

Original Research Paper

# Translation Quality Assessment of Online Translation Systems in Translating English to Thai on Phuket Tourism

Kitsiri Chochiang, Thanakrit Thongkhamdee and Lalita Sathansat

College of Computing, Prince of Songkla University, Phuket Campus, Thailand

## Article history

Received: 24-06-2020

Revised: 05-11-2020

Accepted: 09-11-2020

Corresponding Author:

Kitsiri Chochiang

College of Computing, Prince of Songkla University, Phuket Campus, Thailand

Email: kitsiri.c@phuket.psu.ac.th

**Abstract:** This research aims to compare four English-to-Thai translator applications on translating the comments posted on social media. Four translator applications include Google Translate, Baidu Translate, Yandex.Translate and Microsoft/Bing Translate. The comments in this study are on three main popular attractions in Phuket, Thailand. The attractions include a temple (Wat Chalong), a viewpoint spot (Promthep Cape) and a beach (Patong beach). Fifty comments on each attraction are collected from GoogleReviews and TripAdvisor. The total of 300 comments are translated by four translators making the final translated comments of 1,200 comments. All translated comments are evaluated by three language experts on three categories including the grammar structure, the vocabulary and idiom and the overall translation. The evaluation results show that Google Translate is found to be the most suitable translator for translating English comments to Thai comments in this study at 89.33%. Google Translate receives the average suitable score of 4.20 out of 5.00. However, Google Translate is weak at word choices to generate a beautiful sentence. Baidu Translate is very superb at generating a beautiful sentence similar to human languages. However, Baidu Translate only gets a good result on a specific domain or situation. As a result, Baidu Translate is lost to Google Translate in most cases. Even though this research presents the results on Thai language and specific domains, the techniques used in this study can be applied to any language on any domain.

**Keywords:** Google Translate, Baidu Translate, Yandex.Translate, Microsoft/Bing Translate

## Introduction

Tourism industry contributes significantly to the Thai economy. Therefore, any improvement to facilitate a tourism industry will have a significant impact on the Thai economy as a whole. With rich natural resources in all regions, Thailand is a country suitable for tourism especially, Phuket. Phuket has been known as a pearl of Andaman consisting of many beautiful beaches, historical sites, cultural and religion sites. Beaches in Phuket have been certified as one of the most beautiful beaches in the world by many international awards (Polnyotee and Thadaniti, 2015). Many activities can be conducted in Phuket such as shopping, scuba diving, jet skiing and swimming.

Nowadays, the Internet becomes essential to many tourism industry revolutions and changes. Tourists' behaviors are changing due to the influence of Internet technology. With the information available on the

Internet, tourists can gain more pleasure, comfort and flexibility during their trips (Chochiang, 2017). Social media takes a great part of providing such information to tourists. The tourists can search for recommendations of the place and the activities or compare the rating and reviews; finally, the tourists can also share their experiences online (Hajli, 2014). The customer feedbacks via the reviews on the social media is the key importance factor. A good feedback will attract more customers. Moreover, the tourism industries can also build their attractive social profile from their positive customer feedback (Ba and Pavlou, 2002).

Typically, the customers provide their reviews in a text-based format and most traditional statistical methods do not have an analytical capability to analyze such information (Taecharungroj and Mathayomchan, 2019). Feedback on the tourism industry is typically either negative or positive based on the customer experiences.

According to the records on TripAdvisor-one of the travel websites for travelers looking for data-there are 859 million people conducting user reviews and user opinions in 2019. While, the number of user reviews and user opinions based on <https://www.statista.com> only 200 million people in 2014. Therefore, the increasing of the number shows that tourists are eager to share their information online more often and such feedback can be either negative or positive according to their experiences.

Each language contains its culture of the language owner. By speaking only one language a loss of cultural diversity occurs (Müller *et al.*, 2016). Thailand is an Asian country that has its own identity with many tourism characteristics including language, traditions and culture. Thai language is an official language in Thailand without any secondary language. Therefore, Thai people speak Thai as their main language. However, Phuket receives many international tourists. English language can be an obstacle to tourism industry workforce in Thailand because not everyone has an opportunity to be trained and has suitable English language skills. Multiple solutions exist to overcome this language barrier. Simultaneous translation is one of the promising solutions. Having a good translator will help you to learn and to practice understanding of vocabulary or sentence patterns even more. Also, a good translator can reduce the time it takes to search for information.

Thus, this research aims to analyze and compare many existing English-to-Thai language translators on the tourism industry related reviews. As a result, the employees, entrepreneurs, business owners can better understand the meaning that the foreign tourists communicate via their feedback or reviews. These data can be further analyzed in many ways such as sentiment analysis, analysis of strengths and weaknesses of the tourist spots or hotels or suggestion of improvements provided by reviews. This study is conducted as the first step to our final objective in developing an automatic review translation system for tourism industries in Thailand.

The remaining of this study is organized as follows. Next, the review of literature is given. Then, the design choice and the tools of the proposed application is

described. The resulting application is then discussed. Last, the conclusion is given.

## Literature Review

### *Quality of Translation Assessment*

Translation is an action to convey the meaning of messages from one language to another. Figure 1 shows the translation process. Translation is an understanding of the meaning of the text and later creating an equivalent message; "translation" is for communicating the same message in other languages (Kull, 1998).

There exist many important criteria for determining the quality of the translation (Intakosum, 2014) such as the emphasis on the meaning in which the translation must convey the meaning and the same flavor as the original. The interpretation of the meaning must be complete and unambiguous. In most cases, translation errors are related to grammar structure errors, in which the translator does not understand what the sentence structures or words do or how they relate to each other. Furthermore, any mistake in vocabulary and expression used in which the translator is not careful then polysemy or homographic can also be observed.

