

Article

Comparative assessment of Bing Translator and Youdao Machine Translation Systems in English-to-Chinese literary text translation

Linli He, Mozghan Ghassemiazghandi*, Ilangko Subramaniam*

School of Languages, Literacies and Translation, Universiti Sains Malaysia, Penang 11700, Malaysia

* **Corresponding authors:** Mozghan Ghassemiazghandi, mozghan@usm.my; Ilangko Subramaniam, Ilangko@usm.my

CITATION

He L, Ghassemiazghandi M, Subramaniam I. (2024). Comparative assessment of Bing Translator and Youdao Machine Translation Systems in English-to-Chinese literary text translation. *Forum for Linguistic Studies*. 6(2): 1189. <https://doi.org/10.59400/fls.v6i2.1189>

ARTICLE INFO

Received: 12 March 2024

Accepted: 27 March 2024

Available online: 22 April 2024

COPYRIGHT



Copyright © 2024 by author(s). *Forum for Linguistic Studies* is published by Academic Publishing Pte. Ltd. This work is licensed under the Creative Commons Attribution (CC BY) license. <https://creativecommons.org/licenses/by/4.0/>

Abstract: This study explores the performance of machine translation of literary texts from English to Chinese. The study compares two machine translation systems, Bing Translator and Youdao Machine Translation, using selected texts from the novel “Nineteen eighty-four” by George Orwell. The data collection includes the original source texts, their machine-generated translations by Bing Translator and Youdao Machine Translation, and comparisons with human reference translations to assess the performance of these systems. The research’s focal point is to evaluate the accuracy, fluency, and appropriateness of translations generated by these two machine translation systems, while also analyzing the post-editing effort required to enhance the quality of the final machine-translated product. The study revealed that despite the presence of flaws in both machine translation systems, Youdao Machine Translation demonstrated superior performance, especially in accurately translating technical terms and idiomatic expressions, making it the more effective option overall. Nevertheless, the translations from Youdao Machine Translation required more substantial post-editing efforts to improve fluency and readability. Conversely, Bing Translator yielded more fluent and natural-sounding translations, albeit with a need for improved accuracy in translating technical terms and idiomatic expressions. The study concludes that while machine translation systems are capable of generating reasonable translations for literary texts, human post-editing remains essential to ensure the final output’s accuracy, fluency, and appropriateness. The study underscores the importance of selecting the appropriate machine translation system based on the nature of the text being translated. It also highlights the critical role of post-editing in refining the quality of machine-translated outputs, suggesting that while machine translation can provide a solid foundation, human intervention is indispensable for achieving optimal accuracy, fluency, and overall readability in literary translations.

Keywords: comparative analysis; human-machine translation collaboration; literary text translation; machine translation; neural machine translation

1. Introduction

In the context of a burgeoning global economy and the inexorable trend of globalization, countries worldwide are becoming increasingly interdependent and regularly engaging in exchanges. As the quantity of translated materials grows, there is a concomitant increase in demand for rapid, high-quality translation services. This heightened demand for efficiency and cost-effectiveness is propelling technological advancement, further accelerating the development of machine translation (MT). While MT exceeds human capabilities in terms of memory capacity and storage, it may not be as flexible or adept at understanding the nuances of emotional discourse, especially when translating Chinese literature into English. One of the foremost challenges confronting machine MT presently lies in the processing of creative texts,

such as literature (Guerberof-Arenas and Toral, 2020). Consequently, the quality of machine-translated literature heavily relies on the post-editing expertise of the human translator. Multiple studies (Rivera-Trigueros, 2022; Li and Chen, 2019) have demonstrated that MT is inferior to human translation in terms of quality. Nevertheless, using MT saves substantial time (Yang et al., 2023), as translators need only modify the machine-generated translation. Yang and Wang (2023) proved that post-editing entails a multifaceted cognitive process involving the examination of the source text, revising the machine-generated translation output, and generating the ultimate target text. Notably, literary texts necessitate more post-editing time than other text types. Language serves as the conduit of culture and communication across nations, and it is a critical factor in translation (Ebrahimi, 2020). With the swift evolution of artificial intelligence technology, the market share of traditional human translation in the translation industry is increasingly claimed by MT. The mainstream choice combines MT and post-editing, as it produces high-quality translation results (Yang et al., 2023).

This study aims to investigate the performance of MT systems in rendering literary texts from English to Chinese. Specifically, it focuses on comparing the effectiveness of two prominent MT systems, Bing Translator and Youdao Machine Translation, by utilizing selected excerpts from George Orwell's (1949) seminal novel "Nineteen eighty-four". Through an evaluation of the accuracy, fluency, and appropriateness of translations generated by these systems, the research seeks to illuminate the strengths and weaknesses of MT in handling literary texts, while also underscoring the challenges and opportunities associated with MT in this domain. Furthermore, the study aims to emphasize the importance of selecting the appropriate MT system based on the nature of the text being translated and the critical role of human post-editing in refining the quality of machine-translated literary works. In doing so, this research contributes to the ongoing discourse on the integration of MT and human expertise in the translation industry.

2. Literature review

This section provides an overview of MT and computer-assisted translation (CAT) tools. The concept of MT is introduced, and its various types and characteristics are discussed. An overview of CAT tools is presented, including their functionalities and features. A discussion of the application of MT technology, along with pertinent research findings, is provided, elucidating the insufficiencies in the study of Bing Translator and Youdao Machine Translation within the domain of literary translation. This section explores the different types of post-editing that can be employed to improve MT output, such as light, full, and human-in-the-loop post-editing. Additionally, post-editing in literary translation is explored, which presents unique challenges due to the artistic nature of the text. The potential benefits and drawbacks of employing post-editing in literary translation are discussed. Finally, a literature review is conducted on the post-editing of MT in the English-Chinese pair, exploring studies and experiments evaluating the effectiveness and efficiency of different post-editing approaches and the impact of human involvement in the process.

MT is a technology that combines linguistics, computer science, artificial

intelligence, translation, and statistical research methods (Mondal et al., 2023; Mohsen et al., 2023). This innovative technology has gained significant attention from academics who have explored its capabilities. MT is defined as a computer process that performs both initial translation and any subsequent alterations, resulting in a final translated text. This distinguishes MT from translator tools, and the term “machine translation” is commonly used to describe a fully automated high-quality translation (FAHQT) by some authors (Bergen, 2004, p. 142). The intricate nature of literary texts, characterized by nuances, ambiguities, metaphors, and cultural references, presents formidable obstacles for automated translation systems, notwithstanding advancements in AI technology (Škobo and Petričević, 2023). Literary translation demands not only linguistic proficiency but also a profound comprehension of the cultural milieu of the source language, alongside a discerning grasp of the author’s stylistic, tonal, and artistic inclinations (Rexroth, 2024; Škobo and Petričević, 2023). These nuances can significantly impact the performance of translation systems.

