ISSN 2959-6122

The Translation Quality Assessment of Google Translate's Toward A More Critical Approach To Bio-events

Ziting Zhou[1]

[1]Beijing University of Technology, Beijing, China
*Corresponding author: 23152206@emails.bjut.edu.cn

Abstract:

There are few Chinese translation resources in paleontology, and literature reading relies heavily on machine translation, yet the translation quality cannot be guaranteed. Taking *Towards a More Critical Approach toward Bio-events* as a sample, this study aims to discuss the effectiveness of Google Translate in English-to-Chinese translation of paleontology literature. By comparing the translated text by Google Translate with that in *Selected Papers of Theoretical Paleontology*, the study assesses the translation from the perspectives of accuracy in word, sentence usage, meaning expression, logical accuracy, and readability. The assessment results show that the English-to-Chinese translation by Google Translate can accurately translate the deductive reasoning logic of the original text, and its coherence is basically consistent with the original text. However, there are paucity such as inaccurate word usage and deviations in sentence structure, leading to poor readability in some passages. This study provides insights into the use of machine translation strategies for paleontology researchers.

Keywords: Translation Quality Assessment; Machine Translation; Scientific Literature

1. Introduction

Paleontology is a significant basis and component of historical biology research. In paleontology studies, English databases are dominant [1], and many Chinese scholars need to obtain research materials from these English core databases. Meanwhile, the cost of manual translation is relatively high. Therefore, numerous researchers choose to use machine translation widely. To this status quo, the accuracy of machine translation can directly affect the cross-language transmission of paleontological academic resources.

Presently, a relatively complete TAQ (Translation Quality Assessment) system has been gradually established globally, including two methods: manual evaluation and automatic evaluation [2]. Manual evaluation requires high level of professional knowledge and experience from translators. Evaluators will comprehensively and meticulously judge the translation methods of direct, relative, and referenceable assessment. Manual evaluation has high reliability and accuracy, but it is costly and time-consuming. Automatic evaluation uses machine learning and natural language processing technology to evaluate translations, checking with indicators such as n-gram matching and METEOR. However, it cannot fully capture subtle differences at the semantic level.

Many researchers have previously conducted comparative and quality assessments of machine and human translations, mostly from the perspectives of literature and aesthetics, focusing mainly on close reading of texts, but rarely conducting Paleontology as a significant basis and component of historical biology research. In paleontology studies, English databases are dominant [1], and many Chinese scholars need to obtain research materials from these English core databases. Meanwhile, the cost of manual translation is relatively high. Therefore, numerous researchers choose to use machine translation widely. To this status quo, the accuracy of machine translation can directly affect the cross-language transmission of paleontological academic resources.

This study aims to explore the effectiveness of machine translation in the English translation of paleontological literature, taking Google Translate and *Towards a More Critical Approach toward Bio-events* [3] as the sample. *Towards A More Critical Approach toward Bio-events* focuses on how to use more scientific and detailed methods to accurately evaluate biological events, which is of great significance to paleontological research. Are Google Translate's translations of paleontological literature accurate? How should paleontological researchers correctly use machine translation to read literature? Answering these two questions can help paleontological researchers

correctly understand machine translation, effectively utilize it, and thereby promote the vigorous development of theoretical paleontology, while also looking into the future research directions of machine translation.

2. Literature Review

In 1964, Nida proposed the translation theory of dynamic equivalence [4]. She pointed out that translation should enable target language readers to obtain similar psychological feelings as source language readers. Based on this theory, she constructed a preliminary framework for translation assessment. The framework emphasizes the following evaluation criteria: whether the translation accurately conveys the information and intention of the original text and the readability of the translation. House emphasized that translation assessment should focus on whether the translation achieves its communicative function of the original text and adapts to the linguistic norms of the target language. In 1977, he proposed a translation quality assessment model based on system-functional linguistics [5]. Toury proposed the methodology of descriptive translation studies in 1980 [6], elucidating the importance of sociocultural backgrounds and historical conditions in translation activities.