### *Translation Technologies*

The development of computerized translation tools has not stopped while the cloud-based translation technologies are emerging. With cloud-based translation technologies, a complex and highly digitized environment for the latest translation technology can be accessible without the need to own an expensive computer with a fast processor and a large amount of disk space (Muegge, 2012). A software that allows translators to translate written text from one natural language (original language) into a text in another natural language (target language) is known as a translation tool (Kastberg and Andersson, 2012). Existing translator tools that are well known to translate English language to Thai language include Baidu Translate; Microsoft/Bing Translate; Google Translate; Yandex.Translate.

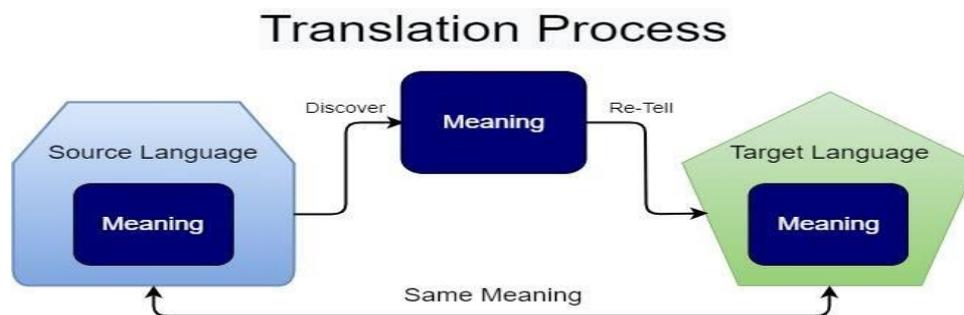


Fig. 1: Translation process (Door43, 2020)

Baidu Translate is a software application search engine service provider from China. Baidu machine translation system was released in June 2011. Currently, Baidu supports 27 languages on multiple platforms including personal computers, mobile devices and a set of APIs for developers (He, 2015). The underlying technology used by Baidu is deep learning technologies. The machine translation tool processes approximately 100 million requests every day. Similar to Google Translate, Baidu translate supports Optical Character Recognition (OCR) and Baidu can also "translate" the object. By pointing the camera at a plant, the plant will be labeled as "plant" in the user preferred language (Sun, 2014).

Microsoft/Bing Translate is a free translator supported by multiple devices running iOS and Android. The underlying technology used by Microsoft/Bing Translate is a statistical machine translation. More than 60 languages are supported by Microsoft/Bing Translate. Various formats are supported such as text, voice, conversations, camera, photos and screenshots. Cutting-edge technology that powers Microsoft/Bing Translate is also powering many other world renounce businesses such as Office, Bing, Skype, Internet Explorer, Yelp, eBay and WeChat.

Microsoft/Bing Translate and Google Translate have been compared in translating English to Spanish language in (Martin, 2017) and Microsoft/Bing Translate has been found to produce more errors in its translation than that of Google Translate. Also, Microsoft/Bing Translate was reported to have an issue in translating specialized text. On the other hand, the comparison of Google Translate and Microsoft/Bing Translate in translating English to Hindi and vice versa in (Dhakar *et al.*, 2013) found that Microsoft/Bing Translate produced a better translation quality.

Google Translate was published at the end of 2008 by Google. The development of PHP and AJAX allows Google Translate to produce a media of translation called translator. "Google Translate" is a medium of online translation service. Google Translate is supporting more than 50 languages (Ulfah, 2015). The underlying technology of Google Translate is also statistical machine translation which is a process of seeking patterns in a large number of texts in order to find a sentence close to a human language. The technology is based on the training of statistical models from the large corpora of human translation. This technology has an advantage of fast training if there is a corpora available when comparing to the rule-based system. The technology has been known to be quite good at solving vocabulary problems (Vidhayasai *et al.*, 2015). Google Translate has been shown to be feasible in translating medical articles in many cases and the process is not a resource-intensive (Balk *et al.*, 2012). However, the translation quality relied heavily on the article's original language. In particular, Roman languages have a higher level of agreement than that of Asian languages and Hebrew.

In early 2011, Yandex implemented a machine translation system. The underlying technology of Yandex.Translate is not based on language rules. Yandex.Translate also uses statistical method and supports major European languages. Yandex.Translate is also supporting Thai language via Yandex Translation API. Yandex.Translate supports both free-form texts and the whole web pages.

## Methodology

The performance of four translators including Google Translate (denoted Translator\_A), Microsoft/Bing Translate (denoted Translator\_B), Yandex.Translate (denoted Translator\_C) and Baidu Translate (denoted Translator\_D), on translating English language reviews to Thai language reviews are evaluated in this study.

### Workloads

To evaluate the four translators, the datasets of Phuket main attractions were collected from GoogleReviews and TripAdvisor. GoogleReviews and TripAdvisor are online-based websites that allow users to review their experience from the real attractions. Both GoogleReviews and TripAdvisor are big and popular travel websites for tourists to visit in order to search for data and to plan their trips. It had been shown in numerous previous studies on traveler behavior that reviews clearly influenced travelers' booking decisions (Xie *et al.*, 2016; Nilashi *et al.*, 2018).

Three domains of attractions including Temple (Wat Chalong), Viewpoint (Promthep Cape) and Beach (Patong beach) are selected in this study. Each domain is very distinct and unique in itself. Fifty comments were collected from each source for each attraction domain as shown in Table 1. Only English comments were used in this study. The data sampling method in this study is selecting fifty comments dated back from January 2020 until reaching 50 comments on each attractions and each source; only English comments were collected. As a result, the reviews of Temple were collected from August 2019 to January 2020. The reviews of Viewpoint were collected from March 2019 to January 2020 and the reviews of Beach were collected from September 2019 to January 2020.

According to the dataset characteristics shown in Table 1, the average number of sentences and words of the whole dataset is at 174.50 sentences and 2,126.83 words. While, the average number of sentences and words in each review is at 3.49 sentences and 42.54 words. For the characteristics of TripAdvisor datasets, the average number of sentences and words is at 178 sentences and 2,204.67 words. For the characteristics of GoogleReviews datasets, the average number of sentences and words is at 171 sentences and 2,049 words.

