Reducing Redundancy in Japanese-to-English Translation: A Multi-Pipeline Approach for Translating Repeated Elements

Qiao Wang
Waseda University
Tokyo, Japan
judy.wang@aoni.waseda.jp

Yixuan Huang
Waseda University
Tokyo, Japan
yixuan.huang@moegi.waseda.jp

Zheng Yuan
King's College London
London, UK
zheng.yuan@kcl.ac.uk

Abstract

This paper presents a multi-pipeline Japanese-to-English machine translation (MT) system designed to address the challenge of translating repeated elements from Japanese into fluent and lexically diverse English. The system was developed as part of the Non-Repetitive Translation Task at WMT24, which focuses on minimizing redundancy while maintaining high translation quality. Our approach utilizes MeCab, the de facto Natural Language Processing (NLP) tool for Japanese, to identify repeated elements, and Claude Sonnet 3.5, a Large Language Model (LLM), for translation and proofreading. The system effectively accomplishes the shared task by identifying and translating in a diversified manner 89.79% of the 470 repeated instances in the test dataset and achieving an average translation quality score of 4.60 out of 5, significantly surpassing the baseline score of 3.88. The analysis also revealed challenges, particularly in identifying standalone noun-suffix elements and occasional cases of consistent translations or mistranslations.

1 Introduction

In the Japanese language, repetition at the word and phrasal levels is frequently employed (Fujimura-Wilson, 2007). One reason for this is that Japanese is a topic-prominent language, where the topic of the sentence is often explicitly stated and reiterated to ensure clarity and prominence (Tsujimura, 2013). Additionally, Japanese is highly context-dependent and typically omits subject pronouns, relying on the repetition of key nouns and verbs to maintain coherence (Maynard, 1997). Specifically for personal names, repetition is commonly used instead of pronouns to convey politeness and respect (Mogi, 2000).

In contrast, English typically favors variety and succinctness to maintain reader engagement (Hinkel, 2002; Halliday, 1994). Research in translation studies emphasizes the importance of lexical variety to ensure fluency and readability in translated texts (Baker, 1992; Newmark, 1988). Therefore, effectively translating repeated elements from Japanese to English may require the use of more diverse expressions while ensuring consistency and clarity.

The Non-Repetitive Translation Task at WMT24 addresses the challenge of translating repeated elements from Japanese into English (Kinugawa et al., 2024). This task aims to develop machine translation (MT) systems capable of identifying repeated expressions in Japanese text and translating them into lexically diverse and fluent English sentences. Participants are provided with training and test datasets comprising Japanese-English parallel corpora, in both raw and annotated formats with repeated targets tagged. Systems are evaluated on their ability to minimize redundancy while maintaining high translation quality.

Our contribution includes the development of a multi-pipeline MT system that effectively avoids redundancy in translating repeated words and phrases from the source Japanese text. Specifically, we utilized MeCab (Kudo, 2005) for tokenization and lemmatization of Japanese sentences to identify repeated elements and adopted the Large Language Model (LLM) Claude Sonnet 3.5 (Anthropic, 2024) for translation and proofreading. When compared with the baseline system provided by the task organizers, our system achieved an average translation score of 4.60, significantly higher than that of the baseline system at 3.88; and a BLEU metric of 24.4

Proceedings of the Ninth Conference on Machine Translation, pages 1047–1055
November 15-16, 2024 ©2024 Association for Computational Linguistics

compared with human benchmark translation.

2 Related work

The identification of repeated elements in Japanese poses unique challenges due to the language's agglutinative nature (Tsujimura, 2013). Unlike Indo-European languages where word boundaries are clearly delineated (Baker, 1992), Japanese requires sophisticated tokenization methods to accurately segment text into meaningful units. As far as we are aware, no previous NLP studies have specifically focused on identifying repeated tokens in Japanese, although some have explored the identification of repetition at the semantic level, not necessarily of the same words (Kawamoto, 2022). However, the repeated elements in this shared task need to be identical, the most straightforward method is to adopt an Natural Language Processing (NLP) tool designed specifically for the Japanese language. Previous studies have recommended MeCab as the de facto text segmentation library capable of part-of-speech (POS) tagging, tokenization and lemmatization for Japanese (Kudo, 2005). Even with MeCab, challenges in identifying repeated elements persist. For example, for "国文学" (Japanese literature) and "漢文学" (Chinese literature), humans may easily detect the repeated element "文学" (literature). However, for MeCab, "国文学" is regarded as one token, while "漢文学" is regarded as two tokens, "漢" and "文学". As such, no repetition can be detected as the tokens do not match.

