

Title	画像生成による誤答の示唆性を活用した語彙学習支援システム
Author(s)	杉田, 一樹
Citation	
Issue Date	2023-03
Type	Thesis or Dissertation
Text version	author
URL	http://hdl.handle.net/10119/18317
Rights	
Description	Supervisor: 長谷川 忍, 先端科学技術研究科, 修士(情報科学)

修士論文

画像生成による誤答の示唆性を活用した語彙学習支援システム

杉田 一樹

主指導教員 長谷川 忍

北陸先端科学技術大学院大学
先端科学技術研究科
(情報科学)

令和5年3月

Abstract

With the development of globalization, there is a need to use a common language to communicate with each other. English occupies the status of a common language in the world. English is spread worldwide in almost all fields such as science, engineering, technology, medicine, and education. Vocabulary learning has long been taught as a basis for L2 English language learning. To acquire vocabulary effectively, it should be learned through context.

However, learning through context requires the learner must learn through tons of reading. This method is effective for vocabulary learning, but for this reason, it is difficult for L2 learners. Another hand, a widely known method for learning English vocabulary is based on translations. Learners acquire English vocabulary through translation, a method that has been taught in school education for many years. However, such translation-based methods are not suitable. This is because, for example, native English speakers and native Japanese speakers have different ways of interpreting the same thing or situation. The same thing or situation is interpreted differently in each culture.

One of the methods of vocabulary learning that does not rely on these translated texts is to use visual clues. There is some research proposing a system utilizing image captioning, which generates captions from images. They regard the caption and images as the context of learning vocabulary.

In vocabulary learning in English by L2 learners, errors are perceived to occur even among learners at a high proficiency level. The similarity of the spell is one of the main difficulties of learning vocabulary, 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 looking at the spell. The reason why spelling errors occur regardless of proficiency level has been attributed to the phonological awareness of native Japanese speakers.

Such causes of error in vocabulary learning can be the serious problem of fossilization. Fossilization makes errors a part of natural language and makes it difficult for L2 learners to improve their errors. Therefore, we need to consider the lack of suggestions for errors to be a serious problem for L2 learners.

Horiguchi et al. proposed an Error-Based Simulation focusing on the implication of error and its effectiveness for learning a physics task. In this study, the difference between the learner's prediction and the correct phenomenon is visualized by simulation and linked to error correction. This effectively helps the learner understand the cause of the error. We hypothesized that applying this visualization of learners' errors to vocabulary learning would prevent fossilization in English vocabulary learning.

In this research, we developed a vocabulary learning support system with image generation, LVIG(Learning Vocabulary with Image Generation) that automatically generates images corresponding to the learner's errors and shows these images to the learner to encourage reflection. 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.

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. This environment allows learners to correct their mistakes.

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. It is realized by the image generation model. To build such a system, we defined three tasks. The first is a Question Generation task, the task is for making learning material automatically from the MS-COCO dataset. To accomplish it, we define the vocabulary list COCO-word list by utilizing Natural Language Processing. We also made the module to generate a question automatically. The task is to extract the target word from the sentence from MS COCO and replace it with the blank. The module also utilizes Natural Language Processing. The second is an Answering Support task, which supports the learner in answering the question by extracting the candidate's word by similar spelling. It enables us to learn vocabulary easily through the candidate's words. The third is an Image Generation task. Utilizing the image generation model DALL-E2 which realizes generating images corresponds to learners' errors.

In addition, we conducted the experiment with 25 subjects. The results of the experiment were analyzed for 22 of the 25 subjects. The reason for this is that we excluded subjects whose data were missing due to a system error. The subjects in the experiment were graduate students with a high level of proficiency in English. The experimental setting consisted of two models. The first model includes image generation. When the learner makes an incorrect answer, the image generated corresponds to the sentence with an error. While the second model is a conventional model that does not include image generation. The vocabulary for the experiment consisted of a total of 100 high school to college-level questions. Eight parts of speech were used. In order to evaluate whether repetitive errors could be prevented, questions consisting of 20% error-only choices were included. The same subject tried on the two conditions and repeated them twice. Each condition is assigned 50 questions, 100 questions per term, and this is repeated twice for a total of 200 questions. Subjects completed a questionnaire when answering the questions. 1. The questions asked whether the images in the COCO dataset were associated with a sentence, 2. Whether the images generated during the wrong answer were associated with a sentence, 3. Whether the images generated during the wrong answer were impressive. In order to eliminate the effects of the order of the questions and the characteristics of the subjects, we adopted a counterbalancing technique. In this case, the learner solves the same problem twice.

To evaluate the learning effect, we measured whether the learner made the same mistake repeatedly when solving a problem in the second term as they did in the first term. To evaluate the learning effect, we compared the two models based on the number of vocabulary that the learners repeatedly got wrong. The Wilcoxon signed-rank test was used as the test method for the evaluation. The Wilcoxon signed-rank test is a non-parametric statistical hypothesis test used either to test the location of a population based on a sample of data or to compare the locations of two populations using two matched samples. We also analyzed the effects of associativity and impressions of images on vocabulary learning, based on questionnaires obtained from subjects. To evaluate the model, we assessed whether the images represented the content of the sentences based on associativity collected from the subjects' questionnaires.

Our experiment result showed that the proposed method can prevent repetitive wrong answers by comparing the proposed method with a model that includes image generation and a conventional method. We also found that the images in the existing dataset and the images generated by the image generation model are of high quality. On the other hand,

due to its high accuracy, the proposed model has the potential to induce learners to learn vocabulary incorrectly when the learning content contains errors. Although these problems occurred because of the high reproducibility of the model, these experiments showed that our proposed method effectively prevents repetition errors and is effective in English vocabulary learning.

In 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. We believe that the ability to reproduce concepts through the application of an image generation model has the potential to replace the experiential learning of our native language by enabling the conceptual acquisition of a second language. We aim to develop a vocabulary learning support system that can help L2 learners learn English vocabulary more effectively by using image generation to understand English concepts.