### Evaluation Methods

Figure 2 shows the evaluation method in this study. There are fifty comments of each attraction from two sources making the total comments of 300 English comments. Each of the four translators translates all 300 English comments in Thai comments resulting in 300 translated Thai comments for each translator. Thus, the total Thai translated comments from all four translators is 1,200. The translation performance of all four translators are evaluated by three language experts. Each expert is a respective faculty member of a Thai University. Each expert has expertise in Educational and Applied Linguistics (English as an International Language).

Each expert must evaluate all 1,200 Thai translated comments. For fairness reasons, the real identity of the translator is not known to the experts. Thus, the experts will be given the comments with the information of the alias names of the translators. Thus, Google Translate alias name is Translator\_A; Microsoft/Bing Translate alias name is Translator\_B; Yandex.Translate alias name

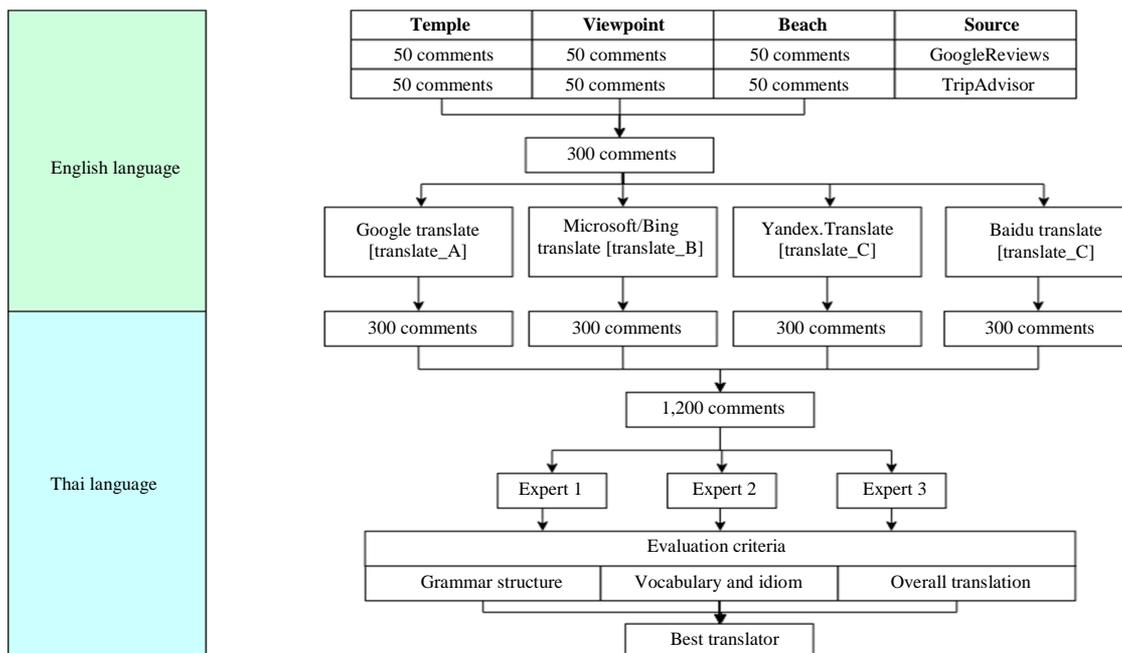
is Translator\_C; Baidu Translate alias name is Translator\_D. Therefore, each expert will evaluate the same set of comments-1,200 Thai translated comments.

The expert must evaluate the translated Thai comments on three categories, including grammar structure, vocabulary and idiom and overall translation. Each category consists of several qualities. These qualities are compiled from translation issues and good features of translation assessments in the literature (Intakosum, 2014; Chanwaiwit, 2016). The evaluations include accuracy, appropriateness and cohesion and focus on examining common grammar structure, vocabulary and idiom problems.

The first category is the grammar structure which consists of 3 qualities including (1) meet the grammar structure completely, (2) has consistency of words and has a coherent relationship in each sentence and (3) be able to communicate appropriately. These three qualities will enable the researcher to evaluate the grammatical structure covering word consistency, interpretation and the complete grammar structure.

**Table 1:** The detail of the datasets

Sources	Domain	Number of sentences	Number of words
TripAdvisor	Temple (Wat Chalong)	165	1915
	Viewpoint (Promthep Cape)	183	2406
	Beach (Patong beach)	186	2293
GoogleReviews	Temple (Wat Chalong)	179	2280
	Viewpoint (Promthep Cape)	170	2109
	Beach (Patong beach)	164	1758
	Total	1047	12761
Average/dataset		174.50	2126.83
Average/comment		3.49	42.54



**Fig. 2:** Evaluation method

The second category is the vocabulary and idiom which consists of 2 qualities including (1) choose a correct vocabulary and (2) is clear and complete content. These two qualities will enable the researcher to know that the chosen vocabulary matches the meaning and complete with no missing parts.

The last category is the overall translation which consists of 2 qualities including (1) is the same or similar to the original and (2) give a feeling, mood, pleasure in reading equal to the original. Both qualities will focus on the translation that must still have the same meaning and clarity as the original.

The evaluation score of each expert on each comment is on a scale of 1 to 5 with the following meaning:

1. Means the text is clear and does not require corrections
2. Means the text contains minor errors
3. Means the meaning of the text is reduced to «gist» but is understandable
4. Means the text is barely comprehensible
5. Means the text is incomprehensible

## Results and Discussion

The comments in this study were collected from three attractions in Phuket including a temple (Wat Chalong), a viewpoint (Promthep Cape) and a beach (Patong beach). Each place consists of two dataset from two sources including GoogleReviews and TripAdvisor. Each data set contains 50 comments. For each comment, there are four translated comments resulting from four translators (i.e., Translator\_A, Translator\_B, Translator\_C and Translator\_D). There are three experts (P1, P2 and P3). All three experts must evaluate all 1,200 translated comments on three categories including grammar structure, vocabulary and idiom and overall translation. Then, the expert will select one of the translators in which he/she found to best translate the comments.

