). In summary, MTPE quality measured in terms of the segment average scores surpassed HT quality. As shown in Table 4, the MTPE scores for each participant exceeded HT scores except on two occasions (E3-P2 & E1-P5). The average MTPE evaluation scores also indicated superior quality (see Table 4).

	E1		**E2**		**E3**	
	HT	**MTPE**	**HT**	**MTPE**	**HT**	**MTPE**
P1	2.52	3.50	3.04	3.88	3.62	3.79
P2	3.57	3.79	3.29	3.88	3.81	3.38
P3	3.05	3.25	3.17	3.79	3.57	3.83

P4	3.76	3.75	3.33	3.83	3.48	3.88
P5	3.67	3.54	3.46	3.79	3.48	3.71
P6	3.29	3.63	3.29	3.88	3.67	3.75
P7	3.05	3.67	3.00	3.71	3.67	3.58

Table 3. HT and MT Quality (Segment Average Scores)

AVE	**P1**	**P2**	**P3**	**P4**	**P5**	**P6**	**P7**
HT	3.06	3.56	3.26	3.52	3.54	3.42	3.24
MTPE	3.72	3.68	3.62	3.82	3.68	3.75	3.65

Table 4. HT and MT Average Scores

The inter-rater agreement in HT quality evaluation was stronger than MT quality evaluation. 95% confidence interval (CI) of interclass correlation coefficient (ICC) for HT evaluation was 0.289 to 0.907 (ICC = 0.720) with a p-value of 0.004. The results suggest that the quality of the post-edited legal texts was not inferior to that of HT, but there was a lack of consensus on the MTPE output quality among the three evaluators.

In terms of the average segment scores, the raw MT output was rated 3.15. Compared with the seven participants' HT and MTPE average scores, MT quality was perceived to be worse than HT except for P1, whereas all the seven MTPE outputs got much higher average points than the raw MT output (see Table 4). That means, DeepL's Korean-English statute MT was inferior to HT, and MTPE could improve the result, largely better than HT, in a time-efficient manner.

Correction rates[2] also demonstrated MTPE quality improvement over the MT output. As shown in Table 5, most of the 26 errors identified by the evaluators were corrected by the seven participants. Except for P3, the other six participants corrected 23-26 out of the total 26 errors, which resulted in an average correction rate of 89.61%. The high correction rate led to higher MTPE average segment scores than the raw MT average segment scores, meaning that post-editing did enhance MT quality.

	P1	**P2**	**P3**	**P4**	**P5**	**P6**	**P7**	**AVE**
Number of errors corrected	26	24	18	24	24	24	23	23.28
Correction rate (%)	100	92.30	69.23	92.30	92.30	92.30	88.46	89.61

Table 5. Number of Corrected Errors and Correction Rates

4. Conclusion

This paper investigated the quality and productivity of MT of legal texts based on DeepL Korean-English MT of statutes. In addition to raw MT output quality, we examined the productivity and quality of HT and MTPE outputs produced by seven professional translators. The BLEU score indicated 29.92, which was not that high, but the MTPE results suggested that the raw MT was good enough to produce publishable quality through post-editing.

[2] In this study, the correction rate refers to the ratio of post-editors' corrections to the errors identified in the raw MT (Kim 2022b).

To examine the productivity of MTPE, we analyzed temporal productivity and technical productivity (TER). The participants tended to spend significantly less time post-editing (57.29 minutes) than translating from scratch (78.14 minutes). The average productivity growth of 21% appears to be smaller than the average productivity of MTPE of non-legal texts in the previous studies. Still, it may be argued that MTPE productivity is better than HT in legal texts, too. As for technical productivity, the average TER was 0.237, which means that an average of 23.7% of MT were edited. However, TER and MTPE time did not tend to correlate.

The overall MTPE quality measured in terms of the segment average scores was superior to HT quality evaluation scores. Regarding the correction rate, most of the errors identified by the evaluators were corrected during MTPE with an average correction rate of 89.61%. As a result, the MTPE outputs got higher points than the original MT output.

