

Study on Post-editing for Machine Translation of Railway Engineering Texts

Yuting Li and Xiuying Lu*

Research Center for Applied Translation of Transportation and Engineering, School of Foreign Languages, East China Jiaotong University, Nanchang 330013, P.R.China

Abstract. With rapid development of China's railways, there are more overseas construction projects and technical exchanges in the field of railway engineering, which have generated widespread demands for translation. To meet the increasingly growing demands for translation of railway engineering texts, the mode of machine translation plus post editing (MTPE) has been frequently applied besides traditional human translation (HT) for the combination of translation quality and efficiency. Through the case study of post editing in the machine translation of China's High-Speed Railway by Google Translate, this paper discusses the common error types of machine translation in railway engineering translation, and puts forward corresponding post editing strategies, so as to provide references for MTPE of railway engineering translation in the future. It is hoped that research on post editing in the mode of machine translation for railway engineering texts may improve the translation quality and efficiency, thus helping speed up the process of China's railway going global.

Keywords: Railway engineering texts, Machine translation, Post editing, strategies.

1 Introduction

Since the first railway was built and put into use in Britain in 1825, railway has gradually become an important mode of transportation in human society and has made great contributions to the development of human society and economy. According to the *Initiative of Vision and Action on Jointly Building Silk Road Economic Belt and 21st-Century Maritime Silk Road* issued in March 2015, the connectivity of infrastructure has been listed as the priority under the Belt and Road Initiative. As the significant part of the Belt and Road Initiative, China's railway, especially high-speed railway has ushered in a golden opportunity. Railway construction is one of the infrastructure industries that has developed rapidly in China in recent years, and many projects have been contracted abroad. The overseas exchange of railway technologies has been increasingly frequent. Demands for related translation are on the rise, including the translation of code for design and construction of railway, project tenders, training of professionals, related academic works etc.

The process of "China's railway going global" requires active participation of translation, during which, the mode of machine translation plus post editing (MTPE) has been

* Corresponding author: luxy0203@163.com

increasingly involved in. FENG Quangong and ZHANG Huiyu [1] pointed out that “with the development of machine translation technology and the increase of translation demand, the role of machine translation plus post editing in the language service industry has been widely recognized, especially by some large corporations”. LI Mei and ZHU Ximing [2] stressed that “due to the unreliability of machine translation, post editing has become an indispensable link in machine translation to improve translation quality, which directly determines the quality, speed and cost of translation”. Therefore, it can be found that post editing, as an important part of machine translation, plays a key part in improving translation quality and efficiency. To some extent, the mode of MTPE provides an efficient way for railway engineering translation.

This paper tries to focus on the post editing through the case study of machine translation for China’s High-Speed Railway by Google Translate, probing into common error types of machine translation in the field of railway engineering translation, and proposing corresponding post editing strategies, so as to provide references for MTPE of railway engineering translation in terms of translation quality and efficiency.

2 Post editing in the mode of machine translation plus post editing

2.1 The mode of machine translation plus post editing

The concept of machine translation was first put forward by an American scientist Warren Weaver. Machine translation (MT) is the automatic translation of the text from one language to another or multiple natural languages by the computers [3]. It can be roughly divided into three types according to the technology used: rule-based translation, corpus-based translation and neural machine translation, among which, the most popular machine translation system is neural machine translation. It is believed that machine translation has the advantages of high speed, low cost and consistency of terminologies [4]. Although great achievements have been made in MT technology, on the whole, the quality of output of MT is still not comparable to that of HT [5]. At the 9th MT Summit IX in United States, John Hutchins [6], a famous British linguist and information scientist, pointed out that “there is no substantial improvement has been made in the quality of machine translation, and the problems unsolved 50 years ago are still exist”. Hence, the mode of MTPE has been frequently applied by language service providers (LSP) for translation practice. The combination of machine translation and post editing develops rapidly, which gives full play to both the efficiency of machine translation and the quality of human translation. This mode not only satisfies the demands of the rapid developed translation market, but also promotes the improvement of translation technology, as well as the exchange and cooperation between the academic and industry. In addition, LIU Yanli [7] pointed out that the model of MTPE is very suitable for the translation of technical and practical texts, which can effectively improve the translation efficiency. Plitt and Masselot [8] found in their experiment that the mode of MTPE greatly improved the translators’ productivity by 74% and the time consuming reduced by 43% on average. In addition, many research on post editing have proved that the mode of MTPE is more efficient than human translation [8-10].