目次

第 1 章 はじめに	1
1.1 研究の背景	1
1.2 目的	2
1.3 論文の構成	3
第 2 章 関連研究	4
2.1 視覚的情報を取り入れた英語語彙学習に於ける研究	4
2.2 誤答の示唆性を活用した研究	6
第 3 章 提案手法	8
3.1 概要	8
3.2 システム概観	8
3.3 質問生成タスク	9
3.3.1 COCO-Word list	10
3.3.2 空所付き問題生成タスク	14
3.4 回答支援タスク	15
3.5 画像生成タスク	18
第 4 章 プロトタイプの実装	20

4.1 概要.....	20
4.2 開発環境.....	20
4.3 システムの3つの構成要素に於ける課題.....	20
4.4 システムの機能.....	24
第5章 ケーススタディ	26
5.1 目的.....	26
5.2 実験方法.....	26
5.2.1 実験条件	27
5.2.2 実験手順	30
5.3 実験の結果	31
5.4 学習効果に対する評価	31
5.4.1 繰り返しの誤りの回避に対する評価	31
5.4.2 定性評価による生成画像の特性と繰り返しの誤りの回避に対する評 価	35
5.4.3 遅延テストに於ける誤った語の理解度に対する評価	38
5.5 モデルの精度に対する評価	43
5.5.1 MS-COCO データセットに含まれる画像, 及び生成画像に対する評価	43

5.5.2 画像生成モデルに依る生成画像に対する考察	46
第6章 まとめと今後の課題	48
対外発表	50
謝辞	51
参考文献	52
付録	55

目次

図 1-1 学術論文に於ける記述言語の割合[1].....	1
図 2-1 WDIARY 操作画面例[11]	4
図 2-2 Show and Tell モデルを適用した語彙学習支援システム学習例[12]5	
図 2-3 Error-Based-Simulation による学習イメージ[16]	7
図 3-1 提案システムユースケース図	9
図 3-2 質問生成タスクフロー	10
図 3-3 教育用例文コーパス SCoRE データ例 [17].....	11
図 3-4 COCO Word list 生成モデル図.....	12
図 3-5 COCO Word list の CEFR-J word list に対する網羅率.....	14
図 3-6 空所付き問題生成タスクモデル図	15
図 3-7 回答支援タスクモデル図	17
図 3-8 Edge Connect による対象物体の消去, および画像合成のイメージ [20].....	18
図 3-9 DALL-E による画像生成例 [21]	19
図 4-1 学習用問題集構築フロー	21
図 4-2 学習用問題集データ	22

図 4-3 画像生成済み学習用問題集データ	22
図 4-4 ラーニングログ	23
図 4-5 実験用プロトタイプユースケース図.....	24
図 4-6 システム操作例 1（誤答を選択した場合）	25
図 4-7 システム操作例 2（正答を選択した場合）	25
図 5-1 実験参加者の英語習熟度別内訳.....	26
図 5-2 学習対象語品詞別内訳	27
図 5-3 被験者実験に於ける条件，順序.....	28
図 5-4 被験者によるリッカートスケールを用いたアンケート	29
図 5-5 実験実施環境	31
図 5-6 Condition1,2 に於ける繰り返しの誤答数に対する被験者数	34
図 5-7 画像の想起性と繰り返しの誤答が発生した語彙の相関関係.....	35
図 5-8 画像の印象と繰り返しの誤答が発生した語彙の相関関係	36
図 5-9 繰り返し誤答が発生した語彙に対する生成画像例①.....	37
図 5-10 繰り返し誤答が発生した語彙に対する生成画像例②.....	37
図 5-11 遅延テストのアンケートの一部抜粋.....	39
図 5-12 特徴消失が発生した生成画像例.....	42
図 5-13 アンケートの結果をタグ付けしたラーニングログ	43

図 5-14 各初期画像の想起性の平均値	44
図 5-15 MS-COCO データセットから抽出した初期画像の想起性.....	45
図 5-16 画像生成モデルによって生成された画像の想起性	46
図 5-17 学習者に誤った認識を与える生成画像例	47

表目次

表 3-1 COCO Word list 及び, CEFR-J word list のレベル・品詞別単語数	13
表 3-2 Levenshtein 距離による 2 文字列間の距離の導出例.....	16
表 5-1 各 Condition に於ける被験者の繰り返し誤り回数.....	32
表 5-2 被験者の繰り返しの誤りの回数に於ける検定表.....	34
表 5-3 各 Condition に於ける誤った語の理解度の評価値.....	40
表 5-4 誤った語の理解度の評価値に於ける検定表.....	41

第1章 はじめに

1.1 研究の背景

近年の国際化により，様々な文化，言語を有する人々の互いのコミュニケーション手段としての共通言語の必要性が高まっている．Parupalli はサイエンス，技術，産業の分野に於いて英語は、国際関係を維持するための共通言語としての役割（the role of English as a global language）を担っていると述べている[1]．又，英語は医学，コンピュータなどに於ける専門家らにとって，必須のスキルとされ，特にサイエンス分野に於ける重要性について言及している．図1では，学術論文に於ける記述言語の割合を示しており，その大半が英語で記述されたものであることを示している．又，Parupalli は国際的に影響を与え，知識を共有する上で，論文を英語で出版する必要性を主張している．

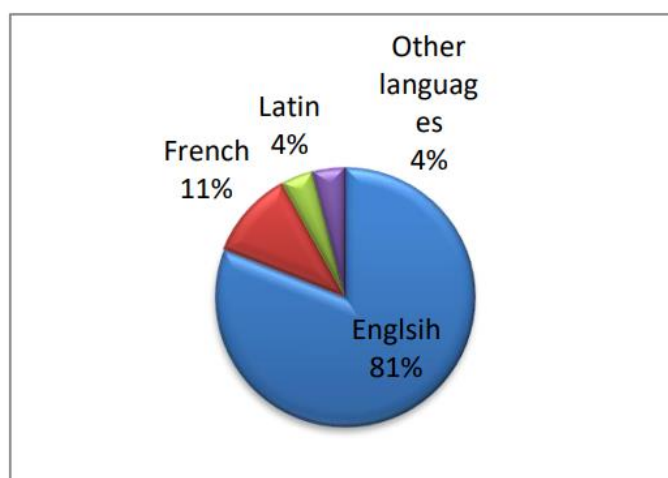


図 1-1 学術論文に於ける記述言語の割合[1]