Regarding the translation evaluation of scientific and technological texts, Li Xi [7] found that the translation of such articles requires precision, objectivity, and simplicity. She proposed an evaluation method covering words, sentences, logic, and rhetoric, as well as the purpose of the translation. Paleontologist Rong Jiayu [8] mentioned that scientific translation must be rigorous, concise, truthful, and poetic.

The n-gram algorithm is a statistical language model-based algorithm. It captures the structural information of the text by segmenting the text into continuous fragments composed of several words. This algorithm is widely used in the fields of language processing, translation modeling, and neural language research.

The control group translation of *Towards a More Critical Approach toward Bio-events* is included in *Selected Papers of Theoretical Paleontology* [9] and was translated by paleontologists with high scientific and rigorous standards.

3. Framework & Methodology

3.1 Framework

A Review of the Introduction and Application of Translation Quality Assessment Theory in China [7] forms an assessment criterion for micro, macro, and super-macro perspective. In this paper, scientific and technological translations are evaluated from the following perspectives: concise, easy-to-understand, and the usage of translation strategies. According to the *Report of The Story of Life in 25 Fossils: On Translation Strategies of Paleontology Terminologies* [10], this study proposes the following translation assessment standards as Table 1: TAQ Model for Scientific and Technological Texts, which also assess the translation from the perspective of micro and macro:

Table 1. TAQ Model for Scientific and Technological Texts

Standard	Perspectives	Interpretation
Micro perspective	Correctness of word and sentence usage	To evaluate the collocation of words, and whether the use of function words comply with the norms of the target language.
	Grammatical accuracy	To check if there is any grammatical error in the sentence structure.
	Accuracy of meaning expression	To ensure that the translated sentence accurately conveys the meaning of the original text.
	Application of translation techniques	To evaluate whether translation techniques such as amplification or omission have been appropriately applied.
Macro perspective	Logic in text	To assess the logical relationships in the translation, including the consistency of concepts, judgments, and reasoning.
	Cohesion and readability	To assess whether the translation correctly uses cohesive devices and sentences to form a coherent discourse.

3.2 Methodology

For scientific and technological texts, this study assesses the texts from the perspectives of correctness in word and sentence usage, logical accuracy, and readability. The methodology involves measuring 1-gram overlap and adopting quantitative analysis.

(1). Word and Sentence Usage: Measuring 1-gram Overlap

1-gram (the number of 1-gram overlaps / the total number of 1-grams) is a method used to quantify the frequency of words appearing in a text segment. By using the human-translated version as a reference, researchers can quantify the precision of word choice. Paleontological papers contain numerous theoretical concepts. It requires precise scientific identifications and translations with precision and conciseness. The version in *Selected Papers of Theoretical Paleontology* is produced by paleontological researchers which ensures its specialization. Furthermore, paleontological terms are unique so words in the Google Translate version should align with those in the human translation. The 1-gram overlap measurement can quickly identify potential mistranslations such as omissions. It can also verify the professionalism of Google Translate and assessing whether the translation results could mislead readers.

N-Gram Based Paraphrase Generator from Large Text Document [11] mentioned that it is insufficient to solely compare the target translation with the reference translation when evaluating translation quality. This study conducts a quantitative analysis of both the reference translation and Google Translate to ensure the professionalism and rigor of the translation. Researchers focuses on the accuracy and consistency of translating proper nouns and the faithful delivery of qualitative descriptions, such as definitions.

(2). Logic and Readability: Quantitative Comparative Analysis

The researcher employs a method of extracting logical chains from the paragraphs to further visually compare their consistency. To do so, the researcher sorted out the arguments, and formed a clear logical framework. Subsequently, this framework is compared with the original text and the reference translation. It can observe whether the logical chain is complete, coherent, and free from logical jumps or contradictions. Through this visual comparison, researchers can accurately determine whether the machine-translated text is logically faithful to the original.

