



EVALUATION OF AI TRANSLATION OF ENGLISH MARITIME DOCUMENTS INTO JAPANESE AND THE IMPACT OF USER-PREPARED KNOWLEDGE BASE IN ChatGPT

Naoyuki Takagi*

Abstract. Takagi and Kimura (2024) evaluated the translation accuracy of Maritime English documents into Japanese by ChatGPT 4.0 (free version), Google Translate and DeepL, and found that all of these applications made errors related to English words and phrases, grammatical interpretation of English sentences, and to producing natural, grammatical Japanese translations. There was not much difference among the three applications in terms of translation accuracy or the error types. In this paper, the same set of Maritime English documents were translated into Japanese by Google Translate and DeepL. There was no improvement over time with DeepL, whereas with Google Translate, there was a 15% reduction in the number of errors, but these improvements came from only one document. ChatGPT-4o allows its users to compile a list of English Maritime words and phrases and their Japanese counterparts as a knowledge base, and to incorporate it in translation. A list of about 700 Maritime English words and phrase covering both deck and engine English including phrases that appear in the COLREGs was used as a knowledge base. With or without this knowledge base, the translation by ChatGPT-4o led to about a 48% reduction in errors compared to ChatGPT 4.0 (free version) used about a year ago. The impact of the knowledge base, however, was limited.

Keywords: AI Translation; ChatGPT; Google Translate; DeepL; COLREGs; Translation into Japanese.

Tokyo University of Marine
Science and Technology,
Faculty of Marine Engineering,
2-1-6 Etchujima, Koto-ku,
Tokyo, Japan

* **Corresponding author:**
takagi@kaiyodai.ac.jp

<https://doi.org/10.7225/imla30.22>

1. INTRODUCTION

In a paper presented at IMEC 29, Takagi and Kimura (2024) evaluated the translation accuracy of Maritime English documents by the free version of ChatGPT 4.0, Google Translate, and DeepL, which were available in July 2024. They found that all three applications made errors involving translation of Maritime English words and phrases into Japanese (Type A-1 errors), those involving translation of ordinary English words and phrases (A-2 errors). In some cases, those programs ignored some original English words or phrases in translation (A-3 errors). There were cases where the grammatical structure of original sentences was not interpreted properly (Type B errors). There were cases where output Japanese translations were not natural or grammatical (Type C errors). Of these error types, the most frequent was Type A, with A-1 errors accounting for about 62 % of the total number of 292 errors observed, with A-2 and A-3 errors about 16% and 10%, respectively. Type B and Type C errors accounted for about 9 and 3%, respectively. There was not much difference among the three applications in the number or the types of translation errors.

OpenAI has launched a new model, ChatGPT-4o, which allows the incorporation of a knowledge base created by its users. In order to evaluate the possibility of this new option, the author created his own Maritime English vocabulary list, which contains 114 COLREG related words and phrases, 346 deck related and 228 engine related vocabulary items. The list was based on the author's Maritime English class materials for deck and engine cadets. Each line of the list contains a English word/phrase followed by its Japanese translation, for example, "stand-on vessel 保持船."

The same set of bench mark passages used in Takagi and Kimura (2024) were used in the present paper, and they were translated into Japanese by ChatGPT-4o, both with and without the knowledge base, by Google Translate and DeepL. Thus, for ChatGPT, the comparisons were between the output of the free version of ChatGPT 4.0 as of July 2024 and those of ChatGPT-4o with and without the knowledge base. For Google Translate and DeepL, their improvement over time was evaluated between July 2024 and March 2025, when the translations for the present paper were performed.

2. METHODS

A total of 10 short passages (2612 words in total) used in Takagi and Kimura (2024) were translated into Japanese. They were chosen from various sources such as COLREGs (IMO, 1972), SOLAS (IMO, 1974), an

engineering article available on the Internet about Scavenging (Jääskeläinen and Khair, 2018), accidents report and a safety bulletin by Marine Accident Investigating Branch, which include a safety warning after explosion (MAIB, 2017), a report on dragging anchor and subsequent collisions (MAIB, 2018), and a collision report (MAIB, 2024). Also part of this bench mark document set was a passage from sailing directions (United Kingdom Hydrographic Office, 2022) and a passage from a textbook (Van Dokkum, 2013). COLREGs and SOLAS were chosen since there are official Japanese translations by the Ministry of Land, Infrastructure and Transport and Tourism. Other documents were chosen in order to represent a wide range of Maritime topics covering both deck and engine matters.

Each passage was translated into Japanese by ChatGPT-4o with and without the knowledge base, Google Translate and DeepL. When prompting translation for ChatGPT 4.0 (old free version) and ChatGPT-4o without the knowledge base, the prompt asked ChatGPT to translate the following into Japanese. With the knowledge base, the prompt asked ChatGPT-4o to refer to the knowledge base file and list up the items referred to in the text, and to indicate on which line the item appeared in the knowledge base file. This allowed us to find out which item was used, or referred to, in the knowledge base.

