CSCanada
Studies in Literature and Language
Vol. 18, No. 3, 2019, pp. 61-67
DOI:10.3968/11087

ISSN 1923-1555[Print]
ISSN 1923-1563[Online]
www.cscanada.net
www.cscanada.org

A Comparative Study of Productivity and Quality Gain Between Post-Editing and Translating From Scratch

HUANGFU Wei[a],*

[a]Associate Professor, School of Foreign Languages, North China Electric Power University, Beijing, China.
*Corresponding author.

Received 8 February 2019; accepted 26 May 2019
Published online 26 June 2019

Abstract

Using machine translation (MT) input represents a fundamental change in translators' work mode. The issue of efficacy of MT uses is worth investigating since it is at the heart of understanding translators' choices in post-editing MT results or translating from scratch. This study focuses on a comparative study of the impact of post-editing MT on productivity and translation quality of student translator subjects with different levels of translation experiences. This study also looks into the influence of translators' translation experiences on their performances. The keylogging experiment results show that MT input contributes positively to productivity gain and time savings with some variations caused by translation experiences, and that the overall final text quality is significantly affected when translating with or without MT input though to a varying degree of quality gain. These findings suggest a positive role of post-editing MT in translator training.

Key words: Machine translation; Post-editing; Productivity; Quality; Translation experience

Huangfu, W. (2019). A Comparative Study of Productivity and Quality Gain Between Post-Editing and Translating From Scratch. *Studies in Literature and Language, 18*(3), 61-67. Available from: http://www.cscanada.net/index.php/sll/article/view/11087
DOI: http://dx.doi.org/10.3968/11087

INTRODUCTION

Today's translation industry is witnessing the ongoing trend of using machine translation (MT) system to help human translators. The automated translations from the current MT system are far from being perfect and further MT uses in translation process is still under heated discussion within the translation industry. This area of research, regarding the usefulness of MT, becomes urgent need when translators now frequently receive from their clients' translation task with some text segments pre-translated by MT. It is translators' doubts whether the MT input benefits their work and have any positive influence on the translation quality, since clients will normally ask for reduction of prices for those segments with MT translation suggestions.

On the one hand, productivity and quality become the core issue for understanding the using MT and its impact on translators' work. It is, therefore, necessary to carry out experimental researches among different levels of translators as to whether the MT input affects their translation quality and changes their productivity. Attempts have been made in the study of MT input and its influences in professional or community translation settings, such as those by Flournoy and Duran (2009) and Garcia (2011). Both studies reported the finding of positive influence of MT input on professional translators. So far, very little quantitative researches are concerning students' translators, especially in the language direction of English to Chinese. This will be the focus of this present study to examine the role of MT in the work of different levels of Chinese student translators when comparing their translation from scratch with full post-editing.

Moreover, the new era of technological innovations for translation also pushes for the need of using MT and PE. The translation industry is looking into the necessity to pre-feed their translations with post-edited MTs to seed their translation memory for the no segment matches. Translators also have to consider whether it is a good choice to follow such a trend in their own work. The present study here may shed some light on such issues.

1. LITERATURE REVIEW

1.1 MT and Post-Editing MT

Machine translation has undergone several evolutions, from the direct machine translation to the rule-based, example-based, statistical, neural MT or hybrid systems. The MT system used in this study is Google Translate, which uses its own SMT approach, boosted with the hybrid use of EBMT and NMT for enhancing the efficiency and compensating for the drawbacks of individual approach. Ever since the emergence of MT systems, they have been perceived as important productivity tools by commercial organizations and government agencies and various levels of education institutions. In spite of the controversy revolving its negative constraints, MT has been well-accepted and still been considered quite helpful for its speed increase and efficiency boosting, and in some cases accuracy assurance (Garcia, 2010). With the rising state-of-art SMT and NMT, a further application of MT can supplement human translators' work and prove to be more effective than without it (Stix, 2006). Hutchins(2007) stated that professional translators could use MT systems as an assistance to increase speed in technical translation of short life-span documents for assimilation purpose and on-the-spot translation for cyber interaction. MT systems are far from being close to human performance in terms of quality, but should be regarded as facilitating tools to speed up translation work. As noted by Van der Meer (2003), MT is not perfect, but it is economically necessary where there is increasing demand for translators to work much faster and maintain higher quality standards.