Meanwhile, the readability is assessed based on the length of phrases and the adjustment and retention of word order. Reasonable control of phrase length contributes to enhancing the fluency of reading, as excessively long or short phrases can negatively impact the reader‘s experience.

4. Results & Analysis

4.1 1-gram Results and Analysis

The 1-gram overlap rate = (the 1-gram overlaps between the two translations)/ (the total number of 1-grams in the Google Translate version). The entire text comprises 5986 words, with 3680 words overlapping. Thus, the average 1-gram overlap rate = 3680/5986 ≈ 0.603.

The result shows that the 1-gram overlap of this paragraph is relatively high. It indicates that the original text and the translation have a high degree of correspondence at the lexical level, which is a basic indicator for evaluating the quality of translation.

On a paragraph level, the maximum overlap rate is observed in the paragraph „innovation events...in dynamic feedback systems.“, which contains 50 words, with 35 overlapping, resulting in an 1-gram overlap rate of 0.780. In contrast, the minimum overlap rate is found in the paragraph „A...equilibrium vs. gradualism, phyletic vs. phylogenetic, macroevolution vs.“, with 60 words and only 8 overlapping, yielding an 1-gram overlap rate of 0.133.

These figures indicate that over half of the vocabulary used by Google Translate are identical to those in the human-translated version. However, significant differences still exist in word choice or phrasing. This is potentially attributed to the limited understanding of contextual nuances, cultural backgrounds, or specific idiomatic expressions. The substantial variation in overlap rates could stem from following factors: the density of technical terminology, distinct handling of specific linguistic structures or sentence patterns, and varying translation strategies applied to different contexts. The following analysis will delve into a quantitative examination of these 1-gram results to enhance their readability.

4.2 Quantitative Comparative Analysis

This study selects some representative paragraphs and data for quantitative analysis, and lists the Table 2. Quantitative Comparative Analysis Form:

Table 2. Quantitative Comparative Analysis Form

result section	Valuable 1-gram overlap rate	Quantitative Comparative Analysis
(1) proper noun Original text: lnnovation events...in dynamic feedback systems. Manual translation: 革新事件 (innovation-event)... 生态系出现之前的动力反馈系统。 Google translation version: 创新事件 ... 生态系统。	1-gram overlap rate =35 / (50 + 50 - 35)≈ 0.78。	The result shows that the 1-gram overlap rate in this paragraph is high. It indicates that the two translations have a strong correspondence at the lexical level. Notably, terms such as „Cambrian“ „Phanerozoic“ „Metazoans“ and „Cephalopods“ are translated with high accuracy. However, the translation of phrases like „innovation events“ is inaccurate. The accurate words, such as „cephalopods“ and „metazoan structures,“ are paleontological jargon with clear definitions and fixed translations. This means Google Translate is good at translating words with definitive meanings and standard translations. Furthermore, these terms are relatively straightforward and frequently encountered in popular science readings. In contrast, more specialized vocabulary like „Palmatolepis gigas“ later in the text remains untranslated, presented directly in English, suggesting that some biological terms are not included in the translation system‘s vocabulary database. The translation of „innovation events“ and some of the other words is not accurate. The meaning of „innovation events“ varies in different fields. It can refer to innovative activities, which have nothing to do with paleontology. This leads to the mistranslation. Here, it is translated as „ 创新事件 ” indicates that Google Translate has less awareness of recognizing the context. The accurate translation of polysemous words is highly depending on the specific context, so Google Translate still has certain limitations in understanding and processing complex contexts. Therefore, when encountering such words, machine translation often fails to accurately judge their true meanings, leading to mistranslation.
(2)Translation of proprietary terms Original text: A... equilibrium vs. gradualism, phyletic vs. phylogenetic, macroevolution vs. microevolution. Manual translation: 近几年来 ... 点断平衡论对渐变论 ; 种系演化对分枝系统演化 ; 宏演化对微演化。 Google translation version: 几年后 ... 渐进主义 , 系统发育与系统发生 , 宏观进化与微观进化。	1-gram overlap rate = 8 / 60= 0.1333	Google Translate translates‘ the pendulum swing ,directly as‘ 钟摆摆回 ‘, which is different from the , 复出现 ‘ in human translation. The hanging swing refers to the periodic movement of ideas, beliefs, and other things in opposite directions. At the same time, this phrase can also represent the swing of the pendulum and the alternation of victory and defeat among political parties in election. This demonstrating that Google Translate is prone to errors in translating words and phrases with multiple connotations. It is also a misunderstanding of the deeper meaning of the original sentence. The periodic changes contained in phrases are completely lost in literal translation, reflecting the limitations of Google Translate in dealing with abstract concepts. Google Translate has translated „phytic vs. phylogenetic“ into two words with similar connotations, reflecting the shortcomings of Google Translate in distinguishing subtle differences in professional terminology. Phylogenetic refers to a small group of closely related species with a common origin, while phylogenetic is a species system in which the organizational structure system is divided into several branches. The former focuses on the phylogenetic relationships and common origins between species, while the latter emphasizes the phylogenetic and branching evolution during evolution. Translating as‘ 系统发育与系统发生 ’ erases this research distinction and has a significant impact on meaning communication.