3. RESULTS

3.1. DeepL and Google Translate

With DeepL, no improvement was found. With Google Translate, only one document (MAIB accident report about a collision) saw a number of improvements, whereas the others led to almost no improvement. (There was only one case of correct translation elsewhere.) The improvements included 3 corrections out of 8 type A-1 errors, and 3 corrections out of 4 A-3 errors (non-translated items), and 3 corrections out of 5 type B errors. Thus, out of 85 former Google Translate errors, only 13 cases (15%) led to correct translations. The following are 2 examples of A-1 errors involving Maritime English words and phrases which were mistranslated into Japanese last time but were translated properly this time.

- which was fine on Scot Explorer's port bow and 5nm ahead

Before: 左舷船首と5海里前方で順調であったが
"port bow and 5 nautical miles ahead"

After: 左舷船首5海里先に (Correct)

- with the master alone on the bridge

Before: 船長一人で橋の上を (Here, the Japanese indicate the master was on a bridge over a river.)

After: : 船長一人がブリッジにいて (Correct)

In the first example, the word “fine” was mistranslated as “順調, or in good condition” before, but now the word is ignored in translation. In the second example, the word “bridge” was mistranslated as a bridge over a river before, but now it is correctly translated as “bridge of a vessel.” It is not clear why this document alone led to so many correct translation cases.

3.2. ChatGPT

The following table shows the number of translation errors by ChatGPT 4.0 for each error type when the translation was conducted about one year ago. For ChatGPT-4o (without Knowledge Base) and ChatGPT-4o_K (with Knowledge Base), the entries are the number of cases where each application was able to achieve correct translations of the cases that were mistranslated. For example, of all 61 type A-1 errors by ChatGPT 4.0, ChatGPT-4o was able to achieve correct translations in 27 cases.

Improvements were found across all types of errors. Of the total 95 errors, 46 cases led to proper translations by ChatGPT-4o. Adding the knowledge base of approximately 700 Maritime English words and phrases, however, achieved only 45 correct translations. Even when we concentrate on the errors involving Maritime English words and phrases (Type A-1), where the knowledge base is expected to have the largest impact, the inclusion of the knowledge base showed only a small advantage (27 vs. 29). This came as a surprise.

There were certainly cases where the knowledge base items appeared to have contributed to proper translation. For example, “drag anchor, 走錨する” was in the knowledge base, and the English phrase, “dragged its anchor” was mistranslated by ChatGPT 4.0 as “錨を流出した” and by ChatGPT-4o in the

same way. By referring to the knowledge base, however, the correct translation of “走錨した” was achieved. The knowledge base contained “stand-on vessel” and “give-way vessel” and their Japanese counterparts. ChatGPT 4.0 failed to reach the correct translations for both, and by referring to the knowledge base, ChatGPT-4o achieved proper translations. However, even without using the knowledge base, ChatGPT-4o came up with the correct Japanese translations in both cases. Thus, we cannot be sure whether this was due to the contribution of the knowledge base or the expanded language model of ChatGPT-4o.

There were cases where items in the knowledge base did not contribute to proper translation. Let us go over two such cases here. The knowledge base contained the following items related to “close-quarters situation”:

- “avoid a close-quarters situation
他の船舶に著しく接近することを避ける”
- “result in another close-quarters situation
別の船舶に著しく接近することになる”

ChatGPT-4o did not include these entries in the list of items that were referred to when it translated the following English sentence:

“At 1042, a crew member, who had been working on deck, ran to the bridge and alerted the master to the developing close-quarters situation.”

Now for the entry “vessel not under command 運転不自由船” in the knowledge base, the exact phrase appeared in the following sentence that was translated:

“At 1018, Happy Falcon came to a stop and started to drift while repairs were being made, but the required ‘vessel not under command’ signals were not displayed.”

Despite this exact match, ChatGPT-4o failed to refer to this item and failed to correctly translate the phrase “vessel not under command signals” into Japanese.