Recent advancements in LLMs have prompted us to explore them for identifying repeated elements in Japanese text. Prior to building our system, we experimented with GPT-3.5[1] (fine-tuned on the WMT24 training dataset), GPT-4[2] (used direct prompting), and Claude Sonnet 3.5[3] (also used direct prompting) to assess their ability to detect repeated tokens. However, none of these LLMs consistently identified repetitions. This indicates that rule-based approaches using MeCab remain necessary.

Machine translation (MT) between Japanese and English has long been challenging due to the significant linguistic differences between the two languages (Wang et al., 2022). While Neural Machine Translation (NMT) systems, particularly those based on Transformer models (Vaswani et al., 2023), have demonstrated success in handling structured text, they face limitations when dealing with informal language, idiomatic expressions, and culturally specific references. Meanwhile, LLMs such as GPT-3.5 and GPT-4 have shown considerable promise in improving translation quality in Japanese-English (JA-EN) tasks (Vincent et al., 2024). These models benefit from extensive training on diverse datasets, which allows them to generate more contextually appropriate translations, particularly in cases where traditional supervised NMT systems may struggle (Siu, 2023).

Previous studies have compared NMT systems and LLMs in translating high-resource and low-resource languages and found that LLMs such as GPT series perform better in high but not low-resource languages (Robinson et al., 2023). Since Japanese is not considered a low-resource language, we expect LLMs to perform better in this shared task. As thus, we piloted with Google Translate[4], GPT4o and Claude Sonnet 3.5 with the training dataset and found the Claude Sonnet 3.5 performed the best in translating the Japanese texts.

In addition to translation, we aimed to incorporate a proofreading pipeline to enhance overall translation quality. This idea was inspired by the translation-review-revise methodology, which emphasizes iterative improvement of translated content through multiple stages of refinement (Ng, 2024). In line with this methodology, we aimed to incorporate a proofreading process with LLMs as well. In the following, we'll describe in detail the system design and implementation results of our system for the shared task.

3 System description

Our system comprises of four pipelines: 1) the preprocessing pipeline for identifying and subsequently assigning IDs and reference numbers

[1] https://platform.openai.com/docs/models/gpt-3-5

[2] https://platform.openai.com/docs/models/gpt-4

[3] https://www.anthropic.com/index/claude

[4] https://translate.google.com/

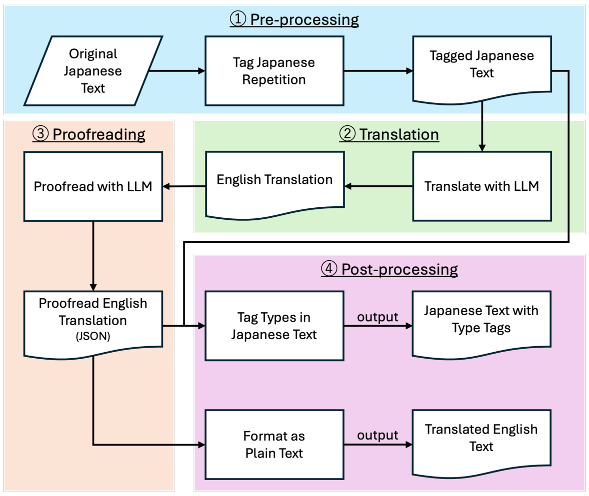

Figure 1: Workflow of the system

(ref numbers) to repeated elements, i.e., targets, in Japanese sentences; 2) the translation pipeline for translating the Japanese sentences into English while trying to reduce redundancy; 3) the proofreading pipeline for revising the translated sentences; and 4) the post-processing pipeline for appending the types of strategies, i.e., substitution or reduction, to the IDs and ref numbers of the repeated targets in the Japanese sentences. The input of the system is the raw Japanese sentences while the output includes the Japanese sentences with repeated targets assigned IDs, ref numbers and types, and raw translated English texts. Figure 1 shows the workflow of the four pipelines. The source code of the developed system can be found at our GitHub repository[5].