Based on the results, it may be argued that for statute translation, MTPE tends to be more productive than HT in terms of task time and number of processed words, and that professionals' MTPE process enhanced MT quality to human parity. All of the seven participants got higher points for MTPE than HT, which pointed to the better quality of MTPE than HT. Nonetheless, the two texts used for this study were not identical, and due to the small sample size, it is impossible to generalize the current research findings. Despite its limitations, however, the findings suggest that in Korean-English legal statute translation, MTPE by professional translators may be more productive and of better quality than translation from scratch. Further research is needed to investigate the merits of MTPE of legal texts in comparison with HT and to explore the cognitive effort involved in MTPE of legal texts.

References

Aranberri, N., Labaka, G., Diaz de Illaraza, A. and Sarasola, K. (2014). Comparison of post-editing productivity between professional translators and lay users. In *Proceedings of the 11th Conference of the Association for Machine Translation in the Americas,* 20-30, Vancouver, Canada.

Cadwell, P., Castilho, S., O'Brien, S. and Michell, L. (2016). Human factors in machine translation and post-editing among institutional translators. *Translation Spaces*, *5*(2): 222–243.

Choi, H., Lee, C. and Lee, J.-H. (2023). 자동화된 기계학습(AutoML)을 활용한 특허 특화 번역엔진의 영한번역 성능(Evaluation of Patent English-Korean Machine Translations by a Patent-Specific NMT Engine Using AutoML). *The Journal of Translation Studies*, 23(5): 101-130.

Daems, J., Vandepitte, S., Hartsuiker, R. and Macken, L. (2015). The impact of machine translation error types on post-editing effort indicators. In S. O'Brien and M. Simard (eds) *Proceedings of the 4th Workshop on Post-Editing Technology and Practice*, 31-45, location, country.

Dames, J., De Clercq, O. and Macken, L. (2017). Translationsese and post-editese: How comparable is comparable quality? *Linguistica Antverpiensia,* 16: 89-103.

Guerberof Arenas, A. (2009). Productivity and quality in MT post-editing. *XII MT Summit Workshop: Beyond Translation Memories*, 26-30.

Guerberof Arenas, A. (2020). Pre-editing and post-editing. In E. Angelone, M. Ehrensberger-Dow and Gary Massey (eds) *The Bloomsbury Companion to Language Industry Studies*. London: Bloomsbury Academic, 333-360.

Jia, Y., Carl, M., and Wang, X. (2019). How does the post-editing of neural machine translation compare with from-scratch translation? A product and process study. *The Journal of Specialised Translation,* 31: 60-86.

Killman, J. and Rodriguez-Castro, M. (2022). PE vs Translating in the legal context: Quality and time effects from English to Spanish*. Revista de Llengua I Diret/Journal of Language and Law,* 78: 56-72.

Kim, J. (2022a). 한영 포스트에디팅 과정에서의 노력 탐색 - 시간, 기술적 노력, 검색을 중심으로(An investigation into temporal, technical, and web-searching efforts during Korean-to-