Table 1. Numbers of Translation Errors and Corrected Cases

	ChatGPT 4.0	ChatGPT-4o	ChatGPT-4o_K
Error Type	Errors	Corrected	Corrected
A-1	61	27	29
A-2	10	5	6
A-3	11	6	3
B	6	4	4
C	7	4	3
Total	95	46	45

One of my graduation thesis students, Ms. Wakana Shimonagane, who is exploring ChatGPT-4o's Maritime English translation possibilities by introducing a knowledge base under my supervision, suggested to me (personal communication) that the knowledge base data size may have a negative effect when the system refers to the knowledge base, as it sometimes stopped searching for possible matches. Given the limitation on the allocated computational resources, the data file size may have some impact. Thus, the present author created another file that contained only three items presented above, namely: "avoid a close-quarters situation," "result in another close-quarters situation," and "vessel not under command" and their Japanese counterparts. The two English sentences mentioned above were translated by ChatGPT-4o by referring to this much smaller knowledge base file, and the results were highly satisfactory. Both sentences were translated properly as follows:

10 時42分、甲板上で作業していた乗組員がブリッジに駆け込み、発生しつつある船舶に著しく接近する状況を船長に警告した。

1018 時、ハッピーファルコンは停止し、修理作業が行われている間に漂流し始めたが、必要とされる「運転不自由船」の信号は表示されなかった。

This suggests that the inclusion of the present large knowledge base did not lead to a sizable improvement in translation accuracy since ChatGPT-4o could not fully utilize the knowledge base information, and this may well be due to the computing power limitation.

4. SUMMARY AND DISCUSSIONS

Using the same set of bench mark Maritime English documents, the translation accuracies of DeepL, Google Translate and ChatGPT-4o were evaluated against their own performances about a year ago (DeepL and Google Translate) and ChatGPT 4.0 (an older free version). DeepL did not show any improvement over time and the improvement by Google Translate came from only one document, and why this was the case is unknown.

Since ChatGPT-4o allows its users to incorporate knowledge base files in making translations, the present author created a file that contained about 700 deck and engine English words and phrases and their Japanese counterparts and translated the bench mark files into English with and without referring to this file. Compared to ChatGPT 4.0, which was available for free about a year ago, ChatGPT-4o led to about 50% reductions in errors both with and with-

out the use of the knowledge base. This is a sizable improvement and ChatGPT-4o is much better than Google Translate and DeepL as far as translating Maritime English documents.

Surprisingly, the inclusion of the knowledge base did not lead to a higher reduction in translation errors, but this appears to be a limited computational power that is allocated to making a full appraisal of knowledge base information since by reducing the number of items in the knowledge base to minimum, the system was able to attain proper translations of identical sentences. At present, simply expanding the knowledge base in terms of its entry will not improve the translation accuracy by ChatGPT-4o. It remains to be seen, however, whether a large knowledge base can be successfully incorporated in translating English documents into Japanese as the ChatGPT-4o system becomes more powerful in terms of computational power.

Acknowledgements: This research was funded by a Grant-in-Aids for Scientific Research granted to the corresponding author (2024-2026). The author would like to thank Prof. Dr. Xiangen Hu for providing technical guidance in the use of ChatGPT.

Conflict of interest: There is no conflict of interest.

REFERENCES

- International Maritime Organization, 1972. The International Regulations for Preventing Collisions at Sea 1972. Available at: <https://doi.org/10.18356/9362d77f-en-fr>
- International Maritime Organization, 1974. The International Convention for the Safety of Life at Sea 1974. Available at: <https://doi.org/10.4324/9781315774695-86>
- Jääskeläinen, H. & Khair, M. K., 2018. Scavenging in Two-Stroke Engines. DieselNet Technology. Available at: https://dieselnet.com/tech/air_scavenge.php
- Marine Accident Investigation Branch, 2017. Safety warning after explosion on container vessel Manhattan Bridge results in 1 fatality and 1 serious injury. Available at: <https://www.gov.uk/maib-reports/safety-warning-after-explosion-on-container-vessel-manhattan-bridge-results-in-1-fatality-and-1-serious-injury>
- Marine Accident Investigation Branch, 2018. Dragging anchor and subsequent collisions by general cargo vessel Celtic Spirit. Available at: <https://www.gov.uk/maib-reports/dragging-anchor-and-subsequent-collisions-by-general-cargo-vessel-celtic-spirit>
- Marine Accident Investigation Branch, 2024. Collision between general cargo vessel Scot Explorer and gas carrier Happy Falcon. Available at: <https://www.gov.uk/>

maib-reports/collision-between-general-cargo-vessel-scot-explorer-and-gas-carrier-happy-falcon

Takagi, N. & Kimura, K., 2024. Evaluation of English Maritime Document Translation Accuracy into Japanese by ChatGPT, Google Translate and DeepL. Proceedings of the 29th International Maritime Lecturers Association Conference. Available at: <https://imla29.org/Abstract-Submission.aspx#abstract>

United Kingdom Hydrographic Office, 2022. Admiralty Sailing Directions: Japan Pilot Vol.1 (NP42B). Available at: <https://archive.ukho.gov.uk/records/SAR/1/42B/15>

Van Dokkum, K., 2013. Ship Knowledge, 8th Edition, The Netherlands: Dokmar Maritime Publishers B.V. Available at <https://www.scribd.com/document/814654563/Ship-Knowledge-Klaas-Van-Dokkum>