Jiang, B. et al. (Eds.) (2025). Proceedings of the 33rd International Conference on Computers in Education. Asia-Pacific Society for Computers in Education

Web Application to Help Learn English Using Medical Bilingual Corpus - Updates to Improve Learners' Writing Skills -

Riku YASUDA[a*], Yoshinori MIYAZAKI[b], Motoko ASANO[c] & Miho FUJIEDA[c]
[a]*Graduate School of Informatics, Shizuoka University, Japan*
[b]*College of Informatics, Shizuoka University, Japan*
[c]*Faculty of Medicine, Osaka Medical and Pharmaceutical University, Japan*
*yasuda.riku.20@shizuoka.ac.jp

Abstract: In the field of medical education, a deep understanding of domain-specific expressions is essential for reading and writing academic papers in English. To address this challenge, we are developing an English learning support web application with a bilingual corpus of medical texts. This paper presents two enhancements to the system designed to support learners' academic writing skills. First, we introduced a writing support function that suggests similar sentences based on user-generated English input. Second, we enhanced the accuracy of grammar-based categorization by integrating a BERT-based grammatical detection algorithm that leverages contextual understanding. These improvements aim to provide a more effective and domain-sensitive writing support experience for users engaging with medical English materials.

Keywords: Medical English, bilingual corpus, web application, reading and writing aid

1. Introduction

In the field of medical education, the ability to effectively utilize English is particularly important for acquiring up-to-date medical knowledge, such as through clinical conferences or reading recent academic papers in English. This reflects the domain of English for Medical Purposes (EMP), a specialized branch within English for Specific Purposes (ESP). Based on this background, our research group developed MEESUS (Medical English Education SUpport System) to support EMP learning among medical English learners, using a bilingual corpus of medical research papers (titles and abstracts) (Asano et al., 2022). The development process involved not only the provision of data by English instructors from a medical university but also incorporated ideas proposed by faculty members in the fields of medicine, engineering, system development, English education, and practicing physicians. As a result, the project was shaped by a multidisciplinary task force that integrates diverse expertise across academic domains. The current version of the system is designed with a focus on supporting reading and writing in English; however, this study aims to expand its functionality to include writing support, with the goal of providing more comprehensive assistance for EMP learning.

2. Literature Review

Related studies to this research include previous work on concordancers and section-based language support, which are key features of the proposed system.

AntConc, developed by L. Anthony, is a free text analysis tool designed for researchers in linguistics and literary studies, and it includes functions such as concordance searches for words and phrases (Anthony, n.d.). JECPRESE is a software tool that places the search term in the center of a keyword-in-context (KWIC) display when retrieving expressions from a corpus. It is designed for transcribed oral presentation data and allows users to search both English and Japanese sentences depending on the speaker's communicative intent (Kunioshi

et al., 2012). AWSuM is a writing support system that utilizes rhetorical “moves” (i.e., communicative intentions) to suggest frequently used word bundles by section of a research article. While it also includes concordancer functions and provides search capabilities, it does not support bilingual (English-Japanese) output (Mizumoto et al., 2017).

3. Medical English Learning System

Figure 1 shows the data structure of the bilingual corpus, which consists of the titles and abstracts of medical research papers. The corpus includes 3,920 such documents (title and abstract pair) and a total of 49,884 English sentences, all aligned with Japanese translations (NEJM Group, n.d.; Nankodo Co., Ltd., n.d. https://www.nejm.jp/). All corpus data has been reviewed and verified by medical faculty experts to ensure accuracy and reliability. Unlike generative AI tools, this expert-validated dataset provides trusted, domain-specific expressions without such risks. Each English document is segmented into six sections: title, background, methods, results, conclusion, and conclusion (research information). By utilizing section-specific information linked to each sentence, learners can study expressions within distinct rhetorical contexts. In Figure 1, for example, the sentence with ID = 1900236 appears as the fourth sentence of the fourth document and belongs to the “conclusion” section.

