IIAI Open Conference Publication Series
IIAI Letters on Informatics and Interdisciplinary Research
Vol. 003, LIIR078
DOI: https://doi.org/10.52731/liir.v003.078

Error-Driven Learning: Development of Vocabulary Learning Support System that Utilizes Suggestibility of Error by Image Generation

Kazuki Sugita *, Wen Gu *, Shinobu Hasegawa *

Abstract

Vocabulary learning has long been taught as a basis for L2 English language learning. The use of visual clues is a widely recognized method in vocabulary learning. Several studies have applied automatic image caption generation, invented with computer vision and natural language processing, to English vocabulary learning in recent years. However, these vocabulary learning systems mainly use correct answers and their corresponding images. Another study focuses on the effectiveness of making errors, utilizing simulation to visualize the learner's error, and its suggestibility contributes to the learning effectiveness of the physics task. We focus on the effectiveness of error with image generation implying suggestibility and hypothesize that errors encourage the students to memorize the vocabulary effectively. We propose a vocabulary learning support system with image generation corresponding to learner's error for L2 English language learners. To accomplish the development, we defined three tasks and researched its possibility.

*Keywords: v*ocabulary acquisition, image generation, question generation, error-driven learning

1 Introduction

With the development of globalization, there is a need to use a common language to communi-cate with each other. English is the world's most spoken language in global communication [1]. The importance of learning vocabulary is well known in second language learning [2]. To ac-quire vocabulary effectively, it should be learned through context. Nagy insists on its importance, and they also mentioned it is more difficult learning through the context for the L2 learner [3]. The learner has to learn it through tons of reading [3]. There is some research proposing the system utilizing image captioning, it generates captions from images. They regard the caption and images as the context of learning vocabulary [4, 5]. In vocabulary learning in English by Japanese, errors are perceived to occur even among learners at a high proficiency level [6]. Laufer insisted that word similarity is because of form and sound making errors, and they found it difficult to compensate for this due to their EFL teaching experience [7]. Japanese learners tend to figure out their meanings by looking at the words because Japanese words are ideograms. Another hand, English words are phonograms the meaning of the word cannot be understood by

* Japan Advanced Institute of Science and Technology, Ishikawa, Japan

looking at the spell [8]. The reason why spelling errors occur regardless of proficiency level has been attributed to the phonological awareness of native Japanese speakers [9]. Krishan found that repetitive errors cause fossilization in his experiment with the Japanese [10]. Fossilization causes the mistakes to become part of the natural language, making it difficult for L2 learners to remedy them [10]. As a result, we consider the absence of suggestions for errors as a serious problem for L2 learners and propose a method to eliminate the same errors by impressing sug-gestions for errors to overcome the tendency of the error caused by the L2 learner. We hypothe-size that the images and the error suggestibility are effective in learning vocabulary. We aim to develop a learning support system that enables learning by generating images corresponding to incorrect answers suggesting errors to the learner. Our method is novel in that it utilizes image generation technology for vocabulary learning to generate images containing suggestive ele-ments that can determine the difference between correct and incorrect answers. In the next sec-tion, we present previous research on vocabulary learning using visual information and discuss error-based learning support systems. The third section introduces our proposed system over-view and methodology regarding three tasks. In the fourth section, we mentioned the discussion of our research. Finally, we conclude this article.

2 Literature Review

2.1 Effectiveness of Image for Vocabulary Acquisition

Hasnine et al. mentioned it is difficult to describe the learning context for each word. They use Show and Tell model which generates such context automatically from the image [4, 11]. Regina et al. compare the effectiveness of learning vocabulary using the 2D image and S3D image. They insist that much information does not need an efficient way of memorizing [12]. This fact indi-cates that depending on the quality of the information, the learning effect will be reduced. Wammes et al. found that describing vocabulary was effective for vocabulary learning, even if its accuracy was low [13]. The limitation of them is utilizing correct answers and corresponding images. However, overcoming the error caused by L2 learners requires learning similar spells. To accomplish this, we develop the system to automatically create the context for learning by utilizing the existing data set and showing a similar word.