3.1 Preprocessing pipeline

For identifying and subsequently assigning IDs and ref numbers to repeated targets in Japanese sentences, we adopted MeCab. The POS tags in MeCab are structured into a maximum of three hierarchical layers. For example, a three-layered POS tag can be "名詞-普通名詞-副詞可能" (noun-common-adverbial), with the top layer indicating the token is a noun, the second layer further explaining it is a common noun, and the third layer showing that the noun can also be used as an adverb.

[5]https://github.com/judywq/non-repetitive-translation

Step 1: POS tagging We performed POS tagging on all the Japanese sentences in the training dataset, creating a pool of tokens with their POS tags.

Step 2: POS screening From the pool, we first did a simple token match to find repeated tokens. Then, we compared the repeated tokens with the tagged repeated targets in the training dataset to decide what POS tags should be included and what should not in identifying repeated tokens. The result was a 'whitelist' of POS tags at different layers. In the first layer, we focused only on content words, including nouns, verbs, adjectives, adverbs and prefixes, while excluding function words such as auxiliaries, conjunctions and particles. For the POS tags from the second layer on, we maximized the coverage of POS tags found in the training dataset while reducing noises. Table 1 shows the POS tags in the 'whitelist'. Blanks in the third layer indicate that all the third-layer tags have been selected.

Step 3: Identifying targets When identifying the repeated targets, the easiest way is to do exact match. However, for some POS tags, special treatment was necessary. These include Verb, Noun-Suffix, and Prefix-Noun connection.

For Verb, we adopted their lemmatized forms using MeCab. This is due to the fact that Japanese verbs are rich in inflections. For example, the verb "食べる" (*taberu*, meaning "to eat") can appear in various forms depending on the tense, politeness level, and grammatical context, such as "食べた" (*tabeta*, past tense "ate"), "食べます"(*tabemasu*, polite form "eat"), or "食べられる" (*taberareru*, potential form "can eat").

For Noun-Suffix and Prefix-Noun, as they are suffixes or prefixes, they should be bound to another token. We thus added a rule where when tokens with the two POS tags are repeated, their neighboring tokens, i.e., the token before the suffix and the token after the prefix, should also be bound together with them. If the bound elements still match, then they are valid targets. Otherwise, they will be dismissed. For example, with the Noun-Suffix "者" (person), if it is preceded twice by the verb "容疑"(suspect), the compound noun

Layer 1	Layer 2	Layer 3
Adverb (副詞)	General (一般)	
Verb (動詞)	Independent (自立)	
Noun (名詞)	Suffix (接尾)	Adjectival noun stem (形容動詞語幹)
		Personal name (人名)
		Area (地域)
		Special (特殊)
	Dependent (非自立)	Adverbial (副詞可能)
	Suru verb (サ変接続)	
	Nai adjective stem (ナイ形容詞語幹)	
	General (一般)	
	Proper noun (固有名詞)	
	Adverbial (副詞可能)	
	Adjectival noun stem (形容動詞語幹)	
Adjective (形容詞)	Independent (自立)	
Prefix (接頭詞)	Noun connection (名詞接続)	

Table 1: POS tags in the ‘whitelist’

“容疑者”(a suspected person) will be the valid target. However, if it is preceded by “容疑” once and by “被爆”(be bombed) the second time, they will be dismissed as “容疑者” does not match “被爆者”(an atomic bomb victim).

Step 4: Assigning IDs and ref numbers We assigned IDs and ref numbers to targets based on their order of occurrence in the Japanese sentence. The output of step 4 are Japanese sentences with repeated targets tagged.