During the evaluation process, each expert will give a score on each translator ability in translating each comment. The score is ranged between 1 and 5. The score will be given in three categories, include grammar structure, vocabulary and idiom and overall translation. Moreover, the expert will also give the detail analysis on his/her decision. To analyze further, the expert choice on each comment is then studied. To calculate the best achievable performance to be used as the baseline for analysis, the performance of the best translator on each comment is presented as the best achievable performance. However, the winning translator must be selected. Thus, the winning translator for each comment is selected based on the voting of three expert choices. That is, if two or more experts select Translator\_A for that comment, Translator\_A will be selected as the winning translator for that comment. The results of the winning translator for each data set is also presented and discussed.

**Table 2:** The average score of the best achievable performance on temple datasets

Topics	GoogleReviews	TripAdvisor
1. Grammar structure	4.20	4.25
2. Vocabulary and idiom	3.96	3.91
3. Overall translation	4.10	4.11
Average	4.09	4.09

### Evaluation Results on Temple Dataset

For the temple (Wat Chalong) dataset, the average number of sentences for each comment is 3.44 while the average number of words for each comment is 41.95. Mostly, the comments are related to the beauty of Buddha statue, the temple history and the dress code of the visitors. Table 2 shows the average score of the best achievable performance on three categories given by three experts on the temple dataset. According to the results shown in Table 2, the score on the vocabulary and idiom is lower than the other two categories. The reason behinds this result is that the temple dataset is related to a specific place which has a specific set of vocabularies. For example, “Moderate dress is required” has been translated by all translators to “สวมใส่ชุดปานกลาง” (suamsai chut panklang) meaning “wear medium set” which is clearly incorrect.

Table 3 shows the average score of the best achievable performance on each category of temple datasets. According to the results shown in Table 3, the translated comments mostly met the grammar structure completely. Also, the translated comments have a consistency of words and have a coherent relationship in each sentence. As a result, the translated sentence gives the correct meaning. For example, “Wat Chalong is one of the most impressive and most popular temples in Phuket.” is translated to “วัดฉลองเป็นหนึ่งในวัดที่น่าประทับใจและเป็นที่ยอมรับมากที่สุดในภูเก็ต” (wat chalong pen nueng nai wat thi na prathapchai lae pen thi niyom mak thisut nai phu ket) which is “Chalong Temple is one of the most impressive and popular temples in Phuket.”. As can be seen that the translation is correct and complete. The translated sentence has a consistency of words including various connecting words. The reader can understand the translated sentence.

Table 4 shows the results of each expert on his/her selection of the translator on the temple datasets. There are 50 English comments from GoogleReviews and TripAdvisor each. The expert must select the best translator for each of the 50 comments. According to the results shown in Table 4, the results of all three experts are quite unanimous. That is, Translator\_C is not selected by any expert. All three experts select Translator\_A the most followed by Translator\_D and Translator\_B. For GoogleReviews dataset, the experts select Translator\_A 72.00% (i.e., 108 comments), Translator\_D 24.67% (i.e., 37 comments) and Translator\_B 3.33% (i.e., 5 comments). For TripAdvisor dataset, the experts select Translator\_A

89.33% (i.e., 134 comments), Translator\_D 8.67% (i.e., 13 comments) and Translator\_B 2.00% (i.e., 3 comments). In total, Translator\_A is selected for 242 comments or 80.67% which is significantly larger than the other three translators.

In conclusion, the number of comments winning by Translator\_A-the winning translator selected by the experts-for the GoogleReviews temple dataset is 38 out of 50 comments or 76%. The number of comments winning by Translator\_A for the TripAdvisor temple dataset is 48 out of comments or 96%. The number of comments wining by Translator\_A for both datasets 86 out of 100 comments or 86%.

Furthermore, there are two comments in which Translator\_A is not selected due to the translation error made by Translator\_A. Both comments are perfectly translated by Translator\_D. For example, “Completely free of cost!” is being translated to “เสียค่าใช้จ่ายโดยสิ้นเชิง” (sia khachaichai doisinchoeng) by Translator\_A which means “waste the cost altogether”. On the other hand, Translator\_D translates the sentence to “ฟรีทั้งหมด” (fri thangmot) which means “free of charge”. As can be seen that Translator\_D is correctly translating the sentence.

#### Evaluation Results on Viewpoint Dataset

For the viewpoint (Promthep Cape) dataset, the average number of sentences for each comment is 3.53 while the average number of words for each comment is 45.14. Mostly, the comments are related to landscape, sunset and views. Table 5 shows the average score of the best achievable performance on three categories given by three experts on viewpoint dataset. Similar to the temple dataset, the average scores on grammar structure category is the highest while the average score on the vocabulary and idiom is the lowest. However, the score on the vocabulary and idiom of the viewpoint dataset is

higher than that of the temple dataset. The reason behind this result is that the viewpoint dataset do not contain as much specific vocabularies.

Table 6 shows the average score of the best achievable performance on each category of the viewpoint dataset. According to the results shown in Table 6, the translated comments mostly met the grammar structure completely. Also, the translated comments are the same or similar to the original comments. The average scores on the viewpoint dataset are higher than their counterpart on the temple dataset (Table 2). The reason behind this result is that the vocabularies of the viewpoint dataset are mostly general vocabularies. Unlike the temple dataset, there is no specific set of vocabularies in the viewpoint dataset. Mostly, the comments describe the beauty of nature which is general.