2.2 Effectiveness of Error Suggestibility

Horiguchi et al. proposed Error-Based Simulation focusing on the implication of error and its effectiveness for learning for a physics task. In their research, the counterexample must contain sufficient information which leads a learner to correct understanding. It states that traditional models have difficulty in showing learners why they are wrong. It states that it is important for teaching to link the understanding of principles and discusses the feasibility of Error-Based Simulation [14].

3 System Design

3.1 Overview

This study aims to construct a system that automatically generates images corresponding to the learner's errors and shows these images to the learner to encourage reflection. Assuming that the reason for second language speakers' errors is spelling mistakes, we create an environment in which learners are more likely to cause errors. We hypothesize that the difference between the image of an error and the image of a correct answer will be retained as an impressive memory. To utilize such memory from the error, we define the learner's model as automatically defining the level of the learner through the process from their learning log. To build those systems, we de-fined three tasks. The first is a Question Generation task, the second is an Answering Support task, and the third is an Image Generation task. Referring the Figure 1, First, we estimate the initial learner's level, the target word selected corresponds to the level and part-of-speech. Next, corresponding to the target word, the question is generated and shown to the web browser by the Question Generation task. The candidate's words are also shown at the same time. It is selected by the Answering Support task. Then, the learner selects the word referring to the candidate word. If the answer is wrong, the image is changed to correspond to the wrong answer by the Image Generation task. And the image data generated by the module and the sentence including the wrong answer is stored as the learning log. In the end, the individual learning logs are re-ferred to generate the next question.

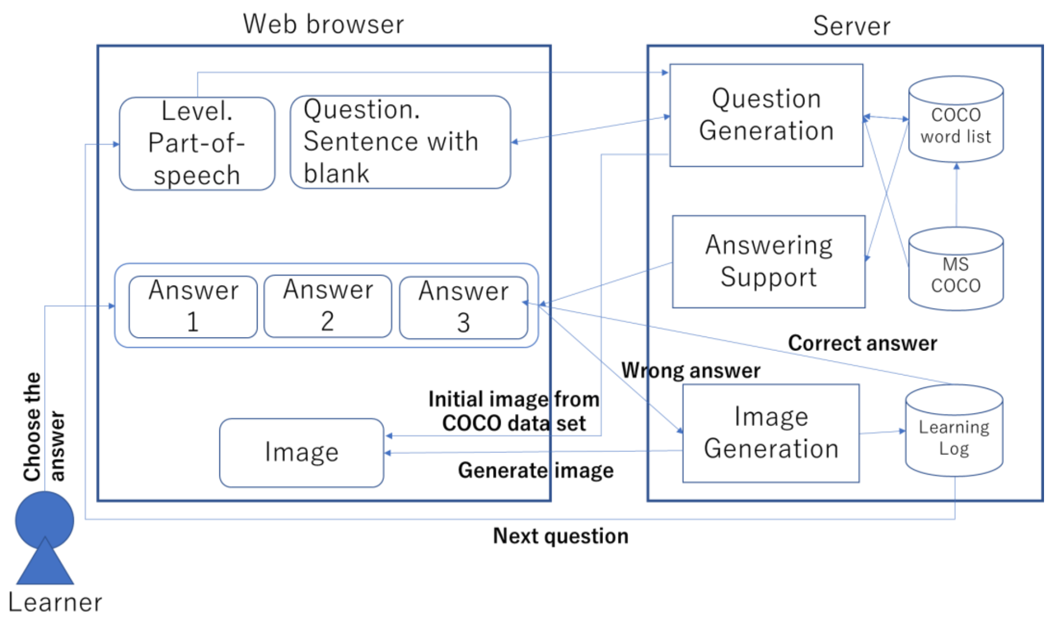

Figure 1: The image of the learning system

3.2 Question Generation Task

The question generation task automatically generates questions for the learner to study. It is generated based on the learner's proficiency level and consists of English sentences containing blanks. To accomplish the task, we define the step below.

