A PLAIN JAPANESE TRANSLATION APP USING OCR TO BOOST READING COMPREHENSION SKILLS

Teppei MIURA* [a]

[a] National Institution of Technology (KOSEN) Toyota College, Toyota City, Japan

Teppei MIURA* miura.teppei.work@gmail.com

Comprehension is an important aspect of reading. One effective method to improve reading comprehension skills is reading plain language text. Plain language writing produces documents that are clear, concise, and easy to understand. Language learners can boost their reading comprehension skills, including vocabulary and grammar, by reading documents that have been translated into plain language from complex text. We develop a web application for plain Japanese language translation using OCR for input text. OCR converts images of text into computer-encoded text. The proposed application allows learners to easily obtain plain Japanese text, simply by taking a photo of original text. In the evaluation, we conduct a reading comprehension test using the Japanese Language Proficiency Test with six international students. Participants complete test set using three different methods: the proposed application, the existing system that requires manual input, and without using any assistance. The results indicate that the proposed application helps students improve their comprehension of evaluation tests.

Keywords: *reading comprehension skills, plain Japanese language, OCR, web application*

Introduction

Comprehension is an important aspect of reading. People read to understand and extract meaning, gaining better awareness of what a document conveys (Herrity (2025)). One effective method to improve reading comprehension skills is reading plain language text. Plain language writing produces documents that are clear, concise, and easy to understand (Bricks Co., Ltd. (2020)). Language learners can boost their reading comprehension skills, including vocabulary and grammar, by reading documents that have been translated into plain language from complex text.

Japanese language is one of the most complex languages in the world. Reading and writing pose a significant challenge for non-native learners because Japanese uses three types of characters -*hiragana*, *katakana*, and *kanji*- with *kanji* alone comprising over 2,000 characters. Although some plain Japanese translation systems have been developed in recent years to be available improving reading comprehension skills, these systems still need better usability, as they require users to manually input complex Japanese text.

We develop a web application for plain Japanese language translation using OCR, which means “optical character recognition” technology, for input text. OCR converts images of text into computer-encoded text from a photo of a document. The proposed application allows learners to easily obtain plain Japanese text to boost their reading comprehension skills, simply by taking a photo of original text, without the need to manually input complex characters.

Materials and Methods

We describe the application design based on the following implementation functions:

- Optical Character Recognition
- Plain Japanese Translation
- Application Design

Optical Character Recognition

We compare three implementation approaches to consider the use of OCR. The first is “Tesseract”, an open-source library that is compatible with multiple operating systems. The second is “EasyOCR”, a Python-based library. The third is “Vision AI”, a web API service provided by Google.

We prepare an image containing Japanese text to compare the performance of each OCR tool. Each tool processes the image and output the text data. As a practical use case, we take a photo of the document using a smartphone; the captured image is shown in Figure 1. The image is in JPG format, with a resolution of 2419x3225 pixels and a file size of 1.6MB. The output results from each OCR tool are indicated in Table 1.

Vision AI indicate the highest accuracy among the OCR tools, followed by Easy OCR and Tesseract, as indicated in Table 1. Additionally, Vision AI successfully recognizes and outputs nearly all the characters in the image. Therefore, we adopt Vision AI as the implementation tool.

Plain Japanese Translation

We use the ChatGPT API provided by OpenAI for translating into plain Japanese text. ChatGPT is a service that generates natural language based on a given text.

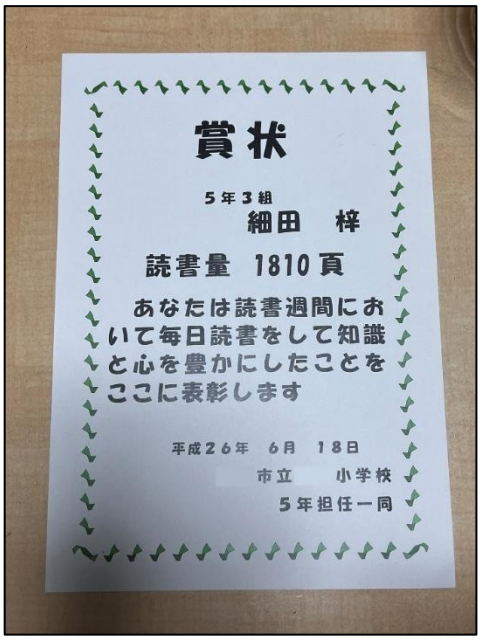


Figure 1. An image to compare OCR performance.