3.2 Translation pipeline

The translation pipeline is responsible for translating the Japanese sentences into English while trying to adopt diversified expressions for the repeated targets tagged in the Japanese sentences; singling out the translation for each occurrence of the targets; and deciding which type and subtype of strategy was used. It should be noted that to facilitate understanding of output from the pipeline, we introduced two new types: “first occurrence” and “consistency” and another subtype: “literal translation”, to the original types (substitution and reduction) and subtypes from the official website of the task (WMT24, 2024). First occurrence is assigned to translation of a target where the translation appeared for the first time and thus there is no need to reduce redundancy. Consistency refers to situations where the target is translated into the same English expressions across multiple occurrences. The subtype “literal translation” is added to complement “non-literal translation” original included in the examples from the official website.

For this pipeline, we adopted Claude Sonnet 3.5 with few-shot in-context learning prompting techniques. In our prompt, we included the explanation of the shared task and the examples of reduction and substitution from the task’s official website. Then we asked the translation pipeline to translate the sentences while trying its best to adopt diversified expressions for the repeated targets tagged in the Japanese sentences, single out the translations for the targets and decide which type and subtype of strategy was used.

For example, for the input sentence:

> RCEP では、7 月 1 日に東京で閣僚 <target id=0 ref=0> 会合 </target> が開かれ、妥結に向け 11 月下旬にシンガポールで首脳 <target id=0 ref=1> 会合 </target> が開催できるよう、交渉に注力する方針を確認。

the output from the translation pipeline is as follows:

```
{
"en_translation": "For RCEP, a ministerial
```

```
meeting was held in Tokyo on July 1,
confirming the policy to focus on negotiations
with the aim of holding a summit in Singapore
in late November to reach an agreement.",
  "targets":
  [
    {
      "id": "0",
      "ref": "0",
      "ja_element": "会合",
      "en_element": "meeting",
      "type": "f",
      "subtype": "lt"
    },
    {
      "id": "0",
      "ref": "1",
      "ja_element": "会合",
      "en_element": "summit",
      "type": "s",
      "subtype": "syn"
    }
  ]
}
```

3.3 Proofreading pipeline

For the proofreading pipeline, we also adopted Claude Sonnet 3.5. Few-shot in-context learning prompt was also designed for this pipeline as for the translation pipeline. The proofreading pipeline receives the Japanese sentence and the translated text, the translations for each occurrence of repeated targets and the types and subtypes of strategies used. It is asked to check if the translation is faithful to the Japanese sentence and if redundancy can be further reduced. If changes are not necessary, it returns

```
{"changed": "No"}
```

Otherwise, it returns

```
{"changed": "Yes"}
```

followed by a revised output in the same format as the translation pipeline. A sample output from the pipeline is shown below:

```
{
  "changed": "Yes",
  "en_translation_updated": "Toshiba stated
  that there is no change to its previous
  projection, as the reversal is already
  incorporated into the full-year earnings
  outlook for the fiscal year ending
  March 2019.",
  "targets_updated":
  [
    {
      "id": "0",
      "ref": "0",
      "ja_element": "予想",
      "en_element": "outlook",
      "type": "f",
      "subtype": "lt"
    },
    {
      "id": "0",
      "ref": "1",
      "ja_element": "予想",
      "en_element": "projection",
      "type": "s",
      "subtype": "syn"
    }
  ]
}
```

3.4 Post-processing pipeline

With the proofread sentence translation and translations of each occurrence of repeated targets together with the types and subtypes of strategies, the post-processing pipeline first appends the types of the translations to the ID and ref number of each occurrence of repeated targets in the Japanese text. Then it replaces the two types we added with the two official types, reduction or substitution. Specifically, in the case of first occurrence, the type of same ID but the following/previous ref number target will be adopted. For example, if for a target A where it appears twice in a sentence, it has two occurrences: ID=0, Ref=0, Type= first occurrence; and ID=0, Ref=1, Type= substitution. The type in ref=0 is replaced by the type in ref=1, substitution. For consistency, which means our system deems it unnecessary to reduce redundancy, we remove the IDs and ref numbers of the targets in the Japanese text. This is based on the official dataset where only the targets that require redundancy reduction are tagged with IDs and ref numbers.

After the post-processing pipeline, the system outputs the Japanese sentences with IDs, ref numbers and types tagged for repeated targets and the raw English translation without any tags.

We also considered situations where our system may fail to identify any repeated targets in the pre-processing pipeline. In such cases, the raw Japanese sentences will be translated into English by Claude Sonnet 3.5 and the system will directly output the raw Japanese sentence and its English translation.