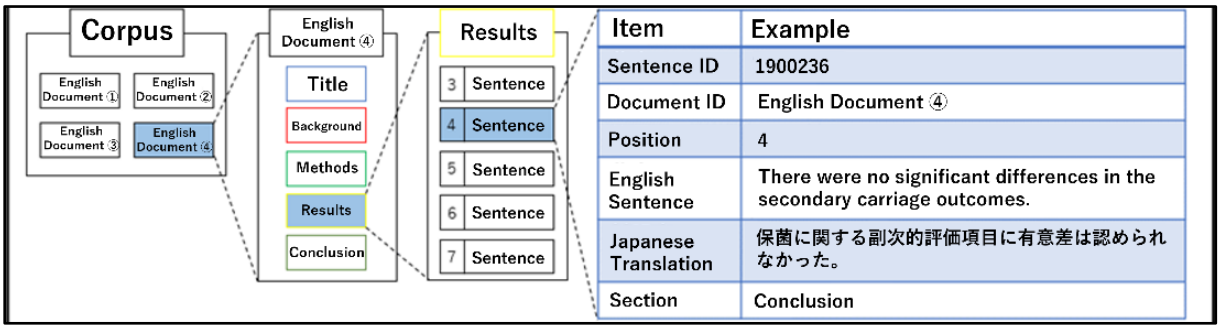

Figure 1. Data Structure and Examples

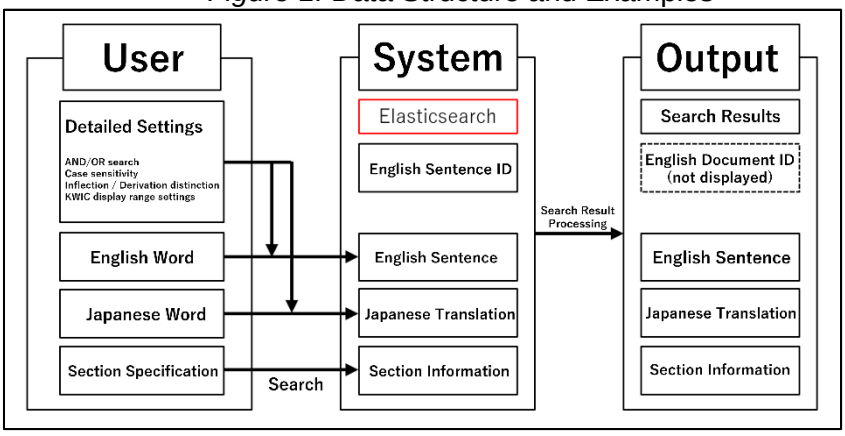

Figure 2. System Architecture

An overview of the system is presented in Figure 2. The search engine is built using Elasticsearch (Elasticsearch, n.d.). User input is converted into an Elasticsearch query, which is then used to retrieve relevant data from the corpus structured (Figure 1). The system offers various search methods, including single search, combined search, and cross-lingual search that allows users to input a mix of Japanese and English in each search field. Search options include specifying the document section (e.g., background or methods), setting the number of words for KWIC display, and adjusting parameters for variations in expressions (Figure 3). The system also implements a suggestion feature. This feature uses Levenshtein distance to recommend English words that are similar to the user's input. For example, if a user mistakenly enters “hospitul,” the system will display the interface shown in Figure 4. By clicking the “Did you mean” prompt, the system suggests possible correct spellings. This function enables

users to receive appropriate word suggestions even when typographical errors are present in the search terms. Search results are initially displayed in KWIC format, with each hit shown at the sentence level. Users can then access additional features, such as viewing the full English sentence with its Japanese translation, or displaying the entire document to which the hit sentence belongs. An example of these functions is shown in Figure 5.

