

Chất lượng dịch tiêu đề film tiếng Anh sang tiếng Việt của Google Translate

Trần Thị Thanh Huyền^{1,*}, Lê Nhân Thành²

¹Trường THPT dân tộc nội trú Đông Gia Lai, thị xã An Khê, tỉnh Gia Lai, Việt Nam

²Khoa Ngoại ngữ, Trường Đại học Quy Nhơn, Việt Nam

Ngày nhận bài: 09/04/2021; Ngày nhận đăng: 22/08/2021

TÓM TẮT

Google Translate (GT) - một công cụ dịch trực tuyến miễn phí - được sử dụng ngày càng phổ biến nhưng chất lượng dịch của GT chưa thật sự tốt. Bài báo này trình bày kết quả nghiên cứu chất lượng dịch của GT khi dịch 130 tiêu đề phim tiếng Anh sang tiếng Việt. Nghiên cứu này sử dụng phương pháp phân tích nội dung để phân tích dữ liệu theo hướng định tính và định lượng. Kết quả nghiên cứu cho thấy rằng mặc dù GT không mắc phải bất cứ lỗi hình vị nào, nhưng mắc 4 loại lỗi dịch thuật theo cách phân loại lỗi của Farrús và cộng sự⁹ với các mức độ rất khác nhau khi dịch tiêu đề phim. Trong tổng số 130 tiêu đề phim, GT chỉ dịch được 16 tiêu đề phim chính xác và mắc lỗi khi dịch 114 tên phim. Trong 4 loại lỗi, lỗi ngữ nghĩa chiếm nhiều nhất. Vị trí thứ hai và thứ ba lần lượt thuộc về các lỗi từ vựng và lỗi cú pháp. Lỗi hình thức đứng ở vị trí tiếp theo. Kết quả này cho biết rằng GT dịch tiêu đề phim tiếng Anh sang tiếng Việt không tốt. Cuối bài báo có các gợi ý và đề xuất dành cho những người sử dụng GT, những người phát triển GT và những người nghiên cứu về chất lượng dịch GT.

Từ khóa: Tiêu đề phim tiếng Anh, tiếng Việt, Google Dịch, lỗi dịch thuật, chất lượng dịch.

*Tác giả liên hệ chính.

Email: tranthithanhhuyenglai@gmail.com

The quality of Google Translate's Vietnamese translations of English film titles

Tran Thi Thanh Huyen^{1,*}, Le Nhan Thanh²

¹*Dong Gia Lai ethnic boarding high school, An Khe Town, Gia Lai province, Vietnam*

²*Department of Foreign Languages, Quy Nhon University, Vietnam*

Received: 09/04/2021; Accepted: 22/08/2021

ABSTRACT

Google Translate (GT), a free online translation tool, is increasingly used, but the translation quality of GT is not really good. This article presents the results of a study on the quality of GT in translating 130 English film titles into Vietnamese. The study used content analysis to analyze the data, both qualitatively and quantitatively. The results show that GT did not make any morphological errors, but it committed orthographic errors, lexical errors, semantic errors, and syntactic errors at very different rates according to Farrús et al.'s framework⁹ of translation errors. Out of 130 English film titles, GT transferred only 16 film titles correctly, and GT made 141 errors in the remaining 114 film titles. Among four error types, semantic errors are the most dominant. The second and third positions belong to lexical errors and syntactic errors, respectively. Orthographic errors take the last place. The result of the study indicates that GT fails to translate English film titles into Vietnamese ones. In other words, the translation quality of GT in translating film titles from English into Vietnamese is low. Implications for GT users, GT technicians, and other researchers are suggested.

Keywords: English film title, Vietnamese, Google Translate, translation error, translation quality.

1. INTRODUCTION

1.1. Introduction to GT

According to Wikipedia,¹ in 2006, GT was introduced as a statistical machine service. Because GT used the United Nations and European Parliament's transcripts as data, accuracy was not appreciated. Then GT switched to a new version of the system for machine-assisted language translation - Google Neural Machine Translation, which allows whole sentences to be translated with more diverse contexts. After that, GT arranges and adjusts the data to find the most suitable translations. Currently, GT uses the Neural Machine Translation system for most language pairs, and this system gives

more accurate results than other ones. However, the translation quality of GT depends on the source of the documents entered into the system. Accordingly, GT only has standard translation when it contains the data related to the requested translation contents. Therefore, to evaluate the translation quality of GT accurately, researchers need to study the translation quality of GT in many different fields. This study investigates the translation quality of GT in translating English film titles into Vietnamese.