Table 7 shows the results of each expert on his/her selection of the translator on the viewpoint dataset. There are 50 English comments from GoogleReviews and TripAdvisor each. The expert must select the best translator for each of the 50 comments. According to the results shown in Table 7, the results of all three experts are quite unanimous. That is, all three experts do not select Translator\_C. Similar results are observed here. That is, all three experts select Translator\_A the most followed by Translator\_D and Translator\_B. For the GoogleReviews dataset, the experts select Translator\_A 78.00% (i.e., 117 comments), Translator\_D 14.67% (i.e., 22 comments) and Translator\_B 7.33% (i.e., 11 comments). For the TripAdvisor dataset, the experts select Translator\_A 81.33% (i.e., 122 comments), Translator\_D 16% (i.e., 24 comments) and Translator\_B 2.67% (i.e., 4 comments). In total, Translator\_A is selected for 239 comments or 79.67% which is significantly larger than the other three translators.

**Table 3:** The average score of the best achievable performance on each category of temple datasets

Topics	GoogleReviews	TripAdvisor	Average
Grammar structure			
1. Met the grammar structure completely	4.38	4.62	4.50
2. Had consistency of words, Had a coherent relationship in each sentence	4.25	4.21	4.23
3. Able to communicate appropriately	3.98	3.91	3.95
Vocabulary and Idiom			
1. Choose a correct vocabulary	3.87	3.71	3.79
2. Is clear and complete content	4.04	4.11	4.08
Overall translation			
1 .Is the same or similar to the original	4.13	4.17	4.15
2 .Give a feeling, mood, pleasure in reading equal to the original	4.07	4.04	4.06

**Table 4:** The results of each expert on his/her selection of the translator on both data sets

No.	Temple_GoogleReviews					Temple_TripAdvisor				
	P1	P2	P3	Total	%	P1	P2	P3	Total	%
Translator_A	48	29	31	108	72.00	50	44	40	134	89.33
Translator_B	0	2	3	5	3.33	0	2	1	3	2.00
Translator_C	0	0	0	0	0.00	0	0	0	0	0.00
Translator_D	2	19	16	37	24.67	0	4	9	13	8.67
Summary	50	50	50	150	100.00	50	50	50	150	100.00

**Table 5:** The average score of the best achievable performance on viewpoint dataset

Topics	GoogleReviews	TripAdvisor
1.G rammar structure	4.35	4.34
2. Vocabulary and idiom	4.06	4.13
3. Overall translation	4.20	4.29
Average	4.20	4.25

**Table 6:** The average score of the best achievable performance on each category of the viewpoint dataset

Topics	GoogleReviews	TripAdvisor	Average
Grammar structure			
1. Met the grammar structure completely	4.66	4.6	4.63
2. Had consistency of words, Had a coherent relationship in each sentence	4.37	4.37	4.37
3. Able to communicate appropriately	4.01	4.04	4.03
Vocabulary and Idiom			
1. Choose a correct vocabulary	3.88	3.95	3.92
2. Is clear and complete content	4.24	4.30	4.27
Overall translation			
1. Is the same or similar to the original	4.29	4.36	4.33
2. Give a feeling, mood, pleasure in reading equal to the original	4.10	4.21	4.16

**Table 7:** The results of each expert on his/her selection of the translator on both viewpoint sets

No.	Viewpoint_GoogleReviews					Viewpoint_TripAdvisor				
	P1	P2	P3	Total	%	P1	P2	P3	Total	%
Translator_A	48	44	25	117	78.00	50	38	34	122	81.33
Translator_B	1	3	7	11	7.33	0	2	2	4	2.67
Translator_C	0	0	0	0	0.00	0	0	0	0	0.00
Translator_D	1	3	18	22	14.67	0	10	14	24	16.00
Summary	50	50	50	150	100.00	50	50	50	150	100.00

In conclusion, the number of comments winning by Translator\_A-the winning translator selected by the experts-for the GoogleReviews viewpoint dataset is 45 out of 50 comments or 90%. The number of comments winning by Translator\_A for the TripAdvisor viewpoint dataset is 42 out of comments or 84%. The number of comments wining by Translator\_A for both datasets is 87 out of 100 comments or 87%.

Furthermore, there are two comments in which Translator\_A is not selected due to the translation error made by Translator\_A. Both comments are perfectly translated by Translator\_D. For example, “but never stopped us from squeezing to the cliff front to have a good view” is being translated to “แต่ไม่เคยหยุดเราจากการปีนหน้าผาเพื่อให้ได้มุมมองที่ดี” (tae mai khoei yut rao chak kan bip napha phuea hai dai mum mong thi di) by Translator\_A which means “But never stopped us from squeezing the cliff for a good view”. On the other hand, Translator\_D translates the sentence to “แต่ไม่เคยหยุดเราจากการเข้าถึงหน้าผาเพื่อเพลิดเพลินกับทัศนียภาพที่สวยงาม” (tae mai khoei yut rao chak kan khaothueng napha phuea phloetphloen kap thatsaniyaphap thi suai-ngam) which means “But never stopped us from reaching the cliffs to enjoy the beautiful scenery”. As can be seen that Translator\_D is correctly translating the sentence with a set of beautiful words.

Another example, “Very honestly, it’s a brilliant sunset point and that’s about it” is being translated to

“ตรงไปตรงมามากมันเป็นจุดชมพระอาทิตย์ตกที่สวยงามและที่ก็ยากช้องกับมัน” (trong pai trong ma mak man pen chut chom phra athit tok thi suai-ngam lae thi kiaokhong kap man) by Translator\_A which means “Very straightforward, it was a beautiful sunset point and related to it”. While, Translator\_D translates the sentence to “บอกตามตรงมันเป็นจุดที่พระอาทิตย์ตกที่ยอดเยี่ยมแค่นั้น” (bok tam trong man pen chut thi phra athit tok thi yotyiam khaenan) which means “To be honest, it was just that wonderful sunset point”. Even Though Translator\_A is correctly translated the sentence but Translator\_D produces more pleasure in reading the sentence than that produced by Translator\_A.

#### Evaluation Results on Beach Dataset

For the beach (Patong Beach) dataset, the average number of sentences for each comment is 3.50 while the average number of words for each comment is 40.51. Mostly, the comments are related to the beach, location and activities, with both positive and negative comments.