2.1. History of machine translation

The concept of MT has been around for almost 80 years, with its roots in the first half of the 20th century. However, it was not until after World War II that MT experienced substantial growth, with significant innovations such as the first machine capable of translating between languages by Petr Petrovich Troyanskii in 1947 and proposals by Warren Weaver in 1949 to turn MT into a reality. During the Cold War, the American government invested heavily in computational linguistics to translate between Russian and English, leading to positive outcomes and highly publicized effort. However, the complex nature of MT led to a decline in government spending in this area after the Georgetown experiment report in 1966. The emergence of SYSTRAN, a business that provides high-quality MT solutions, significantly influenced the development of MT, and it still thrives today. Individuals can also use SYSTRAN’s free online MT tool to translate texts (Brooks, 2020). As of 2016, the transition to neural machine translation (NMT) had materialized, leading major free online providers, including Google and Microsoft, to progressively adopt NMT for a substantial number of language pairs (Rothwell et al., 2023).

Koehn (2009) discusses the *Météo* system, an early MT system for translating weather forecasts, which demonstrated the potential for machines to produce high-quality translations in specialized domains. MT technology has advanced since then, and computer-aided translation (CAT) tools have become popular, making the translation process semi-automatic. Today, it’s practically impossible for human translators to work without CAT software. As demand for fast and affordable translation increases, many companies invest in MT development. However, language complexity and nuances make it challenging to produce translations indistinguishable from those produced by humans, so ongoing research and development are needed.

2.2. Comparative review

With the rapid development and continuous advancement of MT technology, the quality of machine-translated texts has significantly improved. Translation technologies are widely applied to meet the demands of the language services market,

with translation platforms such as Youdao Machine Translation and Bing Translator gaining favor among users.

Zhang (2020) conducted his comparative analysis of machine translators' efficacy in legal texts, encompassing Google, Bing, Oulu, and Tencent Translate. He discerned certain deficiencies pertaining to language comprehension and logical interpretation in the machine-rendered translations of legal documents. Tang and Chen (2020) compared the efficacy of seven online MT platforms (Baidu, Sogou, Google, Bing, Jinshan, Youdao, and Tencent) in the translation of medical texts by data processing techniques and quantitative analysis, revealed that none of the evaluated platforms exhibited the capability to produce translations conforming to established medical standards consistently. Dun and Liu (2022) examined the translation quality of Google and Youdao, employing a self-generated Chinese-to-English translation corpus focused on political publicity and technical texts, revealing the respective strengths of the two translation engines. All the investigations underscore the challenges various MT systems encounter across diverse textual genres, necessitating subsequent post-editing interventions. However, a conspicuous lacuna exists in examining English-to-Chinese translation performance between Bing Translator and Youdao Translation within the domain of literary texts—our research endeavors to address and bridge this identified gap in scholarly inquiry.

As Zhong and Liu (2023) pointed out, most Chinese EFL (English as a foreign language) learners favored Youdao Machine Translation owing to its practical search functionality and personalized features. Lan and Zhao (2021) also demonstrated that Youdao Machine Translation excels in rendering distinctive Chinese expressions more effectively, benefiting from its reliance on a Chinese Internet-based database. Fitria (2021) stated that Bing Translator, offered by Microsoft, introduces functionality that enables users to contribute input to the translation process. This translation service operates as a cloud-based platform, supporting multilingual capabilities. Jibreel's (2023) study substantiated that Bing Translator consistently yielded optimal and precise cultural and communicative equivalents in English, even upon back-translation. The Bing Translator exhibits proficiency in translating over 100 languages, offering practicality, high speed, and extensive applicability. Meanwhile, Youdao excels in specific expressions within the Chinese language, while Bing Translator provides versatility across various languages and contexts. These two systems exhibit distinctive competencies. Our investigation seeks to authenticate their efficacy within literary translation, thus furnishing an empirical foundation for enhancing MT processes in the context of literary texts, particularly pertinent to Bing Translator and Youdao Translation.

2.3. Post-editing of machine translation

Machine translation post-editing (MTPE) is reviewing and revising the machine-translated text to improve its readability and accuracy, including identifying errors and revising sentences. Moorkens and O'Brien (2017) highlighted that post-editing machine translation output involves distinct challenges compared to revising translations suggested by a translation memory (TM) or translating without prior suggestions, as it requires addressing errors unique to machine-generated text. The

specifics of post-editing may vary depending on the translation project. The process requires multilingual language processing skills to ensure the resulting text is high quality. The title “post-editor” is often used to describe individuals who perform this task. To ensure the highest quality translation, it is essential to incorporate post-editing tasks into the translation process. Ali (2020) suggested that while MT applications can be employed for obtaining a broad understanding of a source text when translating from English into Arabic, a comprehensive and meticulous post-editing procedure appears indispensable to complete and precise comprehension of the machine-translated output in English to Arabic context. Through post-editing, translators can refine machine-generated translations, enhancing their readability and correcting errors that may have occurred during the MT process. By combining the strengths of both human and MT, the resulting output can achieve the desired level of quality for a given translation project.

People have been focused on MT’s quality improvement since it was created. A complex series of corrections, reformulations, and polishing must be applied to the MT output in order for it to match the quality of human translation, which it currently needs to catch up on. These days, MT and post-editing go hand in hand and have become a fundamental component of the translating process as Pondělíčková (2022) mentioned that there are two types of post-editing: light post-editing and full post-editing. The first kind of light post-editing consists of little adjustments to the MT result. Light post-editing is useful at “a lower level where translation products including stylistic issues may be deemed ‘good enough’ and be fit for purpose,” according to Breyel and Grass (2020, p. 12). Correcting spelling errors is an example of light post-editing, which is just intended to make the output comprehensible. There are no significant adjustments to the output that the post-editor must make.