1.2. Research into the translation quality of GT

Up to now, studies on the translation quality of GT have not been carried out in various fields.

*Corresponding author:

Email: tranthithanhhuyenglai@gmail.com

One of the studies on the quality of GT was conducted by Luong Kim Hoang.² The researcher investigated the common errors in Vietnamese - English translation of labels and captions in tourist attractions in Ho Chi Minh City, Vietnam. The researcher used the translation error classification of Dastjerdi and Abdolmaleki³ to examine GT's errors when it randomly translated 450 selected labels and captions. The findings show that 96.6% of the translation versions are incorrect and that there is no consistent pattern in the most common translation errors.

Another study investigating the quality of GT when translating English metaphors into Vietnamese was conducted by Huynh Ha Mi.⁴ The data were collected from the novel *Kafka on the Shore* by Murakami Haruki. The researcher employed the theory of Lakoff and Johnson to identify metaphors and the framework of Nord⁵ to discover the occurrences of translation errors. It can be seen from the findings that GT translates orientational metaphors better than other metaphor types, including structural metaphors and ontological metaphors.

Besides, Lu et al.⁶ translated ten common anesthetic pre-assessment questions in the medical and anesthetic history, and assessment of the airway from English into ten languages (Arabic, Filipino, French, German, Greek, Hindi, Italian, Polish, Spanish and Vietnamese) by using GT. They concluded that Spanish gets the most accurate translations of the questions with 80%, and Vietnamese gets the worst translations, with an accuracy rate of only 10%.

In general, studies on the translation quality of GT in different fields give different conclusions on the translation quality of GT. Therefore, GT's translation quality research should be done in a more diverse range of text to get an accurate overview of GT's translation in general. Accordingly, research on GT's translation quality in the entertainment industry, namely film titles, is needed to enrich the findings

of GT's quality and to enable GT's developers to have a more specific orientation in improving the quality of GT in a variety of fields.

1.3. Translation error classifications

Many researchers have been interested in translation errors. As a result, different translators have suggested various translation error frameworks.

One of the most popular frameworks of detailed error taxonomy in Machine Translation (MT) is proposed by Stymne and Ahrenberg⁷ with 10 error types: (1) ER - Error rate, (2) Ling - Linguistic categories, (3) GF - grammatical and function words, (4) Form, (5) POS+ - part-of-speech, (6) FA - fluency, adequacy, (7) Ser - serious, (8) Reo - reordering, (9) Index, (10) Other. Each error type includes sub-types with clear descriptions. However, this framework is not suitable for identifying film title translation errors. Film titles contain simple words, phrases, or sentences, so the framework with too many error types is too complicated to apply.

Another framework that also works with an inter-annotator agreement is Multidimensional Quality Metric.⁸ This framework concerns with accuracy and fluency and consists of many complex sub-types. Therefore, it is not easy to apply this framework to identify film title translation errors.

The linguistic-based evaluation criteria for identifying statistical MT errors put forward by Farrús et al.⁹ contain orthographic, morphological, lexical, semantic, and syntactic errors. *Orthographic errors* are the errors of punctuation, capitalization, and spelling. *Morphological errors* are related to the forms (i.e., inflections, often suffixes) of verbs, nouns, and others (adjectives and adverbs). *Lexical errors* include two sub-types: extra words and missing words. *Semantic errors* occur when the wrong meaning of a target word is chosen to render a source word. *Syntactic errors* have five sub-types, namely conjunction, preposition,

article, syntactic element reordering, and category errors. Due to the different characteristics of languages, it is not easy to find any framework suitable for all kinds of contexts. The framework suggested by Farrús et al.⁹ is not perfect, too. It does not contain pragmatic errors. However, Farrús et al.'s framework⁹ explains translation error types in detail and is easy to apply at the level of simple words, phrases, or sentences. Therefore, this study used this framework to examine translation errors in GT's translations of the English film titles into Vietnamese.