Table 8 shows the average score of the best achievable performance on three categories given by three experts on the beach dataset. One interesting finding from the beach data is that the average score on GoogleReviews dataset is higher than that of the TripAdvisor dataset. The cause of this result is that the comments on TripAdvisor dataset contain incomplete sentences and slangs.

Table 9 shows the average score of each category in detail for both beach datasets. According to the results shown in Table 9, the translated comments mostly met the grammar structure completely. However, the word choice seems to trouble the translators in this case. The main reason is that many words can have various meanings. Selecting the vocabularies that are suitable for the context is the key to success. Moreover, some words may be in a metaphorical manner. For example, all translators translate “The water isn’t really clear” to “น้ำไม่ชัดเจน” (nam mai chatchen) which means “water is not obvious”. Another example, “Local sellers are heavy talkers and sticky” is being translated to “ผู้ชายในพื้นที่เป็นนักพูดหนักและเหนียว” (phukhai nai

phuenthi pen nak phut nak lae niao) by Translator\_A which means “Local vendors are heavy speakers and gummy”. While the sentence is being translated to “ผู้ชายท้องถิ่นเป็นคนช่างพูดเหนียว” (phukhai thongthin pen khon chang phut niao) by Translator\_D which means “Local seller is a talkative gummy person”.

**Table 8:** The average score of the best achievable performance on beach dataset

Topics	GoogleReviews	TripAdvisor
1. Grammar structure	4.47	4.26
2. Vocabulary and Idiom	4.26	4.02
3. Overall translation	4.33	4.28
Average	4.35	4.18

**Table 9:** The average score of each category in detail for both beach datasets

Topics	GoogleReviews	TripAdvisor	Average
Grammar structure			
1. Met the grammar structure completely	4.62	4.44	4.53
2. Had consistency of words, Had a coherent relationship in each sentence	4.49	4.27	4.38
3. Able to communicate appropriately	4.31	4.07	4.19
Vocabulary and Idiom			
1. Choose a correct vocabulary	4.11	3.85	3.98
2. Is clear and complete content	4.41	4.18	4.30
Overall translation			
1. Is the same or similar to the original	4.35	4.29	4.32
2. Give a feeling, mood, pleasure in reading equal to the original	4.26	4.26	4.26

**Table 10:** The results of each expert on his/her selection of the translator on both beach sets

No.	Beach_GoogleReviews					Beach_TripAdvisor				
	P1	P2	P3	Total	%	P1	P2	P3	Total	%
Translator_A	45	46	38	129	86.00	50	50	34	134	89.33
Translator_B	0	0	2	2	1.33	0	0	1	1	0.67
Translator_C	0	0	0	0	0.00	0	0	0	0	0.00
Translator_D	5	4	10	19	12.67	0	0	15	15	10.00
Summary	50	50	50	150	100.00	50	50	50	150	100.00

**Table 11:** The performance of Translator\_A against the performance of the best achievable performance

Datasets	Topics	No. Of comments	Grammar structure			Vocabulary and idiom		Overall translation	
			Q.1	Q.2	Q.3	Q.4	Q.5	Q.6	Q.7
Temple_Googlereviews	The best answer	150	4.38	4.25	3.98	3.87	4.04	4.13	4.07
	Translator_A	108	4.31	4.17	3.85	3.73	3.94	4.05	3.99
	Difference	<b>42</b>	<b>0.07</b>	<b>0.08</b>	<b>0.13</b>	<b>0.14</b>	<b>0.10</b>	<b>0.08</b>	<b>0.08</b>
Temple_Tripadvisor	The best answer	150	4.62	4.21	3.91	3.71	4.11	4.17	4.04
	Translator_A	134	4.64	4.22	3.91	3.69	4.12	4.17	4.03
	Difference	<b>16</b>	<b>-0.02</b>	<b>-0.01</b>	<b>0.00</b>	<b>0.02</b>	<b>-0.01</b>	<b>0.00</b>	<b>0.01</b>
Viewpoint_Googlereviews	The best answer	150	4.66	4.37	4.01	3.88	4.24	4.29	4.1
	Translator_A	117	4.68	4.36	4.02	3.88	4.26	4.27	4.09
	Difference	<b>33</b>	<b>-0.02</b>	<b>0.01</b>	<b>-0.01</b>	<b>0.00</b>	<b>-0.02</b>	<b>0.02</b>	<b>0.01</b>
Viewpoint_Tripadvisor	The best answer	150	4.6	4.37	4.04	3.95	4.3	4.36	4.21
	Translator_A	122	4.64	4.35	4.01	3.93	4.36	4.37	4.20
	Difference	<b>28</b>	<b>-0.04</b>	<b>0.02</b>	<b>0.03</b>	<b>0.02</b>	<b>-0.06</b>	<b>-0.01</b>	<b>0.01</b>
Beach_Googlereviews	The best answer	150	4.62	4.49	4.31	4.11	4.41	4.35	4.3
	Translator_A	129	4.65	4.46	4.26	4.07	4.41	4.35	4.26
	Difference	<b>21</b>	<b>-0.03</b>	<b>0.03</b>	<b>0.05</b>	<b>0.04</b>	<b>0.00</b>	<b>0.00</b>	<b>0.04</b>
Beach_Tripadvisor	The best answer	150	4.44	4.27	4.07	3.85	4.18	4.29	4.26
	Translator_A	134	4.50	4.29	4.08	3.88	4.22	4.33	4.30
	Difference	<b>16</b>	<b>-0.06</b>	<b>-0.02</b>	<b>-0.01</b>	<b>-0.03</b>	<b>-0.04</b>	<b>-0.04</b>	<b>-0.04</b>

Table 10 shows the results of each expert on his/her selection of the translator on the beach dataset. There are 50 English comments from GoogleReviews and TripAdvisor each. The expert must select the best translator for each of the 50 comments. According to the results shown in Table 10, the results of all three experts are quite unanimous. That is, all three experts do not select Translator\_C. All three experts select Translator\_A the most followed by Translator\_D and Translator\_B. For GoogleReviews beach dataset, the experts select Translator\_A 86.00% (i.e., 129 comments), Translator\_D 12.67% (i.e., 19 comments) and Translator\_B 1.33% (i.e., 2 comments). For TripAdvisor dataset, the experts select Translator\_A 89.33% (i.e., 134 comments), Translator\_D 10.00% (i.e., 15 comments) and Translator\_B 0.67% (i.e., 1 comments). In total, Translator\_A is selected for 263 comments or 87.67% which is significantly larger than the other three translators.