(3)Translation of definitions Original text 1: Every occurrence is an event. Manual translation 1: 每件发生的事情都可以称为事件。 Google translation version 1: 每次发生都是一个事件。 Original text 2: Global event means, that it is detectable worldwide. Manual translation 2: 全球事件是指可以在全球范围内识别出来的事件。 Google translation version 2: 全球事件意味着它可以在全球范围内检测到。	1-gram overlap rate = 7/11≈ 0.636。	As definitions, these two sentences have not been translated fully into standard sentence structure. Definitions, as a special sentence expression, aims to accurately explain the essential characteristics of a concept or thing. Therefore, translation is required not only to accurately convey the meaning of vocabulary, but also to maintain the standardization and logicality of sentence structures. The standard sentence pattern for defining is the „specific concept plus differential „, which is mostly expressed in Chinese by condensing the essential characteristics of things and using the sentence pattern of „... 是 ...“; In English, it means „... is defined as...“, „... is...“, or „... means...“. The first two structures can be directly translated as “... 被定义为是 ...”. while “means” have a lot of meanings. It can be translated as “ 意味着 ....”, “ 有 ... 的目的 ” as well, which is far from defined properties. This phenomenon once again proves that Google Translate cannot determine the genre of an article through commonly used words. It would obviously fail to fully understand the stylistic requirements and contextual constraints behind it, unless there are genre-specific vocabularies in the sentence. This shows that Google Translate lacks the ability to judge the genre of articles and lacks flexibility in responding to the diversity of human language.
(4)Logic accuracy Original text: In undisturbed ecosystems... without changing the genetic machinery. Manual translation: 在一个未受干扰的生态系中 ... 换句话说 , 就是不要求基因机制的变化。 Google translation version: 在未受干扰的生态系统中 ... 无需改变遗传机制。		The logic of this translation is complete and same with the original text, with no missing elements. The logic of the three texts is: The extreme environment causes species which cannot adapt this climate to extinct. The survived species tend to overgrow and experience increased genetic variation due to the lack of compete. Every individual survived, and there is no natural selection, therefore these species are not evolving. Meanwhile, this process provides more possibilities for organisms to adapt to new environments because large population provides more genetic terms. Therefore, these species have more possibilities to adapt to new environments if extreme climate appears again.
(5)readability Original text: But it was not until ...the Palmatolepisgigas zone (the „upper“ KW horizon). Manual translation: 只是在最近几年 , ... 它指示一种典型的缺氧或黑色页岩事件。 Google translation version: 但直到几年后 , ... 该地平线位于 Palmatolepis gigas 区最上部 (“ 上部 ”KW 地平线) 的顶部。		Researchers found the differences in translation strategy in the two versions. Google Translate tends to reflect the grammatical structure and vocabulary of the original text directly and faithfully. Therefore, it unconditionally retains the attributive sentence structure in this paragraph. This strategy can ensure the accuracy of the translation and avoid misunderstandings or information loss caused by changing the structure. However, from the perspective of Chinese readers, this method may make the sentence structure appear stiff, requiring extra effort to parse and understand. When dealing with such sentences, human translation adopts a more flexible strategy. They often make the text closer to the expression habits of target language. In this sentence, the translator switched the order of the attributive clause, in order to restore the word order of Chinese. It describes the positional relationship of things. Restoring the Chinese word order would make the short sentence too lengthy. Retaining the original order would make the sentences between commas shorter, making the translation more fluent and easier to read and better conveying the information and intention of the original text.