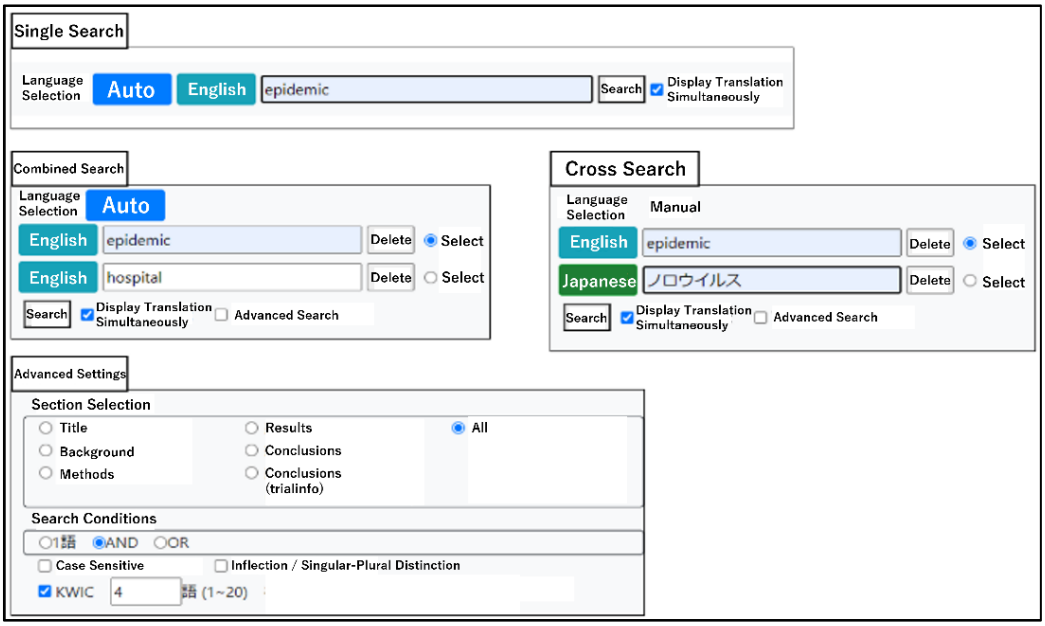

Figure 3. Various Search Methods

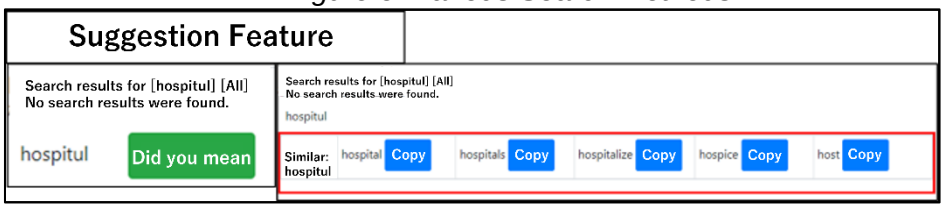

Figure 4. Suggestion Feature

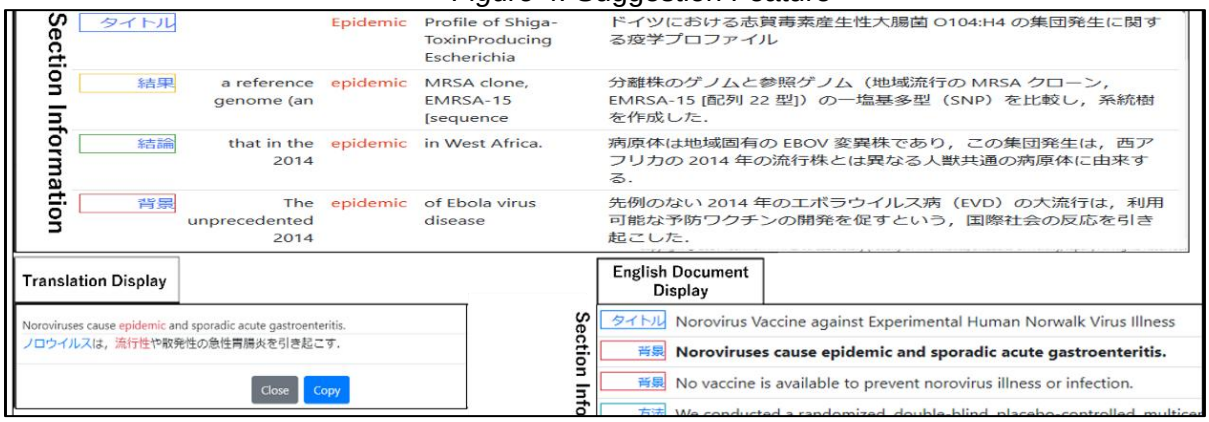

Figure 5. Search Results

4. Additional Features

4.1 Similar Sentence Search Function

In order to support academic writing more effectively, the system includes a function that presents similar sentences based on either search results or user-inputted English sentences. This feature was developed in response to user feedback.

The authors incorporated the EWSS system (Miyazaki et al., 2014) to add a feature that searches for sentences similar to either user-inputted text or search result sentences. By clicking the "Similar Sentence Search" button or selecting a sample sentence from the search

results, a modal window appears. The Similar Sentence Search function allows users to toggle between general English and medical-specific English, and the input sentence is automatically populated in the search field, making it easy to conduct a similarity-based search. Users can also view their search history for repeating a previously searched sentence.

In terms of system operation, the input sentence is analyzed using the morphological analyzer TreeTagger to extract the base forms and part-of-speech information of each word. Synonyms are also utilized to generate n-grams—sequences