Comparison Between Human Translation and Machine Translation

-- Take Lolita as an Example

Zhihan Ke[1,a,*]

[1]*Zhongnan University of Economics and Law, 182 Nanhu Street, Wuhan, China*
a. 1828533213@qq.com
*corresponding author

Abstract: Taking the excerpt of *Lolita* as the research corpus, this paper compares the differences between the manual translation and the machine translation, and finds that the machine translation system has some problems such as inaccuracy in semantic understanding and grammatical errors. In order to improve the quality of machine translation and promote cross-cultural communication and information flow,a classification system of translation errors is established and the errors in machine translation are summarized in detail. The proposed classification system will facilitate a comprehensive understanding of the limitations inherent in machine translation, enable an in-depth analysis of existing challenges, and serve as a valuable reference for enhancing the performance of machine translation systems, thereby elevating the quality and precision of automated translations.

Keywords: Lolita, manual translation, machine translation system, neural machine translation

1.   Introduction

With the process of globalization, the demand for cross-language communication is increasing. As an important natural language processing technology, machine translation is crucial to promote cross-cultural communication and information exchange. At present, machine translation technology has brought great changes to the field of translation, greatly improving the accuracy and speed of translation, and further developing the direction of literary translation. In particular, the emerging neural machine translation (NMT), which mimics the neural network structure of the human brain to predict the possibility of word sequence in translation, often modeling entire sentences in a single integrated model, has high automation and learning capabilities. Its latest advances have led to greater acceptance of the technology of machine translation, even among professional translators.

Nevertheless, although the technology has achieved some success in practical application, its application in the field of literary translation still faces a series of challenges. Machine translation has the characteristics of automation, machinery, sentence as the translation unit, second degree imitation and limited context constraints [1], and cannot take the polysemy and ambiguity of language into account, cultural differences between different countries, the complexity of grammar and language rules, the author's writing intention and the literariness of the text. Once the sentence structure is complex and the implied meaning increase, the quality of the machine translation is greatly

compromised. Since literary texts have greater referential cohesion, the integration of textual features above the sentence level is an important indicator of literary translation, which makes the application of machine translation in the literary translation extremely difficult. At present, the number of literary works successfully translated by machines is relatively small, and domestic and foreign scholars lack the research on machine translation of literary works, as well as the research experiment on machine translation of literary works. Therefore, it is very important to focus on this field and explore the translation strategy and optimized path of machine translated literary works.

As a translator, translation is not only a simple language conversion, which converts words word for word and sentence for sentence, but also involves the understanding and expression of context, culture, emotion and other aspects, so as to accurately convey the meaning and expression of the original text and achieve cross-language and cross-cultural communication. On the premise of maintaining the meaning of the original text, the translator should choose the appropriate vocabulary, grammar and expression according to the language rules and habits of the target language, so as to make the translation flow naturally and conform to the custom and understanding of the target reader. target readers. The biggest difference between literary translation and other text translation is that literary works often have rich expressions and strong artistry, which requires the translator to accurately convey the author's intention, while retaining the artistry of the work, that is, the author's emotion and the beauty of the text, which are difficult to be accurately understood and translated by the machine.

Humbert, the narrator of *Lolita*, is a sensitive, paranoid, extremely imaginative and individualistic artist, who is pathologically obsessed with Lolita. Because he suffers from mental illness, his large segments of fragmented self-statements and monologues are more strongly emotional, and often hallucinates to express his inner pain and repentance. Such a character sets the novel's strong lyrical tone and draws the reader into Humbert's narrative deception. Nabokov's language is blurred, magical and beautiful, with gorgeous artistry. He excels at elaborating events by describing the details of the scene, which is cryptically described with a large amount of environmental rendering and the protagonist's inner monologue. Instead of describing a complete plot clearly, he hides the events in the scene, implicitly suggesting them to the reader, thus giving the reader a sense of being there. In particular, the plot of the third and fourth chapters of the book on sex contains a large number of metaphors, describing the sex scene in extremely obscure language, which can greatly reflect the difference between the manual translation and the machine translation.

Therefore, due to the extremely obscure and ambiguous language and the use of a large number of metaphors, the translator's translation is very different from the machine translation which is loyal to the original text, making this book a typical model for studying the differences between machine translation and human translation.

The Chinese version of *Lolita* is translated by the translator Zhuwan. Mr. Ye Zhi (September 1924 - May 2004), pseudonym Zhuwan. Since the 1950s, Mr. Ye has been engaged in the translation of British and American literature, and he is indifferent to fame and wealth. He is good at translation and has translated *Great Expectations, Basset County Chronicle, Lawrence Short Stories, Dreiser short stories, Lolita* and so on. Because of his translated works of American writers Dreiser, Saul Bellow, Eudora Welty and others, he won the Thornton Niven Wilder Award from the Translation Center of Columbia University in the United States, becoming the first person in China to win this award. Compared with other translations, the Chinese version translated by Zhuwan is more rigorous and maximizes the meaning of the original. At the same time, Zhuwan also makes a detailed explanation of the differences between Chinese and Western cultures involved in the article, which helps to eliminate readers' reading obstacles caused by different backgrounds.