2.2 Error types and procedures for post editing

The development of post editing has always been closely associated with machine translation, which is the basis of post editing. As defined by Allen [11], post editing is a process to edit, modify and correct the texts that have been translated from one language to another language

by an MT system. According to CUI Qiliang [5], post editing in a broad sense is “a process in which professional editors manually review or partially revise the MT output according to the specific quality standard in order to achieve high quality and translation efficiency in an integrated translation environment”. FENG Quangong and LI Jiawei [12] defined post editing as “the process of processing and modifying the original MT output according to specific purposes, including correcting translation errors and improving the accuracy and readability of MT output.” This concept is more specific, and points out the specific aspects of post translation processing and revision, which has a strong guidance for the future research on the concept of post editing.

2.2.1 Error types of machine translation

CUI Qiliang and LI Wen [13] concluded 11 error types of machine translation, namely, under translation, over translation, terminology mistranslation, form errors, format errors, addition or omission of meaning, redundancy, errors in parts of speech, clause mistranslation, inappropriate word order and constraint of sentence structure. YAN Qingjia and YAN Wenpei [14] suggested that the errors produced by machine translation can be classified into several common types, which mainly come from “the errors caused by redundant words or disjunctive words, mistaken words selection, syntactic structure conversion, misuse of specific words in a certain language, and wrong forms, etc.” FENG Quangong and LI Jiawei [12] concluded that the scope of the post editing mainly includes: checking whether the names of people and place and terminologies are used correctly and consistently; whether the word order needs to be adjusted; whether the word translation is accurate and unambiguous; whether it is in line with the target language syntax and expression practice; whether there are additional or missing translations; and whether there are cultural or ideological conflicts, etc.

MT can only play the role of “primary translation”. To ensure the quality of translation, it’s necessary to combine with the key part of post editing. Post editing of machine translation is the operation of editing and modifying the MT output, which can improve the quality and efficiency of translation to a certain extent. In today's era, with the rapid growth of information, the total amount of texts and data to be translated is beyond the competence of human translation. Therefore, as a post editor, the error types in machine translation including mistranslation, phrase collocation, conjunction, lack of subject, predicate or other components of the sentence, repetition, disagreement of subject and verb, errors in part of speech and format must be noted, for the correct and quick recognition of the errors in MT output can lead to a sound modification.

2.2.2 Competence and strategies for post editing

2.3 Error types of machine translation

The competence of post editing requires the comprehensive mastery of editing and translation, involving the strong competence of both source language and target language, recognizing of subject knowledge, application of software and cross-cultural communication. According to Kruger [15], editing ability includes the ability of reading and writing the target language, focusing on the text level and details, being highly sensitive to the text, the author, the context and readers, and the understanding and mastery of editing process and result. Therefore, strategies taken are closely related to the various capabilities for post editing, which are conducted at lexical, syntactic and discourse levels by linguistic analysis, semantic

understanding, terminology management, error recognition and checking, stylistic refinement, etc.

Railway engineering translation is a kind of translation demand arising with the acceleration of “China’s railway going global”. The application of MTPE in the practice of railway engineering translation will greatly improve the translation speed and quality and speed up the process of China’s railway “going global”.

3 Research and analysis

Neural network machine translation (NMT) created by Google is one the best machine translation engines at present. It takes each sentence as an independent neuron, breaking the phrase-based translation barrier. In this chapter, some contents from the first two chapters of *China’s High-Speed Railway* [16] is selected and processed by Google translate. The MT output is compared with the final version after PE word by word, and the typical MT error type of railway engineering translation is analyzed, and then corresponding strategies are put forward.

3.1 Classification of MT errors

Based on the related researches on error types reviewed above and the analysis of MT output of the excerpt, 4 typical MT error types of railway engineering translation are concluded, namely, lexical, syntactic, discourse and other errors, and then each category is subdivided. Finally, the typical types of MT errors in railway engineering translation are demonstrated in table 1. Understanding common errors in MT can help editors quickly recognize errors so that the editors can quickly deal with them according to corresponding strategies.

Table 1. Categories of MT Errors(C-E) for Post Editing.

Lexical errors	Syntactic errors	Discourse Errors	Other Errors
Mistranslation	Inappropriate translation of passive voice	Lack of coherence	Punctuation problems
Omission	Mistranslation of complex sentences		
Verb collocation	Inappropriate word order		

The excerpt selected has a total of 3892 words in Chinese. Based on the above error types, there are 129 errors in MT output, and the frequency and ratio of four error types are showed in figure 1.