Post-editing is one of the approaches, if not the only short-cut, to bring this defective technology to successful deployment in real translation market. Post-editing (PE) refers to the process of correcting and modifying machine-generated translation. Post-editors can work both on MT proposals or TM fuzzy matches or from no matches. The purpose of post-editing can be either for dissemination or assimilation. If the translations are used for gist, the current main MT systems (Google Translate, Bing Translator and Baidu Fanyi, etc.) can satisfy this need. If the input is pre-edited according to controlled language rules, the MT results are particularly helpful, or even up to the similar quality to professionally translated ones. If for dissimilation purposes, full post-editing should be applied and may involve more temporal, technical and cognitive efforts (Krings, 2001).

On the other hand, when dealing with translation aided by TM in CAT tools, MT can be used when some segments have no matches from the TM. In such translation projects, translators are expected to achieve greater productivity without a compromise on quality. The reason is believed to be that translators are working on fuzzy matches either generated by TM or MT instead of starting from scratch, and that the overall quality is expected the same for the whole translation projects instead of different standards for those machine-generated segments. If this is scientifically valid, the new decade will see the increasingly greater focus on MT and PE in translation assisted with TM.

1.2 Student Translators as Post-Editors

Though Fulford (2002) reported 53% of the surveyed translators were interested in PE of MT or TM on real work for clients, those interests were from novice or freelance translators rather than from professional translators. In Fulford's survey, experienced translators preferred to work by retranslation instead of post-editing because it was difficult for them to accept translation suggestions below an expected quality. On the contrary, novice translators thought it was easy and time-saving and efficient for them to work by PE of MT or TM. Yamada (2015) found PE contributed to an average of 20% increase of productivity by novice translators compared with their translation only by human, and resulted in fewer revisions of MT input in contrast with those by professional translators. It is seemingly true that PE is not a job for trained specialists, professional translators, market or product experts, because they hold their prejudice towards PE jobs.

Then, as a result, novice translators, some of whom are students, become the suitable candidates for both rapid and full PE jobs, which differ in the final quality but still involve the work with editing MT input. Researchers have confirmed productivity increase by experienced translators (Arenas, 2008), but it is still an ongoing debate whether this can be still scientifically proved when subjects are student translators, in particular working on English-Chinese language direction. In addition, the final quality of post-edited texts by student translators is also worth investigating.

1.3 PE Errors and Translation Quality

Traditionally, the translation quality is determined by the number of translation errors. In post-editing, translators are to discover deviations in MT from authentic translations. Due to the substantial amount of human involvement in PE, approaches to evaluate human translation should be applied to them as opposed to the automatic translation evaluation by computers. In general, there are two directions of translation quality assessment (TQA), one from the appraisal of the goodness of translation, and the other from that of badness of translation. Error analysis, which assures the translation quality by counting the number and types of error in the final texts, falls into the second category. On the contrary, TQA models that assesses translation quality in terms of accuracy and intelligibility (or readability and clarity) refer to the first category. The present study is to evaluate translation quality from the perspective of PE error analysis. This type of research is based on the assumption that the number of different types of errors in translation with or without PE indicate deviations from the expected target language accuracy and intelligibility. Error analyses tend to be more likely to produce objective evaluation results by human evaluators than those TQA models that rate the goodness of translation.

It involves different degree of efforts to correct different types of errors in translation. Lexical, formatting and grammatical errors are easy to identify and correct, while errors in word order, clauses, over/under translation and omission/amplification are difficult to handle. One way to deal with this issue is to classify errors into various types and give different weighting values to errors based on types of errors and difficulty levels to correct errors. One study by Daems et al. (2017) used both coarse-grained and fine-grained TQA approach, which classifies errors into two main categories (adequacy and acceptability) and corresponding thirteen sub-categories (word sense, adequacy others, other meaning shift; agreement, verb form, structure, word order, grammar, coherence, lexicon, wrong collocation, spelling, style).

Researchers tend to apply different error typologies in the studies, which causes the difficulties to compare research results, although some researchers provided similar results. It is heavy workload for evaluators to choose from a large number of error types, since the more choices mean the greater probability of having ambiguity and confusion to classify and rate errors. For this consideration, the present research employs the Pym's (1992) classifications, which put errors into binary and non-binary ones. Binary errors mean wrong translations, while non-binary errors mean the translation is not wrong but with better replacements. According to Pym, the former errors are the result of the lack of language proficiency and the latter errors are due to the lack of translation competence. With only two types of errors, the difficulty in evaluation will be noticeably lowered.