Post-editing in its entirety is the second type. The MT output entails significant and more profound modifications. Breyel and Grass (2020, p. 12) stated that it is employed “where post-edited products need to be indistinguishable from a human translation carried out from scratch”. It requires, for instance, syntax or formatting adjustments and is more time-consuming and labor-intensive than light post-editing. Literary works, including poetry and prose, are imbued with the artistry of language. Translating such works can be challenging since capturing the same effect as the original is only sometimes feasible. Literary translation emphasizes the ambiguity of language and the uncertainty of meaning, which contrasts with non-literary translations, such as those of scientific, technological, legal, and medical texts. Literary translation is arguably the most challenging of all types, regardless of context, rhetoric, or stylistic style. In a sense, it is a hybrid of “creation” and “translation.” For literary translation, the quality of the final product primarily depends on the bilingual skills and attitude of the translator, even though the quality of MT can, to some extent, affect the process. As translators, we must possess strong bilingual abilities, creativity, and excellent translation skills to handle various texts. Literary texts have distinct structural features that vary according to the genre, and post-editing strategies should be chosen according to the subject matter’s characteristics to achieve the best outcome.

Huang and colleagues (2019) researched the post-editing of machine-translated output from English to Chinese using an online tool. Their results indicated that post-editing effectively improved the quality of MT output, as evidenced by a notable

reduction in errors from 13.1 to 6.3 per 100 words on average. Jia and colleagues (2019) discovered that post-editing of NMT resulted in decreased temporal, technical, and cognitive effort when contrasted with phrase-based statistical machine translation (PBSMT) and human translation (HT). Compared to translating from scratch, post-editing yields fewer errors in the ultimate output, substantially enhancing translators' productivity when dealing with metaphorical expressions (Jia and Lai, 2022). Therefore, the authors recommend using post-editing to improve the quality of machine-generated output in cases where the quality of translation is crucial. The post-editing process can be implemented manually or via automated tools to streamline the process. Additionally, Huang and Xue (2021) analyzed the effects of post-editing on the productivity of professional translators who work with MT systems. Their results show that post-editing significantly enhances professional translators' productivity, with an average reduction in the time required for translation by 32%. Furthermore, the authors discovered that post-editing is particularly valuable for complex documents that require extensive editing.

3. Materials and methods

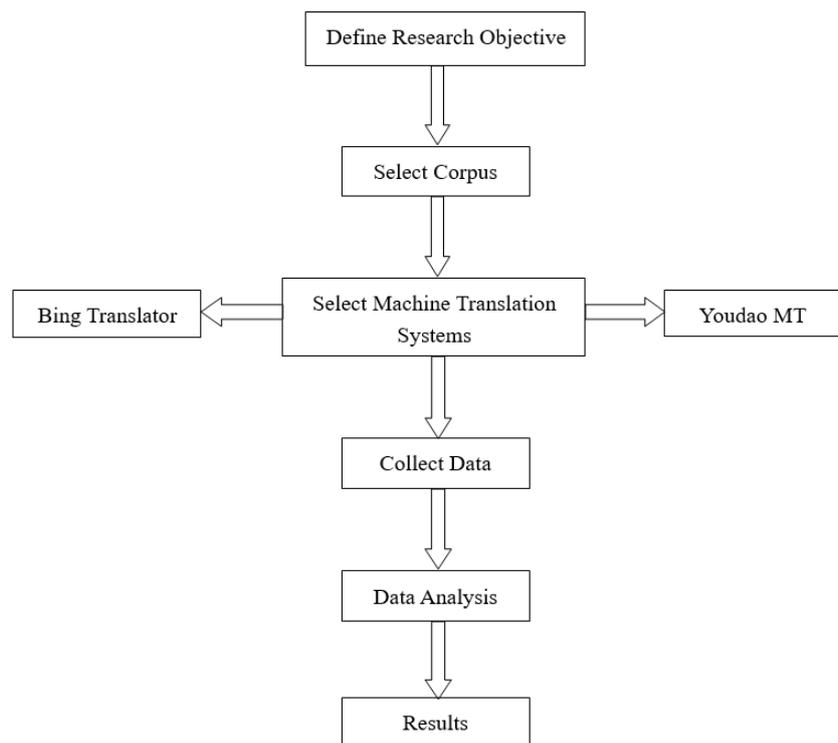


Figure 1. Flowchart of the methodology.

This section outlines the methodological approach employed to explore the performance of Bing Translator and Youdao Machine Translation of English-to-Chinese literary texts. This case study uses the human evaluation metric instead of automatic evaluation for the reason that automatic evaluation metrics may not fully capture the fidelity and fluency of literary translation, particularly in addressing cultural, emotional, and other nuanced aspects. As a descriptive case study, we briefly describe the corpus selection process, followed by an overview of the chosen MT

systems. Subsequently we outline the data analysis techniques employed to compare and assess translation quality between the two MT systems. The experiment engaged two professional literary translators, one of whom demonstrates proficiency in both English and Chinese languages, along with a profound comprehension of literary nuances and cultural subtleties. Participants volunteered their expertise without financial remuneration. To ensure the consistency of evaluations, we conducted an inter-rater reliability test. Using a series of standardized translation samples, raters were instructed to independently evaluate these samples based on predefined evaluation criteria and within a specified score range. Subsequently, we calculated inter-rater agreement measures to ensure the reliability and consistency of the ratings. A flowchart has been designed, as illustrated in **Figure 1**, to clearly delineate the research methodology.

3.1. Corpus selection

In the methodology of this study, we focused on George Orwell's "Nineteen eighty-four" as our corpus, selecting 100 translation units that offer a rich examination of English-to-Chinese literary translation. In determining the selection criteria for the 100 sentences, the study's aim was to ensure comprehensive representation of the linguistic complexities, cultural nuances, and stylistic elements inherent in George Orwell's "Nineteen eighty-four". These determined criteria for the selection of 100 translation units were devised to encompass a spectrum of challenges encountered in the translation process of English-to-Chinese literary texts and to reveal the strengths and weaknesses of the two MT systems. The selection criteria are detailed as follows: Sentences are chosen based on their ability to highlight potential translation challenges, which include the use of specialized language, cultural references, figurative expressions, or intricate syntactic patterns. Each sentence is rigorously evaluated to ensure it presents unique difficulties encountered in the translation process. Moreover, the chosen sentences feature a wide range of themes, narrative techniques, and linguistic nuances found throughout the novel. This variety facilitates an extensive examination of different translation strategies within varied contexts and structural complexities. Furthermore, we have strived to achieve a balance among the types of sentences selected, including dialogues, descriptive passages, narrative sections, and excerpts rich in cultural references. This equitable selection aims to thoroughly illustrate the array of translation challenges inherent in "Nineteen eighty-four".