3.2.1 Methodology of making the word list COCO-word list

This research utilizes MS-COCO [15], a large dataset of images with captions. We defined the COCO word list extracted from the 21,463 captions given to the 5,000 images that are part of MS COCO. The list is divided into part-of-speech and four levels of difficulty. For the extracted COCO captions, the natural language processing library spaCy [16] is used to split them into words, convert them into their original forms, and convert them from uppercase to lowercase. Then we refer to the Vocabulary List by difficulty, CEFR-J word list [17]. CEFR-J word list is based on the English textbooks used at primary and secondary schools in nearby Asian countries [18]. The COCO word list was defined at four levels of difficulty, from A1 to B2, and classified into eight part-of-speech. Table 1 shows the COCO word list that is classified by difficulty level A1-B2, part-of-speech, and the number of COCO word lists indicates the number of words in-cluded in the list. The number of CEFR-J indicates the number of words included in the CEFR-J word list. Figure 2 shows the coverage of the COCO word list to the CEFR-J word list.

Table 1: The number of the word in the COCO word list and CEFR-J word list

Level	Part-of-Speech	Number of COCO Word List	Number of CEFR-J		Level	Part-of-Speech	Number of COCO Word List	Number of CEFR-J
A1	adjective	123	148		B1	adjective	175	512
	adverb	64	75			adverb	38	157
	conjunction	10	10			conjunction	5	13
	determiner	29	31			determiner	3	4
	noun	504	631			noun	628	1267
	preposition	30	30			preposition	8	16
	pronoun	37	47			pronoun	4	8
	verb	114	134			verb	240	464
A2	adjective	128	243		B2	adjective	105	591
	adverb	69	122			adverb	28	197
	conjunction	7	8			conjunction	2	6
	determiner	7	10			determiner	0	1
	noun	528	773			noun	493	1421
	preposition	16	19			preposition	4	11
	pronoun	17	24			pronoun	1	4
	verb	127	204			verb	193	547