English post-editing process compared with from-scratch translation). *Interpreting and Translation Studies*, 26(2): 1-24.
Kim, J. (2022b). 한영 포스트에디팅에서 정확성 오류의 수정 양상 고찰 (An investigation into correction of accuracy errors in post-editing). *The Journal of Translation Studies*, 23(5): 91-116.
Kim, S. M., Lee, J. H. and Shin, H. S. (2019). 번역학계와 언어서비스업체(LSP)간 산학협력연구: '포스트에디팅 생산성'과 '기계번역 엔진 성능 비교' (A university-industry joint study on machine translation post-editing productivity and MT engine error rate). *The Journal of Translation Studies,* 20(1): 41-76.
Krings, H. (2001). *Repairing texts: Empirical investigations of machine translation post-editing processes.* Kent, Ohio: Kent State University Press.
Lee, J. and Choi, H. (2023). A case study on the evaluation of Korean-English legal translations by generic and custom neural machine translation engines. *Interpretation and Translation*, 25(1): 75-98.
Lee, J. and Choi, H. (2022). 인공신경망 특허 기계번역 성능에 관한 연구: Patent Translate 와 WIPO Translate 한영 번역 결과물의 누락과 통사 오류 분석을 중심으로 (A study on the quality of patent neural machine translation: A comparison of omission and syntactic errors in the Korean-English translations by patent-specialized Patent Translate and WIPO Translate). *T&I Review,* 12(2): 129-154.
Lee, J.-H. (2021a). 한영 포스트에디팅 노력 예비연구: 트랜스로그 **II** 를 활용한 한영 인간번역과 포스트에디팅의 차이 분석 (A pilot investigation into Korean to English post-editing efforts). *The Journal of Translation Studies,* 22(5): 271-298.
Lee, J.-H. (2021b). 한영 포스트에디팅, 누구나 수행할 수 있는가? 포스트에디팅 수업 설계를 위한 예비 연구 (Can anybody perform Korean to English post-editing tasks? A pilot study for mtpe module design). *The journal of Translation Studies,* 22(1):171:197.
Lee, J.-H. (2022). 법률 특화 번역엔진 성능 평가-한영 계약서 번역을 중심으로 (A study on performance evaluation of a specialized machine translation engine in the legal domain: Focusing on the Korean-to-English translation of legal contracts). *T&I Review,* 12(1): 162-192.
Lee, J.-H. and Kim, S. M. (2022). 풀 포스트에디팅에 대한 고찰—풀 포스트에디팅 생산성에 영향을 주는 요소를 중심으로 (An investigation into English to Korean full post-editing: Factors affecting productivity of full post-editing). *The Journal of Translation Studies,* 23(5): 119-146.
Lee, S. M. (2021). 기계번역을 활용한 법령번역의 실제와 과제 (Practices and challenges in translating statutes using machine translation). *T& I Review*, 11(1): 35–56.
Lee, S.W. and Lee, S. B. (2021). 학부번역전공자의 인간번역과 기계번역 포스트에디팅 품질 비교 (Comparing human translation and post-edited machine translation: A case study of Korean undergraduate students). *T&I Review*, 11(2):101-123.
O'Brien, S. (2017). Machine translation and cognition. In J. W. Schwieter and A. Ferreira (eds) *The Handbook of Translation and Cognitio*n, 311-331. Wiley-Blackwell.
Popović, M., Lommel, A., Burchardt, A., Avramidis, E. and Uszkoreit, H. (2014). Relations between different types of post-editing operations, cognitive effort and temporal effort. In *Proceedings of the Seventeenth Annual Conference of the European Association for Machine Translation*, 191-198, Dubrovnik, Croatia.
Roberts, B. (2022). Machine vs. Human translation: When to use which for legal translation. Retrieved July 10, 2023, from https://www.attorneyatwork.com/machine-translation-vs-human-translation-when-to-use-which-for-legal-translation/
Seo, B. H. and Kim, S, (2020). A case study on the influence of translator's experience in translation education on the quality of post-editing results. *The Journal of Translation Studies*, 21(3): 63-91.
Snover, M., Dorr, B., Schwartz, R., Micciulla, L. and Makhoul, J. (2006). A study of translation edit rate with targeted human annotation. In *Proceedings of the Association for Machine Translation in the Americas,* 223-231, Boston: AMTA.
Temnikova, I. (2010). Cognitive evaluation approach for a controlled language post-editing experiment. In *Proceedings of the 7th International Conference on Language Resources and Evaluation*. European Language Resources, 3485-3490, Valletta, Malta.
Tsai, Y. (2017). Linguistic evaluation of translation errors in Chinese–English machine translations of patent titles. *Forum,* 15(1): 142-156.