This novel was chosen for its complex linguistic structures and the translation challenges it presents, notably in its use of parallelism, irony, and culturally nuanced expressions. The selection process focused on units that embody the novel's political significance, showcase linguistic intricacies, and represent a range of genres and styles within the text. These units were carefully chosen to highlight the complexities and subtleties in translating a politically and culturally significant work, providing a focused yet comprehensive lens for analyzing translation practices in literary contexts. In addition to the original English text, an essential part of the corpus is the Chinese version of "Nineteen eighty-four", translated by Dong (1985), who was the first translator to translate "Nineteen eighty-four" into Chinese, and his translation is the

most widely circulated. The inherent advantages of literal translation also influenced the choice of Dong Leshan's version. This translation method gives readers a genuine and rich sensory experience while maintaining the novel's vibrant narrative and stream-of-consciousness style.

3.2. Machine translation systems

This study uses two MT systems: Bing Translator and Youdao Machine Translation. They are widely adopted as practical MT tools, given their extensive user bases and translation datasets. By juxtaposing these two MT systems, we aim to comprehensively evaluate the strengths and weaknesses of general-purpose MT in handling literary texts, as well as its real-world performance.

3.2.1. Bing Translator

Bing Translator, a widely used MT system developed by Microsoft, was chosen as one of the translation engines for this study (Microsoft, 2009). As a website offering translation services for text passages and entire web pages, the system utilizes statistical machine translation (SMT) technology to translate text between English and Chinese.

3.2.2. Youdao Machine Translation

Youdao Machine Translation, developed by NetEase, was selected as the second MT system for comparison purposes (NetEase, 2007). Youdao employs state-of-the-art neural machine translation (NMT) techniques to translate text from English to Chinese. The primary architecture adopted by Youdao NMT is the Transformer model. By employing adversarial training and scheduled sampling, the model's robustness to noise on both input and output ends is enhanced. During training, multiple supervisory objectives are concurrently introduced to optimize the model. Additionally, transfer learning is leveraged to improve the translation performance across multiple languages (Jiqizhixin, 2018).

3.3. Evaluation metrics

To evaluate and score the MT output, the evaluation criteria were adopted from White et al. (1994), Church and Hovy (1991), Blanchon and Boitet (2007), and Araújo and Aguiar (2023). The evaluation involved subjective assessments of three criteria: accuracy, fluency, and appropriateness. **Table 1** below presents the human evaluation criteria derived from these sources, with scores ranging from 1 to 5, where 1 indicates poor performance, and 5 indicates excellence. **Tables 2–4** provide detailed descriptions for each score from 1 to 5 for every criterion, offering nuanced assessments of translation quality. Researchers utilize this detailed rubric to assess and score each translation unit, thereby determining the average score for each MT system used in this study. These four tables serve as rubrics outlining the definitions of each evaluation criterion. Researchers use this guideline to assess and score each translation unit, thereby determining the average score for each MT system used in this study.

Table 1. Translation evaluation criteria rubric.

Criteria	Definition
Accuracy	How well the translation captures the original material’s meaning, intent, and content is referred to as accuracy. When a translation accurately conveys the original meaning without changing the tone or meaning of the text or adding or removing any significant details, it is deemed accurate.
Appropriateness	How closely a translation matches the translated text’s context, goal, and intended audience is referred to as appropriateness. In order to ensure that the translation is acceptable for its intended application and audience, an appropriate translation accurately conveys the message while considering linguistic, cultural, and situational elements.
Fluency	The readability and naturalness of the translated material are referred to as fluency. With appropriate language, syntax, and phrasing that facilitate reader comprehension, a fluent translation makes sense and flows easily. It should flow naturally without any uncomfortable or stiff expressions, just as if it were initially written in the target language.

Table 2. Accuracy evaluation criteria.

Score	Definition
1	The translation is entirely erroneous and fails to convey the original meaning.
2	The translation contains substantial errors or distortions, making the original meaning difficult to comprehend.
3	While the translation has some errors, it still manages to convey some aspects of the original meaning.
4	The translation is generally correct but may contain some inaccuracies or incomplete expressions.
5	The translation is entirely accurate, with almost no errors, effectively conveying the original meaning.

Table 3. Appropriateness evaluation criteria.

Score	Definition
1	The translation completely mismatches the context, goals, and intended audience, rendering it entirely unsuitable for its intended application.
2	While the translation partially aligns with the context, goals, and intended audience, it still contains significant mismatches or deviations.
3	The translation generally fits the context, goals, and intended audience but may exhibit some minor discrepancies or cultural insensitivities.
4	The translation closely matches the context, goals, and intended audience, with only minor discrepancies or cultural nuances.
5	The translation perfectly aligns with the context, goals, and intended audience, demonstrating thorough consideration of linguistic, cultural, and situational elements.

3.4. Data analysis

In the data analysis of this study, the researcher evaluates and scores each translation unit based on these criteria, then compares the results/scores of each MT system to determine which system performed better when processing literary texts. Furthermore, researchers will select specific samples for more detailed analysis to provide insights into the translation quality and performance across various criteria. Through the analysis of these samples, researchers will gain deeper insights into the strengths and weaknesses of each MT system, thus offering more specific

recommendations for improving MT technology.

Table 4. Fluency evaluation criteria.

Score	Definition
1	The translation is very awkward or incomprehensible, making it challenging for readers to understand.
2	The translation lacks fluency and reads disjointedly, hindering smooth comprehension.
3	While the translation is generally readable, it may contain some unnatural or awkward phrasing.
4	The translation flows smoothly and is easily readable, with only minor instances of awkwardness.
5	The translation is exceptionally fluent, reading naturally and seamlessly as if originally written in the target language.

4. Results

The outputs from Bing Translator and Youdao Machine Translation are compared based on evaluation criteria adopted from White et al. (1994), Church and Hovy (1991), Blanchon and Boitet (2007), and Araújo and Aguiar (2023). **Table 5** below displays the results of this comparison.

Table 5. Human evaluation criteria.