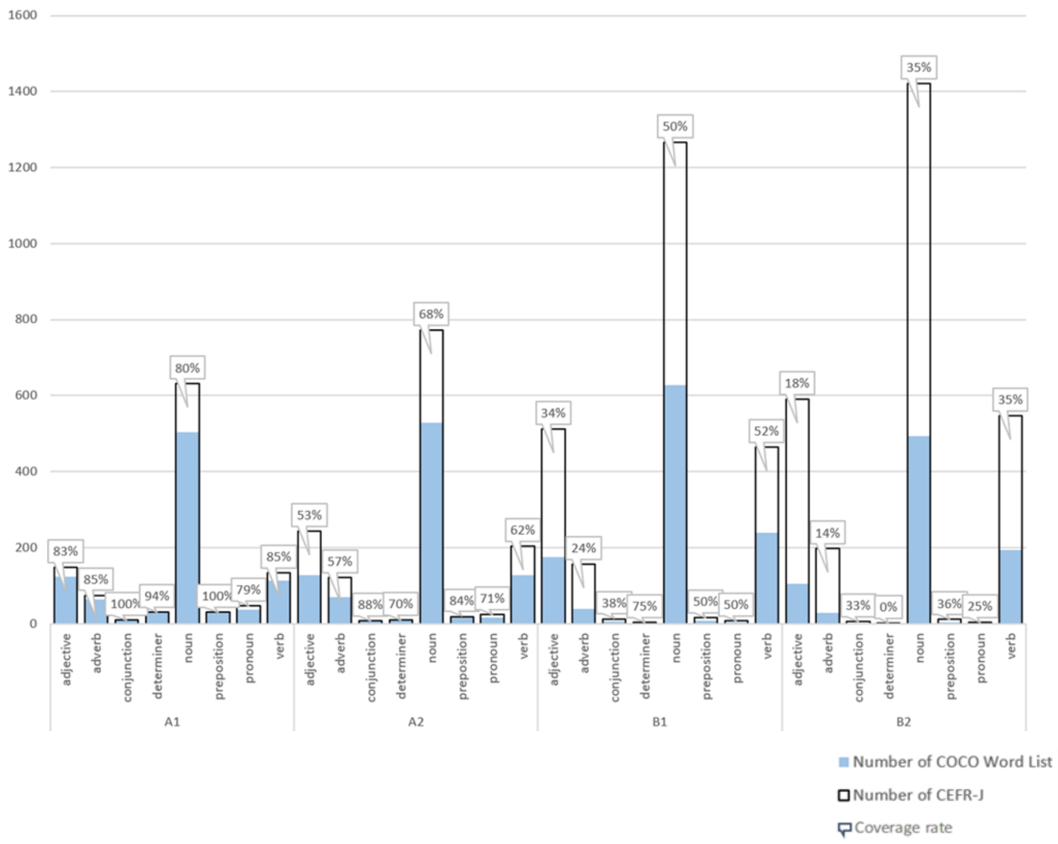

Figure 2: Coverage rate of COCO word list

The figure indicates lower level A1 is almost covered. The data set seems to cover the important words that must be learned in elementary education. This suggests that the dataset is composed of words that are commonly used in daily life. This is probably because MS COCO is composed of sentences generated from photographs of everyday scenes. The figure shows indicate the rate is decreasing as the difficulty level increases. However, our proposed list COCO word list is created by one of MS COCO. So, there is a possibility to increase the coverage rate if we apply all of the captions from MS COCO.

3.2.2 Methodology of Generate a question with a blank

The target word is extracted from the COCO word list defined earlier and corresponds to the learner's level and part-of-speech, and the task is to extract the target word from the sentence from MS COCO and replace it with the blank. The task is also like the previously mentioned COCO word list creation order. For the captions extracted from MS COCO, the natural language processing library spaCy is used to split them into words, convert them into their original forms, convert them from uppercase to lowercase, and label part-of-speech information. Then, if the target word matches the word processed in the previous step, replace the word with a blank.

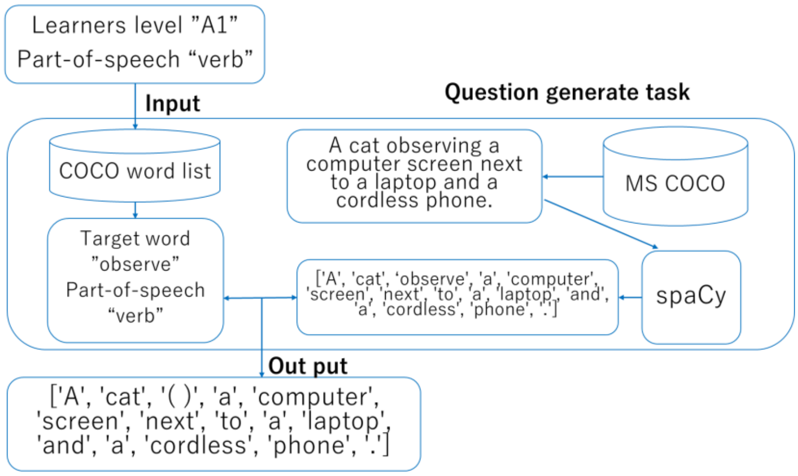

Figure 3: How to generate the question by the proposed model

3.3 Answering Support Task

In the Answering Support task, words that learners tend to answer incorrectly are presented as candidates. Most existing approaches answer such questions with a blank. The reason why we apply the method of selecting candidate words is that for novice learners it is hard to answer without candidate words. The task uses the target word we defined before in the question generation task as input. The candidate answers are searched from the COCO word list using the Levenshtein distance for words with similar spellings. The motivation for using Levenshtein distance is to measure the similarity among each spell. Generally, the distance of each word is measured by its meaning of similarity. However, our proposed method needs to indicate the differences between the error and correct images by image generation. And L2 learners tend to make mistakes by not meanings but similar spells. To accomplish our purpose, we use Levenstein distance to measure the spelling similarity. Figure 4 shows how to extract the candidate words.