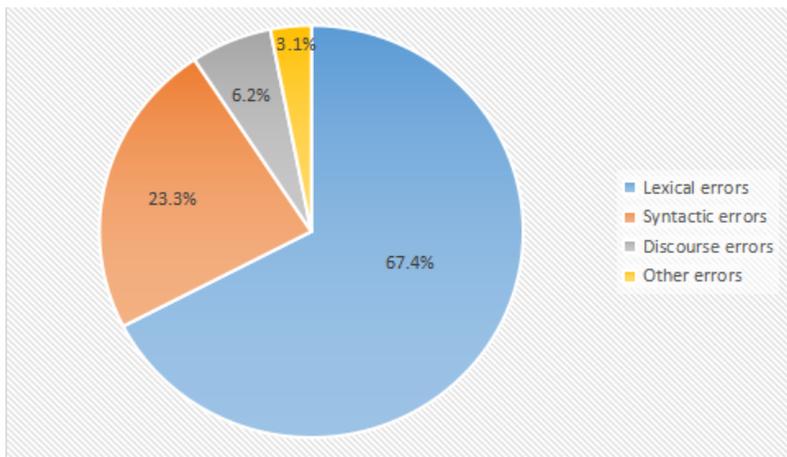


Fig. 1. The frequency and ratio of four error types of MT output of railway engineering text

According to statistics, there are 129 errors in the Chinese-English translation of 3892 words, including 87 lexical errors, accounting for 67.4% of the total errors, 30 syntactic errors, 23.3% of the total errors, 8 textual errors, 6.2% of the total errors, and 4 other errors, accounting for 3.1% of the total errors. From figure 1, it's clearly found that these four types of errors decrease in order of vocabulary, syntax, text and other aspects. Therefore, the editors need to pay attention to the types of errors in the process of PE, focus on the most frequent type of error and make corresponding modifications. Admittedly, the translation quality of Google translate is better than that of some traditional MT engines, but there still many errors in the MT output that need to be modified by human translators. In this chapter, some examples are taken from the MT output to analyze different types of errors in machine translation of railway engineering text and illustrate the corresponding strategies of post editing.

3.2 MT errors in railway engineering translation

3.2.1 Lexical errors

As analyzed above, it can be found that lexical errors appear most frequently in Google translate, accounting for 67.4% of all errors. The main errors are mistranslation, omission and inappropriate use of verb.

Mistranslation. We all know that in translation, the meaning of the word depends on the related context, but sometimes MT engine can't identify the subtle difference between synonyms and choose the inappropriate expression. In the analysis, it is found that mistranslation of nouns mainly lies in the mistranslation of proper nouns (terms) and common nouns. There are many special terms in the field of railway engineering translation. It's easy for Google engine to make mistranslations because of the lacking of professional knowledge.

Example 1: Including the addition of an open hole at the entrance of the tunnel when necessary.

In the above example, the underlined part is actually a terminology in railway engineering, but MT engine translates it literally into "open hole" which is obviously not correct. The correct translation should be "cut-and cover tunnel".

Example 2: The traction power supply system is the high-speed railway. Charger provides sufficient energy for high-speed trains.

Example 3: The main building is built to maintain the stability of the tunnel.

The mistranslation of common nouns is due to the inability of MT to distinguish synonyms according to different contexts. In example 2, the correct translation of the underlined part should be “electricity”. Although “energy” has the meaning of the capacity of physical system to do work, but according to the context, it can be judged that the meaning of word in the original sentence refers to “electric power” required by train operation; and in example 4, “building” generally refers to formal buildings, mansions, etc., while “structure” focuses on the structural form. According to the context, the underlined part of original text refers to the composition of high-speed railway tunnel, so the use of “structure” is more appropriate here.

Omission. Omission means that some contents of source text is dropped off by MT engine, thus the information can't be completely conveyed. To deal with this problem, the translators need to supplement the information omitted in the target language to make the information complete.

Example 4: The Chinese standard EMU has a traction power of about 10,000 kilowatts at a speed of 350 kilometers per hour.