Evaluation criteria	Bing Translator	Youdao Translation
Fluency	3.25	3.9
Appropriateness	2.8	3.9
Accuracy	3.15	3.8

The results indicate that Youdao Translation outperforms Bing Translation in terms of fluency, appropriateness, and accuracy. Specifically, Youdao Translation achieved higher scores in fluency, presenting translations that are more natural and coherent. Additionally, its scores for appropriateness and accuracy are also notable, aligning more closely with the target audience’s expectations and the contextual demands of the literary text. In comparison, Bing Translation exhibited relative weaknesses in these areas, with instances of disfluency and inaccuracy. The comparison of translations from Bing Translator and Youdao Translation reveals the following scores: Bing Translation received a fluency score of 3.25, indicating reasonably smooth translations but occasionally marred by awkward phrasing or disruptions in readability. Its appropriateness score of 2.8 suggests that, in some instances, the translations may not completely resonate with the intended context or audience, highlighting areas for improvement. For accuracy, Bing Translation scored 3.15, suggesting that while it generally conveys the essence of the original text, it is also prone to errors or inaccuracies. Conversely, Youdao Translation scored 3.9 for both fluency and appropriateness, and 3.8 for accuracy. These high scores across all indicators suggest that Youdao Translation ensures smooth and natural language flow, maintains alignment with the context, and accurately reflects the source material. The analysis underscores the varying strengths and weaknesses of the two translation engines in different aspects of translation quality. Youdao Translation’s higher fluency

score of 3.9 indicates its translations are more natural and coherent, enhancing readability and comprehension. This score reflects a translation that flows smoothly, avoiding awkward or jarring language. The appropriateness score of 3.9 suggests that Youdao Translation is more effectively tailored to the intended audience and context, offering a translation that more accurately interprets the nuances of the original text. Additionally, with an accuracy score of 3.8, Youdao Translation demonstrates a higher fidelity in conveying the original text's meaning, with fewer errors or discrepancies in translation. Overall, this comparison underscores the significance of selecting the appropriate MT tool. In this case, Youdao Translation performs better in fluency, appropriateness, and accuracy. These findings emphasize the value of using objective data and empirical evidence to assess translation quality. Such an approach enables researchers to make informed decisions and provide recommendations on the post-editing work necessary for a high-quality translation. This systematic evaluation highlights the critical role of choosing the right translation tool to meet specific translation needs and quality standards.

5. Discussion

To illustrate the comparison between Bing Translator and Youdao Translation, elucidate theoretical propositions and viewpoints through concrete cases, and enable readers to grasp the research content and conclusions more intuitively, here are some examples showcasing their respective translations. The six examples have also been categorized accordingly, with corresponding analyses provided. These analyses will highlight differences in accuracy, fluency, and appropriateness, providing a clearer understanding of the strengths and weaknesses of each translation tool in practical use. In each example, a provision of back translation for all Chinese sentences is included to ensure comprehensibility for non-native Chinese language speakers. Additionally, the term “Dong’s version” serves as a reference point for comparison, denoting the human-translated rendition of the text.

5.1. Grammatical structure

In this example as shown in **Table 6**, the study found that Youdao's output achieved high accuracy, appropriateness, and fluency. It correctly conveys the information in the original text, including the situation of the chess problem and the white pieces' task, and performs well in terms of grammatical structure. In contrast, Bing Translator's output has some inaccuracies and ambiguities in terms of word usage and expression, especially in using the word “mating”, which may cause misunderstanding. The translation of this version does not apply to Chinese readers, and some words in the translation may need to be clarified and more apparent to readers. In summary, the translated version of Youdao is the closest to the original meaning and the most readable option, suitable for both academic and general readers.

Table 6. Example 1 of the book “Nineteen eighty-four”.

Original text	He examined the chess problem and set out the pieces. It was a tricky ending, involving a couple of knights. ‘White to play and mate in two moves.’
Bing Translator Output	他检查了国际象棋问题并列出了棋子。这是一个棘手的结局，涉及几个骑士。“白棋两招交配。”
Back Translation	He examines the chess problems and lists the pieces. It’s a tricky ending that involves several knights. “Two moves of white to mate.”
Youdao Output	他仔细研究了这道象棋题，并把棋子摆了出来。这是一个微妙的结局，涉及到几个骑士。白棋出，两步将子
Back Translation	He studied the chess problem carefully and set out the pieces. It’s a delicate ending that involves several knights. White is out. Two moves
Dong’s Version	他看了一下报纸上的那局难棋，就把棋子摆了出来。这局棋结局很巧妙，关键在两只相。“白子先走，两步将死。”
Back Translation	He took a look at the challenging chess game in the newspaper and then rearranged the pieces. The endgame of this chess match is quite clever, with the key move involving two bishops. “White moves first, and in two moves, it’s checkmate.”

Table 7. Example 2 of the book “Nineteen eighty-four”.

Original text	Down at street level another poster, torn at one corner, flapped fitfully in the wind, alternatively covering and uncovering the single word INGSOC.
Bing Translator Output	在街道上，另一张海报被撕裂了一角，在风中摇曳，交替地遮住了一个字，露出了 INGSOC 这个词。
Back Translation	On the street, another poster was torn in a corner, swaying in the wind, alternately obscuring a word, revealing the word INGSOC
Youdao Output	在街道上，另一张海报的一角被撕破了，在风中断断续续地拍打着，时而盖上，时而揭下 INGSOC 这个词。
Back Translation	On the street, another poster had a torn corner, flapping intermittently in the wind, covering and peeling off the word INGSOC.
Dong’s Version	在下面街上有另外一张招贴画，一角给撕破了，在风中不时地吹拍着，一会儿盖上，一会儿又露出唯一的一个词儿“英社”。
Back Translation	On the street below is another poster, with one corner torn and fluttering in the wind intermittently. Sometimes covered, and sometimes revealing the sole word “INGSOC”.

In **Table 7**, it is evident that the original text features complex sentence structures, containing several parallel structures and verb phrases. Both Youdao’s and Bing Translator’s translations maintain the structure of the original text better in terms of grammatical structure. Bing Translator has a high accuracy in translation, but its expression is a little stiff and the sentence structure needs to be more natural, which affects the reading experience. The version performed better in terms of accuracy, successfully conveying the meaning of the original, including details such as the poster being torn, swaying in the wind, alternating covering and revealing the word INGSOC. The language expression is clear, in line with English expression habits, but some places are a little blunt, its language style is not in line with the reading habits of Chinese readers. Youdao’s translation accurately conveys the meaning and emotion of the original text, while the language is smooth and natural, in line with readers’ reading habits. Bing Translator’s translation is accurate, but the language expression is not natural.

