

Building a Multilingual Support System with Generative AI and RAG for Students

Zelan Shi* Yasuhiro Sudo* Kaunori Matsumoto*

Abstract

This paper addresses the challenges faced by international students with limited Japanese proficiency in understanding materials in Japanese during classes. We then describe in this paper the design and preliminary experimental results of a support system leveraging Generative AI and Retrieval-Augmented Generation (RAG). Our goal is not only to support international students in understanding the content of their classes, but also to simultaneously aim for improving their Japanese language proficiency. Thus, simple translation tools are insufficient to achieve this objective. The core feature of our system is its ability to adaptively enhance Japanese academic materials by adding translated terms in the students' native languages and further adding concise explanations. These supplementary aids are tailored based on each student's Japanese language proficiency, ensuring personalized and effective support. These functionalities are implemented by utilizing the capabilities of Generative AI, as well as those provided by RAG. This study serves as an example of how Generative AI can be applied in educational settings.

1. Introduction

The number of international students in Japan is increasing[1]. Many of these students often have insufficient Japanese language proficiency. At the authors' university, lectures are conducted in Japanese with few exceptions. Textbooks and other reference materials for lectures, and examinations are also conducted in Japanese. Then students with limited Japanese skills face difficulties in understanding the lectures. The university provides Japanese language support through its International Office. However, while the staff possess linguistic skills, their knowledge in specialized professional academic fields is often insufficient, making their support for such fields may become insufficient. It can be said that many universities across Japan are in similar situations, although to varying degrees.

Many students actively use automatic interpretation or translation tools, many of them now run in real time. Although these tools are helpful for understanding the contents in lectures, they are not designed to improve Japanese language proficiency. As a result, continued reliance on these tools possibly pose the issue of hindering the improvement of students' Japanese language

skills[2]. It is therefore necessary to develop and provide tools from the perspective of supporting international students.

AI technology has significantly advanced with the emergence of generative AI. Subhankar Maity and Aniket Deroy [3] provide surveys recent progress in fields such as automatic task generation, grading, and tutoring. Generative AI is already being utilized, while numerous challenges remain to be addressed. Attempts to utilize generative AI for language education have also been proposed [4][5], and these methods hold promising potential for language learning among international students. However, from the perspective of supporting international students, research and development of assistance technologies in specialized professional education areas remain limited, and are still at an early stage. Among the technologies surrounding generative AI, RAG has garnered attention, and its development in the education sector has commenced, and can compensate for the shortcomings of Large Language Models (LLMs), which often lack domain-specific information[6]. For instance, studies have proposed systems[7] that recommend the most suitable educational materials tailored to users' characteristics from a diverse range of resources. RAG has the potential to be utilized for supporting specialized education for international students; however, research in this area remains limited and is still in its early stages. This paper represents pioneering research aimed at advancing this purpose.

In this paper, we first conduct a rough evaluation of tools commonly used by students, such as translation and automatic summarization tools, and highlight their limitations. Next, we explain the proposed system and present a general evaluation of its results. To simplify the discussion, this paper focuses on online courses delivered through on-demand videos.

2. Evaluation of Traditional Tools

International students often use tools such as translation and automatic summarization[8][9][10] for on-demand learning materials. We present briefly here investigations of the tools through experiments. The experiments use materials from lectures on data science and AI for first-year students, as well as from course guidance sessions. As mentioned above, this paper focuses on online on-demand classes.

* Department of Information and Computer Sciences, Graduate School of Engineering, Kanagawa Institute of Technology

2.1 Reliability of Automatic Translations

We conducted an experiment simulating a situation where Chinese international students attend classes taught in Japanese. First, we utilized YouTube's automated subtitling feature to generate Chinese subtitles and subsequently assessed their accuracy. The subtitle generation process necessitated an evaluation of the system's ability to accurately transcribe the instructor's Japanese speech. Therefore, we transcribed the Japanese audio into text and evaluated its accuracy.

From each 40-minute lecture video in Japanese, 20 frames containing essential content were extracted and assessed by evaluators who are native Japanese speakers. The results are shown in Table 1. These videos were recorded under good conditions, with minimal noise and an overall favorable environment. Additionally, the instructor's Japanese pronunciation was clear, with no parts difficult for typical Japanese listeners to comprehend.

Next, the Chinese subtitles underwent scrutiny by international students whose native language is Chinese. Table 1 presents the outcomes of both experiments, with the table's values denoting error rates in speech-to-text recognition or translation. The accuracy of Chinese translation is inferior to that of Japanese recognition, potentially leading to adverse effects on comprehension.

