WEP Warwick Evans Publishing

Transactions on Social Science, Education and Humanities Research
ISSN: 2960-1770, eISSN: 2960-2262 | Volume 11
CEPC 2024

Exploring Failures and Possible Remedies in AI and Human Translation of English Idioms

Tianyao Li *

Faculty of Arts and Humanities, University of Macau, Macau, China

* Corresponding Author: AC11765@um.edu.mo

Abstract. This paper explores the characteristics, similarities and differences between Chinese EFL students and AI when translating English idioms into Chinese. Based on the data collected from the questionnaire on English idioms in English-Chinese translation by Chinese EFL students and AIs, the distribution of human and AI translation accuracies and the focus of translation were analysed using the stereotyped heterogeneous ratio. The results of the survey showed these points: the overall accuracy of translations of Chinese EFL students is lower than that of AI, and the distribution of human translation accuracy among all categories of English idiom was similar to AI translation accuracy; Chinese EFL students tend to retain the form of the proverbs, while AI translators tend to translate the exact meaning; and AI is not able to fully utilise the context to facilitate the understanding of proverbs in the same way as human beings do. Based on the results, this paper expects to provide a reference for future research in this domain.

Keywords: EFL Students; Artificial Intelligence; Translation; Idioms; Language transfer.

1. Introduction

Within the process of second language learning, pragmatic translation often reflects the proficiency in the target language of the learners. Recently, with the advancement of machine translation and artificial intelligence technology, AI translation based on human dialogue modeling has become the preferred translation tool for many people in an application scenario for a second language. However, AI often fails to achieve high-quality language transfer in many cases, resulting in wrong translation output that differs from human translation. To provide feasible measures for further improving the quality of AI translation, the investigation of this problem is deemed worthwhile.

1.1. Theoretical Framework of Human Translation

Firstly, according to the Functional Equivalence Theory, as discussed by Guo Jianzhong, translation is defined as Copying information from the source language to the target language using the most appropriate, natural and reciprocal language in terms of semantics and style [1, 2]. The implementation of "equivalence" is reflected in four aspects: lexical equivalence, syntactic equivalence, discourse equivalence, and stylistic equivalence [2]. The functional equivalence of meaning is considered a priority over the form. The form of the source language may not accurately express the meaning it conveys in the source language due to its high correlation with the cultural background of that language when migrating to the target language. Skopos Theory proposed by Hans Vermeer also emphasizes three principles that translators should adhere to: Skopos Rule, Coherence Rule, and Fidelity Rule [3]. Among them, the Skopos Rule is considered to be the most fundamental and important, which means translation should function in a way that can be understood and accepted by the target language audience in the context and culture of the target language [4]. Both theories emphasized the fundamental principles of translation: the meaning in the target language should play a role consistent with the expression of the source language in its cultural context, before achieving coherence and consistency in the form of the target language.

Humans often share some commonalities in their first and second languages. According to Jung's prototype theory, even between two different human cultural groups, there are highly similar images with significant semantic functions, which reflect universal primitive memories and the essential

needs of all human beings [5]. Based on Chomsky's Universal Grammar (UG) theory, the existence of Language Acquisition Device (LAD) and innate grammar knowledge enables all human beings to possess a certain degree of knowledge of grammatical structures. Therefore, theoretically speaking, it is feasible for humans to translate L1 into L2. The writer attributes the failure of translating L2 into L1 by humans to pragmatic knowledge, namely the inadequate knowledge of fundamental vocabulary and grammar and the insufficient knowledge of the cultural background of the source language. This aligns with the classification of failure in human translation mentioned by Thomas [6]. The author selected English idioms that are not well-known among Chinese learners to be the translation objects from English to Chinese. According to Song, English idioms are highly concise, with literal meanings that are not exactly equivalent to their metaphorical meanings and are highly related to the cultural background of English language users [7]. This feature fits the purpose of my research. Furthermore, providing context expands the cultural and background knowledge of idioms' normal usage, which can promote the accuracy of human translation of English idioms.