of n consecutive words—from the sentence. Based on this information, the system calculates sentence similarity in a vector space, allowing it to retrieve example sentences that closely match the user's intended meaning. Figure 6 shows the search results screen. The input sentence is displayed at the top, with content words (nouns, verbs, adjectives, and adverbs) highlighted. Synonyms for each content word in the input sentence are also shown, based on WordNet. By clicking the checkbox labeled "Add," users can save selected sentences to "My List."

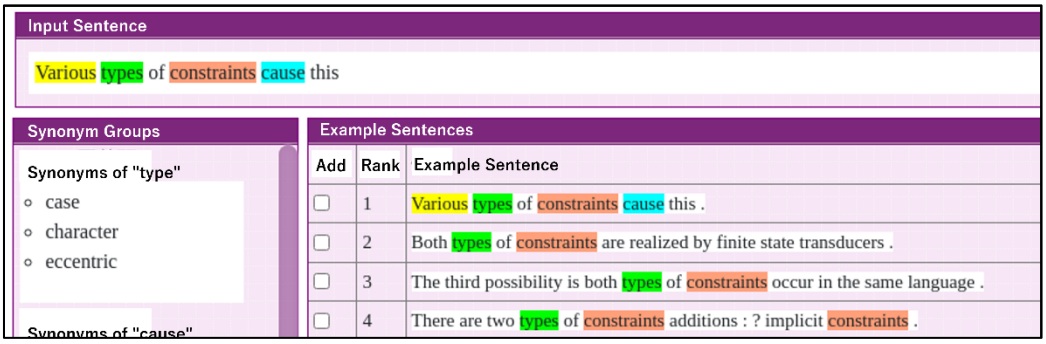

Figure 6. Search Results from the Similar Sentence Search Function

For the current system, we adapted the dataset used in MEESUS to specifically support learning in the domain of EMP. Simply adapting the dataset was not sufficient to present users with the medical-specific English sentences they were seeking. Therefore, we applied customized weighting to selected medical terms. To improve the scoring mechanism, we utilized a predefined list of medical terms extracted from domain-specific resources. When a sentence in the dataset contains a medical term that also appears in the user's input, the similarity score for that sentence is specially increased. By utilizing this feature, users are expected to enhance their writing skills in specialized medical contexts.