国際関係を築く上で重要とされる英語の学習に於いて，特に語彙学習の重要性が広く認知されており，語彙の習得は文脈から学習することが効果的であるといわれている[2]．一方で，英語語彙学習では，英語の習熟度と語彙の正確さは必ずしも一致しないとし，小林は高い習熟度の学生に於いても語彙の間違ひが多く見受けられる事を明らかにした[3]．Nagy らは文脈から学習する方法の

重要性について認めつつも、これらの手法は英語を第二言語とする学習者である L2 学習者に於いては困難であることを示している[4]。その理由として、文脈から学ぶ方法は膨大な量の読書が求められる点を述べている。一方で語彙学習では、画像を用いた研究が多く存在し、画像から対応する文を自動で生成するイメージキャプションを活用した語彙学習支援システムが提案されている[5]。ここでは、文に対応する画像を文脈としてみなすことが可能だとし、膨大な読書量に対する画像の代替の可能性を示している。Laufer は、L2 学習者にとって、スペルと発音の近似が、語彙の間違いを誘発するとしている[6]。又、英語のスペルに於ける近似が学習を困難にさせることは、日本人のみならず、英語を第二言語とする L2 学習者全般の共通課題として認識されている。日本人学習者を対象としたとき、今井は、日本人学習者の英語と日本語の完璧な対応への過信が学習者の理解を阻害していると述べ、対訳に頼る学習方法を懐疑的にとらえている[7]。田中は対訳に依る弊害について、前置詞を例にとり対訳の限界について言及している[8]。例として”机に本がある”や”川にたくさん魚がいる”といった文章に対し、英語では、”There’s a book on the desk.”と”There is a lot of fish in the river.”となる。ここで、両方とも日本語では英文の前置詞に対応する語を”に”と表現している点に日本語の曖昧さがあると述べている。一方、英語ではそれらを明示的に示す必要があるとされている。これらは、英語の語彙習得に於いて、対訳に依る学習の限界を示している。

日本人は、漢字に代表される形から意味の類推が可能な、表意文字を主に用いているが、英語は表音文字である。それ故に、英語は形から意味を類推することが困難である[9]。一方、英語学習に於いて、先述した近似したスペルなどを誤った語の意味として記憶し、繰り返し語彙を誤って用いることにより、誤りを自然言語として獲得してしまう、Fossilization がある。一度自然言語化したものは、修正が困難であるとされている。Krichan らは、日本人を対象とした実験に於いて Fossilization を観測した[10]。これらは、L2 学習者にとって、誤答への認識の重要性を示しており、学習者の語彙に対する誤りを効果的に認識させることが必要であると考えられる。

1.2 目的

本研究では、英語語彙補充問題を基本とし、学習者の誤答を画像生成により可視化させることにより学習者の誤答に対する内省を促す、語彙学習支援システ

ム LVIG (Learning Vocabulary with Image Generation) を提案する。学習者の誤答に応じて画像を生成させることにより学習者の内省を促し、学習における効果的な印象を与えると仮定する。ここでは、学習者に効果的に誤答を誘発する為、効果的に誤答をする傾向が高い環境を作り出す。提案する語彙学習支援システムの構築に向け、以下の3つのタスクを実現させる。

- ① 学習用問題を既存のデータセットより自動で生成する質問生成タスク
- ② 学習者が誤答しやすい解答候補を提示する回答支援タスク、
- ③ 学習者の誤答に応じた画像を生成する画像生成タスク

また、提案するシステム LVIG を用いた学習により、正しい画像と誤答によって生成された画像の差異が学習者の内省を促し、効果的に誤答を印象付け、繰り返しの誤答によって生じる Fossilization を防ぐとの仮説に基づき、誤答時の印象、学習効果について評価を行う。

1.3 論文の構成

本論文では、第2章において、関連研究に於ける取り組みである、画像を用いた英語語彙学習、誤答の示唆性に注目した研究を紹介する。第3章では提案手法について述べ、第4章に於いて提案手法であるシステムの設計、および実装について述べる。第5章では評価実験に於ける提案手法の効果について考察を行う。第6章では本論文のまとめと今後の課題を述べる。

第2章 関連研究

本章では、視覚的情報を活用した英語語彙学習、および誤答時の示唆性を活用した関連研究について紹介する。

2.1 視覚的情報を取り入れた英語語彙学習に於ける研究

視覚的情報を取り入れた英語語彙学習の先行研究は多く存在する。菊池らは、実体験に基づいた写真オブジェクトに英単語をタグ付けする英単語学習アプリケーション WDIARY(W:Word D:Diary I:Image A:Addict RY:memoRY)を提案した[11]。WDIARYでは過去の記憶情報と英単語を結びつけることで、「情報のつながり」を構築し、学習後の記憶検索の向上に対する評価を行った。WDIARYでは、学習者が写真に英単語のタグ付けを行う「日記を書く」、復習のための「学習単語帳を復習」、学習内容の評価である「評価テスト」の3つの機能を実装した。図2-1では、1つ目のプロセスである「日記を書く」プロセスを示している。

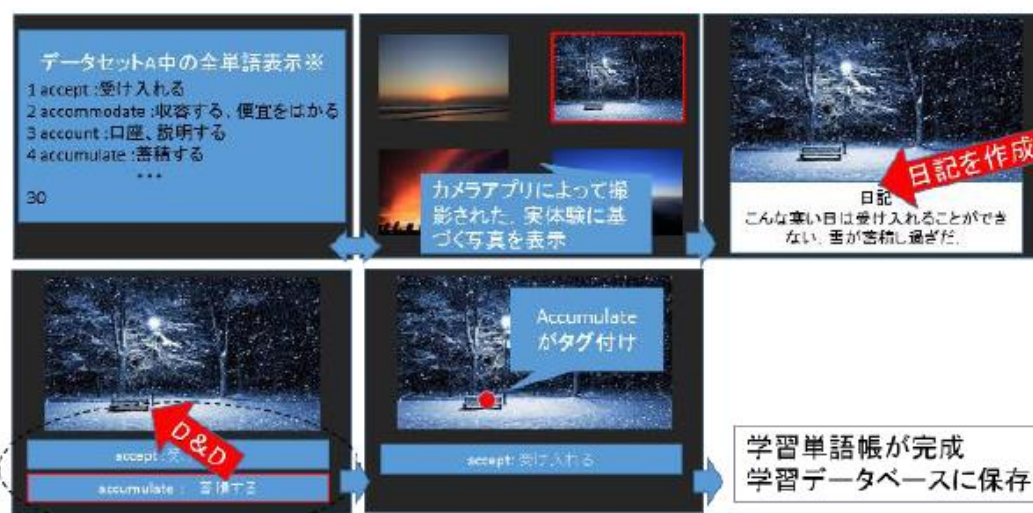


図 2-1 WDIARY 操作画面例[11]