1.2. AI in Translation

Andrew Lampinen believed that choices of language models regarding their own functionality and judgment reflect important issues about human capabilities [8]. Lang Chen argued that we should acknowledge that one of the most representative natural language processing technologies in AI, ChatGPT based on LLM, has learned human language; this is because humans cannot accurately distinguish between outputs generated by AI and those generated by humans [8]. Pinker stated that ChatGPT demonstrates that language is learned through experience, it is incapable of understanding the purpose of language, so the phenomenon that it makes common sense mistakes is not unexpected [9]. Everett pointed out that ChatGPT can learn language by using a large amount of data rather than innate grammar or language rules, which challenges the UG theory [10]. However, ChatGPT4.0 showed limitations in its cognitive involvement with spatial folding concepts in October 2023. For example, the output indicated that ChatGPT considered the ability to load water into a bulb with the bottom removed higher than such ability of a cup with the bottom removed. Some scholars argue that due to the lack of a human body, ChatGPT cannot understand certain spatial concepts and cannot truly learn a language. The writer believes that translation errors made by large data models differ from those made by humans, despite LLM being extracted from human conversations. The writer thinks that AI translation does not suffer from deficiencies in terms of vocabulary and grammar in the pragmatic aspect, but it can still produce translation results that do not conform to the coherence of the target language. In particular, with idioms that are relatively uncommon in current social environments, even with provided context, AI translation results may still present illogical errors. In other words, it is believed that providing context cannot improve AI translation accuracy as it does for humans because AI lacks the ability to infer the meaning of idioms based on context during the translation process. If AI outputs incorrect results during the translation of idioms in isolation, it will continue the same failures in context-based translation.

This research aims to explore the differences in the failure between AI and human translation of English idioms and compare the types of failure in AI translation with those in human translation, in order to propose reasonable suggestions for improving AI translation.

2. Methodology

As mentioned above, the formation of idioms is closely related to the imagery and agreed metaphors accumulated over a long period of time in the culture related to that language, which is considered difficult to be translated by non-native speakers. At the same time, the roles of artificial intelligence and various machine translation applications on translation have been increasingly emphasized, and many people believe that machine translation will eventually replace human translation. However, this study will focus on the following issues:

1. Do machines surpass humans in translating colloquialisms? What are the similarities and differences in the focus of translation between the two?

2. Does the provision of context play the same role for human and AI translators?

The purpose of this study is to compare the similarities and differences between EFL students and AI translators and to make longer-term suggestions for improving both. The author predicts the results of the experiment as follows:

H1: Machine translation has a higher overall accuracy rate than human translation.

H1a: The vocabulary of machines and the vocabulary of the corpus exceeds that of human beings, so the accuracy of machine translation is generally higher than that of human beings:

H1b: The focus of machine translation may be different from that of human translation. When translating idioms, humans may pay more attention to reducing them into an idiom, and their idiomatic form reduction is greater than translating the meaning from the source language; while machines may better reduce semantics, prioritizing over form.

H2: the provision of context will contribute to both human translation and AI translation accuracy. However, human translators may be facilitated more by context.

First of all, this study extracts English idioms from the Oxford Dictionary as a corpus, filters out the idioms that are too well-known within the scope of Chinese English learners, and finally selects twenty idioms. The author categorized the selected idioms into four categories based on their relevance to Chinese idioms (e.g., Table 1).

Table 1. Categories of idioms in the questionnaires

Category	Identical idioms	Similar idioms	Contradict idioms	English-culture-bound idioms
Number of items	3	5	6	6
Idioms	kill two birds with one stone get to the bottom of something add insult to injury	bite the bullet speak of devil hit the nail on the head barking up the wrong tree on the same page	cutting corners once in a blue moon cost an arm and a leg hold your horses sleep on it up in the air	let the cat out of the bag break a leg have a cow stick to your gun call it a day bed of roses

In this study, two questionnaires were developed, one containing only a single idiom, and the other questionnaire providing a brief dialogue context including the idiom to observe the auxiliary effect of the context on the participants.

In the process of investigation, the author selected English majors whose first language is Chinese or college students and graduate students with all-English teaching conditions as the human survey subjects. These students are not prone to translation errors caused by a lack of vocabulary or basic grammar knowledge when translating, so their mistakes are more meaningful in the study of negative language transfer: errors in the understanding of the metaphorical meaning of idioms [10].

The two questionnaires were combined and published on the WeChat platform, a well-known social media platform in China, through the "Questionnaire Star" app, and the answers were collected online. In the questionnaire for human survey subjects, this study collected the language, major, and education information of the survey subjects, and finally selected 30 valid questionnaires. This study selected six well-known translation software and artificial intelligence in China as the AI survey

subjects, and directly entered the two questionnaires into the input boxes provided by the corresponding artificial intelligence and asked them to translate the full text.