1.4. The characteristics of the film titles

According to Ailan,¹⁰ a film title exhibits linguistic, cultural, and aesthetic properties. The linguistic characteristics guarantee the clarity, accuracy, and direct disclosure of the film's content without too many words. The cultural characteristics shown in the film title must reveal the unique culture of each ethnic group. The aesthetic characteristics are reflected in the harmonious beauty of image, creativity, artistry, rhyme, and tone. Together with accuracy, the translation versions must ensure the characteristics of film titles.

2. METHODOLOGY

To obtain the data for the present study, 130 film titles in English were collected from 32 websites. Among these websites, <https://vi.wikipedia.org/wiki/> is the website where 43.8% of the film titles were collected. Then these film titles were translated into Vietnamese by GT, and the Vietnamese translations were analyzed to investigate GT's translation quality.

To prepare the data for analysis, a 4-column table with the following contents was drawn: column 1 for the ordinal number, column 2 for the English film titles, column 3 for GT's Vietnamese translations of the English film titles, and column 4 for suggested Vietnamese translations of the English film titles. Besides, the sources of the film titles were also included in the table.

To analyze the translation errors in the Vietnamese translations, the linguistic-based evaluation criteria for identifying statistical MT errors introduced by Farrús et al.⁹ were used. The errors were coded as follows.

In *orthographic errors*, *O1* stands for *punctuation*, *O2* for *capitalization* and *O3* for *spelling*.

In *morphological errors*, *M1* is *verb*, *M2* is *noun*, while *M3* is *other errors*.

In *lexical errors*, *an extra word* is coded as *L1*, and *a missing word* is coded as *L2*.

SE represents *a semantic error*.

In *syntactic errors*, *S1* stands for *a conjunction error*, *S2* for *preposition error*, *S3* for *article error*, *S4* for *syntactic element reordering error*, and *S5* for *a category error*.

Some translations had no errors. As a result, one more code, *NE*, was added for *No Error* cases.

The error types identified after the analysis were recorded in the fifth column. The following is an illustration of the data storage table.

Table 1. Data preparation table

No	English film title	GT's Vietnamese translation	Suggested Vietnamese translation	Error type
1	Blue Valentine	Valentine xanh	Lẽ tình nhân buồn	SE
Source: https://anninhthudo.vn/chieu-phim-de-cu-giai-oscar-le-tinh-nhan-buon-post423622.antd				
2	12 Years a Slave	12 năm nô lệ	12 năm nô lệ	NE
Source: https://laodong.vn/giai-tri/diem-lai-nhung-bo-phim-doat-giai-oscar-hay-nhat-trong-lich-su-659048.ldo				

3. RESEARCH RESULTS

The analysis of GT's Vietnamese translations of 130 English film titles showed that GT correctly translated 16 English film titles, accounting for 12.3%. This means that GT committed errors in the translations of 114 English film titles, making up 87.7%.

The Vietnamese translations of 114 English film titles had 141 translation errors because many of GT's Vietnamese translations had more than one translation error.

Specific numbers and frequencies of translation errors are presented in the following table.

Table 2. Numbers and frequencies of translation errors

No	Error types	Number of errors	Percentage (%)
1	Morphological errors	0	0
2	Orthographic errors	6	4.3
3	Syntactic errors	11	7.8
4	Lexical errors	20	14.2
5	Semantic errors	104	73.7
Total		141	100

Table 2 indicates that four broad error types exist, with a total of 141 instances. The error types take up very different rates. Among the errors, semantic errors are dominant, with 104 instances, accounting for 73.7%. The second and third positions are of lexical errors and syntactic errors with 20 (14.2%) and 11 instances (7.8%), respectively. Accordingly, lexical errors are more common than syntactic errors (20 versus 11 errors, respectively). Orthographic errors take the next place, with 4.3% (6 instances). GT does not commit any cases of morphological errors. This can be explained by the fact that there is no inflection in the Vietnamese language.

The results prove that GT is not successful in translating English film titles into

accurate Vietnamese ones because four types of translation errors, including orthographic errors, syntactic errors, lexical errors and semantic errors, appeared in its Vietnamese translations at high percentages. Following is a more detailed report of the findings.