In this case, although the basic meaning of the original sentence has been translated, but a modifier is left out. The subject in this sentence has two attributives: “formation of 8 vehicles” and “with the speed of 350km/h”, but it's obvious that the translation of the first attributive can't be found in MT output, so the information of the original text is not completely conveyed and the scope of China's standard EMU is enlarged. To solve the problem, the editors need to add the omitted information.

Verb Collocation. Word is used flexibly in English, which is mainly reflected in the selection of verb because subtle differences may produce a number of different verbs. When translating Chinese into English, the choice of verbs, especially synonym discrimination, is even not easy for human translation, let alone machine translation. For instance, to explain something, MT engine often translates it into “expand” or “elaborate”. In fact, both the words have the meaning of “detailed description”. However, the verb “elaborate” is used more in professional articles, so it should be used in the translation of scientific and technological texts. In a word, MT engine can't think like human translators and choose the correct verb according to different context.

Example 5: Its function is mainly to carry the lines on the bridge.

“Carry” and “bear” both have the meaning of “support and bear the weight or pressure”, but “carry” lays particular emphasize on the moving items. In this example, what needs to be supported is a road, so the use of the verb “bear” is more appropriate here.

Example 6: Sufficient specific power to pull the train at high speed.

In example 6, “pull” means “the act of pulling” or “applying force to move something toward or with you”, it often refers to the pulling of trigger or switch. Therefore, it's improper to use the word in this sentence. It's more appropriate to use the expression “enable trains to run”.

3.2.2 Syntactic errors

According to statistics, the number of syntactic errors is second only to lexical errors, accounting for 23.6% of the total errors. Based on the study of typical errors types in MT output of professional automobile technology, LI Mei and ZHU Ximing [2] concluded that “among the syntactic errors, the order problem appears most frequently, followed by passive and infinitive problems, and the error rate is directly proportional to the length of the sentence”.

Inappropriate Translation of Passive Voice. One major feature of English scientific and technical texts is the high frequency use of passive voice. However, in contrast with it, passive voice is seldom used in Chinese. On the contrary, the structure of active voice with

omitted subjects is often used in Chinese. MT engine often translates the text in accordance with word order of the original Chinese texts, which is against the normal expression in English.

Example 7: Shanghai will use the Hongqiao hub as another new engine for Shanghai's economic development.

According to the source text, what the sentence wants to convey is that "Hongqiao Transportation is a new engine for Shanghai's development". However, MT engine directly translates the sentence in an active voice with "Shanghai" as its subject. But the subject of the verb "use" is usually human beings, "Shanghai" is a city, so the problem occurs. It's better to omit the original subject "Shanghai" and convert the voice into passive one, which has no effect on the meaning and more conform to common English practice.

Mistranslation of Complex Sentences. One of the characteristics of MT is: "for some simple sentences, there is no serious problems in MT output; but for some long sentences or sentences with a slightly complex structure, the quality of the MT output is not satisfactory, and sometimes it is even unreadable". To translate a long sentence, it is necessary to make a deep analysis of grammar, figure out the relationship between each part of the sentence, and then generate the translation according to the language rules of the target language. But it's difficult for MT engine to do this because machine translation tends to translate the sentence according to the grammar rules of the original text, connect each part into a sentence without any pause, thus resulting in the confusion of sentence structure and unclear meaning, which seriously affects the readability.

Example 8: The definition of my country's high-speed railway is: a newly-built passenger-dedicated railway with an EMU train of 250 km/h and above, and an initial operating speed of not less than 200 km/h.

In the above example, the definition of China's high-speed railway includes two parts: the newly-built multiple unit trains at the speeds of 250 km/h or above and the passenger dedicated railway lines in the early stage of operation carrying high-speed trains at the speeds of 200km/h or above. "The multiple unit trains" and "the passenger dedicated railway lines" have two modifiers respectively. However, MT engine mixes the two categories together and translates the sentence into "a newly-built passenger-dedicated railway with an EMU train", and the latter half of the MT output does not conform to the grammatical rules, which doesn't convey the original meaning accurately.

Inappropriate Word Order. Inappropriate word order refers to the phenomenon that very single word meaning is translated correctly by MT engine, but the order of the words is not conform to English practice, which will affect the readability of the translation.

Example 9: There is a good road, without a good car, nothing can be said about it.

Example 10: Without a good road, a car cannot drive, no matter how good it is.

In the above two examples, machine translation follows the order of the original text and translates the sentences word by word. However, the final translation does not conform to the common English practice. The structure of Chinese expression is relatively loose and short sentences are often used. However, English sentences are closely related according to the systematic grammatical structure, with many modifying and supplementary elements, and generally longer in length. According to the authentic expression of English, the sentences need to be reconnected.