Table 1. Output from OCR of the image in Figure 1.

library	Output
Tesseract	引いプ NN へへへへへへへへ YY へ>。軸較納着 3| いいんき/Ne 開寺杖」記 2 に ANN1 昌間|有馬き人 Ss 95 年 3 組」し。さか）も|和んど 1 細田害 8] 和を 0 読書量 1810 具 8 -、、貞人 KK ロ 2C で 5. 三き「毎日読書をして知識>還|ん(1 て写日読書 Bk| |てとを豊かにしたことを較剛 WCS 用 LA でみ Sa| 上蘭間いこごごに表彰します>国(知っかーデシ旦をず 2 も替わり NG を。[続い 5 玉成 26 年 6 月 18 日ひみ 2 市立由学検の隊、年担任一同生還/ ままままままる 4 机| 上 Mg
easyOCR	賞状 5 年宮組細田梓読書量 1810 肩|ョ三西成口島年 6 月 7 昌日市立小学校 5 年担任一同う
Vision AI	**^^^^^^^^^^^^ 賞状 5 年 3 組細田梓読書量 1810 頁あなたは読書週間において毎日読書をして知識と心を豊かにしたことをここに表彰します平成 26 年 6 月 18 日市立| 小学校 5 年担任一同

Table 2. ChatGPT plain Japanese translation examples.

Original	ただ、所々丹塗の剥げた、大きな円柱に、蟋蟀が一匹とまっている。
Plain	しかし、大きな丸い柱の一部分の赤い塗りがはがれているところに、コオロギが 1 匹とまっています。
Original	必ず、かの邪智暴虐の王を除かなければならぬと決意した。
Plain	あの悪い王を絶対に取り除かなければならないと決めた。
Original	吾輩は猫である。名前はまだ無い。
Plain	私は猫だよ。まだ名前はないんだ。

We will continue to improve the application usability based on user feedback. We can easily adjust the level of plain text by providing prompts such as “Translate into plain Japanese that elementary school students can understand” because of high scalability of ChatGPT.

We first assess the ChatGPT’s capability to translate text into plain Japanese, specifically using the GPT-4 model. We refer to the passages as examples of complex text from following Japanese literature:

- *Rashomon* by Ryunosuke Akutagawa
- *Run, Melos!* by Osamu Dazai
- *I Am a Cat* by Soseki Natsume

We input each passage into ChatGPT with the prompt “Please rewrite the following text in plain Japanese”. The outputs are indicated in Table 2. we confirmed that the sentences were successfully translated into plain text however the outputs include a mix honorific and colloquial expressions.

Application Design

Users take a photo of a Japanese text to be translated into plain language using the device’s camera. Then the application displays both the original text and translation results on the screen. We show the display design in Figure 2. We describe the data flow between the application device and cloud-based services such as Vision AI API and ChatGPT API in Figure 3. Additionally, we indicate the process flow below.

1. Activate the camera.
2. Call Vision AI API with a captured image.
3. Receive the original text from Vision AI API.
4. Display the original text.
5. Call ChatGPT API with the original text.
6. Receive the plain text from ChatGPT API.
7. Display the plain text.

We use a Python-based web application framework to develop the plain Japanese translation system, as both the Vision AI and ChatGPT API are compatible with the Python language. Although Django and Flask are

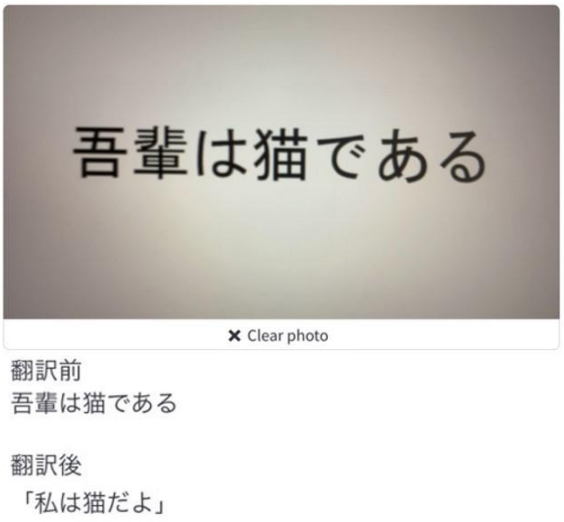


Figure 2. Web application execution screen.