2. METHODOLOGY

2.1 Research Questions

In this study, the time needed to complete the translation with the use of automatic MT and that from scratch was calculated among two groups of student subjects, advanced group student translators (AG) and beginner group student translators (BG). The accuracy of their translations was evaluated by counting the binary and non-binary errors to see: (1) how the translators working with MT input would change their productivity and quality, and (2) how the performances of translators with different translation experiences would be affected by MT input. And if the positive answers are confirmed, the amount of the gain as well as contribution resulting from MT uses will be determined from the experimental results.

2.2 Participants

This study selected a total of twelve students, six of whom was senior MTI (Master of Translation and Interpretation) students in a university from China's "211" project key institutions, another six junior BTI (Bachelor of Translation and Interpretation) students from the same university. The six BTI students were marked as beginner group (BG) and six MTI students as advanced group (AG). The AG subjects all held the third level CATTI (China Accreditation Test for Translators and Interpreters) certificate and completed their one hundred thousand words translation tasks in the practice bases. The BG subjects did not attend the CATTI exam and finished total ten thousand words translation assignments in classes. Both AG and BG subjects learned the CAT course, which taught how to work with computer-aided translation tools (SDL Trados, MemoQ), but nothing particularly related to post-editing in their CAT classes.

2.3 Instruments and Data Collection

In order to measure the translation time for each segment in the translated text, both groups of subjects worked on the computer with SDL Trados 2017 (30-day trial version) with the post-edit version add-on, which recorded translation time, segment max-characters, editing distances, post-editing machine percentage and translators' revisions, etc. The source text was taken and revised from a guide manual of smart power grid, the used version of which contained 752 words in 52 segments (twenty-six segments in NO Match with 374 words and another 26 segments in MT Match with 378 words). The machine-translated segments were pre-translated by Trados's Google Translate add-on before the experiment and stored in the TM but provided to translators during the experiment. A glossary list of 1825 entries with 85 entries of core terminologies was made into a Trados Multiterm term-base file and then used to provide term suggestions to translators.

The final work was assessed by two reviews, who were familiar with the CAT class and trained in advance with the Pym's binary and non-binary error classification model. They were asked to judge whether each type of errors was spotted in every segment. For every segment, a total three points were assigned if no errors were found. Two points were taken from the total score for binary errors and one point deducted for non-binary errors. Under this evaluation method, the same error type in each segment was counted only once. Thus, the highest score a translator could get for segments with NO match was 78, and for segments with MT match was also 78, and for a translated pierce of work was 156 in total.

3. RESULTS AND DISCUSSIONS

3.1 Productivity Gain and Time Savings

This section describes the experiment results and conducts some discussions of productivity gain and time savings. From the results, the researcher wants to clarify whether translators working with MT are roughly as productive as working from scratch or even better. Table 1 is the data of productivity gain and time savings based on the match category and processing time in minute as well as the processing speed in terms of word per minute. Table 2 shows the descriptive results of mean processing time and processing speed, and also the SPSS paired sample t-test results.

Table 1
Data of Productivity Gain and Time Savings by Match Category

Group	Match type	Time (min.)	WPM (word per min.)	Productivity gain (%)	Time savings (%)
AG1	MT	27.31	13.84	65.95	39.74
AG1	NO	44.84	8.34		
AG2	MT	31.11	12.15	75.58	43.05
AG2	NO	54.05	6.92		
AG3	MT	30.36	12.45	74.13	42.57
AG3	NO	52.31	7.15		
AG4	MT	23.12	16.35	65.32	39.51
AG4	NO	37.82	9.89		
AG5	MT	28.15	13.43	45.82	31.42
AG5	NO	40.61	9.21		
AG6	MT	29.03	13.02	82.10	45.08
AG6	NO	52.31	7.15		
BG1	MT	33.72	11.21	74.88	42.82
BG1	NO	58.35	6.41		
BG2	MT	42.33	8.93	39.10	28.11
BG2	NO	58.26	6.42		
BG3	MT	28.40	13.31	70.20	41.25
BG3	NO	47.83	7.82		
BG4	MT	33.90	11.15	41.32	29.24
BG4	NO	47.40	7.89		
BG5	MT	40.95	9.23	46.28	31.64
BG5	NO	59.27	6.31		
BG6	MT	31.11	12.15	28.84	22.39
BG6	NO	39.66	9.43		