JPO Adequacy	<NON-REP>	<REP>	<INCORRECT>	Total
[5,5]	127	20	0	147
[4,5)	280	17	3	300
[3,4)	15	1	7	23
[2,3)	0	0	0	0
[1,2)	0	0	0	0
Total	422	38	10	470

Table 2: JPO adequacy and translation style

4 System implementation and results

Our system was applied to translate 470 sentences from the test dataset provided by the Non-Repetitive Translation Task at WMT24. The output from our system was rigorously evaluated by three human raters assigned by the task organizers. Each rater independently reviewed the translation of the repeated targets and assigned each target a translation quality score ranging from 1 to 5 based on criterion for translated patent documents from the Japanese Patent Office (JPO), with 5 representing the highest quality (JPO adequacy). The raters also labelled each translation target with one of the following translation styles: "c" (consistent/repetitive translation), "s" (substitution), "r" (reduction), or "m" (mistake in translation). The final evaluation score for each sentence was determined by averaging the scores given by the three raters and the label was determined by a majority vote.

In the 470 sentences, there are a total of 489 repeated targets tagged by the organizers in the dataset, meaning that some sentences contain more than 1 repeated target. When there are multiple targets in one sentence, the evaluation of all targets is aggregated to one by the organizers, resulting in 470 evaluation instances for the 470 sentences in total. Results suggest that our system produced 38 instances of repetitive translations, 422 instances of correct non-repetitive translations, and 10 incorrect translations. This shows a correct non-repetitive translation rate of 89.79% for our system.

Table 2 shows the detailed evaluation results including the instance counts of translation styles (non-repetitive, repetitive and incorrect translation) and translation quality (JPO adequacy).

The average JPO adequacy of our system is 4.60, significantly higher than that of the baseline system at 3.88 (t=14.09, p<0.00). To view the balance between translation quality and style, the JPO adequacy score for each instance are converted to 0 if its style is not '<NON-REP>', i.e., correct non-repetitive translation. The average of this filtered JPO scores is 4.13. This is significantly higher than the baseline system at 2.13 (t=16.60, p<0.00).

For reference purposes, our system's performance was also evaluated using the BLEU metric (Papineni et al., 2002). The BLEU score for our system was 24.4, indicating moderate similarity to the human benchmark translations provided by the organizers. The verbose BLEU score breakdown shows a precision of 58.3% for 1-grams, 30.0% for 2-grams, 18.0% for 3-grams, and 11.3% for 4-grams. No Brevity Penalty (BP) was applied, as the length of the system's output (15,700 words) closely matched that of the benchmark (15,579 words), with a length ratio of 1.008.

5 Discussions

Our system demonstrated high performance in the shared task, effectively combining multiple NLP and LLMs pipelines to achieve impressive results. However, some issues were observed that highlight the challenges of this task and the limitations of our approach.

One challenge our system faced was in certain tagging targets deemed repetitive by human raters, particularly those related to standalone noun-suffix elements. For example:

> **Japanese:** 専門家の 1 人は、鑑定した 110 個の遺骨の中で日本人の DNA<target_1> 型 </target_1> は 5 個、フィリピン人の <target_1> 型 </target_1> が 54 個だったと報告。
>
> **System Translation:** One of the experts reported that among the 110 bone samples examined, 5 had DNA patterns matching Japanese individuals, while 54 had patterns matching Filipino individuals.

In this instance, our system failed to identify the target "型" (type), a noun-suffix, as a repeated element and subsequently the translation style was consistent. According to Step 3 in the pre-processing pipeline, noun-suffixes are only identified as repeated targets if their preceding tokens also match. This rule was implemented to reduce noise; however, it led to the negligence of independent suffixes tagged by humans in the benchmark dataset. A similar issue occurred with the word "量" (quantity) in the following sentence:

> **Japanese:** 国際貨物 <target_1> 量 </target_1> は 10%減の 16 万 8510 トン。|| ジェット燃料給油 <target_1> 量 </target_1> は 2%減の 37 万 3805 キロリットルだった。
>
> **System Translation:** International cargo volume decreased by 10% to 168,510 tons. Jet fuel supply volume fell by 2% to 373,805 kiloliters.