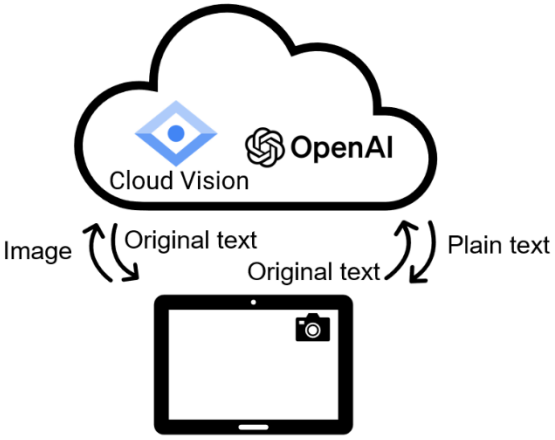
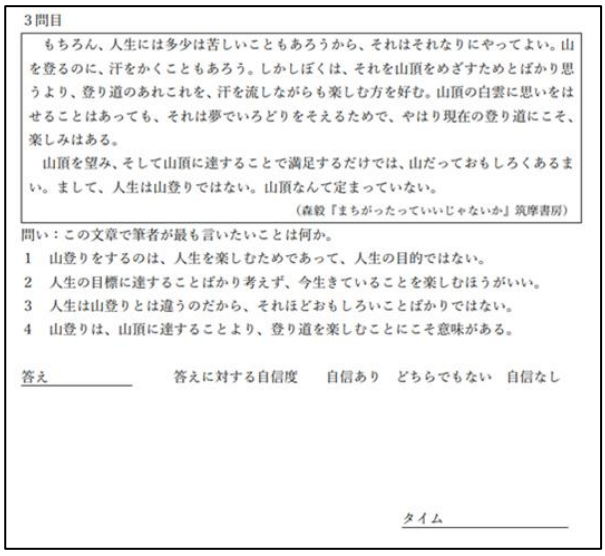


Figure 3. Data flow between the device and the web services.


3問目

もちろん、人生には多少は苦しいこともあろうから、それはそれなりにやってよい。山を登るのに、汗をかくこともあろう。しかしぼくは、それを山頂をめざすためとばかり思うより、登り道のあれこれを、汗を流しながらも楽しむ方を好む。山頂の白雲に思いをはせることはあっても、それは夢でいろどりをそえるためで、やはり現在の登り道にこそ、楽しみはある。

山頂を望み、そして山頂に達することで満足するだけでは、山だっておもしろくあるまい。まして、人生は山登りではない。山頂なんて定まっていない。

（森毅『まちがったっていいじゃないか』筑摩書房）

問い：この文章で筆者が最も言いたいことは何か。

1 山登りをするのは、人生を楽しむためであって、人生の目的ではない。
2 人生の目標に達することばかり考えず、今生きていることを楽しむほうがいい。
3 人生は山登りとは違うのだから、それほどおもしろいことばかりではない。
4 山登りは、山頂に達することより、登り道を楽しむことにこそ意味がある。

答え＿＿＿＿　答えに対する自信度　自信あり　どちらでもない　自信なし

タイム＿＿＿＿

Figure 4. An example of a reading comprehension test at the N1 level.

popular Python frameworks, they require the use of OpenCV to implement camera functionality, which can be cumbersome. Therefore, we adopt Streamlit as the framework. Streamlit is a framework for developing web applications in Python and is widely used in the fields of machine learning and data science. It allows easy integration of camera functionality within a web application.

Results and Discussion

Evaluation Design

We conduct reading comprehension tests (Method1 to Method3) based on the Japanese Language Proficiency Test (JLPT) with six international students at National Institute of Technology (KOSEN), Toyota College.

In Method1, the participants answer a reading comprehension test without any support tools. In Method2, participants manually input a part of Japanese text into their smartphones to translate it into plain language using the existing system provided by Alfasado Inc. (2025). In Method3, participants use the proposed application to take a photo of the Japanese text with a smartphone camera and translate it into plain language.

A reading comprehension test consists of three questions, each corresponding to the difficulty levels of N3, N2, and N1 of the JLPT. Participants complete one test for each method: Method1, Method2, and Method3. An example of a reading comprehension question at the N1 level is shown in Figure 4. We collect additional information from participants, including confidence in each answer, the response time, and feedback on their experience.

Results and Discussion

We indicate the overall results in Table 3. We indicate JLPT level and years of Japanese learning for each participant A to F. The results of the reading comprehension tests are indicated as True (1) or False (0). The participants' confidence levels in each answer are also indicated as High (1), Middle (0), or Low (-1). Additionally, the response time for completing the three questions in each test is also indicated.