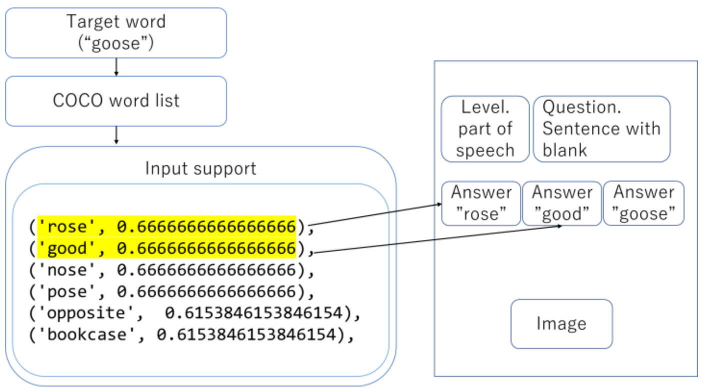

Figure 4: How to extract the candidate word by the proposed model

3.4 Image Generation Task

For each new sentence that contains an incorrect answer, the image generation model generates an image corresponding to that sentence. In this research, we applied DALL·E2 as the model [19]. The model is available as Application Programming Interface. The motivation to use the image generation model is the difficulty to make images correspond to the error. Using image generation from the sentence model, it is possible to make images correspond to the error. This approach also has the possibility to make images that do not exist in the world. Figure 5 shows that generated photographs represent that the input text is reproduced correctly.

Figure 5: The image generated corresponds to the input sentence

4 Discussion

4.1 The Quality of the Generated Question

We mention the problem we currently face. In this study, we introduce the method to create the question automatically with blank utilizing MS COCO. Such a data set is useful for making learning material. However, we need to evaluate the accuracy of the sentence. We found at least 200 misspellings of noun words in 21,463 captions in the MS COCO dataset. To prevent the learner from memorizing the wrong answer, we need to remove such sentences include not only misspellings but also grammatical errors, and need to find a method how to remove them automatically. In addition, the time required to automatically generate questions depends on the total number of data used, so the structure of the model would need to be modified to apply it to larger data sets.

4.2 The Suggestibility of the Generated Image

Our proposed method generates an image that corresponds to a sentence by learner's error. It suggests the error to the learner by changing the image and it must represent the differences between the initial image and generated image. The image must include sufficient information and need to make learners understand the mistake. Figure 6 indicates it is difficult to show the differences among these sentences. We need to consider how to indicate the differences.

Figure 6: The limitation of the image-generated model

5 Conclusion

Recent technology contributes to the learning support system. A large amount of the data set also has a high potential for adopting education. Our proposed method utilizing MS COCO and sentence-to-image generation model has the possibility to enhance the learning vocabulary. In this study, we proposed a method consisting of three tasks to build a vocabulary learning support system that can suggest errors and learn from them by automatically generating images corresponding to the error. We described the possibility of using existing data resources to create word lists by difficulty level. We also described how these data resources could be used to generate questions with blanks for use in vocabulary learning automatically. Currently, our proposed system is under development. Hence, we used a part of MS COCO. When the system is complete, we apply the original data to the system. We hope the coverage of the word is improved corresponds to the size of the dataset. In the image generation task, we have been limited in our use of DALL·E2 until recently. That is the model was offered as a web-like service, and the text as an input needs to be manually entered to obtain the generated image. Therefore, we have planned to use an alternative model like DALL·E mini [20]. However, Now DALL·E2 is available for use in the system as an Application Programming Interface, it is possible to apply DALLE·2 to our proposed model. Now that we are so close to completing our system, we would like to conduct experiments with subjects. We hope our proposed method enhances the effectiveness of learning vocabulary for L2 learners and overcomes their difficulties.

References

[1] Michaud, Matthew & Colpitts, Bradley. (2015). English as a Lingua Franca: Globalization, Ownership, and the Diversification of English.

[2] Laufer, Batia. (1989). A factor of difficulty in vocabulary learning: Deceptive transparency. AILA Review. 6. 10-20.