Table 2. Participants

Background	Human	AI
Participants	Chinese EFL Students	Baidu Translate GPT-4.0 Turbo ChatGPT 4.0 ERNIE Bot DeepL Youdao
Years of learning English	13(28 people) 14(2 people)	NA

After obtaining the translation results, the author used Excel sheets to count all the answer results, and analyzed the accuracy of the two questionnaires and the accuracy of each of the four types of colloquialisms. In addition, all translated answer types for a single colloquial phrase are also listed at the end of the text.

In the data analysis, the author will synthesize the above statistical data and discuss the experimental results.

3. Results

3.1. AI Translation and Human Translation Idiom Accuracy

According to results from both Questionnaire 1 and Questionnaire 2, the accuracy of artificial intelligence-translated results was higher than human-translated results.

Table 3. The accuracy of translated results (average)

Categories	Human Q1	Human Q2	AI Q1	AI Q2
Identical idioms	76.6%	94.4%	94.4%	100%
Similar idioms	34.6%	81.3%	76.7%	93.3%
Contradict idioms	34.4%	82.7%	72.2%	94.43%
English-culture-Bound idioms	28.8%	73.2%	61.1%	87.5%
Mean	40%	80.66%	73.3%	92.91%

According to Table 3, the accuracy of AI-translated results was never lower than 61.1%, while the accuracy of human's Q1 results hardly reached 40%. When AI translators do a separate idiom questionnaire, the accuracy rate of the translation results has been nearly twice that of humans. In Q2, which provides context for idiom use, although the accuracy rate of human translation has increased significantly, it is still lower than that of AI translation overall.

However, even though AI translations are significantly more accurate than human translations, they still cannot be 100% accurate. Moreover, the accuracy of human translation and AI translation accuracy in the four categories listed were almost identical: the accuracy of Identical idioms was the highest, the accuracy of Similar idioms was similar to that of contradict idioms, and the distribution of high and low was opposite in Questionnaire 1 and Questionnaire 2, and English-culture-bound Idioms translations have the lowest accuracy rate. In future research, it is essential to further analyze the possible reasons for the different distribution of different data.

3.2. Comparison of the Similarities and Differences Between Human Translation Results and Artificial Intelligence Translation Preferences (Structure or Accuracy)

The author counted the number of all the different translation results corresponding to each colloquial phrase in the corpus (as shown in Table 4).

Table 4. Type of all answers

Idioms	All answers	Category	Idiom form (%)	Literal Meaning (%)
1. Cutting corners	Human 15 AI 2	Contradict Idioms	Human 40 AI 100	Human 20 AI 0
2. Once in a blue moon	Human 16 AI 3	Contradict idioms	Human 56 AI 33	Human 38 AI 33
3. Bite the bullet	Human 8 AI 3	Similar idioms	Human 62 AI 100	Human 38 AI 0
4. Barking up the wrong tree	Human 17 AI 6	Similar idioms	Human 35 AI 8	Human 35 AI 66
5. let the cat out of the bag	Human 15 AI 4	E-C-B idioms	Human 53 AI 50	Human 27 AI 50
6. cost an arm and a leg	Human 20 AI 5	Contradict idioms	Human 30 AI 40	Human 20 AI 20
7. kill two birds with one stone	Human 4 AI 4	Identical idioms	Human 75 AI 75	Human 25 AI 25
8. hold your horses	Human 17 AI 6	Contradict idioms	Human 65 AI 33	Human 18 AI 67
9. break a leg	Human 18 AI 5	E-C-B idioms	Human 22 AI 40	Human 28 AI 60
10. have a cow	Human 12 AI 5	E-C-B idioms	Human 25 AI 40	Human 58 AI 60
11. stick to your gun	Human 18 AI 3	E-C-B idioms	Human 78 AI 67	Human 16 AI 33
12. sleep on it	Human 16 AI 5	Contradict idioms	Human 31 AI 0	Human 31 AI 100
13. call it a day	Human 16 AI 5	E-C-B idioms	Human 25 AI 25	Human 25 AI 50
14. bed of roses	Human 19 AI 5	E-C-B idioms	Human 42 AI 50	Human 53 AI 50
15. speak of devil	Human 13 AI 3	Similar idioms	Human 46 AI 33	Human 38 AI 67
16. on the same page	Human 20 AI 4	Similar idioms	Human 30 AI 25	Human 30 AI 75
17. get to the bottom of something	Human 21 AI 6	Identical bottom	Human 38 AI 67	Human 29 AI 17
18. hit the nail on the head	Human 15 AI 3	Similar idioms	Human 67 AI 67	Human 27 AI 33
19. add insult and injury	Human 6 AI 1	Identical idioms	Human 83 AI 100	Human 17 AI 0
20. up in the air	Human 11 AI 4	Contradict idioms	Human 55 AI 25	Human 36 AI 75