5.2. Semantic clarity and precision

Through the analysis of Bing Translator and Youdao Translation in **Table 8**, we

can see the characteristics of various translations. Although Bing Translator captures the concept in the original text, it is a little clumsy in expression and fails to convey the mood and feeling of the original text fully. In lexical selection, the translation of the original “palimpsest” to “一纸空文” is largely true, but the expression is stilted and less natural. The accuracy of the translation is high, and the meaning of the original is successfully conveyed, but it may be a little rigid. Youdao’s translation performed better in terms of accuracy and fluency and was able to capture the meaning and emotion of the original text most accurately. Bing Translator’s translation, while accurate, is a bit stilted.

Table 8. Example 3 of the book “Nineteen eighty-four”.

Original Text	All history was a palimpsest, scraped clean and reinscribed exactly as often as was necessary.
Bing Translator Output	所有的历史都是一纸空文，被刮得干干净净，并在必要时重新刻字。
Back Translation	All history is a dead letter, scraped clean and, if necessary, reinscribed.
Youdao Output	所有的历史都是一本重写本，只要有需要，就擦得干干净净，重新书写。
Back Translation	All history is a palimpsest, to be wiped clean and rewritten whenever necessary.
Dong’s Version	全部历史都像一张不断刮干净重写的羊皮纸。
Back Translation	All of history is like a parchment constantly being scraped clean and rewritten.

Table 9. Example 4 of the book “Nineteen eighty-four”.

Original Text	They pretended, perhaps they even believed, that they had seized power unwillingly and for a limited time, and that just around the corner there lay a paradise where human beings would be free and equal.
Bing Translator Output	他们假装，也许他们甚至相信，他们已经不情愿地在有限的时间内夺取了权力，而且就在拐角处有一个天堂，人类将在那里自由和平等。
Back Translation	They pretended, perhaps they even believed, that they have unwillingly seized power for a limited time, and that just around the corner there is a paradise where mankind will be free and equal.
Youdao Output	他们假装，也许他们甚至相信，他们是在不情愿的情况下，在有限的时间内掌握了权力，而在不久的将来，一个人类自由平等的天堂就会出现。
Back Translation	They pretended, perhaps they even believed, that they had held power for a limited time against their will, and that in the near future a paradise of human freedom and equality would emerge.
Dong’s Version	他们假装，或许他们甚至相信，他们夺取权力不是出于自愿，只是为了一个有限的时期，不久就会出现一个人都自由平等的天堂。
Back Translation	They pretend, or perhaps they even believe, that seizing power is not voluntary, but merely for a limited period, and soon a paradise of freedom and equality for all will emerge.

As shown above in **Table 9**, from the perspective of accuracy, Youdao output demonstrates commendable performance in terms of accuracy. However, Bing Translator’s translations exhibit some issues in accuracy and appropriateness. For instance, the phrase “拐角处有一个天堂” is not accurately conveyed by Bing Translator, failing to capture the intended meaning of the original text. Additionally, Bing Translator output employs expressions such as “他们已经”，which slightly implies past tense, whereas Youdao output utilizes a closer rendition to the original text with the past perfect tense. Such subtle differences might impact the reader’s understanding of the original meaning. Regarding fluency, both Bing Translator output and Youdao output perform reasonably well, albeit occasional instances of stiffness or unnaturalness may occur, such as the use of “已经” in Bing Translator

output, which might appear somewhat verbose.

5.3. Cultural appropriateness

A conclusion can be drawn based on the evaluation of Bing Translator and Youdao Translation in **Table 10**. In terms of accuracy, Youdao output performs better, and its translation is closer to the original text’s meaning, especially in terms of vocabulary selection and expression. While Bing Translator adds unnecessary emotional elements (冷漠) in terms of word choice, reducing accuracy. In terms of appropriateness, Youdao output is also the best choice, which considers Chinese readers’ language habits and understanding without introducing unnecessary emotional factors. Bing Translator added a mood of apathy. In terms of fluency, both Youdao Translation and Bing Translator perform well, and the language is smooth, natural, and easy to understand. In summary, Youdao output is excellent in accuracy, appropriateness, and fluency. Bing Translator works well in some areas but take care to add emotional elements.

Table 10. Example 5 of the book “Nineteen eighty-four”.

Original Text	There was something that he lacked: discretion, aloofness, a sort of saving stupidity.
Bing Translator Output	他缺乏一些东西：谨慎，冷漠，一种拯救的愚蠢。
Back Translation	He lacks something: prudence, indifference, a stupidity of salvation.
Youdao Output	有一种东西是他所缺少的：谨慎、超然和一种救世的愚蠢。
Back Translation	There was one thing he lacked: caution, detachment, and a kind of saving stupidity.
Dong’s Version	赛麦有着他所缺少的一些什么东西：谨慎、超脱、一种可以免于患难的愚蠢。
Back Translation	Saimai lacks something: caution, detachment, a kind of foolishness that can be free from adversity.

As demonstrated in **Table 11**, Bing Translator’s output has significant grammatical errors and lexical selection issues, impacting the accuracy, fluency, and appropriateness of the translation. It fails to capture the cultural nuances of the original text, resulting in a loss of emotional resonance. For example, the phrase “But the smiles an’ the tears across the years” is translated as “但笑容和泪水却在岁月的流逝中流逝”，lacking the poetic and emotive quality of the original English text. In contrast, Youdao Translation performs better in accuracy, appropriateness, and fluency. It successfully conveys the meaning and emotional depth of the original text while maintaining cultural relevance. For instance, the translation of “But all these years of smiles and tears” as “但这些年的微笑和泪水” accurately reflects the emotional journey conveyed in the original text.

The study results showed that Youdao Translation performed better in translating literary texts, possibly due to its more flexible and accurate handling of complex sentence structures and literary styles. The study found that the time and effort involved in post-editing largely depended on the MT quality. High-quality translations require less post-editing, while low-quality translations require more adjustments and revisions. Therefore, when choosing a MT tool, the impact of its translation quality on the later editing workload should be considered. Despite the potential of MT technology in processing literary texts, post-editing remains a critical step in ensuring

translation quality. Post-editing can help correct errors in MT, making the results more accurate and smoother.

Table 11. Example 6 of the book “Nineteen eighty-four”.