4.2 Grammar-Based Display Enhancement

Interrogative Form	0%	To-Infinitive	9%	Modal Verb	3%
Passive Voice	6%	Bare Infinitive	100%	Relative Pronoun	45%
Subjunctive Mood	100%	To-Infinitive + Bare Infinitive	11%	Existential "There" Construction	17%
Progressive Aspect	22%	Comparative	0%		
Perfect Aspect	0%	Gerund	83%		

Figure 7. Error Rates in Grammatical Category Detection by Previous Studies

To further enhance the usefulness of the Similar Sentence Search function, we focused on improving the accuracy of grammar-based categorization. In follow-up studies to Miyazaki et al. (2014), grammatical item detection methods using regular expressions in Python and BACT were employed. However, these approaches frequently misidentified complex grammatical constructs such as the subjunctive mood, bare infinitives, and gerunds (see Figure 7). These misclassifications were primarily due to the reliance on simple pattern matching, which lacked contextual and semantic understanding. To significantly improve the accuracy of grammatical item detection, we combined advanced Python libraries such as spaCy and NLTK with BERT to analyze English sentences. Specifically, spaCy's dependency parsing functionality was used to clarify syntactic relationships within sentences, while the machine learning model enabled context-aware classification of grammatical elements. In the case of gerund detection in particular, the approach goes beyond simple sentence pattern identification by taking into

account the overall meaning of the sentence and its relationship with verbs. This allows for more nuanced distinctions that were difficult to achieve with previous methods. Furthermore, to enhance detection accuracy, we adopted a supervised learning approach and trained the model using a large-scale corpus. This approach is expected to achieve high detection accuracy even for grammatical items such as the subjunctive mood and bare infinitives, which have previously yielded limited success with traditional methods.

4.3 Comparative Evaluation of the Grammar-Based Display Accuracy

Figure 8 presents the results of a comparison of grammatical item detection accuracy between the proposed method and previous studies. The blue bars represent the results from prior research, while the orange bars show the performance of the proposed method. For the comparison experiment, 200 sentences were randomly selected from a technical literature corpus, following the same procedure used in previous studies. To ensure a fair evaluation, each sentence was manually annotated with grammatical items, and detection accuracy was assessed based on the agreement between these annotations and the model outputs.

	Previous Studies	Proposed Method			Previous Studies	Proposed Method
Interrogative Form	100%	100%		To-Infinitive + Bare Infinitive	89%	96%
Passive Voice	94%	100%		Comparative	100%	100%
Subjunctive Mood	0%	100%		Gerund	17%	94%
Progressive Aspect	78%	100%		Modal Verb	97%	100%
Perfect Aspect	100%	100%		Relative Pronoun	55%	91%
To-Infinitive	91%	95%		Existential “There” Construction	83%	100%
Bare Infinitive	0%	100%				

Figure 8. Comparative Accuracy of Grammatical Category Detection

The results confirm that the proposed method significantly improved detection accuracy for grammatical items that had been challenging in previous studies, particularly the subjunctive mood, bare infinitives, and gerunds. This substantial enhancement is primarily attributed to BERT’s ability to understand context, allowing the model to consider relationships between surrounding words for more accurate classification. Overall, detection accuracy also improved across other grammatical categories. The average detection accuracy increased from 77% in previous studies to 96% with the proposed method. Although this experiment was conducted using technical literature, BERT's pretraining on a wide variety of text genres suggests a certain level of generalizability to the medical domain. However, to achieve even higher accuracy in medical contexts, fine-tuning with additional domain-specific data will be necessary, and this remains an area for future work.

5. Experimental Evaluation and Analysis

Undergraduate and graduate students from the Faculty of Informatics at a certain university were invited to use the system, and we collected both usage logs and questionnaire responses. The experiment involved 10 participants, and based on the results, we analyzed the effectiveness of the newly added features. Each participant completed four tasks using the system, followed by a post-experiment usability questionnaire. The task details are:
Task 1. Writing about the topic of a research paper, Task 2. Investigating a research paper, Task 3. Paraphrasing task, and Task 4. Error correction task.