In this example, "量" was not identified as a repeated target, resulting in the translation being repeated instead of diversified.

Moreover, there were instances where our system correctly identified repeated targets but still opted for consistent translations. This occurred in cases where the system judged consistent translation to be preferable or where it misinterpreted derivatives or inflected forms of a word as non-repetitive. For example:

> **Japanese:** 同氏は <target_1> 五輪 </target_1> には 04 年アテネ <target_1> 五輪 </target_1> から 4 大会連続で出場した。
>
> **System Translation:** He competed in four consecutive Olympic Games, starting with the 2004 Athens Olympics.

In this case, our system treated "Olympic Games" and "Olympics" as substitutions. However, human raters considered these terms to be consistent translations, as the word "Olympic" and its inflected form "Olympics" are essentially the same. Though we stated specifically in our prompts to Claude that derivatives and inflected forms are consistent translations, it failed to perform the translation/proofreading task adequately in some cases.

In the 10 instances where the system produced incorrect translations, these errors occur because the translations are too flexible and non-literal, compromising the "faithfulness" of translation. An example of such an error is:

> **Japanese:** 一方、66 歳以降も働きたいと答えた人が挙げた理由は、「<target_1> 経済的 </target_1> にゆとりある生活を送りたい」が 28.9%、「働き続けないと生活費が足りないと思う」24.9%などで、<target_1> 経済的 </target_1> な理由が半数を超えた。
>
> **System Translation:** On the other hand, among those who expressed a desire to continue working beyond the age of 66, over half cited financial reasons. The most common reasons were 'wanting to maintain a comfortable lifestyle' (28.9%) and 'believing that living expenses would be insufficient without continued employment' (24.9%).

In this example, "the most common reasons" should be interpreted as "the most common comments given by those who cited financial reasons". However, the translation assumes that readers can infer from the text, but the same word "reasons" makes the sentence confusing.

6 Conclusions

In conclusion, the proposed multi-pipeline Japanese-to-English machine translation system successfully addresses the challenge of translating repeated elements from Japanese into fluent and varied English. By integrating MeCab for accurate tokenization and Claude Sonnet 3.5 for translation and proofreading, the system achieved a high rate of correct non-repetitive translations, with a translation quality score that significantly exceeded the baseline. However, certain challenges remain, particularly in identifying and translating standalone noun-suffix elements and in cases where consistent translation is deemed preferable. Additionally, the study highlighted the limitations of current human evaluation processes, where inter-rater reliability was low, affecting the consistency of the evaluation results. Future work could explore more advanced language models and refined evaluation methodologies to further enhance the system's performance and address these challenges.

Limitations

In our translation pipeline, we compared the performance of Claude with Google Translate and GPT-4 before selecting Claude as the translation model. However, it is important to acknowledge that more effective LLMs may emerge in the future, which could offer improved performance. Besides, one of the inherent issues of relying on commercial LLMs like Claude is the issue of token limits, which can pose challenges in large-scale projects where the tasks requires days to complete.

Furthermore, the inter-rater reliability among the human raters was relatively low. We noticed that one of the raters was conspicuously more severe in their evaluations compared to the other two raters. The inter-rater reliability analysis also revealed only a slight agreement among the raters for JPO Adequacy, with an average Weighted Kappa (Cohen, 1968) of 0.161. The Fleiss' Kappa (Fleiss, 1971) for Style was -0.204, suggesting that the agreement among the raters was not only poor but worse than what would be expected by chance. This means that the evaluation results may have differed if the inter-rater reliability was higher. To illustrate, the raters did not reach consensus on some mistranslations. The following shows an example:

> **Japanese:** JAXA によると、<target_1> クレーター </target_1> は直径 10 メートル規模と推測され、小惑星への人工 <target_1> クレーター </target_1> 生成に成功したのは世界で初めてだという。
>
> **System Translation:** According to JAXA, the crater is estimated to be about 10 meters in diameter, marking the world's first successful artificial impact on an asteroid.