3.1. No errors

Sixteen English film titles were translated into Vietnamese by GT with absolute acceptability, and these translations match the Vietnamese film title style. With the title "12 Years a Slave", GT delivered a perfectly accurate translation of "12 năm nô lệ". Or the title "Sorry to bother you" has an acceptable translation of *Xin lỗi đã làm phiền bạn*". One more example for a no-error case is the correct translation "Cuốn theo chiều gió" from "Gone With The Wind".

3.2. Morphological errors

Of all 141 translation errors made by GT in 114 inaccurate film title translations, there was no case of morphological errors. This result suggests that GT was free of morphological errors in its Vietnamese translations.

As Nguyen Phu Hoang Nhu¹¹ argued, there is a big morphological difference between English and Vietnamese as two languages. In English, suffixes are used to change the form of a word, often nouns, verbs, and adjectives. In contrast, Vietnamese does not have suffixes. As a result, GT and human translators do not commit morphological errors in their Vietnamese translations.

3.3. Orthographic errors

There are six orthographic errors, accounting for 4.3% of all the errors. All these six errors are capital errors. GT made no errors in punctuation and spelling. With this rate, orthographic errors come at the fourth position among five error types.

For example, "Jurassic Park" was translated into "công viên kỷ Jura". In this example, the letter *c* in "công" and *k* in "ky" do not follow the capitalization rule in Vietnamese.

GT's Vietnamese translation “*Người phụ nữ xinh đẹp*” from “*Pretty Woman*” does not obey Vietnamese standard capitalization rule. Only the letter *N* in “*Người*” in capitalization is enough for accuracy. “*Người phụ nữ xinh đẹp*” is the best translation for this title.

3.4. Syntactic errors

Syntactic errors were made at the third-highest rate out of five error types. GT did not make any mistakes when translating prepositions, conjunctions, and articles. However, syntactic element reordering errors accounted for 8 (5.7%) out of 11 cases (7.8%). And the category errors were present in only 3 instances (2.1%).

“When Harry Met Sally...” was rendered into “*Harry gặp Sally khi nào...*”. This Vietnamese translation sounds confusing, and it does not have a high aesthetic value. A better translation should be “*Khi Harry gặp Sally...*”.

Another example for syntactic element reordering errors is with the title “*People on Sunday*”. GT rendered it into “*Những người vào Chủ nhật*”. This Vietnamese version sounds odd to the ears of Vietnamese natives because of an erroneously syntactic ordering. A more appropriate translation should be “*Chủ nhật của mọi người*”.

When translating “10 Things I Hate About You” into “*10 điều tôi ghét về bạn*”, GT made a category error. *Bạn* and *tôi* do not sound as appropriate as *em* and *anh*.

3.5. Lexical errors

With 20 lexical errors, making up 14.2%, this error type ranked second among the error types committed by GT. Specifically, the rate of missing word errors was 4 times higher than that of extra word errors (16 errors vs 4 errors or 11.4% and 2.8%, respectively).

The title “*10 Things I Hate About You*” should be translated into “*10 điều khiên em ghét anh*”, but GT translated it into “*10 điều tôi ghét về bạn*”. The extra word “về” is used in this case, making the Vietnamese translation sound unnatural.

GT's Vietnamese translation “*Hoàng hôn*” from “*Sunset Boulevard*” indicates that GT did not translate “*Boulevard*”. The Vietnamese translation should be “*Dại lộ Hoàng hôn*”. In this case, a missing word error was committed by GT.

3.6. Semantic errors

Semantic errors were made by GT at the highest rate of 73.7%. The percentage of semantic errors are many times higher than those of the other error types. The translation of a film title requires much consideration of the content and context of the film as a whole. However, GT – a kind of translation machine – can not know these elements. Understandably, GT committed many errors of this type.

GT's translation of the film title “*The Godfather*” illustrates this error kind. According to the Oxford Learners' Dictionary at <https://www.oxfordlearnersdictionaries.com/>, Godfather has the meanings: a male godparent - (*cha đỡ đầu*) and a very powerful man in a criminal organization, especially the Mafia - (*bố già*). GT translated this film title into “*Cha đỡ đầu*” though, in this situation, it should be translated into “*Bố già*”, considering the film content.