学習者自らが撮影した写真と学習したい単語をつなげることにより、記憶の探索効率を向上させる菊池らの研究に対し、本研究では、学習者が誤答を行うこ

とによって画像を作り出す点に於いて、学習者が体験的に視覚的情報と英単語を結びつける点で共通している。しかし、学習者が能動的に学習したい単語を選定するコストを鑑みたとき、必要英単語を網羅的に学習することに於いては、困難が生じると考えられる。本研究では、難易度別語彙集に準拠した学習コンテンツを自動で生成される点に於いて異なる。

現実に存在する画像を用いた英語語彙学習では、Hassine らは、各語彙に応じた学習用教材を用意することの困難さについて言及し、画像から自動で対応する文を生成する、Show and Tell モデルを適用した語彙学習支援システムがある[12]。Hassine らの提案する語彙学習支援システムでは、学習者が既に学習済みの単語をラーニングログとして記録する。学習者が写真に語彙を付与したものをシステムにアップロードすることで、ラーニングログとアップロードされた写真を基に、既に学習された単語が含まれる新たな語彙学習の為のコンテンツを自動生成することを実現させた。この研究では、3つの目的を掲げ、それぞれ、1、画像から文を生成するモデルによりイメージから得られる情報から学習用コンテンツの自動生成を実現させる。2、学習者は、単語ごとに得られるイメージに頼らず、一枚の絵から複数の語彙を学習することを可能とする、3、すでに学習済みの語彙を新たな単語を学習する際に提示することで、再び思い出すことを可能とする。図 2-2. では、学習者は写真に語彙“Lunch box”を付与し、システムにアップロード、学習者がすでに学習済みの“食べた(eat)”をラーニングログから抽出し、3つの学習コンテンツの生成を実現していることを示している。

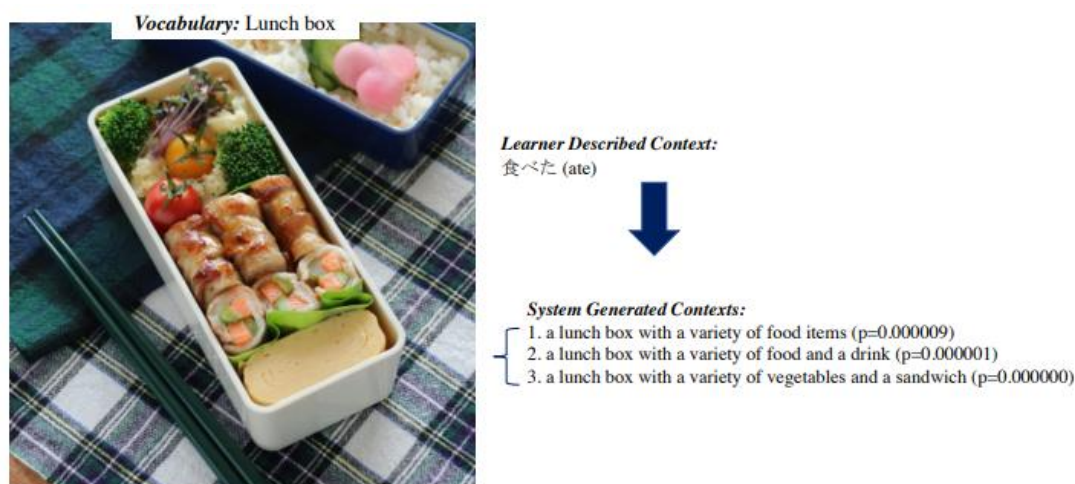


図 2-2 Show and Tell モデルを適用した語彙学習支援システム学習例[12]

Hassine らの研究では、画像から選ばれた学習対象語を対象としている点で、本研究の MS-COCO[13]データセットから問題を自動生成するプロセスと似ているが、学習者の誤答については言及されていない。本研究では、学習者が学習時に誤って覚えてしまうことによる弊害に注目し、繰り返しの誤答を防ぐことを目的にしている点で異なる。

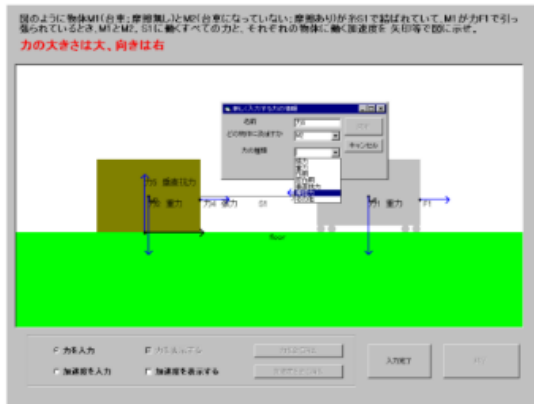
Regina らは S3D の画像を用いた英語語彙学習支援システムの有効性について、2D の画像との比較実験に於いて、2D の画像のモデルの方が学習効果が高いことを明らかにした[14]。情報の増加が必ずしも学習効果に寄与するわけではないことを示した。又、Wammes らは、語彙に対応する絵を描くことが、語彙学習に於いて効果的であることを明らかにし、描写した絵の質が低い場合に於いても効果があることを述べている[15]。これらの研究では、学習に用いられるコンテンツの画像の精度の高低に関わらず学習が有効であることについて述べられ、本研究が適用する画像生成モデルの精度に関わらず学習効果を得られる事が期待できる。ここまでに、画像を扱った研究について紹介したが、視覚的情報を活用した英語語彙学習に於いて、誤答時の対処について言及されていない。