Table 2
Mean Time and WPM With Paired Sample T-Test Results

Group	Match type	Time (min.)	t	p	WPM(word/min)	t	p	Productivity Gain (%)	Time Savings (%)
AG	MT	28.1800	-10.019	0.000	13.5400	17.84	0.001	68.17	40.33
	NO	46.9900			8.1100				
BG	MT	35.0683	-7.475	0.000	10.9967	7.209	0.001	50.00	32.50
	NO	51.7950			7.3800				

It can be seen from table 1 and table 2 that least processing time in minute and most words processed per minute are in AG_MT category (advanced group with machine translation), and next are in the BG_MT category (beginner group with machine translation), followed by AG_NO category and BG_NO category. In order to test whether significant difference exists between category of MT and NO, paired sample t-test was conducted in SPSS 23. For this variable, significant differences are observed ($|t|=|-10.019|$ or $|-7.475|$ or 17.836 or $7.209>2.571$, the critical value at two-tailed 95% confidence interval with the degree of freedom df=5). Thus, it is obvious that MT increases productivity and saves time. This is in line with many previous researches, as are in Arenas (2008) and O'Brien (2010). But in their studies, they mostly reported the average processing speed of word per minute as around 22 words when translators working with MT and in various language directions. In this study, the mean values are far less that this average value. The mean value for WPM in this study across categories are respectively 13.5400 (AG_MT), 8.1100 (AG_NO), 10.9967 (BG_MT), and 7.3800 (BG_NO). This may be caused by subjects' different translation experience and their familiarity with the CAT tools. Those afore-mentioned researches used professional translators as subjects and translators were working mostly with European language pairs. In present study, subjects were student translators working from English into Chinese and having relatively less or no experience to work in real translation projects. This may explain why subjects in this study are slower and their time used in translation with MT may not be comparable to that by professional translators.

The results in table 1 and table 2 also suggest that in average MT PE is 50%-70% faster than translation from scratch. The productivity gain was calculated by taking into account of processing speed of both MT and NO match categories with this formula: Productivity Gain= $(WORD_{MT}/TIME_{MT}-WORD_{NO}/TIME_{NO})/(WORD_{NO}/TIME_{NO})$. As a result, three AG translators and two BG translator show higher gain than 70% in average. Also, only one AG translator and four BG translators show lower gain than 50% in average. The results also suggest that in average translators with MT PE save 40.33% time compared with translation with 32.5% by those with NO matches. The time savings were calculated by using this formula: Time Savings=1-1/(1+Productivity Gain). It can be seen that all the six AG translators invest 30%-50% less time, while only half of the BG translators can do the same. This implies that there are different degrees of performance in terms of productivity gain and time savings. In spite of individual difference, it is estimated that translators are faster if with MT matches because they save considerable time by directly borrowing from the MT input rather than typing words down and figuring out translations. Thus, these research results answer the research questions by establishing a connection between the gain in productivity and MT uses as well as translation experiences.

3.2 Translation Quality Evaluation

Quality was evaluated by counting the errors in the finished texts. Two reviewers checked every translation segment to decide whether binary or non-binary errors could be detected. The aim is not to judge translators' performance, rather it is to find out whether the number and type of errors made by the subjects are correlated with the category of MT and NO matches, and hence influence the overall productivity of translation. In this study, the translation quality was decided not only by the number of error-free segments but also the type and number of errors as well. The error-free segment was marked with 3 points, and 1 point or 2 points were deducted from the total 3 points if binary or non-binary errors were found respectively. So, the highest possible score will be 78 points for a finished translation and the lowest will be zero.