According to the results shown in Table 10, Translator\_A has been selected for almost 100% due to its accuracy in translating the comments in the beach datasets. Especially, Translator\_A gets almost all short sentences translated correctly.

In conclusion, the number of comments winning by Translator\_A-the winning translator selected by the experts-for the GoogleReviews beach dataset is 45 out of 50 comments or 90%. The number of comments winning by Translator\_A for the TripAdvisor beach dataset is 50 out of 50 comments or 100%. The number of comments winning by Translator\_A for both datasets is 95 out of 100 comments or 95%.

#### *Evaluation Results on the Whole Dataset*

In conclusion, Translator\_A is winning for 268 out of 300 comments or 89.33%. For GoogleReviews dataset, Translator\_A is winning for 128 out of 150 comments of 85.33%. For TripAdvisor dataset, Translator\_A is winning for 140 out of 150 comments or 93.33%.

Table 11 shows the performance of Translator\_A against the performance of the best achievable performance. According to the results shown in Table 11, the performance of Translator\_A is very close to the performance of the best answer as can be seen that Translator\_A has been selected for 72 to 89.33% of the times. The outstanding quality of Translator\_A is that Translator\_A produces the resulting translated comments that meet the grammar structure completely with clear and complete content.

On the GoogleReviews temple dataset, Translator\_A has been selected for only 108 out of 150 which is the least among all six datasets. The reason behind this result is that Translator\_D outperforms Translator\_A in producing a result that is a natural language and has more discretion in context. For example, “Very clean and beautiful temple. people follow discipline here” is

translated to วัดสวยสะอาดมาก ผู้คนที่ปฏิบัติตามระเบียบวินัย (wat suai sa-at makphukhon thi ni pati bat tam rabiap winai) by Translator\_D which means “Beautiful temple, very clean people here comply with discipline” while Translator\_A translates the sentence to “วัดที่สะอาดและสวยงามมาก คนติดตามวินัยที่นี่” (wat thi sa-at lae suai-ngam makkhon tittam winai thi ni) which means “Very clean and beautiful temple people trace discipline here”. Another example, “Nice Peachful Temple” is translated to “วัดที่สวยงามและเงียบสงบ” (wat thi suai-ngam lae ngiap sangop) by Translator\_D which means “Beautiful and quiet temple” while Translator\_A translates the sentence to “วัดสันติที่ดี” (wat santi thi di) which means “peace temple that is good”. Translator\_D can translate some comments to natural human language results. Therefore, Translator\_D receives a very high 5 out of 5 score on some comments in the GoogleReviews temple dataset.

In total, there are six datasets. Each dataset contains 50 comments. Thus, there are 300 comments. By counting individual expert result, Translator\_A has been selected 744 times or 82.67%; Translator\_B has been selected 26 times or 2.89%; Translator\_D has been selected 130 times or 14.44%; no expert selects Translator\_C. The quality of the best answer is at 4.21 while the quality of Translator\_A answer is at 4.20 which is very similar. As a result, it can be concluded that Translator\_A outperforms other translators in this study. In this study, the experts do not know the identity of the translator for a fair judgement on the results.

According to the results shown in this study, Yandex.Translate (denoted Translator\_C) is not suitable for translating the comments. The resulting sentence translated by Yandex.Translate is either incomplete or incorrect grammar structure. For example, “good sand” has translated to “สินค้าดี” (sinkha di) which means “Good product”. Another example, “Clean beach” has been translated to “สะอาดชายหาด” (sa-at chaihath) meaning “beach clean” which is incorrect sentence structure in Thai. Or “Peaceful place to spend some time” has been translated to “สงบสถานที่ต้องใช้เวลาหน่อย” (sa-ngop sathanthi tong chai wela noi) meaning “Peaceful place, need a little time”.

For Microsoft/Bing Translate (denoted Translator\_B), some sentences have been translated correctly. However, some resulting sentences translated by Microsoft/Bing Translate are incomplete or have incorrect meaning. For example, “Near Bangla market and good” has been translated to “ใกล้ตลาดบางลา” (klai talat bang la) which is correct but “and good” is missing from the result. Another example, “Lovely beach close to many amenities. Highly recommended” has been translated to “มีสิ่งอำนวยความสะดวกมากมาย 9.6” (mi sing amnuai khwam saduak makmai 9.6) which means “There are many facilities 9.6” which is clearly incorrect.

Baidu Translate (denoted Translator\_D) receives the second rank in this study, especially on the temple datasets. Baidu Translate must improve the translation of specific names such as Wat Chalong has been translated to Wat-Cha-ra which is a common Thai male name. Moreover, Baidu Translate is not consistent on the results. Some comments are well translated while some comments are completely incorrect. The good qualities of Baidu Translate include natural language, good word choices and the same expressions as the original. An example of some good results include “Very nice beach” has been translated to “ชายหาดที่สวยงามมาก” (chaihat thi suai-ngam mak) which means “Very beautiful beach”.

The good qualities of Google Translate (denoted Translator\_A) which is the winning translator in this study include translating all words, choosing standard words, polite languages, preserving the English sentence structure and translating the location name correctly.