Based on these factors, *Lolita* is taken as an example. After fully studying Evgeny Matusov's classification theory on translation errors, my paper collect and sort out the corpus of the excerpts,

and extract the original English text as well as the Chinese translation by manual translation and machine translation. According to whether the translation conforms to Chinese grammar and the error classification system constructed by Matusov, the corpus in machine translation that conforms to the type of translation errors is selected. By comparing the manual translation with the machine translation, the most common errors in machine translation are explored, and the errors are divided and summarized in detail, aiming to further understand the problems existing in machine translation. This thesis also provides practical strategies for machine translation optimization. This will help improve the quality of machine translation and facilitate the ease of communication and information flow across languages.

2. Research on machine translation and classification of translation errors

In this thesis, my paper mainly draw on the article *The Challenges of Using Neural Machine Translation for Literature* by Evgeny Matusov [2]. Matusov made a quantitative analysis of the application of neural machine translation (NMT) in the field of literature, compared the BLEU and TER values of Google and App Tek, and finally proposed a novel classification method for translation errors, including the following ten categories: M1:severe meaning error, M2: minor meaning error, U: unknown word or segmentation error, C: consistency/term translation error, P: pronoun resolution error, L: locution error, O, I, R: omission, insertion, repetition errors, S1: severe syntax error, S2: minor syntax error, T: tone/register error. This theory provides a more comprehensive system to summarize the categories of translation errors, and also provides a theoretical basis for the subsequent analysis of literary translation. In the following paper, Matusov's classification principle is drawn on and the corpus in machine translation that conforms to the type of translation errors according to the criterion of whether the translation conforms to Chinese grammar are selected. Finally, the third and fourth chapters of *Lolita* are selected to compare the manual translation and machine translation (mainly using the machine translation software of NetEase Youdao Translation) and classify the errors in the machine translation. Through classification analysis, there are seven types of errors in the two chapters of *Lolita*, which will be elaborated in detail in the following paper.

3. Corpus analysis

1. M1: severe meaning error. When context is translated into the target language, if it is not clearly explained, its meaning may mislead the reader and cause reading difficulties. If the reader has not seen the original text, it is difficult to understand the true meaning of the original. In most cases, the meaning of these misleading words or phrases is broad and vague, and the original meaning is not clearly expressed.

Example 1: *She would try to relieve the pain of love by first roughly rubbing her dry lips against mine; then my darling would draw away with a nervous toss of her hair, and then again come darkly near and let me feed on her open mouth, while with a generosity that was ready to offer her everything, my heart, my throat, my entrails, I gave her to hold in her awkward fist the scepter of my passion.* [3]

let me feed on her open mouth [4] was translated into Chinese as "*let me eat her open mouth.*" The meaning of *feed* in the dictionary does mean "*feed," "eat," "supply,*" [5] but it does not mean *"kiss.*" However, it is obviously not true in the context of context, because the machine translation does not consider the context and adapt.

Example 2: *Let me therefore primly limit myself, in describing Annabel, to saying she was a lovely child a few months my junior.*

In the sentence, *Let me therefore primly limit myself* is translated as *"qing yunxu wo yansude zhi shuo (请允许我严肃地只说) "*, while machine translation is translated as *"rang wo yange xianzhi ziji. (让我严格限制自己)"* Considering the context, machine translation is very loyal to the original

text but distorts the original meaning so it becomes a serious translation error.

2. M2: minor meaning error. The translated context conveys the original meaning but with a slight deviation. The reader can still understand the meaning of the word or phrase without looking at the context and can think of better alternatives. This type of word or phrase is usually synonymous with a better replacement word, which is closer in meaning, but it is not possible to choose the most appropriate word or sentence in different contexts.

Example 3: *The softness and fragility of baby animals caused us the same intense pain.*

The softness and fragility is translated manually as *"ruanruowuli (weak and powerless)"* and mechanically as *"rouruan he cuiruo. "* The former is more subjective emotion, implying the narrator's pity feelings, while the latter is a neutral, objective word without personal emotional tendencies. However, there is little difference between the two words literally, and it does not affect the reader's reading.

Example 4: *After one wild attempt we made to meet at night in her garden (of which more later), the only privacy we were allowed was to be out of earshot but not out of sight on the p*opulous part of the plage.

The Chinese translation of the *only privacy we were allowed* is *"the only undisturbed situation"* and the machine translation is *"the only privacy allowed."* Compared with English, less passive voice and more active voice are used in Chinese. But the machine translation process does not take the grammatical differences between Chinese and English into account, resulting in slight semantic errors in translation.