Table 3
Data of Error Occurrences and Converted Scores with Accuracy Gain

Group	Match type	Binary	Non-binary	Converted score	B/N Ratio	Accuracy gain (%)
AG1	MT	4	10	60	0.40	9.09
AG1	NO	7	9	55	0.78	
AG2	MT	3	8	64	0.38	10.34
AG2	NO	6	8	58	0.75	
AG3	MT	4	8	62	0.50	14.81
AG3	NO	7	10	54	0.70	
AG4	MT	5	10	58	0.50	13.73
AG4	NO	8	11	51	0.73	
AG5	MT	5	11	57	0.45	14.00
AG5	NO	8	12	50	0.67	
AG6	MT	3	6	66	0.50	8.20
AG6	NO	5	7	61	0.71	
BG1	MT	5	6	60	0.83	27.66
BG1	NO	11	9	47	1.22	
BG2	MT	8	10	49	0.80	36.11
BG2	NO	15	12	36	1.25	
BG3	MT	6	8	56	0.75	24.44
BG3	NO	12	9	45	1.33	
BG4	MT	8	9	50	0.89	61.29
BG4	NO	17	13	31	1.31	
BG5	MT	7	9	53	0.78	70.97
BG5	NO	17	13	31	1.31	
BG6	MT	10	13	42	0.77	44.83
BG6	NO	18	13	29	1.38	

Table 4
Mean Error Occurrences and Converted Scores With Pair-Sample T-Test Results

Group	Match type	Binary	t	p	Non-binary	t	p	Converted score	t	p	B/N ratio	t	p	Accuracy gain (%)
AG	MT	4.00	-17.000	0.000	8.83	-1.581	0.175	61.17	12.810	0.000	0.4550	-7.892	0.001	11.67
	NO	6.83			9.50			54.83			0.7233			
BG	MT	7.33	-11.500	0.000	9.17	-0.277	0.793	51.67	8.765	0.000	0.8033	-9.862	0.000	35.41
	NO	15.00			9.33			38.67			1.6200			

From the results in table 3 and table 4, for one thing, it can be seen from the total error occurrences that there are more errors made by subjects in the group of NO match category than the MT match category. Also, in terms of error types, there are more non-binary errors than binary errors in the AG_MT and AG_NO match groups, and the same with the BG_MT match group, but the opposite is true with the BG_NO match group. In terms of converted score, significant differences also exist between MT and NO match categories for both AG and BG translators $((|t|=|-7.892|$ or $|-22.603| > 2.571)$. The evidences may suggest that the MT matches help translators in reducing errors in total. To be more specific, for advanced group student translators, MT matches help translators avoid making more binary errors, but for beginner translators this accuracy gain effect will be different in that the less translation experience they have the more likely they will make binary errors in translation. For another, the results also prove that translators who make binary errors with MT matches tend to make significantly more binary errors with NO matches ($|t|=|-17.000|$ or $|-11.500| > 2.571$, the critical value at two-tailed 95% confidence interval with df=5). The same assumption cannot be true when it comes to non-binary errors Translators who make non-binary errors tend to make similar amount of such errors with MT matches ($|t|=|-0.277|<2.571$). When comparing the score of MT and No match categories, the mean differences between AG and BG translators becomes smaller from 13 to 6.34. This also means that MT is bridging the gap between translators with different translation experiences.

The above findings are most surprising and seem to partly have incongruity with most related works. Previous studies mostly report no significant difference in quality regardless of working with MT or from scratch with no matches (Garcia, 2010). The present study does not intend to contradict with most findings of related topics by other researchers. Here, it must be acknowledged that the number of subjects is comparatively smaller with only six in each group and their translation experiences are classified by holding CATTII certificate or not and by amount of completed translation tasks. Though these variable settings affect the generalization of the results, they do not influence the reliability and most importantly they make the study less affected by the negative effect of objective variables, such as MT engines, negative translatability indicators in source texts, and also by the cross effect of objective variables, such as translators' attitudes and evaluators' translation preferences, etc. And based on the finding in the previous section, it shows that though in terms of binary errors and converted scores, the significant differences are observed between MT and NO match categories for both AG and BG translators, translators' productivity gain has not put the translation quality in danger. As to non-binary errors, there are no significant differences between BG_MT and BG_NO categories for both AG and BG translators, and it can be attributed to the fact that student translators, especially those of less translation experience, are often confined by the MT suggestions and do not want to put a second thought when working in post-editing mode, or in another word they have adopted an attitude of "if it is not wrong, do not fix it".