2.2 誤答の示唆性を活用した研究

学習者の学習に於ける誤りを修正することに注目した研究では、堀口らの物理の課題に対する研究が存在する[16]。堀口らは、学習者の予測と正しい現象との差異をシミュレーションにより可視化し、誤りの修正に結び付ける **Error-Based-Simulation** を提案した。ここでは、誤答を示す際に十分な情報が含まれており、正答と誤答を明確に区別できる必要性を説いている。又反例の示す事実には、十分な信頼性がある必要がある点について言及している。図 2-3 左では、システムの操作画面を表しており、図 2-3 右は学習者が提示された課題に対して、垂直抗力を誤ったために、ブロックが床に沈み込むシミュレートがされている。

本研究では、英語語彙学習に於いても、誤答時のシミュレートを画像生成モデルの活用による可視化を実現し、それを提示することにより学習者の内省を促し効果的に **Fossilization** を防ぐことが可能であると仮説を立てる。又、語を入力することによって生成される画像を通じ、間違った語彙に対する理解を促す事を目指す。

誤りの修正を支援するシミュレーション環境



a システム画面 (1)

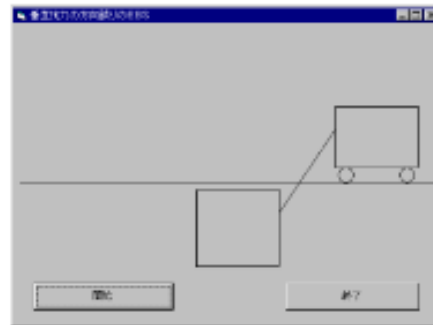


図 8 ブロックが床に沈む EBS



b システム画面 (2)

図 2-3 Error-Based-Simulation による学習イメージ[16]

第3章 提案手法

3.1 概要

本研究では、学習者が誤答を繰り返すことによって生じる Fossilization を防ぐ為、学習者の誤答に応じた画像を提示することによって学習者に内省を促し、繰り返しの誤答の回避につながる語彙学習支援システムの構築を行う。その為に学習者が誤答を起こしやすい環境を構築し、誤答に応じた画像を自動で生成し提示する必要がある。又、手動での学習コンテンツの生成の困難さから、学習に用いる学習用コンテンツを自動で生成する必要がある。

これらのシステムを実現する上で、本章では、初めにシステムを概観し、その後システムを構成する以下の3つのタスクについて論じる。

- ① 問題生成タスク・・・学習コンテンツの自動生成を行う。
- ② 回答支援タスク・・・学習者の誤答しやすい環境を構築する。
- ③ 画像生成タスク・・・学習者の誤答に応じた画像を生成する。

3.2 システム概観

ここでは、先述したシステムの構成する3つの要素を踏まえ、システムの流れについて概観する。以下のユースケース図 3-1 に於ける①～③については、それぞれ先に述べた3つのタスクに対応する。

- ① 学習を行う上で、学習者のレベル、学習対象品詞を選定する。ここでは、継続した学習を行う場合、学習ログを基に選定される。選定したレベル、品詞を基に、COCO-Word list より学習対象語を選定する。次に学習対象語を空所とする空所付き問題を質問生成タスクにより生成し、学習画面に提示する。この時、文に対応した画像を MS-COCO データセットより抽出し表示する。
- ② 回答の候補となる語を、学習対象語に近似したスペルのものを回答支援タスクにより選定し提示する。
- ③ 学習者が誤った回答を行った時、画像生成タスクにより、学習者の誤答

に応じた画像を生成し提示する。これらの生成された画像，空所付き問題は学習用問題集としてストレージに補完され，再利用される。又，同時に学習者の学習ログを取り，それを基に次の質問が生成される。

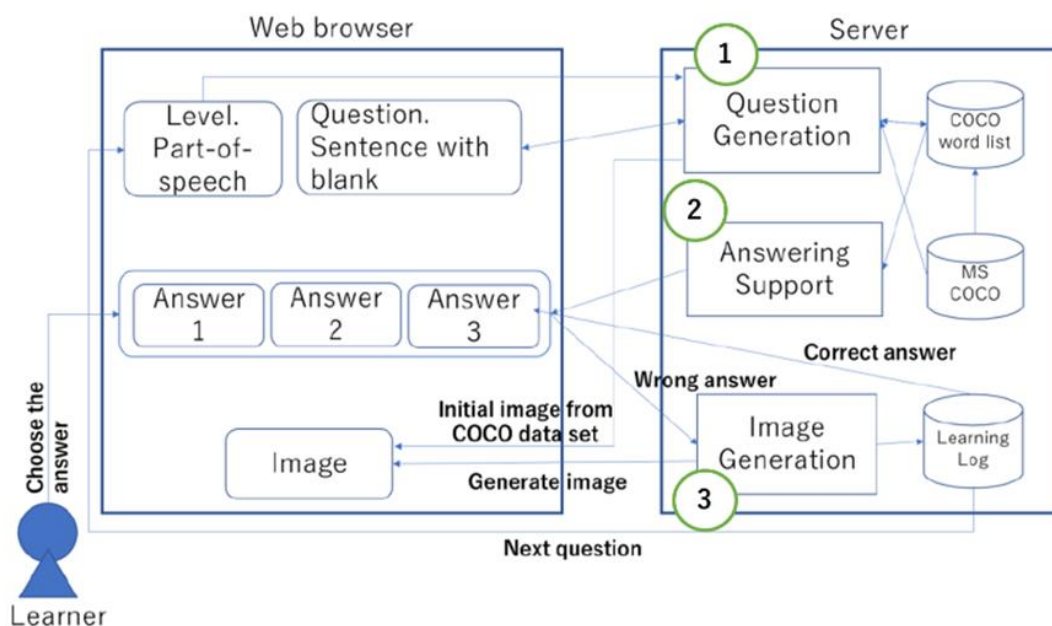


図 3-1 提案システムユースケース図

3.3 質問生成タスク

質問生成タスクでは，学習者のレベル，対象の品詞に応じて学習対象語を選定し，学習対象語を空所とする問題文を自動的に生成する。質問生成タスクを構築する上で必要となる，語彙集 COCO-Word リストの構築手法，並びに空所補充問題の自動生成手法について述べる。又，既存の空所補充問題コーパスを紹介し比較を行う。質問生成タスクのフローは以下の図 3-2 に示す。