## Conclusion

In this research, four popular translators (Google Translate, Baidu Translate, Yandex.Translate and Microsoft/Bing Translate) are compared in translating English to Thai. The data used in the experiment comes from comments on social media related to the 3 major tourist attractions in Phuket, Thailand. This research has been evaluated by 3 language experts from a university in Thailand. The assessment covers 3 major topics including the grammar structure, the vocabulary and idiom and the overall translation. The evaluation results show that Google Translate is found to be the most suitable translator for translating English comments to Thai comments in this study at 89.33%. For GoogleReviews datasets, Google Translate has been selected 85.33% while Bing Translate has been selected 93.33% for TripAdvisor datasets. The overall performance of the best translator on the grammar structure, the vocabulary and idiom and the overall translation is at 4.31, 4.05 and 4.21, respectively. Even though this research presents the results on specific languages and domains, the techniques used in this study can be applied to any language on any domain. Our future work is analysis of word cut translation to extract the important themes in tourism in order to analyze the strengths and any improvement for the attraction. Since Phuket is a place that is popular among foreigners, another future work direction includes a comparison of these translators for other foreign languages such as Chinese, Russian and French to Thai language in order to find a suitable translator.

## Acknowledgement

This work is supported by College of Computing, Prince of Songkla University, Phuket Campus

(CoC63041665). The authors would like to thank AI-TaSI (Andaman Intelligent Tourism and Service Informatics Center) for providing the workload data and we also would like to thank all six members of Chochiang Linguistic Team (CLT) for their strong harmonic cooperation.

## Author's Contributions

**Kitsiri Chochiang:** Defining the research question, designing the research, designing the experiments, conducting the data analysis, discussing and concluding the research results, providing the conclusion, writing the manuscript, contacting the editorial office.

**Thanakrit Thongkhamdee:** Conducting the experiments.

**Lalita Sathansat:** Conducting the experiments.

## Ethics

This article is original and contains unpublished material. The corresponding author confirms that all the authors have read and approved the manuscript and there are no ethical issues involved.

## References

- Ba, S., & Pavlou, P. A. (2002). Evidence of the effect of trust building technology in electronic markets: Price premiums and buyer behavior. *MIS quarterly*, 243-268.
- Balk, E. M., Chung, M., Hadar, N., Patel, K., Winifred, W. Y., Trikalinos, T. A., & Chang, L. K. W. (2012). Accuracy of data extraction of non-English language trials with Google Translate.
- Chanwaiwit, P. (2016). Enhancing English to Thai Translation Skills Using Exercises. Western Languages Department, Chiang Mai Rajabhat University.
- Chochiang, K. (2017, December). L2D: A modified algorithm based on edit distance for searching Thai-English transliterated words. In *Proceedings of the 2017 International Conference on Computer Science and Artificial Intelligence* (pp. 242-246).
- Dhakar, B. S., Sinha, S. K., & Pandey, K. K. (2013). A survey of translation quality of English to Hindi online translation systems (Google and Bing). *International Journal of Scientific and Research Publications*, 3(1), 1-4.
- Door43. (2020). Translation Manual. Door43. [https://door43.org/u/Door43/en\\_ta/dcae73d489/03-translate.html](https://door43.org/u/Door43/en_ta/dcae73d489/03-translate.html)
- Hajli, M. N. (2014). A study of the impact of social media on consumers. *International Journal of Market Research*, 56(3), 387.

- He, Z. (2015, July). Baidu translate: Research and products. In Proceedings of the Fourth Workshop on Hybrid Approaches to Translation (HyTra) (pp. 61-62).
- Intakosum, S. (2014). L'evaluation dan l'enseignement de la traduction. Ramkhamhaeng University Journal, Humanities Edition. 33, 23-50.
- Kastberg, P., & Andersson, T. B. (2012). Machine Translation Tools-Tools of the Translator's Trade. *Communication & Language at work*, 1(1), 34-45.
- Kull, K. (1998). Organism as a self-reading text: anticipation and semiosis. *International Journal of Computing Anticipatory Systems*, 1, 93-104.
- Martin, B. S. (2017). Translation Quality Assessment of Google Translate and Microsoft Bing Translator. Unpublished thesis). Universidad de Valldolid, Spain.
- Muegge, U. (2012). Cloud-based translation memory tools are changing the way translators work and train. *The big wave of language technology*, 1(3).
- Müller, M., Nguyen, T. S., Niehues, J., Cho, E., Krüger, B., Ha, T. L., ... & Waibel, A. (2016, June). Lecture Translator-Speech translation framework for simultaneous lecture translation. In Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Demonstrations (pp. 82-86).
- Nilashi, M., Ibrahim, O., Yadegaridehkordi, E., Samad, S., Akbari, E., & Alizadeh, A. (2018). Travelers decision making using online review in social network sites: A case on TripAdvisor. *Journal of computational science*, 28, 168-179.
- Polnyotee, M., & Thadaniti, S. (2015). Community-based tourism: A strategy for sustainable tourism development of Patong Beach, Phuket Island, Thailand. *Asian Social Science*, 11(27), 90.
- Sun, L. (2014). Comparative analysis of English and chinese passive voice in machine translation (Doctoral dissertation).
- Taecharungroj, V., & Mathayomchan, B. (2019). Analysing TripAdvisor reviews of tourist attractions in Phuket, Thailand. *Tourism Management*, 75, 550-568.
- Ulfah, M. (2015). An accuracy analysis in Indonesian-English translation using "Google Translate" machine translation (translation of "Jurnal Walisongo Abstracts vol. 23 number 1") (Doctoral dissertation, UIN Walisongo).
- Vidhayasai, T., Keyuravong, S., & Bunsom, T. (2015). Investigating the Use of Google Translate in "Terms and Conditions" in an Airline's Official Website: Errors and Implications. *PASAA: Journal of Language Teaching and Learning in Thailand*, 49, 137-169.
- Xie, K. L., Chen, C., & Wu, S. (2016). Online consumer review factors affecting offline hotel popularity: evidence from tripadvisor. *Journal of Travel & Tourism Marketing*, 33(2), 211-223.