Original Text	They sye that time ‘eals all things, They sye you can always forget; But the smiles an’ the tears acrorss the years They twist my ‘eart-strings yet!
Bing Translator Output	他们嘲笑时间‘eall’万物， 他们总是可以忘记的； 但笑容和泪水却在岁月的流逝中流逝 他们拧了我的‘eart-strings’！
Back Translation	They mock time ‘eall’ all things, They can always forget. But smiles and tears passed by the years They screwed my ‘eart-strings!
Youdao Output	他们说时间可以改变一切， 他们说你总是可以忘记； 但这些年来的微笑和泪水 他们还在牵动我的心弦！
Back Translation	They say time changes everything, They say you can always forget. But all these years of smiles and tears They still tug at my heartstrings!
Dong’s Version	他们说时间能治愈一切创伤， 他们说你总能把它忘得精光； 但是这些年来的笑容和泪痕 却仍使我心痛像刀割一样！
Back Translation	They say time can heal all wounds, They say you can always forget about it completely. But the smiles and tears over these years Still make my heart ache as if being cut by a knife!

6. Conclusion

The study revealed varying levels of translation quality by comparing and analyzing the original text of George Orwell’s literary novel “Nineteen eighty-four” with the translations produced by Bing Translator and Youdao Translation. Bing Translator’s output contained significant grammatical errors and inappropriate vocabulary choices, which compromised the accuracy and fluency of the translation. In contrast, Youdao Translation showed superior performance in terms of accuracy, appropriateness, and fluency, effectively capturing the emotion and connotations of the original text. This finding shed light on the actual performance of MT in the context of literary text translation. It highlights the disparities in quality between different translation tools and provides insights for future translation endeavors and the advancement of MT technology. In literary translation, accurately conveying the emotion and meaning of the original text is crucial. Youdao’s translation excels in this regard, preserving the original’s literary style and emotional tone, thus enabling readers to grasp the nuances of the original text more effectively. Conversely, Bing Translator’s output requires further refinement to enhance the quality and readability of the translation. Additionally, the study’s comparison between human and machine

translations concludes that in literary translation, the quality of human translation often surpasses that of MT. Human translators are more adept at capturing and conveying the original text's emotion and meaning, producing translations that are smooth, natural, and faithful to the original's literary quality and emotional depth. Hence, the choice of translation tools and the translator's skill are pivotal, influencing the quality and readability of the final translated work. While Youdao MT demonstrates potential as a MT tool, ongoing research and development are essential to refine the accuracy and efficacy of MT technology. It's also crucial for users of MT tools to recognize their limitations and the irreplaceable value of human expertise in generating high-quality translations.

7. Contribution and limitation of the study

The main contribution of this study lies in the comparative analysis of Bing Translator and Youdao Translation in the translation of English-to-Chinese literary texts, revealing the practical application of MT in the field of literary translation. The research not only provides insights into the performance of these two MT tools in literary translation but also offers important implications for the future development of MT technology and literary translation practices.

Firstly, the findings of this study highlight the quality disparities between different machine translations, particularly in terms of emotional expression and accuracy of conveying the original meaning in literary translation. This provides valuable guidance for users in selecting appropriate MT tools.

Secondly, the research compares human translation with MT, concluding that in the field of literary translation, the quality of human translation often exceeds that of MT. This emphasizes the advantage of human translation in preserving the emotional and literary qualities of the original text, making it an irreplaceable choice for translation tasks that prioritize literary quality and emotional expression.

Lastly, the study's findings have implications for the translation practice of other MT tools and different literary works. By comparing the performance of two MT tools in literary translation, the study offers references for improving the performance of other MT tools and enhancing the quality of literary translation. It also provides guidance for future research on other MT tools and different literary works.

This study conducts a comparative evaluation of Bing Translator and Youdao Translation in the context of literary text translation, aiming to contribute to the practice and research of MT in literature. However, it is important to acknowledge its limitations: the research is based on a limited selection of literary texts, which may not represent the full diversity of literary genres. The analysis is confined to Bing and Youdao, excluding other MT systems, and focuses on the Chinese-English language pair, potentially limiting the generalizability of the findings to other languages due to unique linguistic, cultural, and expressive differences.

Author contributions: Formal analysis, LH; investigation, LH; data curation, LH; writing—original draft preparation, LH; writing—review and editing, MG; supervision, MG; supervision, IS; funding acquisition, IS. All authors have read and agreed to the published version of the manuscript.

Acknowledgments: I extend my sincere gratitude to my supervisor, Mozghan Ghassemiazghandi, for her invaluable guidance, timely advice, and constructive suggestions throughout my research. Additionally, I thank my family for their unwavering encouragement and support.

Conflict of interest: The authors declare no conflict of interest.