3.2.3 Discourse errors

The application of translation strategies and methods of MT engine is limited to sentence level, and few translation methods are adopted at paragraph level or discourse level [18]. There is often no logical connection between sentences in MT output.

Lack of logic. The logical relationship in Chinese text is often hidden, while English is a language of “hypotaxis”, and the logical relationship is more obvious. According to statistics, logical errors in this project account for 6.2% of the total errors, which is one the most common problems appear in MT output.

Example 11: My country has a vast territory, different climates from north to south, a huge road network, uneven population distribution and economic development in various regions, different requirements for long and short distance travel by passengers, and complex transportation needs.

In this case, machine translation roughly expresses the information of the original text. Logically, it seems that there is no problem. However, if the translator makes a deep analysis, it can be found that the logical coherence between sentences isn't conveyed in MT output. This sentence first introduces China's background knowledge to show the reason for “complex transportation demand”. In fact, there is a causal relationship between them, which needs to be supplemented in post editing.

3.2.4 Other errors

Other error is mainly reflected in the misuse of punctuation, which appears less frequently than other error types, taking up 3.1% of all errors. In the output of machine translation, some punctuation marks may be missed or inaccurately used. In view of punctuation errors, the editors do not need to make special corrections, but only need to make the final check after revising the other three types of errors.

4 Strategies for PE

4.1 Lexical level

Choosing Correct Expression of Terminologies. When processing noun mistranslation in MT output, to deal with common nouns, translators need to compare several feasible words to define the connotation and denotation of their meanings accurately, and finally make judgments and choices of the proper expression. For terminologies, relying on the powerful online corpus, MT engines can quickly output the translation of terminologies, thus saving the translators a lot of time for verification. But in the terms of uncommon or special terminologies, the accuracy of machine translation can't live up to the translator's expectation. Therefore, in post editing, it is still necessary for translators to verify the translation of terminologies by checking the professional database or term base.

Choosing Correct Word Meaning According to the Contexts. In the process of post editing, based on the MT output, translators can quickly find words or sentences with inappropriate semantics, and then choose the closest lexical meaning in accordance with the context of the source language. In the cases of a word with many different meanings, translators should consult the online dictionary to determine the most reasonable explanation that can be integrated into the context. The choice of verbs, as important as that of nouns, plays a significant role in the accurate translation. Therefore, translators should pay attention to the synonym discrimination and collocation of verbs in the process of PE. By retrieving large English corpus, such as BNC, COCA, etc., the translators can quickly obtain the information of specific verb register, collocation and semantic prosody, so as to make a quick judgment.

4.2 Syntactic level

4.2.1 Transforming active voice into passive voice

In Chinese to English translation, translators can appropriately convert the active voice of the original text into the passive voice, especially when dealing with the sentence without subject. Over literal translation of MT engine often leads to the phenomenon of no subject in English, which will not only affect the semantic expression, but also bring misunderstanding to the readers. Therefore, experienced translators should consciously convert the sentence without subject into passive one, so as to avoid the phenomenon of no subject in English.

4.2.2 Segmenting and reorganizing the sentence structure

Each component of English sentence is closely related according to the systematic grammatical structure, with a lot of modifications and supplements, and relatively long in length; while in Chinese, the expression is relatively loose and short sentences are often used. When dealing with long sentences in Chinese, MT engine often follows the original text structure and outputs long sentences with mixed patterns. Therefore, when editing the MT output, it is necessary to analyze the sentence structure of the original text, segment and reorganize the sentence structure, and then output the correct expression.

4.2.3 Adjusting the word order

According to FENG Quangong and LI Jiawei [12], it is necessary to rearrange the word order to correct the errors caused by inappropriate word order which makes the translation obscure and can't be understand clearly. There are many obvious differences in word sequence between English and Chinese, and the machine mainly generates the translation according to the word order of the original text. Therefore, when editing MT output, it is necessary to adjust the word order reasonably according to the expression of English practice, so that the translation can convey the information accurately and make the expression more authentic.

4.3 Discourse level

4.3.1 Supplementing logical words

In view of the lack of logic, the translators only need to manifest the logical relationship in the target language in the process of PE. By means of adjusting the sentence structure or adding some logical words and conjunctions to make the logical relationship clear and translation more coherent and compact.