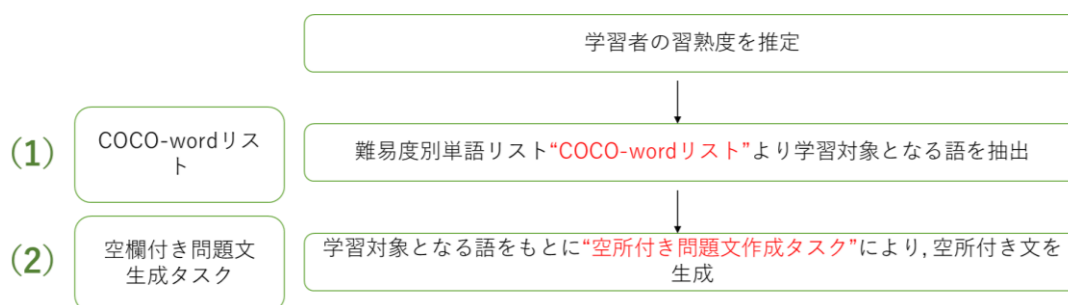


図 3-2 質問生成タスクフロー

質問生成タスクでは、先述した学習システムに於いて、学習者の習熟度を推定したのち、難易度別単語リスト“COCO-word list”より学習対象となる語を抽出する。そののち、学習対象語を空所とする問題を生成する”空所付き問題生成タスク“により、問題を生成する。質問生成タスクでは、大別して、(1)COCO-word list の作成、(2)空所付き問題生成タスクに分けられる。

3.3.1 COCO-Word list

英語語彙学習に於ける学習用コーパスはいくつか存在し、教育用例文コーパス SCoRE (Sentence Corpus of Remedial English) [17]は、自然な英文と日本語対訳を自由に使える「データ駆動型英語学習支援プログラム」である。SCoREには、図 3-3 に示すように空所補充問題用例文が含まれ、対象品詞、対象語、レベル、が充てられる。一方、本研究の視覚としての情報である画像をコンテンツとして含めたい場合、これらのデータセットはこれらの文に対応する画像を集めることが困難である。

大分類	小分類	キーワード	レベル	用例	
代名詞	指示代名詞	that	beginner	The footstep was (that) of a teacher.	その足あとと先生のものは
代名詞	指示代名詞	that	beginner	His voice is louder than (that) of mine.	彼の声は私の声よりも
代名詞	指示代名詞	that	beginner	My room is bigger than (that) of John's.	私の部屋はジョンの部屋
代名詞	指示代名詞	that	beginner	His house is connected to (that) of Linda's.	彼の家はリンダの家と
代名詞	指示代名詞	that	beginner	The voice was (that) of a singer.	その声はある歌手のもの
代名詞	指示代名詞	that	beginner	The smell was (that) of mother's pie.	その匂いは母のパイの
代名詞	指示代名詞	that	beginner	His house is larger than (that) of mine.	彼の家は私の家より大
代名詞	指示代名詞	that	beginner	Her foot is smaller than (that) of his.	彼女の足は彼の足より
代名詞	指示代名詞	that	beginner	His is different than (that) of Miranda's.	彼のものはミランダの
代名詞	指示代名詞	that	beginner	Liam's pie is better than (that) of Kyle's.	リアムのパイはカイル
代名詞	不定代名詞	one	beginner	He bought (one).	彼はそれをひとつ買
代名詞	不定代名詞	one	beginner	Sam got (one).	サムはそれをひとつ
代名詞	不定代名詞	one	beginner	Linda built (one) very quickly.	リンダはそれをとても

図 3-3 教育用例文コーパス SCoRE データ例 [17]

本研究では、画像とキャプション等のアノテーションデータを含む大規模データセットである、MS-COCO を活用し、学習コンテンツの作成をする。画像とその画像に応じたキャプションを含むアノテーションデータで構成される、大規模データセット MS-COCO データセットの一部である 5,000 のイメージに対して付与された 21,463 のキャプションを対象に学習対象語を抽出し、COCO Word list を構築した。構築の流れとしては、図 3-4 より、(a)初めに COCO-API を用い、MSCOCO データセットのアノテーションデータの一つである、キャプションを抽出する。キャプションには、図に示す通りそのキャプションに対応する画像の ID、キャプション個々に与えられたキャプション ID などが含まれる。(b)次に、抽出したキャプションを自然言語処理ライブラリである SpaCy[18]を用いて、単語分割、原型への変換、大文字から小文字への変換を行う。図 3-4 の (b)では、以上のプロセスによって得られた単語を示している。(b)のプロセスによって得られた単語を後述する語彙難易度表である CEFR-J word list を参照し、難易度が付与されたデータを抽出、8 つの品詞、A1~B2 の難易度別語彙集として(c)COCO Word list を構築した。