The usage statistics for the Similar Sentence Search options follows, General English: 32 (times), Medical-Specific English: 5 (times), and Number of My List Usages: 19 (times).

It was found that the general English search option was used more frequently. This is likely because the example sentences for the paraphrasing task were written in general English. The ratio of "My List" usage to the total number of searches was approximately 51%, indicating that the feature was actively used to support task completion. According to the questionnaire responses on feature usage, the Similar Sentence Search function was used

by all participants in both Task 3 and 4. This suggests a strong demand for the feature in the tasks. The responses to the questionnaire item “Do you feel that using the system reduces the difficulty of writing academic texts?” showed that 30% of participants strongly agreed, 30% agreed, and 30% slightly agreed. Only 10% slightly disagreed, and no one strongly disagreed. Tables 3 and 4 present examples of feedback regarding the usability of this feature. 90% gave positive feedback, and the questionnaire responses indicated a 100% usage rate, suggesting that the system's support was highly valued. Below are the concrete comments: “I found it useful for rewriting English sentences,” “It is helpful when I want to express something in a different way,” “I would appreciate it if the system also displayed similar sentences based on grammatical rules, in addition to similar vocabulary.” The feature was particularly well-received for tasks such as paraphrasing, where it effectively assisted in developing writing skills. However, some users noted that the current version merely lists similar sentences, and suggested the system could categorize the results by grammatical rules.

6. Concluding Remark

In this study, we conducted an experiment using a developing medical English learning support system. Based on the feedback obtained, we implemented enhancements focused on supporting academic writing. The Similar Sentence Search function was designed to support writing skills by presenting similar sentences based on either search result sentences or user-inputted text. With some problems detected, we adapted a medical literature dataset and applied customized weighting to medical terms, to provide more domain-specific sentence suggestions. In addition, we aimed to further enhance writing assistance by improving the accuracy of the grammar-based categorization function. By combining BERT’s contextual understanding with syntactic parsing using spaCy, the overall average detection accuracy increased from 77% to 96%. The results of the pilot experiment using the enhanced system showed that the Similar Sentence Search function had a 100% usage rate, and 90% of participants rated it positively in the questionnaire, suggesting that the system successfully offers specialized writing support tailored to users’ needs. As for future directions, we plan to conduct a field-based validation study in actual educational settings to objectively evaluate the effectiveness of the newly implemented features for qualitative and quantitative feedback. This will allow us to assess the effectiveness, usability, and practical value of the system as a learning support tool. The findings will help identify areas for further improvement and reveal new user needs, which will inform the development of the next version of the system.

Acknowledgements

This work was supported by JSPS KAKENHI Grant Number 23K02800.

References

Anthony, L. (n.d.). AntConc [Computer software]. Retrieved August 13, 2025, from https://www.laurenceanthony.net/software/antconc/

Asano, M., Nakano, M., Miyazaki, Y., & Fujieda, M. (2022). Introducing a bilingual corpus database system of medical abstracts for exploring academic connotations of words: A case study of first-year medical students. *Journal of Medical English Education*, *21*(1), 18–26.

Kunioshi, N., Noguchi, J., Hayashi, H., & Watanabe, Y. (2012). An online support site for preparation of oral presentations in science and engineering. *European Journal of Engineering Education*, *37*(6), 600–608. https://doi.org/10.1080/03043797.2012.738171

Miyazaki, Y., Tanaka, S., & Koyama, Y. (2014). Development of a corpus-based web application to support writing technical documents in English. In *Proceedings of E-Learn 2014* (Vol. 2014, No. 1, pp. 1371–1380).

Mizumoto, A., Hamatani, S., & Imao, Y. (2017). Analyzing applied linguistics research articles by integrating moves and lexical bundles: Toward the development of an academic writing support tool. *Journal of English Corpus Studies*, *23*, 21–32. [in Japanese]