We indicate the average correct answer rates for each method and each question in Table 4. The correct answer rates decrease as the difficulty level increases from Q1(N3) to Q3(N1). The proposed method (Method3) achieved a correct answer rate of 0.67, which was lower than that of the method without any support tools (Method1).

We indicate the average confidence levels for each method and each question in Table 5. Similar to the correct answer rates, the confidence levels also decrease as the difficulty increases. However, the proposed method (Method3) achieved the highest confidence level, at 0.56, when compared to the other methods. These results suggest that the proposed method contributed to helping participants' answering with greater confidence.

There is a possibility that participants may answer correctly by coincidence since the reading comprehension test in the JLPT is multiple-choice questions. Therefore, we indicate the correct answer rates based only on the questions where participants understood the content and answered with confidence.

Table 3. The overall results of evaluation.

Participant			A	B	C	D	E	F
JLPT level			N1	None	N2	N4	N1	N1
Years of learning			5 – 6	1 – 2	3 – 4	3 – 4	longer 7	3 – 4
Method1	Q1 (N3)	Correctness	True (1)	True (1)	True (1)	True (1)	True (1)	True (1)
		Confidence	High (1)	High (1)	High (1)	High (1)	High (1)	High (1)
	Q2 (N2)	Correctness	True (1)	False (0)	False (0)	True (1)	True (1)	True (1)
		Confidence	Middle (0)	Middle (0)	High (1)	Low (-1)	Middle (0)	High (1)
	Q3 (N1)	Correctness	True (1)	False (0)	True (1)	True (1)	True (1)	False (0)
		Confidence	Middle (0)	Middle (0)	High (1)	Low (-1)	Low (-1)	Middle (0)
	Time (second)		241	382	622	333	278	228
Method2	Q1 (N3)	Correctness	True (1)	False (0)	True (1)	True (1)	True (1)	True (1)
		Confidence	High (1)	High (1)	High (1)	Middle (0)	High (1)	High (1)
	Q2 (N2)	Correctness	False (0)	False (0)	False (0)	True (1)	True (1)	True (1)
		Confidence	Middle (0)	Middle (0)	High (1)	Low (-1)	High (1)	Middle (0)
	Q3 (N1)	Correctness	False (0)	False (0)	True (1)	True (1)	True (1)	False (0)
		Confidence	Low (-1)	Middle (0)	High (1)	Low (-1)	High (1)	Low (-1)
	Time (second)		580	576	653	387	217	322
Method3	Q1 (N3)	Correctness	True (1)	True (1)	True (1)	True (1)	True (1)	True (1)
		Confidence	High (1)	High (1)	High (1)	High (1)	High (1)	High (1)
	Q2 (N2)	Correctness	False (0)	False (0)	True (1)	True (1)	True (1)	True (1)
		Confidence	Middle (0)	Middle (0)	High (1)	Middle (0)	High (1)	Middle (0)
	Q3 (N1)	Correctness	True (1)	False (0)	False (0)	False (0)	True (1)	False (0)
		Confidence	High (1)	Middle (0)	High (1)	Middle (0)	High (1)	Low (-1)
	Time (second)		450	529	424	586	279	391

Table 4. Average of correctness.

Correctness		Ave.	Ave.
Method1	Q1 (N3)	1.00	0.78
	Q2 (N2)	0.67	
	Q3 (N1)	0.67	
Method2	Q1 (N3)	0.83	0.61
	Q2 (N2)	0.50	
	Q3 (N1)	0.50	
Method3	Q1 (N3)	1.00	0.67
	Q2 (N2)	0.67	
	Q3 (N1)	0.33	

Table 5. Average of confidence.

Confidence		Ave.	Ave.
Method1	Q1 (N3)	1.00	0.33
	Q2 (N2)	0.17	
	Q3 (N1)	-0.17	
Method2	Q1 (N3)	0.83	0.28
	Q2 (N2)	0.17	
	Q3 (N1)	-0.17	
Method3	Q1 (N3)	1.00	0.56
	Q2 (N2)	0.33	
	Q3 (N1)	0.33	

Table 6. Average of correctness when confidence is middle or higher.

Corr. with Conf. (Middle or High)		Ave.	Ave.
Method1	Q1 (N3)	1.00	0.61
	Q2 (N2)	0.50	
	Q3 (N1)	0.33	
Method2	Q1 (N3)	0.83	0.50
	Q2 (N2)	0.33	
	Q3 (N1)	0.33	
Method3	Q1 (N3)	1.00	0.67
	Q2 (N2)	0.67	
	Q3 (N1)	0.33	

Table 7. Average of correctness when confidence is high.