Table 1: experimental result of error rates

No.	Subtitle length (JP)	Error (JP)	Error rate (JP)	Subtitle length (CN)	Error (CN)	Error rate (CN)
1	64	4	6.25%	94	15	15.96%
2	71	12	16.90%	114	22	19.30%
3	87	5	5.75%	166	19	11.45%
4	77	10	12.99%	138	26	18.84%
5	83	13	15.66%	123	24	19.51%
Average		11.94%		16.83%		

2.2 Reliability of Automatic Summarization

We investigated the practical application of generative AI tools for video summarization, specifically NoteGPT. The efficacy of the lecture video summaries was assessed through experimentation. First, the lecturer created a summary for each lecture based on the video content, and then the summaries were compared with the outputs of automatic summarization to examine whether any omissions of critical information or inclusion of extraneous details. The results are presented in Figure 1.

The horizontal axis in the figure represents the types of instructional videos analyzed. The vertical axes display two

evaluation metrics. The left-side bars in each group indicate the evaluation values based on the number of missing or omitted instructional points. A higher value suggests that the AI-generated summaries lack more of the essential content that the instructor intended to convey, whereas a lower value implies a better-quality summary.

The right-side bars represent evaluation scores based on the number of redundant or extraneous elements included in the summaries. Similarly, a lower value in this metric reflects better summarization performance. The line graph indicates the overall evaluation score, where lower values again correspond to higher-quality outputs.

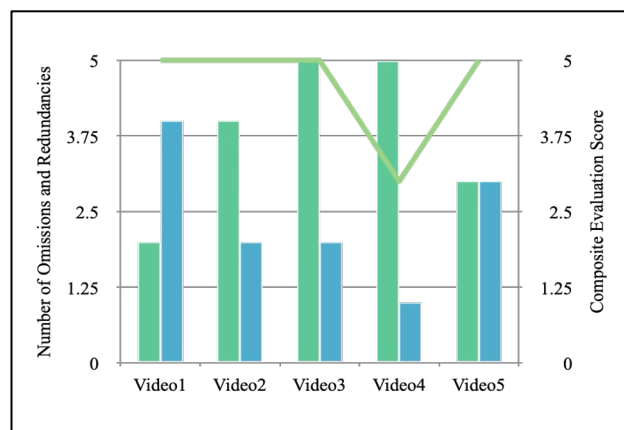


Figure 1: experimental results: comparison of human summary with automatic summary

Across all videos, the AI-generated summaries contain both missing and unnecessary components to varying degrees. However, Video 4 demonstrates notably better performance compared to the others. This is likely due to the nature of its content, it is a seminar guidance video that uses more everyday language and contains fewer technical terms, making it easier for the AI system to process and summarize effectively.

3. Multilingual Support System

3.1 System configuration diagram

This paper proposes a system named MSS-GEAR (Multilingual Support System with Generative AI and RAG), and its conceptual diagram is presented in Figure 1. Following this diagram, we explain the rough functions of its components.

The proposed system consists of two core components: MSS-PERSONA and MSS-RAG. MSS-PERSONA is responsible for constructing and maintaining individualized learner profiles. It continuously accumulates and updates various types of information, including language proficiency levels and domain-specific comprehension metrics. MSS-RAG, on the other hand, functions as an enhanced retrieval engine that supports the generation of personalized learning materials. It receives learner-specific data from MSS-PERSONA and leverages this information to identify and annotate content according to each student's proficiency.

Figure 2 presents a simplified output example. The upper section of the figure illustrates the MSS-PERSONA subsystem, which generates a personalized word database by assessing learners' comprehension of course-related terminology. This process ensures that the resulting vocabulary set corresponds to the learner's current proficiency level.

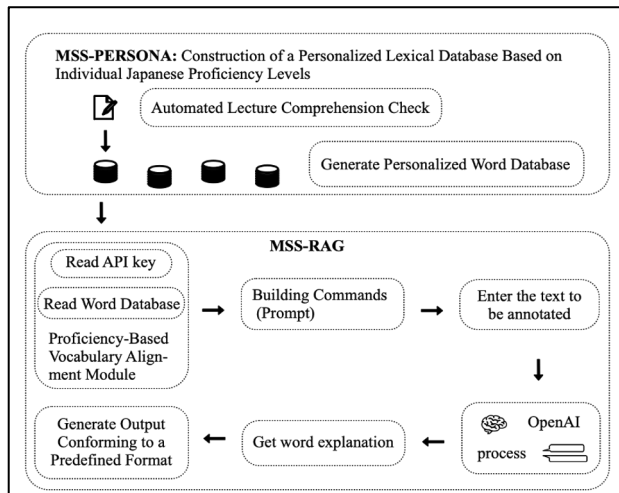


Figure 2: MSS-GEAR System Framework Diagram

The lower section represents the MSS-RAG engine, which references the learner personalized word database to identify unfamiliar or complex vocabulary from new content. It then generates appropriate explanatory annotations to enhance the learner's understanding.