In this example, the word "impact" can be considered a term referring to "a collision between astronomical objects causing measurable effects" (Rumpf et al., 2017) in planetary science, which usually results in the formation of an impact crater. In this sense, it may be a substitution to "crater". Indeed, one rater regarded it as an appropriate substitution and scored it a 5, while the other two raters considered it an incorrect translation. Such disagreement highlights the importance of a more rigorous and standardized human evaluation process in future tasks.

Acknowledgement

We would like to thank the task organizer and their raters who evaluated the performance of our system. Our gratitude also extends to Jiji Press LTD for providing the datasets used in this study.

References

Anthropic. 2024. Claude 3.5. Large language model, retrieved from https://www.anthropic.com/claude.

Mona Baker. 1992. *In Other Words: A Coursebook on Translation*. Routledge.

Jacob Cohen. 1968. Weighted kappa: nominal scale agreement with provision for scaled disagreement or partial credit. *Psychological Bulletin*, 70(4):213–220.

Joseph L. Fleiss. 1971. Measuring nominal scale agreement among many raters. *Psychological Bulletin*, 76(5):378–382.

Kayo Fujimura-Wilson. 2007. Japanese exact repetitions involving talk among friends. *Discourse Studies*, 9(3):319–339.

M. A. K. Halliday. 1994. *An Introduction to Functional Grammar*. Edward Arnold.

Eli Hinkel. 2002. *Second Language Writers' Text: Linguistic and Rhetorical Features*. Lawrence Erlbaum Associates.

Toshiki Kawamoto. 2022. Generating repetitions with appropriate repeated words. *Journal of Natural Language Processing*, 29(4):1302–1307.

Kazutaka Kinugawa, Hideya Mino, Isao Goto, and Naoto Shirai. 2024. Findings of the wmt 2024 shared task on non-repetitive translation. In *Proceedings of the Ninth Conference on Machine Translation (WMT)*.

Taku Kudo. 2005. Mecab: Yet another part-of-speech and morphological analyzer.

Senko K. Maynard. 1997. *Japanese Communication: Language and Thought in Context*. University of Hawai'i Press.

Norie Mogi. 2000. Japanese ways of addressing people. *Investigationes Linguisticae*, VIII. ∗.

Peter Newmark. 1988. *A Textbook of Translation*. Prentice Hall.

Andrew Ng. 2024. Machine learning yearning: Technical strategy for ai engineers, in the era of deep learning. DeepLearning.AI.

Kishore Papineni, Salim Roukos, Todd Ward, and Wei-Jing Zhu. 2002. Bleu: a method for automatic evaluation of machine translation. In *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics (ACL)*, pages 311–318, Morristown, NJ, USA. Association for Computational Linguistics.

Nathaniel R. Robinson, Perez Ogayo, David R. Mortensen, and Graham Neubig. 2023. Chatgpt mt: Competitive for high- (but not low-) resource languages. In *Proceedings of the Eighth Conference on Machine Translation (WMT)*, page 392–418. Association for Computational Linguistics.

Clemens M. Rumpf, Hugh G. Lewis, and Peter M. Atkinson. 2017. Asteroid impact effects and their immediate hazards for human populations. *Geophysical Research Letters*, 44(8):3433–3440.

Sai Cheong Siu. 2023. Chatgpt and gpt-4 for professional translators: Exploring the potential of large language models in translation. Available at SSRN: `https://ssrn.com/abstract=4448091` or `http://dx.doi.org/10.2139/ssrn.4448091`.

Natsuko Tsujimura. 2013. *An Introduction to Japanese Linguistics*. Wiley-Blackwell.

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Łukasz Kaiser, and Illia Polosukhin. 2023. Attention is all you need. *arXiv*, arXiv:1706.03762. Doi: 10.48550/arXiv.1706.03762.

Michael Vincent, Giovanni Sutanto, De Giacomo Gatti, Toshiaki Nakazawa, and Masaru Yamada. 2024. Chatgpt as a translation engine: A case study on japanese-english. In *Proceedings of the Machine Translation Summit XVIII*, pages 95–98.

Haifeng Wang, Hua Wu, Zhongjun He, Liang Huang, and Kenneth Ward Church. 2022. Progress in machine translation. *Engineering*, 18:143–153.

WMT24. 2024. Non-repetitive translation task. Website.