In Table 4, the greyed-out 11 cells represent more than half of the total number of subjects in questionnaire1 for the different types of answers translated by humans (15). The author believes that the diversity of answers can reflect the uncertainty of the answers of the subject group to a certain extent, because about 50% of the grayed-out tables belong to the English-culture-bound idioms with

the lowest translation accuracy rate, and the rest are mostly Contradict idioms, and the translation accuracy rate of the above-mentioned isolated idiom questionnaire ranks second-to-last. However, from the data in the table, it can be seen that most of the translated answers provided by the AI in Questionnaire 1 (13/20) are more than half of the number of AIs tested (3). It can be seen that despite the high accuracy of translation, the content of the AI answers is not completely consistent.

Based on the percentage of the structure of answers in Table 4, the author found that compared with humans, AI is more inclined to translate English idioms that are difficult to directly find alternative idioms in Chinese, such as "stick to your gun". 50% of the AI literally translates it to a Chinese idiom with the meaning of "stick to your point", while the rest translates it as "stand up for yourself" or "stick to yourself" in their true application meaning. In human translation, some translators tend to find an idiom in Chinese for direct replacement, such as "single-minded" or "strictly guarded", maintaining the formal "idiom" characteristics, but causing a certain deviation in semantic translation. Another similar example is the translation of “have a cow”, an idiom highly related to English culture that is translated by multiple AIs as "there is a cow" or "give birth to a cow", while Chinese EFL students often substitute the "cow" in this idiom into the Chinese image of "cow", which is understood as "wealth", resulting in translation errors. The author believes that humans are more likely to migrate from L1 to L2 than artificial intelligence.

3.3. Similarities and Differences in the Impact of Context on the Accuracy of Human and AI Translations

According to Table 3, after providing the application context of idioms, the accuracy of AI and human translation has increased significantly, and the accuracy of human translation has improved relatively significantly. However, when human subjects are exposed to a context that is far away from their lives, such as financial crisis, the promotion effect of context on understanding idioms will be relatively weakened – when the average human translation accuracy of similar idioms in Q2 reaches 0.813, the translation average of the idiom “bite the bullet” with a context related to financial crisis is only 0,566. Despite this, the accuracy rate of human translation of idioms under each category is still significantly improved under the premise of context.

According to all the results in Questionnaire 2, the application context also plays a role in improving the accuracy of AI translation. However, the author believes that this facilitating effect is not the same as the role of context in helping human translation. It can be seen that some AIs have new errors that are not found in Questionnaire 1 in the contextual translation results, and some AIs have output incorrect results with or without context. For example, Baidu Translation translates "break a leg" as "break a leg" in Questionnaires 1 and 2, although Questionnaire 2 gives the context of the actor's upcoming stage.

Applying context sometimes does not correct AI translation errors, and can even interfere with AI's correct translation results when only isolated idioms are given. For example, DeepL translates "let the cat out of the bag" as "leak" in Questionnaire 1, which fits the meaning of this idiom in the source language, and translates the idiom as "let the cat get out" when provided with the following context:

(Lucy's birthday is coming up.)

Student A: Why do we need to plan the party activities once again?

Student B: We planned to give Lucy a surprise, but Jack let the cat out of the bag.

This is difficult for the target language reader to understand and does not conform to the true meaning of the idiom in the source language.

In another example, after ChatGPT obtained the context of the negative definition of "bed of roses", it also added a negative word before the translation of "bed of roses" when it translated it alone: "not a broad road". When the author interviewed it, ChatGPT outputted this as the result of contextual adjustments, although idioms alone do not contain any negative words.

In summary, although the distribution of the accuracy of artificial intelligence and human translation of the same idiom is similar, there is still a difference between human translation and machine translation. The author does not think that AI has the real ability to understand a new language, and the author believes that AI relies on mechanical input data for induction, it lacks the thinking to understand language.

4. Discussion

4.1. Discussion of Results

Overall, the results of this study are in line with expectations H1 and H2.