Another example of semantic errors is GT's Vietnamese translation of “*All the Money in the World*”. “*Tất cả tiền trên thế giới*” is a literal translation of the English film title. The Vietnamese translation sounds natural, but it does not reflect the film's content because GT relies totally on the film title. Considering the film's content, human translators have translated the English film title into “*Vụ bắt cóc triệu đô*”.

4. CONCLUSION

The research results reveal that GT committed four broad error types in the framework introduced by Farrús et al.⁹ The highest rate of errors that GT made fell on the *semantic* errors. The second biggest rate belonged to the *lexical* errors. The *syntactic* errors and the *orthographic* errors were

in the third and fourth positions, respectively. However, GT did not commit morphological errors related to inflection because inflection does not exist in Vietnamese. The sub-types of errors, including spelling and punctuation errors of orthographic errors, prepositions, articles, and conjunctions of syntactic errors, were not found in the data of this present study.

To sum up, although GT did not produce morphological errors and some sub-types of orthographic errors and syntactic errors in translating the English film titles into Vietnamese, the percentage of translation errors was found in 87.7% of the Vietnamese translations. Up to 73.7% of the errors were semantic errors. This is understandable because the understanding of a film title depends a lot on the understanding of the film, but GT relies solely on the words in the title. Moreover, film titles are often phrases rather than full sentences. As a result, the linguistic context, which is often very important for translators to choose appropriate meanings for their translation, is not clear enough for GT to choose contextually proper meanings of the words in the English film titles for its Vietnamese translations.

GT is an online machine translation tool whose translation quality is subjective to the amount of text fed into it. The larger the amount and variety of texts are, the higher the accuracy is. This research motivates GT developers to be more concerned about GT's translation quality in film title translation. Besides, this study raises GT users' awareness in using GT to translate texts whose understanding depends too much on what must be known beyond the texts themselves. Also, researchers should examine GT's quality of English translations of Vietnamese book titles and article titles.

REFERENCES

1. Google Translate (n.d), https://en.wikipedia.org/wiki/Google_Translate, retrieved on September 21, 2020.
2. Lương Kim Hoàng. *Common errors in Vietnamese - English translation of labels and captions in tourist attractions in Ho Chi Minh City, Vietnam*, MA thesis, Ho Chi Minh City University of Technology, 2008.
3. H. V. Dastjerdi & S. D. Abdolmaleki. A study of translation problems of tourism industry guidebooks: An error analysis perspective, *International Journal of Foreign Language Teaching and Research*, 2002, 1(1), 71-82.
4. Huỳnh Hà Mi. *The quality of Google Translate's translations of metaphors in "Kafka on the Shore" by Murakami Haruki into Vietnamese*, MA thesis, Quy Nhon University, 2020.
5. Nord, C. *As a purposeful activity: Functional approaches explained*, Manchester: St. Jerome Publishers, 1997.
6. N. Nguyen-Lu, P. Reide & S. M. Yentis. Do you have a stick in your mouth? - use of Google Translate as an aid to anaesthetic pre-assessment, *Journal of the Association of Anaesthetists of Great Britain and Ireland*, 2010, 65, 94-133.
7. S. Stymne & L. Ahrenberg . *On the practice of error analysis for machine translation evaluation*, the eighth international conference, European Language Resources Association (ELRA), 2012.
8. A. Lommel. *Multidimensional quality metrics (MQM) issue types: draft 2018-10-04*, W3C Community and Business Groups, 2018.
9. M. Farrús, M. R. Costa-Jussà, J. B. Mariño, & J. A. R. Fonollosa. *Linguistic-based evaluation criteria to identify statistical machine translation errors*, 14th Annual Conference, the European Association for Machine Translation, 2010.
10. D. Ailan. A study of film title translation from the perspective of Peter Newmark's communicative translation theory, *CSCanada*, 2016, 13(3), 32-37.
11. Nguyễn Phú Quỳnh Như. *A quick analysis of some typical gaps in phonology and morphology between English and Vietnamese that lead to Vietnamese students' common errors*, <http://nnkt.ueh.edu.vn/wp-content/uploads/2015/06/21.pdf>, retrieved on July 19, 2011 .