In addition, the study results also suggest that MT contributes more to beginner student translators than the advanced student translators. From table 2, it shows that the binary/non-binary ratio is almost 0.8167 smaller between MT and NO match category for the former group, whereas that value is only 0.2683 smaller for the latter group. This indicates that MT is helpful in improving accuracy but to a varied degree for different levels of translators. This also explains why the results show the AG translators show less accuracy gain than the BG translations, with 11.67% gain for the former and 35.41% gain for the latter in terms of mean accuracy gain. From this perspective, this research is not contradictory to those by previous similar studies on the related topics, rather it provides more supportive evidences for the usefulness of MT in training student translators for the globalized market in the information era.

CONCLUSION

In this research, one major finding is that student translators show significant differences in time and productivity between the MT match and NO match category. All subjects in this research invest less time and produce more words per minute when translating with post-editing MT input. This finding is conclusive because the research data clearly prove so, as in section

Therefore, notwithstanding those experienced-related individual differences, this study confirms that MT uses contribute positively to productivity gain and time savings.

The second major finding is that final overall text quality of MT match category is significantly higher than the NO match category, as in section 3.2. To be more specific, MT uses help translators make less errors in total, though to a varying degree of error reduction in terms of binary/non-binary ratio and accuracy gain. Those who makes more binary errors with MT input tend to make those errors much more when translating from scratch. But the same effect cannot be observed with non-binary errors, which can be a surprising result, but can be reasonably explained from the fact that the less experienced translators have adopted an attitude of "if it is not wrong, do not fix it" when translating with MT compared with translating from scratch. So, if considering the different error types, this finding is not conclusive, but in general, this study supports the claim that MT uses can assure translation quality by reducing errors, in particular those more serious or binary errors, or at least by maintaining similar quality in terms of non-binary error reductions.

The last finding, regarding the experience-related benefits from MT, shows the performances of translators with different translation experiences would be affected by MT input. Due to different translation experience, the translators in this study show varying degree of increase in productivity and time savings, ranging from 22.39% the lowest to 82.1% the highest. In this study, the average productivity increase is 68.17% for the advanced group and 50% for the beginner group, and the average time savings is 40.33% for the former group and 32.5% for the latter. Due to different translation experiences, translators in this study have varying degree of accuracy gain, ranging from 9.09% the lowest to 70.97% the highest with the average accuracy gain 11.67% for the advanced group student translators and 35.41% for the beginner group. These experience-related varying benefits from MT provide the evidence that gain in productivity and time savings will not undermine the translation quality and also point out the necessity for training translators with MT uses, especially those of less experience.

ACKNOWLEDGMENTS

This study is funded by North China Electric Power University Teaching Research Project in the 2018-2019 school year.

REFERENCES

Arenas, A. G. (2008). Productivity and quality in the post-editing of outputs from translation memories and machine translation. *Localisation Focus*, *7*(1), 11-21.

Daems, J., Vandepitte, S., Hartsuiker, R. J., & Macken, L. (2017). Identifying the machine translation error types with the greatest impact on post-editing effort. *Frontiers in Psychology*, *8*, 1282.

Flournoy, R., & Duran, C. (2009). Machine translation and document localization at Adobe: From pilot to production. *MT Summit XII: proceedings of the twelfth Machine Translation Summit*, 425-428.

Fulford, H. (2002). Freelance translators and machine translation: An investigation of perceptions, uptake, experience and training needs. In *6th European Association of Machine Translation Workshop* (pp.117-122).

Garcia, I. (2010). Is machine translation ready yet?. *Target. International Journal of Translation Studies*, *22*(1), 7-21.

Garcia, I. (2011). Translating by post-editing: is it the way forward?. *Machine Translation*, *25*(3), 217.

Hutchins, J. (2007). Machine translation: A concise history. *Computer Aided Translation: Theory and Practice*, *13*, 29-70.

Krings, H. P. (2001). *Repairing texts: empirical investigations of machine translation post-editing processes* (Vol. 5). Kent State University Press.

Pym, A. (1992). Translation error analysis and the interface with language teaching. *The Teaching of Translation*, 279-288.

Stix, G. (2006). The Elusive goal of machine translation. *Scientific American*, *294*(3), 92-95.

Van der Meer, J. (2003). At last translation automation becomes a reality: an anthology of the translation market. *Controlled language translation, EAMT-CLAW*, *3*, 180-184.

Yamada, M. (2015). Can college students be post-editors? An investigation into employing language learners in machine translation plus post-editing settings. *Machine Translation*, *29*(1), 49-67.