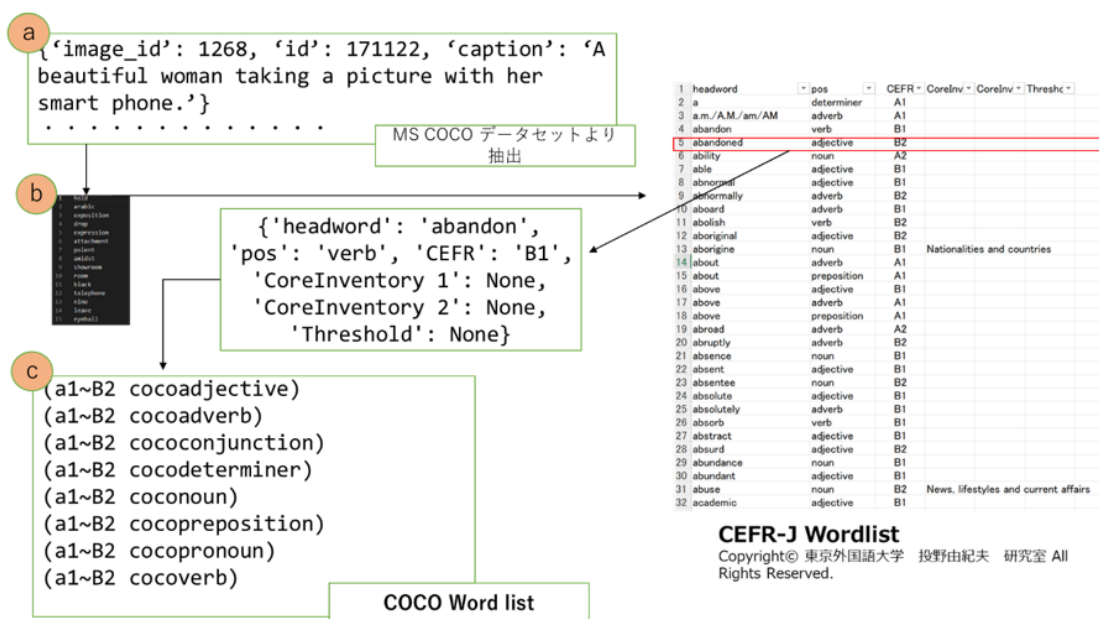


図 3-4 COCO Word list 生成モデル図

CEFR-J[19]は欧州共通言語参照枠 (CEFR) をベースに、教科書コーパス (中国・台湾・韓国の小中高の主力教科書を CEFR 基準に大まかに分類したものをベースに、各国・地域の CEFR レベル・テキストに取り扱われている共通語彙を抽出する、という方法で構築された。それぞれのレベルごとの語彙数は、小学校～中学校1年程度である A1 (1,166words), 中学校2年～高校1年程度である A2(1,411words), 高校2年～大学受験レベル程度である B1(2,445words), 大学受験～大学教養レベルである B2 (2,779words) である。ここで、本研究で定義した COCO Word list の語彙の網羅性について言及する。表 3-1 では、先の手順により定義した COCO Word list の語彙数と CEFR-J word list の語彙数の対応関係を表している。図 3-5 では、グラフにより COCO Word list の網羅性を可視化した。これらの図が示すように、COCO Word list では、初等教育で定める重要な語について網羅していることを表している。一方、高いレベルになるにつれ網羅性は低下していることが伺えるが、今回提案した手法が MS-COCO データセットの一部しか使用していないことから、データの拡張により向上することが予見される。提案手法に於いて、MS-COCO データセットのすべてのデータを適用しなかった理由としては、詳細については後述するが、データセット

を拡張することにより，後述する空所付き問題生成タスクに於ける生成時間が増加する為である．

表 3-1 COCO Word list 及び, CEFR-J word list のレベル・品詞別単語数

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

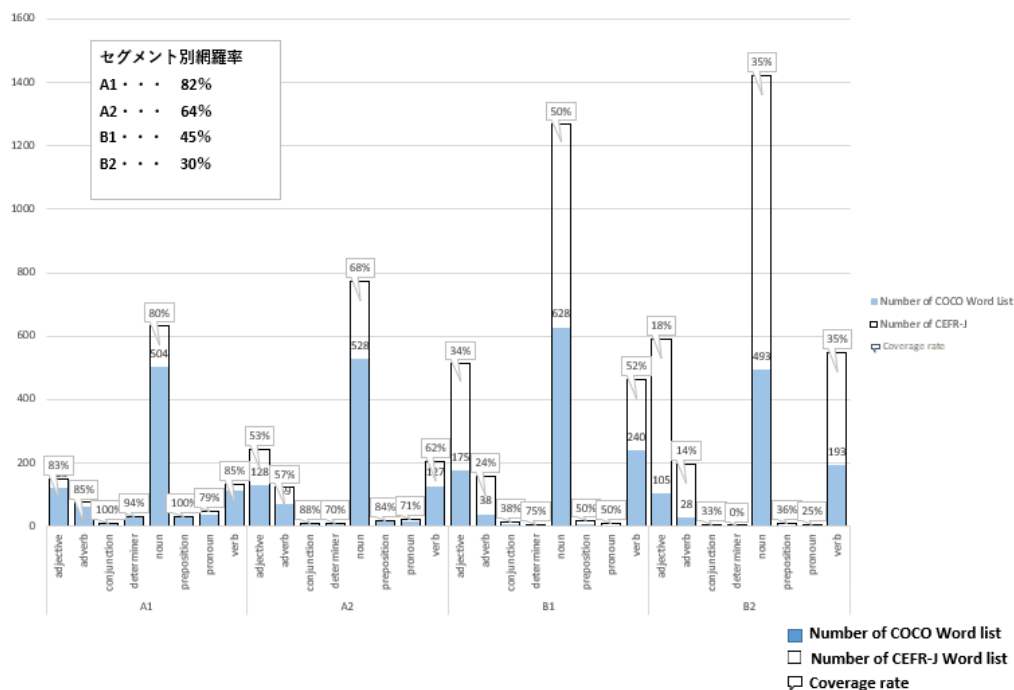


図 3-5 COCO Word list の CEFR-J word list に対する網羅率

3.3.2 空所付き問題生成タスク

空所付き問題生成タスクでは、学習者のレベルに応じた学習対象語を空所とする問題を自動生成する。

図 3-6, (a)では、学習者のレベルに応じ、COCO Word list より学習対象語を選定する。ここでは、学習対象語として“observe”，品詞“verb”が選定されることを示している。(b)では、先に選定された学習対象語を空所とする問題を生成する為、MS-COCO データセットからキャプションにより検索を行う。検索の流れは、初めに抽出したキャプションに対して、自然言語処理ライブラリである SpaCy を用いて、単語分割、原型変換、品詞情報の取得を行う。ここでは、文脈に応じた単語の品詞情報を取得することが重要となる。英語語彙に於いて、スペルは一致しているがその品詞の使われ方により難易度、意味が異なることがある。又、後述する回答支援タスクに於いて解答候補となる語に於いて同一品詞に保つ必要があるためである。以上の処理によってキャプションより得られた単語と、学習対象語を、品詞、スペルにより照合し、合致した初めの語を空所に置換することによって空所付き問題の自動生成を行う。この時、一つの語に対し