3.2 Prototype Implementation

To investigate the feasibility of integrating RAG mechanisms into a personalized multilingual support system, we developed a prototype system that combines semantic retrieval with generative language modeling. The implementation leverages GPT-4 for text generation and a vector-based approach for knowledge retrieval, with the entire pipeline developed in Python. To enhance the retrieval efficiency of vocabulary entries, this system employs FAISS [11], an optimized library designed for fast and scalable nearest neighbor search in high-dimensional spaces. Traditional keyword-based search methods often fail to capture the semantic relationships between words, relying on exact matches rather than contextual similarities. FAISS addresses this limitation by enabling vector-based retrieval, allowing the system to find words that are semantically closest to a given input text rather than merely matching keywords.

The prototype architecture is composed of several key components. First, a user-defined vocabulary database MSS-PERSONA is ingested as a structured dataset, containing lexical entries in Japanese alongside corresponding pronunciations and Chinese explanations. This dataset is preprocessed and transformed into vector representations using a pretrained embedding model. The resulting embeddings are indexed via FAISS, enabling efficient similarity-based retrieval.

This retrieval-augmented framework allows the system to dynamically integrate retrieved vocabulary items tailored to

individual users, thereby aligning the generative output with the learner's linguistic proficiency. The prototype demonstrates the potential of combining efficient vector-based semantic retrieval with large-scale language models to construct interpretable, adaptable, and learner-aware multilingual support systems.

3.3 Optimizing RAG for Language Learning

Technically, MSS-GEAR adopts a RAG framework. This enables generation grounded in user-specific information and facilitates dynamic adaptation to each student's needs. Unlike conventional RAG systems primarily designed for general-purpose knowledge retrieval, MSS-GEAR is specialized for the language learning domain. Specifically, it incorporates the following distinguishing features:

1. Domain-specific focus: The system is optimized for vocabulary comprehension and contextual learning, rather than open-domain information retrieval.

2. Structured lexical knowledge base: It utilizes user-provided Excel vocabulary lists as its core knowledge source, allowing the integration of structured, pedagogically relevant content.

3. Semantic retrieval: By leveraging FAISS for high-dimensional vector search, the system performs semantic-level matching rather than relying on simple keyword overlap. The retrieved vocabulary is then passed to GPT-4 to generate simplified, context-aware explanations, offering greater clarity for learners.

4. Dynamic knowledge base updating: Unlike static RAG architectures, MSS-GEAR supports on-the-fly vocabulary expansion. Users can upload new Excel-based vocabulary datasets, which are automatically embedded and indexed to update the FAISS retrieval system without retraining the entire model.

3.4 System Preliminary Testing

To evaluate the feasibility and effectiveness of MSS-GEAR, we conducted an initial test using specialized course materials. The system was provided with text summaries from on-demand courses in the field of information engineering, and the retrieved vocabulary explanations were compared against the proficiency level of the test subjects. Preliminary results indicate that MSS-GEAR successfully identifies relevant vocabulary items and provides accurate, context-aware explanations, aligning moderately well with the users' expected comprehension level. The subsequent examples illustrate specific instances of input and output.

Input : 今回の授業は、オペレーティングシステムのセキュリティに関する講義です。認証、認可、監査といったセキュリティの基本概念と、コンピュータ資源を様々な脅威から保護するための対策について解説します。また、パスワード認証、生体認証、アクセス制御リスト、暗号化など、具体的な技術についても触れています。

セキュリティの重要性：コンピュータの資源を脅威から守ることが必要である。

認証のメカニズム：ユーザーが誰であることを確認するための手段として ID とパスワードが使われる。

認可の役割：ユーザーに対する権限を設定し、アクセス制御を行う。

アクセス制御リスト：誰が何にアクセスできるかを記述する方法。

暗号化技術：情報を特定の人だけが理解できる形式に変換し、保護するための手段。

Output：以下の難しい単語を見つけました：

- 単語：セキュリティ
- 発音：
- 注釈：指保护计算机系统、网络和数据免受损害或未授权访问的措施
- 単語：オペレーティングシステム
- 発音：
- 注釈：操作系统，计算机的核心软件，用于管理和控制计算机硬件和软件资源，提供各种服务和功能。
- 単語：認証
- 発音：にんしょう
- 注釈：认证，确认用户的身份或数据的完整性的过程。
- 単語：認可
- 発音：にんか
- 注釈：授权，赋予用户或程序访问资源的权利的过程。
- 単語：監査
- 発音：かんさ
- 注釈：审计，对系统的操作和活动进行记录和检查的过程。
- 単語：アクセス制御リスト
- 発音：アクセスせいぎょリスト
- 注釈：访问控制列表，描述谁可以访问什么的方式。
- 単語：暗号化
- 発音：あんごうか
- 注釈：加密，将信息转换为只有特定人才能理解的格式，以保护信息的手段。