First of all, the accuracy rate of AI translation idioms based on big data modeling is indeed higher than that of English college translation, as H1a, but due to the different emphases of machine translation and human translation, it cannot be concluded that machine translation surpasses human translation. As H1b, the focus of AI translation is to directly explain the meaning of the idiom, followed by the form, while Chinese EFL students often focus on finding the corresponding idiom in their native language, ignoring the correct meaning of the English idiom. The language transfer of the mother tongue has had a certain adverse effect on the understanding of English culture-related colloquialisms by Chinese English students, but the similar idiom structures in the two languages should be noted and restored as much as possible. Jia Lyu pointed out that the focus of idiom translation should be to reproduce the cultural factors and cultural significance of the source language in the target language [11].

In addition, as H1b, the role of context in promoting the understanding of English idioms by humans and AI was not exactly the same, and the translation results of AI had even been negatively affected by the context to a certain extent. Therefore, the author speculates that AI does not yet have the ability to learn languages like humans. While people cannot directly deny the ability of AI to transfer between two languages, AI cannot avoid social pragmatic lapses as naturally as humans can.

4.2. Limitation

There are certain limitations to this study. Firstly, the collection of human translation results in the form of online questionnaires has limitations and the possibility of cheating, which will affect the accuracy of the results. Secondly, the number of artificial intelligences selected in this study is not enough, so the research results may not be universal enough, and it is difficult to rule out the possibility that some artificial intelligence may encounter network failures in the translation process. In addition, the number of idioms found by the author for each classification is not completely consistent, which may lead to inaccurate data.

4.3. Further Suggestions

To give some suggestions, Chinese EFL students should pay attention to the cultural background while learning English idioms, especially the different images between Chinese culture and the culture of the English language area. According to Thomas, understanding implicit meanings helps promote mutual understanding among people [6]. Also, if the isolated idiom makes learners feel confused, then putting the idiom into a specific context will help learners to better understand it. As Zhang discussed, the whole process of communicating should be seen as the process of translation [3]. Moreover, AI translation should be trained to restore the original form of the idiom.

5. Conclusion

In conclusion, this study examines the different performances of human and AI translators when translating English idioms into Chinese. According to the correlation between the English idioms and the corresponding idioms in Chinese from strong to weak, the accuracy of the two translations also

shows a distribution trend from high to low; AI translators are more accurate than human translators, but AI focuses on translating the meanings of idioms, while humans are relatively more concerned with maintaining the non-detachable form of the colloquialisms; and humans are more affected by language migration. In addition, context enhances the correctness of both human and AI translations, but humans are better able to use social context to aid comprehension, which AI does not fully possess. This paper can provide a new way to study the "learning" of language by big data models, i.e., to examine whether AIs can perform effective and socio-linguistic language transfer as well as humans do, and to use this as another criterion to judge whether AIs can understand and learn a language.

References

[1] E. A. Nida, (2001). Language and Culture, Shanghai Foreign Language Education Press.

[2] Y. Li, (2004). Reflections on the Evolvement of the Concept of Translation, Journal of Jinan University, 2, 57 - 59.

[3] Y. Zhang, (2014). The Application of Functional Equivalence Theory in English Idiom Translation, Chinese Language Journal – Foreign Language Education and Teaching, 37 - 38.

[4] H. Li, (2018). English Translation of the Network Hot Word " Antediluvian power" from the Perspective of Translation Skopos Theory, Advances in Social Science, Education and Humanities Research (ASSEHR), International Workshop on Education Reform and Social Sciences (ERSS 2018), 544-547.

[5] W. Li, (2018). Analysis of Psychology and Literature, Shijian Tansuo, 12 (1266).

[6] J. Thomas, (1983). Cross-cultural pragmatic failure. Applied Linguistics, 4 (2), 91 - 112.

[7] H. Song, (2023). A Study of English Metaphorical Idioms Based on Embodied-Cognitive Linguistics.

[8] X. Yang, & R. Luo, (2024). ChatGPT is Coming: How Language Science Looks at ChatGPT, Shanghai Education Press.

[9] Information on: https://news.harvard.edu/gazette/story/2023/02/will-chatgpt-replace-human-writers-pinker-weighs-in/.

[10] Information on: https://www.tehrantimes.com/news/483187/Exclusive-Linguist-says-ChatGPT-has-invalidated-Chomsky-s-innate.

[11] J. Lyv, (2011). Exploring the Chinese Translation of English Idioms from a Cultural Perspective, Journal of Language and Literature-Foreign Language Education and Teaching, 7 (1).