[3] Nagy, W. (1995). On the Role of Context in First- and Second-Language Vocabulary Learning. Technical Report No. 627.

[4] Hasnine, M.N.; Flanagan, B.; Akcapinar, G.; Ogata, H.; Mouri, K.; Uosaki, N. (2019) Vocabulary learning support system based on automatic image captioning technology. In Proceedings of the International Conference on Human-Computer Interaction, Orlando, FL, USA, 26–31 July 2019; Springer: Cham, Switzerland, pp. 346–358

[5] Shadiev, Rustam & Wu, Ting-Ting & Huang, Yueh-Min. (2020). Using image-to-text recognition technology to facilitate vocabulary acquisition in authentic contexts. ReCALL. 32. 1-18. 10.1017/S0958344020000038.

[6] C. Kobayashi (2016) "Lexical Errors in the English Compositions of Japanese Learners," Foreign language education: Theory and practice, vol42, pp39-51.

[7] Laufer, Batia. (1988). The concept of "synforms" (similar lexical forms) in L2 vocabulary acquisition. Language and Education. 2. 113-132. 10.1080/09500788809541228.

[8] C.Tsuda, & N.Takahashi (2014). Effects of English phonological awareness on English learning among native Japanese speakers, Developmental Psychology Research, vol. 25, no. 1, pp. 95-106.

[9] M.Kimura, & H.Arao (2018), Vocabulary Learning for Japanese of English, GEN TEFL Journal, no.13, pp. 94-105.

[10] Krishan. K. (2014), Fossilization Errors in Intermediate Japanese L2 Learners of English Speech: A Small-scale Inquiry, The Journal of Nagasaki University of Foreign Studies, no 18, pp161-172

[11] Vinyals, O., Toshev, A., Bengio, S., Erhan, D. (2017): Show and tell: lessons learned from the 2015 mscoco image captioning challenge. IEEE Trans. Pattern Anal. Mach. Intell. **39**(4), 652–663

[12] Kaplan-Rakowski, R., Lin L., & Wojdynski, T. (2021). Learning vocabulary using 2D pictures is more effective than using immersive 3D stereoscopic pictures. International Journal of Human-Computer Interaction, 37(12)

[13] Wammes, J.D., Meade, M.E., & Fernandes, M.A. (2016). The drawing effect: Evidence for reliable and robust memory benefits in free recall. The Quarterly Journal of Experimental Psychology, 69, 1752–1776.

[14] T. Horiguchi & T. Hirashima (2002) Simulation-Based Learning Environment for Assisting Error-Correction *Management of Error-Based Simulation Considering the Cause of Errors*, Transactions of the Japanese Society for Artificial Intelligence, vol. 17, no. 4, pp. 462–472

[15] Lin, Tsung-Yi & Maire, Michael & Belongie, Serge & Hays, James & Perona, Pietro & Ramanan, Deva & Dollár, Piotr & Zitnick, C. (2014). Microsoft COCO: Common Objects in Context.

[16] Honnibal, M., & Montani, I. (2017). spaCy 2: Natural language understanding with Bloom

embeddings, convolutional neural networks and incremental parsing

[17] The CEFR-J Wordlist Version 1.6. Compiled by Yukio Tono, Tokyo University of Foreign Studies. Retrieved from http://www.cefr-j.org/download.html on dd/09/2022.

[18] Tono, Yukio. (2019). Coming Full Circle —From CEFR to CEFR-J and back. CEFR Journal - Research and Practice. 1. 5-67. 10.37546/JALTSIG.CEFR1-1.

[19] Ramesh, A., Dhariwal, P., Nichol, A., Chu, C., and Chen, M.(2022), "Hierarchical Text-Conditional Image Generation with CLIP Latents", arXiv e-prints.

[20] Dayma, B., Patil, S., Cuenca, P., Saifullah, K., Abraham, T., Lê Khắc, P., Melas, L. \& Ghosh, R. (2021), DALL·E Mini. https://github.com/borisdayma/dalle-mini