3. P: pronoun resolution error. This type of error often occurs during machine rollover and can only be avoided by looking at the original context. Pronouns have strong relevance, and tests the reader's understanding of the context. However, the translation is only faithful to the original text, and the pronoun may refer to different objects, causing difficulties for the reader.

Example 5: *Her parents were old friends of my aunt's, and as stuffy as she.*

Manually, it is translated as *"ta de fumu shi wo yima de lao pengyou, ye gen yima yiyang gubanfawei. "* In machine translation is *"ta de fumu shi wo gugu de lao pengyou, he ta yiyang guban."* The second *ta* in machine translation is not clear, which is easy to confuse the reader and cause difficulty in understanding. The manual translation directly refers to the object as *yima*, which will not cause trouble for readers. This also reflects that the meaning of the original English text will become more and more unclear and inaccurate in the process of machine translation and re-translation into Chinese, and the phenomenon of information loss will occur.

4. O, I, R: omission, insertion, repetition errors. Insertion error is quite rare, which means that words are added to the translation that are not present in the original text. Omission errors is to not fully translate the meaning of all the words and phrases in the original text, and this error still occurs when translating longer sentences. Repetition error includes repetition of single words or phrases, conjunctions, and structures.

Example 6: *The same June of the same year (1919) a stray canary had fluttered into her house and mine, in two widely separated countries.*

The machine translation is *"in the same June of the same year (1919), a wandering canary flew into her home and mine, which were far apart."* Apparently, the word *countries* is not translated.

Example 7: *In a nervous and slender-leaved mimosa grove at the back of their villa we found a perch on the ruins of a low stone wall.*

"A nervous and slender-leaved mimosa grove" is translated as *"a forest of mimosa trees with slender branches and leaves"* in machine translation without a translation of the word *nervous.*

5. S1: severe syntax error: this type of syntax error usually manifests as incorrect sentence structure after translation. Even if the meaning of a single word or phrase is correct, the incorrect structure severely distorts the meaning of the original text, making it impossible for the reader to understand.

Example 8: *After one wild attempt we made to meet at night in her garden (of which more later), the only privacy we were allowed was to be out of earshot but not out of sight on the populous part of the plage.*

The particular phrase, *"be out of earshot but not out of sight"*, is manually translated to *"can see us, but can't hear us"* and the machine translation to *"can't see, but can't hear."* It contains serious semantic errors. *"Can't see"* and *"can't hear"* are both negative words and should be prepositions that indicate juxtaposition rather than a preposition that shows a twist.

Example 9: *But that mimosa grove -- the haze of stars, the tingle, the flame, the honey-dew, and the ache remained with me, and that little girl with her seaside limbs and ardent tongue haunted me ever since -- until at last, twenty-four years later, I broke her spell by incarnating her in another.*

"That little girl with her seaside limbs and ardent tongue haunted me ever since" is translated into *"that little girl with bare arms and legs and a hot tongue at the beach has been haunting me ever since"* manually. And machine translation is *"that little girl with the limbs of the sea and the passionate tongue haunted me ever since."* But from a grammatical point of view, the machine translated text does not conform to the Chinese word order, and it reads very strange. This is because machine translation does not take grammatical differences between Chinese and English into account, such as the handling of prepositions, and does not relate to the context to add the appropriate word *"xintou（心头）"* to the *"yingrao（萦绕）"* object. This is a serious syntax error.

6. S2: minor syntax error: this kind of syntax error is of a minor degree. It is usually a local syntax error of a word or phrase, and no major changes are required.

Example 10: *There are two kinds of visual memory: one when you skillfully recreate an image in the laboratory of your mind, with your eyes open (and then I see Annabel in such general terms as: "honey-colored skin," "thin arms," "brown bobbed hair," "long lashes," "big bright mouth"); and the other when you instantly evoke, with shut eyes, on the dark innerside of your eyelids, the objective, absolutely optical replica of a beloved face, a little ghost in natural colors (and this is how I see Lolita).*

The manual translation of *in natural colors* is *"an elf dressed in natural colors"*, and the machine translation is *"a little ghost of natural colors."* There is little difference in the meaning of the two translations. However, after searching in numerous corpora, we have not found the use of the word *natural color* as an attributive modification, and it is only used as a neutral noun. Therefore, it can be concluded that there is grammatical error in the machine-translated phrase *"a little ghost of natural colors"*, but the degree of the error is less, so the reader can still understand the original meaning.

Example 11: *I am convinced, however, that in a certain magic and fateful way Lolita began with Annabel.*

The phrase *Lolita began with Annabel* was translated by manually as *"Lolita started from Annabel"* and by machines as *"Lolita began with Annabel."* However, in Chinese, *"Lolita began"* is rarely followed by a person's name. Instead it is usually followed by something. But even if there is grammatical error in this sentence, the degree is minor and does not affect reading.