In general, google translate is good at dealing with vocabularies with clear definitions and fixed translations, but there are still deficiencies in context understanding, translating professional terminology and stylistic adaptability. It lacks sufficient in contextual knowledge and stylistic awareness. In comparison, manual translation is based on specific context, thereby it can better meet the requirements of translation quality assessment standard.

5. Conclusion

Google translate can translate vocabulary with clear semantic meaning and stable corresponding translations correctly so that paragraphs in the front have higher accuracy and readability. When facing contextual understanding, differentiation of professional terminology and complex grammatical structures, the accuracy is low, which requires critical reading.

When utilizing google translate, researchers can refer to the original text or use a professional dictionary as comparison and seek for manual verification when necessary. Attention should be paid on logic and structural changes in sentences. Reading the original text is essential in comprehending the context. Finally, improving language proficiency is the fundamental way to reduce the reliance on machine translation and enhance reading quality.

Google Translate can further expand its professional terminology library and establish an efficient terminology update mechanism by cooperating with knowledge bases in various countries. At the same time, they can strengthen the recognition of keyword sentences in different contexts and genres, enhance the contextual awareness of translation tools and logical structure of the text as a whole.

This study may be limited by the number and variety of sample text. Different paleontological literature may have significant differences in style, language and terminology, which limited data sample cannot fully reflect the differences. Future research can establish more objective and quantifiable assessment criteria, expand the scope of sample and further improve the machine translation to meet the needs of paleontological researches.

References

[1] Deng Yiying, Fan Junxuan, Wang Yue, Shi Yukun, Yang Jiao & Lu Zhengbo. (2020). The current status of paleontological databases and data-driven scientific research. Geological Journal of China Universities (04)

[2] Wang Junsong, Zhuang Congqian & Wei Yongpeng. (2024). Quality Assessment of Machine Translation: Methods, Applications, and Prospects. Foreign Language and Literature (03)

[3] Walliser,Otto.H.Towards A More Critical Approach To Bio-events.1986.

[4] Nida, E. A. (1964). Towards a Science of Translating.

[5] House, J. (1977). A Model for Translation Quality Assessment.

[6] Toury, G. (1980). In Search of a Theory of Translation.

[7] Li Xi(2010). A Review of the Introduction and Application of Translation Quality Assessment Theory in China. Ethnic Translation, 2010 (04)

[8] Yang Liu & Li Yashu(2012). The Uniqueness of Paleontological Translation: A Conversation with Academician Rong Jiayu. Chinese Science & Technology Translators Journal, 2012(2)

[9] Rong Jiayu, Fang Zongjie, Wu Tongjia. Selected Papers of Theoretical Paleontology (1990)

[10] Jing Xitong(2017). Report of *The Story of Life in 25 Fossils*: On Translation Strategies of Paleontology Terminologies [D]. Beijing Foreign Study University,2017.

[11] Gadag, Ashwini I , and B. M. Sagar . N-gram based paraphrase generator from large text document. IEEE (2016):91-94.