Corr. with Conf. (High)		Ave.	Ave.
Method1	Q1 (N3)	1.00	0.44
	Q2 (N2)	0.17	
	Q3 (N1)	0.17	
Method2	Q1 (N3)	0.67	0.39
	Q2 (N2)	0.17	
	Q3 (N1)	0.33	
Method3	Q1 (N3)	1.00	0.56
	Q2 (N2)	0.33	
	Q3 (N1)	0.33	

We indicate the correct answer rates for responses with confidence levels of "Middle" or "High" in Table 6, and those for responses with a "High" confidence level in Table 7. In both cases, the proposed method (Method3) achieved the highest correct answer rates. These results suggest that the proposed method supported participants' understanding of the question text and helped them to answer with greater confidence.

We indicate the average and standard deviation of the response time for answering the questions for each method in Table 8. The average time for the proposed method (Method3) was 443.17 seconds, which was longer than that of Method1 (no support tools), but shorter than that of Method2 (manual text input). Moreover, the standard deviation for the proposed method was 98.20, the smallest among all methods. These results suggested that the proposed method contributed to reducing the influence of differences in Japanese comprehension skills and Japanese text input abilities on the response time.

On the other hand, three participants (D, E, and F) took more response time to answer using Method3 than Method2 as shown in Table 3. This result can be attributed to two main reasons. First, the proposed method (Method3) requires approximately 20 seconds to complete the translation process after taking a photo. As a result, participants had to wait for the plain Japanese text to be displayed before answering. Second, participants entered varying amounts of difficult Japanese text in Method2. Some participants input only specific difficult words rather than the entire sentence, which shortened the input time.

Finally, we indicate the participants' feedback on using the proposed application in Table 9. Overall, the feedback was generally positive; however, there were some suggestions, such as a request for a "adding *furigana* to *kanji*" and a "selecting the area of text in the image".

Conclusions

We developed a plain Japanese translation system that uses OCR to extract text from images and translates difficult Japanese language into more understandable Japanese. The system employed ChatGPT API provided by OpenAI and Vision AI API provided by Google. We developed the system as a web application to ensure compatibility with various devices.

We conducted an evaluation using reading comprehension tests from the JLPT. The proposed application helps participants better understand difficult text and answer with greater confidence. Furthermore, we indicated that the proposed application reduces the response time to answer compared to the existing system that require manual input.

These results indicate that the proposed application helps participants improve their comprehension of evaluation tests. We need to enhance the application's usability and conduct a long-term evaluation to assess its effectiveness in improving reading comprehension skills.

Table 8. Average and Standard of Time.

Time (second)	Ave.	Std.
Method1	347.33	133.67
Method2	455.83	157.28
Method3	443.17	98.20

Table 9. Results of the survey on participant's feedback.

Feedback on usage
The text was made very easy to understand, which was very helpful when answering the questions.
It didn't take much time, and I could include all the text, so I found it easy to use.
It was helpful. The sentences became somewhat simpler.
It made the text much easier to understand.
I think it became significantly easier to understand compared to the original text.
It was very convenient to be able to use the camera. I appreciated that it simplified difficult N1-level sentences.
Requested features
There were still some difficult *kanji* even in the simplified text, so I think it would be helpful if the readings (*furigana*) were also provided.
I think it would be good if users could select the area of the text to be captured.
It would be helpful if there were a function to select a range and automatically copy and paste the selected text into the translation input field.
Even after translation, there were still some difficult *kanji*, so I think it would be better to either add *furigana* or replace them with simpler words.

Acknowledgements

This work was supported by JSPS KAKENHI 25H00473, 25K16298, and 23K17511.

References

Jennifer Herrity (2025). "7 Simple Strategies to Improve Reading Comprehension". Retrieved from https://www.indeed.com/career-advice/career-development/reading-comprehension-improvement-strategies .

Bricks Co., Ltd. (2020). "多言語翻訳サービス利用における「やさしい日本語」の活用に関する調査研究報告書" in Japanese title. "A Research Report on the Use of Plain Japanese in Multilingual Translation Services" in English title.

Alfasado Inc. (2025). "伝えるウェブ" in Japanese title. "Information-Friendly Web" in English title. Retrieved from https://tsutaeru.cloud .