References

- Ali, M. A. (2020). Quality and Machine Translation: An Evaluation of Online Machine Translation of English into Arabic Texts. *Open Journal of Modern Linguistics*, 10(5), 524–548. <https://doi.org/10.4236/ojml.2020.105030>
- Araújo, S., & Aguiar, M. (2023). Comparing ChatGPT's and Human Evaluation of Scientific Texts' Translations from English to Portuguese Using Popular Automated Translators Notebook for the SimpleText Lab at CLEF 2023.
- Bergen, D. (2004). Translation technology and translator training. In: Tommola J (editor). *Kieli, teksti ja kääntäminen = Language, Text and Translation*. University of Turku. Painosalama Oy. pp. 141–155.
- Blanchon, H., & Boitet, C. (2007). For the external evaluation of MT systems using task-based methods (French). *Trait. Autom. des Langues*, 48(1), 33-65.
- Breyel, C., & Grass, T. (2020). Machine translation and “biotranslation”: an unlikely marriage. *SKASE Journal of Theoretical Linguistics*, 17(4).
- Brooks, R. (2020). A Brief History of Machine Translation Technology. K International.
- Church, K. W., & Hovy, E. H. (1993). Good applications for crummy machine translation. *Machine Translation*, 8(4), 239–258. <https://doi.org/10.1007/BF00981759>
- Dong, L. (1985). *Nineteen Eighty-Four*. Hua Cheng Publishing House.
- Dun, X., & Liu, S. (2022). Corpus-Based Comparative Study of Google and Youdao Machine Translation Quality. <https://doi.org/10.25236/etmhs.2022.040>
- Ebrahimi, M. A. (2020). Cultural value of translation of proverbs and synopsis. *Journal of Research and Knowledge Spreading*, 1(1), e11484. <https://doi.org/10.20952/jrks1111484>
- Fitria, T. N. (2021). A review of machine translation tools: The translation's ability. *Language Circle: Journal of Language and Literature*, 16(1), 162-176. <https://doi.org/10.15294/lc.v16i1.30961>
- Guerberof-Arenas, A., & Toral, A. (2020). The impact of post-editing and machine translation on creativity and reading experience. *Translation Spaces*, 9(2), 255–282. <https://doi.org/10.1075/ts.20035.gue>
- Huang, M., & Xue, Q. (2021). Investigating the effects of post-Editing on the productivity of professional translators. *International Journal of Emerging Technologies in Learning*, 16(1), 135-146.
- Huang, M., Lin, Y., Chen, C., & Liu, Y. (2019). Evaluating the effectiveness of post-editing machine translation output: An empirical study of Chinese–English machine translation. *International Journal of Computational Linguistics Research*, 10(2), 99-107.
- Jia, Y., Carl, M., & Wang, X. (2019). Post-editing neural machine translation versus phrase-based machine translation for English–Chinese. *Machine Translation*, 33(1–2), 9–29. <https://doi.org/10.1007/s10590-019-09229-6>
- Jia, Y., & Lai, S. (2022). Post-Editing Metaphorical Expressions: Productivity, Quality, and Strategies (*Journal of Foreign Languages and Cultures*). 6, 28–43. <https://doi.org/10.53397/hunnu.jflc.202202004>
- Jibreel, I. (2023). Online Machine Translation Efficiency in Translating Fixed Expressions Between English and Arabic (Proverbs as a Case-in-Point). *Theory and Practice in Language Studies*, 13(5), 1148-1158. <https://doi.org/10.17507/tpsls.1305.07>
- Jiqizhixin. (2018). Defeating two renowned international translation engines. Interpreting NetEase Youdao's neural machine translation model. Available online: <https://www.jiqizhixin.com/articles/2018-12-25-22> (accessed on 23 March 2024).
- Koehn, P. (2009). *Statistical machine translation*. Cambridge University Press.
- Lan, M., & Zhao, L. (2021). Contrasting and Analyzing Machine and Human Translation: A Case Study on Red Sorghum. In: 2021 4th International Conference on Humanities Education and Social Sciences (ICHESS 2021). Atlantis Press. pp. 2522-2528. <https://doi.org/10.2991/assehr.k.211220.437>
- Li, H., & Chen, H. (2019). Human vs. AI: An Assessment of the Translation Quality Between Translators and Machine Translation. *International Journal of Translation, Interpretation, and Applied Linguistics (IJTIAL)*, 1(1), 1–12. <https://doi.org/10.4018/IJTIAL.2019010104>

- Microsoft. (2009). Bing Translator. Available online: <https://www.bing.com/translator/> (accessed on 9 January 2024).
- Mohsen, M. A., Althebi, S., & Albahooth, M. (2023). A scientometric study of three decades of machine translation research: Trending issues, hotspot research, and co-citation analysis. *Cogent Arts & Humanities*, 10(1), 2242620. <https://doi.org/10.1080/23311983.2023.2242620>
- Mondal, S. K., Zhang, H., Kabir, H. M. D., et al. (2023). Machine translation and its evaluation: a study. *Artificial Intelligence Review*, 56, 10137–10226. <https://doi.org/10.1007/s10462-023-10423-5>
- Moorkens, J., & O'Brien, S. (2017). Assessing user interface needs of post-editors of machine translation. In: Kenny, D. (editor). *Human issues in translation technology*. Routledge. pp. 109-130. <https://doi.org/10.4324/9781315648934>
- NetEase. (2007). Youdao Machine Translation. Available online: <https://www.youdao.com> (accessed on 9 January 2024).
- Orwell, G. (1949). *The Literature Network*, Jalic LLC. Available online: <https://www.online-literature.com/orwell/1984/> (accessed on 9 January 2024).
- Pondělíčková, A. (2022). Students Versus Machines: A Comparison of Student and Machine Translations.
- Rexroth, K. (2024). *Encyclopedia Britannica*. Available online: <https://www.britannica.com/art/literature> (accessed on 9 January 2024).
- Rivera-Trigueros, I. (2022). Machine translation systems and quality assessment: A systematic review. *Language Resources and Evaluation*, 56(2), 593–619. <https://doi.org/10.1007/s10579-021-09537-5>
- Rothwell, A., Moorkens, J., Fernández-Parra, et al. (2023). *Translation Tools and Technologies*. Taylor & Francis.
- Škobo, M., & Petričević, V. (2023). Navigating the Challenges and Opportunities of Literary Translation in the Age of AI: Striking a Balance Between Human Expertise and Machine Power. *Društvene i Humanističke Studije (Online)*, 8(2(23)), 317–336. <https://doi.org/10.51558/2490-3647.2023.8.2.317>
- Tang, B., & Chen, S. (2020). A Comparative Analysis of Translations of Medical Texts Produced by Online Machine Translation Software. *Chinese Science & Technology Translators Journal*, 33(3), 23-26,49. <https://doi.org/10.16024/j.cnki.issn1002-0489.2020.03.007>
- White, J. S., O'Connell, T. A., & O'Mara, F. E. (1994). The ARPA MT Evaluation Methodologies: Evolution, Lessons, and Future Approaches. *Proceedings of the First Conference of the Association for Machine Translation in the Americas. AMTA 1994*, Columbia, Maryland, USA.
- Yang, Y., Liu, R., Qian, X., & Ni, J. (2023). Performance and perception: Machine translation post-editing in Chinese-English news translation by novice translators. *Humanities and Social Sciences Communications*, 10(1). <https://doi.org/10.1057/s41599-023-02285-7>
- Yang, Y., & Wang, X. (2023). Predicting student translators' performance in machine translation post-editing: Interplay of self-regulation, critical thinking, and motivation. *Interactive Learning Environments*, 31(1), 340–354. <https://doi.org/10.1080/10494820.2020.1786407>
- Zhang, F. (2020). On Machine Translation Technology in Legal Translation. *Technology Enhanced Foreign Language Education.*, 1, 53–58.
- Zhong, S., & Liu, X. (2023). Cultural specificities of online dictionaries for English learners: Evidence from a user survey and a multimodal discourse analysis. *Digital Scholarship in the Humanities*, 38(4), 1740-1763. <https://doi.org/10.1093/llc/fqad055>