5 Conclusion

This paper selects railway engineering text for Chinese-English machine translation. Through manual analysis and comparison between MT output and MTPE output, four error types and their frequency distribution of MT output of railway engineering translation are discussed. Through detailed analysis of translation examples, this study puts forward the corresponding adjustment methods. Based on the research, there are some suggestions for improvement of the efficiency and quality of PE in railway engineering translation: first, proper pre-editing can reduce the errors produced by MT engine and improve the accuracy of machine

translation, thus the workload of post editing can be reduced; second, the translator's own quality and ability is also one of the key factors to ensure the quality of PE, therefore, a qualified post editor of railway engineering translation should be equipped with not only bilingual competence, but also professional knowledge of railway engineering; besides, translators should improve their ability to obtain information in the new era; finally, practiced use of various computer-aided tools with high technology can greatly improve the efficiency of post editing.

The above research results can provide some references for MTPE practice of railway engineering translation, which is conducive to improving the efficiency of Chinese-English translation. However, the deficiency lies in that the samples analyzed in this paper are not supported by powerful theory, and the PE strategies provided can't be applied to all types of texts.

Acknowledgements

This research was financially supported by the Project "Study on Post-Editing of MT for Academic Works of Transportation Engineering" from Research Center for Applied Translation of Transportation and Engineering under Grant JD20017 and Jiangxi Provincial Humanity and Social Science Key Research Base, and Jiangxi Provincial Educational Reform Project "Integration Approaches of Internationalization and Localization in MTI Education" under Grant JXYJG-2018-090.

References

1. Q G Feng, H Y Zhang. To Train Post-Editors in the Background of the Global Language Services Industry. *J. Foreign Language World*, (01), 65-72, (2015)
2. M Li, X M Zhu. New Exploration of English-Chinese Machine Translation with Automatic Post Editing. *J. Chinese Translators Journal*, (04), 83-87, (2013)
3. Z S Li, J M Meng. *A Concise Course of Machine Translation*. M. Shanghai: Shanghai Foreign Language Education Press, (2009)
4. H H Lin. The Combination of Machine Translation and Human Intervention. *J. Journal of Yibin University*, (08), 114-116, (2009)
5. Q L Cui. Study on Post Editing of Machine Translation. *J. Chinese Translators Journal*, (06), 68-73, (2014)
6. J Hutchins. Has Machine Translation Improved. A. *MT Summit IX: In Proceeding of the Ninth Machine Translation Summit*. C. East Stroudsburg, PA: AMTA, (2003)
7. Y L Liu. Feasibility Analysis of "MT+PE" in Translation of Different Text Types. D. Shanghai International Studies University, (2020)
8. M S Plitt. A Productivity Test of Statistical Machine Translation PE in a Typical Localization Context. *J. Prague Bulletin of Mathematical Linguistics*, (93), 7-16, (2010)
9. A Guerberof. Productivity and Quality in MT Post-editing. C. In *Proceedings of MT Summit Workshop on New Tools for Translators*. A. (39), 137-144, (2009)
10. D Groves, D Schmidtke. Identification and Analysis of PE Patterns for MT. A. In *Machine Translation Summit XII*, (39), 429-436, (2009)
11. J Allen. Post-Editing. A. In H. Somers(ed.). *Computers and Translation: A Translator's Guide*. C. Amsterdam: John Benjamins Publishing, (2003)

12. Q G Feng, J W Li. Study on the Post Editing Model of News Translation. *J. Technology Enhanced Foreign Language Education*, (06),74-79, (2016)
13. Q L Cui, W Li. The Character of Error Types of Post-Editing: Perspective of Machine Translation Based on Scientific and Technological Materials. *J. Chinese Translators Journal*, (04), 19-22, (2015)
14. Q J Yan, W P Yan. *Translation Methods of English for Science and Technology*. M. Beijing: Metallurgical Industry Press, (1992)
15. H Kruger. Training Editors in Universities: Considerations, challenges and strategies. A. In John Kearns(ed.). *Translator and Interpreter Training: Issues, Methods and Debates*. C. London: Continuum International Publishing Group, (2008)
16. Z W Feng. *Study on Machine Translation*. M. Beijing: China Translation & Publishing Corporation, (2013)
17. K B Hu, Y Li. Research on the Characteristics of Machine Translation and Its Relationship with Human Translation. *J. Chinese Translators Journal*, (05), 10-14, (2016)