Since the input content involves a lecture transcript in Japanese and the output includes vocabulary annotations in Chinese, a brief explanation is provided here for clarity:

Input: The input consists of a summary of a Japanese on-demand lecture on operating system security.

Output: The output is a list of domain-specific or complex Japanese terms identified from the lecture content. For each term, the system provides: the term itself (in Japanese), its pronunciation (using Japanese kana), and a definition or annotation in Chinese, aimed at supporting native Chinese speakers in understanding the Japanese terminology.

This input-output format supports Chinese-speaking learners by linking Japanese lecture content with annotated vocabulary. Initial review confirms the extracted terms reflect key technical concepts.

However, further validation is required to assess the long-term effectiveness of the system in facilitating language acquisition. Future research will focus on comprehensive evaluation metrics. Additionally, we aim to enhance MSS-GEAR's functionality by integrating real-time vocabulary expansion. Users will be able to input new words, which will be dynamically embedded and indexed without requiring full FAISS retraining. Another planned improvement is adaptive vocabulary recommendations, prioritizing words that users have previously struggled with. These

optimizations will further improve personalization, ensuring MSS-GEAR evolves alongside learners and continuously enhances their multilingual proficiency.

4. Conclusions

International students often face disadvantages due to their limited language proficiency, which can hinder their learning and research in specialized academic fields. This study addresses these challenges through the application of RAG mechanisms in language learning. While existing tools such as machine translation and automatic summarization assist comprehension, they do not actively support language acquisition, making them less effective for personalized learning.

To that end, we developed MSS-GEAR, an experimental system that provides supplementary explanations in the student's native language, calibrated to their current proficiency level. As learners improve, support is dynamically adjusted, ensuring a smooth transition toward independent understanding. Beyond adaptation, MSS-GEAR is designed for continuous enhancement in the future, including:

1. Dynamic vocabulary expansion: Users can input new words, which are instantly embedded and indexed within FAISS, eliminating the need for full retraining.

2. Adaptive learning recommendations: The system can prioritize vocabulary that users have previously struggled with, refining retrieval based on individual progress.

These advancements ensure greater personalization and efficiency, positioning MSS-GEAR as an effective AI tool for multilingual education. Future work will improve retrieval, knowledge expansion, and contextual generation. As an initial report of our study, this paper outlined the system's design and potential to support international students in academic environments.

Reference

- [1] Ministry of Education, Culture, Sports, Science and Technology, "Survey on the Enrollment, Status of International Students and the Number of Japanese Students Studying Abroad," May, (2024).
- [2] P. Novikov and A. Kiseleva, "OVER-RELIANCE ON TECHNOLOGY IN FOREIGN LANGUAGE LEARNING: CASE STUDY OF LLB UNDERGRADUATES," presented at the 18th International Technology, Education and Development Conference, No. 2962–2966(2024).
- [3] S. Maity and A. Dero, "Generative AI and Its Impact on Personalized Intelligent Tutoring Systems," (2024).
- [4] E. Creely, "Exploring the Role of Generative AI in Enhancing Language Learning: Opportunities and Challenges," IJCE, Vol. 1, No. 3, 158–167, Apr. (2024).
- [5] L. Law, et al. "Application of generative artificial intelligence (GenAI) in language teaching and learning: A scoping literature review," Computers and Education Open, Vol. 6, No. 100174, Jun. (2024).
- [6] P. Lewis, "Retrieval-augmented generation for knowledge-intensive nlp tasks." Advances in neural information processing systems 33, No.9459-9474 (2020).
- [7] J. Rao and J. Lin, "RAMO: Retrieval-Augmented Generation for Enhancing MOOCs Recommendations," (2024)
- [8] DeepL Translate. <https://www.deepl.com/en/translator>
- [9] Summarize - ChatGPT. <https://calmai.net/>
- [10] YouTube.
- [11] M. Douze, et al. "The faiss library." arXiv preprint arXiv:2401.08281 (2024).