7. T: tone/register error: the most common form of this type of error is the choice of formal and informal language in different situations, as well as written and spoken language. In addition, some languages have different grammatical rules for different genders, and if they are used incorrectly, they fall into this category.

Example 12: *But there we were, unable even to mate as slum children would have so easily found an opportunity to do.*

The word *mate*, is manually translated as *"jiaohuan（交欢）"*, machine translated as *"jiaopei（交配）."* Although there is no difference in meaning, the former is more subtle and obscure, reflecting the secret ambiguous love between the two children, while the latter is only described from a biological point of view, destroying the ambiguous atmosphere created by the article.

Example 13: *Was my excessive desire for that child only the first evidence of an inherent singularity?*

The manual translation of *the first evidence* is *"earliest signs"*, and the machine translation is *"the first evidence."* In view of the fact that the narrator is a patient with mental illness and rich emotions, the translation of the latter is too cold and objective in the context of the original text, which does not conform to the language style of the original text.

4.   Conclusion

Through classification analysis, in the corpus of the two excerpts of Lolita, there are mainly seven wrong categories in the ten categories of Matusov's theory. The following table lists the specific error types and frequency.

Table 1: Specific error types and frequency.

Error Types	Number of Occurrence	percentage
Severe meaning error	31	40.8%
Minor meaning error	27	35.5%
Pronoun resolution error	1	1.3%
Omission errors	2	2.6%
Insertion errors	1	1.3%
Severe syntax error	2	2.6%
Minor syntax error	6	7.9%
Tone/Register error	6	7.9%

According to statistics, the occurrence times of Severe meaning error and Minor meaning error types are quite high, accounting for most of the times. It can be seen that the errors of machine translation are mostly concentrated in semantic aspects, while the grammar is better. In addition, Tone/Register error type is also worth paying attention to. Machine translation needs to pay more attention to context and there's a long way to go with regards to word selection.

There are some problems and errors in the translation of *Lolita*, which limit the accuracy and fluency of the machine translation system in practical application. This study establishes a classification system of common errors in translation through comparative analysis of *Lolita*. Through the analysis of the error classification, the problems in the machine translation system can be better understood and be targeted optimization. Its significance includes the following three parts:

1. Provide more accurate translation. Optimized machine translation systems can improve understanding of literary language and style by learning a large number of literary works.

2. Maintain the style and characteristics of literary works. Literary works often contain unique styles, rhetorical devices, cultural connotations, etc. The optimized machine translation system can better retain the characteristics of the original after the translation, making the translated context smoother and more natural.

3. Promote cross-cultural communication and understanding. The optimized machine translation system can not only express the exact meaning of the original, but also retain the unique language style, so as to facilitate readers to understand literary works under different cultural backgrounds to a greater extent, and promote cross-cultural communication and understanding.

In conclusion, optimizing machine translation system is very crucial to improve the quality of machine translation and facilitate cross-cultural communication and information flow. Future studies can be done to explore the effectiveness and implementation of various optimization strategies, in order to improve the performance and application value of machine translation systems.

References

[1]   *Hu J B, Li Y. (2016) Research on the characteristics of machine translation and its relationship with human translation [J]. In Chinese Translation, 37(05):10-14. https://libvpn.zuel.edu.cn/piskns.cnki.net/KXReader/Detail?invoice=NSESSI4dh2vjbitmK9esls9Qz2gDMUIqpw0K TwKkLqma14I1I9iVxMPrZi9naVIEIt6%2BXvp2eEjoNoh%2BB8Oezc0oZGSll4xaQ8XqjfIGVa5nOXFXdNR%2F1u GbMwDJ1qnIQ3MLwQ%2Fg2%2FSo6Z9sGGQ63mdJX1E5z9iNIlmga01wwiU%3D&DBCODE=CJFD&FileNa me=ZGFY201605005&TABLEName=cjfdlast2017&nonce=05A0513BA1DE4F97896B0BFD07962923&TIMEST AMP=1694957908676&uid=.*

[2]   *E. Matusov, The Challenges of Using Neural Machine Translation for Literature. Proceedings of the Qualities of Literary Machine Translation, 10-19. https://aclanthology.org/W19-7302.*

[3]   *V. Nabokov, (1955) Chapter 3-4. Lolita. http://s.downpp.com/apk12/com.to.learnenglish_v2.0.2_2265.com.apk*

[4]   *V. Nabokov (2013) Chapter 4. Lolita. Shanghai Translation Publishing House, Shanghai. 44. 9787532738151*

[5]   *A S Hornby, (2018) In: Oxford Advanced Learner's English-Chinese Dictionary. The commercial Press. China. 787. 978-7-100-15860-2.*