て、複数の問題を生成する可能性を有しているが、問題生成では、MS COCO データセットのデータの探索を行う為、生成時間を考慮し初めに合致する文のみを対象とした。同様の理由として、先述したデータセットを 5,000 のイメージに与えられた 21,463 のキャプションに制限している点も生成時間を低減させることを目的としている。現状のシステムでは、問題生成タスクに於ける空所付き問題生成タスクの探索の時間は対象となるデータの大きさに依存する。(c)では、生成された空所付き問題を示している。

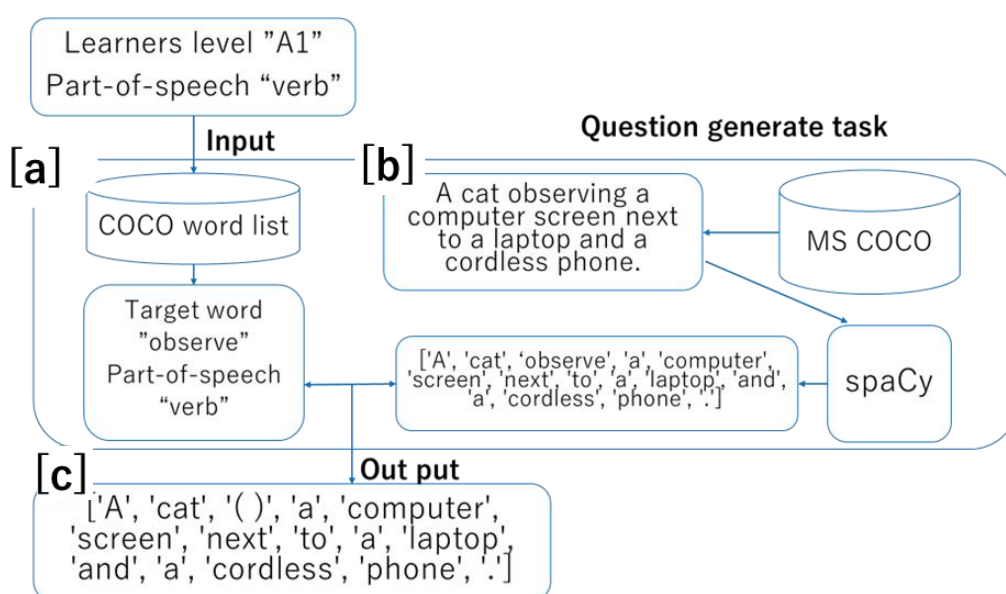


図 3-6 空所付き問題生成タスクモデル図

3.4 回答支援タスク

回答支援タスクでは、学習者が円滑に学習システムで学習を行うための支援を行う。ここでは、学習者が誤答しやすい環境を構築することにより効果的に学習者に誤答への内省を促すことを目的とする。誤答をしやすい環境については、L2 学習者の語彙学習に於ける誤りの一因とされるスペルの近似に注目し、近似したスペルの語彙を提示することにより学習者が誤答しやすい環境を構築した。

英語語彙空所補充問題に於いて、入力方法は学習者がスペルを入力する方法と、正しい語を選択する選択式の問題の二つの方法が考えられる。学習者がスペル

ルを入力する方法では、英語の習熟度が低い学習者に於ける学習の困難さが予想されるため、本研究では、学習者に解答候補を提示する選択式の問題を採用する。学習者に近似したスペルの語彙を提示する際の語彙間の距離を測定する為に、Levenshtein 距離を用いた方法を適用する。ここでは、Levenshtein 距離の測定方法を表 3-2 を基に例示する。

Levenshtein 距離は、文字列に対する挿入、削除、置換、の操作により距離の測定を行う。ここでは、簡単のため各編集コストを 1 と仮定する。マスには、それぞれ、以下に定める 3 つの方法により導出された値の最も小さい値を埋める。

- (a) 補充するマスの上の数値 + 1
- (b) 補充するマスの左の数値 + 1
- (c) 補充するマスの左上の数値 + x(列, 行がともに同一文字の時, x=0, 列, 行が異なる時 x=1)

例えば、“せったい”と“せたい”の文字列の距離を測りたいとき、表 3-2 を埋めることにより導出が可能となる。表 3-2 黄色で示す箇所では、n 文字の文字列から成る N に於ける n と 0 の距離は n となり、黄色で記した一行目と一列目が定まる。次に、先述した 3 つのルールに沿って表を埋める。次に青で記されたマスについては、列, 行がともに等しいため、(c)x=0 の時が最も小さい数値となり、左上である緑マスとの合算により 1 が埋められる。すべてのマスが埋まり右下に埋められた表で記す太文字の数値が 2 文字列間距離となる。

表 3-2 Levenshtein 距離による 2 文字列間の距離の導出例

	□	せ	っ	た	い
□	0	1	2	3	4
せ	1	0	1	2	3
た	2	1	1	1	2
い	3	2	2	2	1

このように、Levenshtein 距離は、動的計画法によって距離が導出され、計算量は、(n+1)(m+1)の二次元配列により、挿入、削除、置換が行われる、この時の

計算量は $O(mn)$ となる。

Levenshtein 距離を用いた python API の一つとして、pyspellchecker [20]がある。Pyspellchecker では、事前に辞書が導入されており、入力したスペルにスペルミスがあった場合に、Levenshtein 距離により近似した語を辞書から抽出し、入力したスペルミスに対し、正しいスペルの候補を提示するものである。本研究では、回答の候補となる単語に対しても学習対象語と見做しており、事前に定義された辞書では学習難度の設定が困難な点から、これらツールを使用せず、提案手法で定義した COCO Word list 内の語を対象に Levenshtein 距離を測定し、候補として提示する独自の機能を実装した。

提案する回答支援タスクのモデルについて、図 3-7 では、学習者の学習対象語に対し、定義した COCO Word list から近似したスペルの抽出を行う。具体的には、先述した Levenshtein 距離を用いて対象となる語に近似した語を抽出する。ここでは、学習対象語“goose”に対して、“rose”、“good”、“nose”などが候補として抽出されている。抽出した候補から一致度が高いものから解答候補とし、正答を含む 3 つの解答候補を生成する。

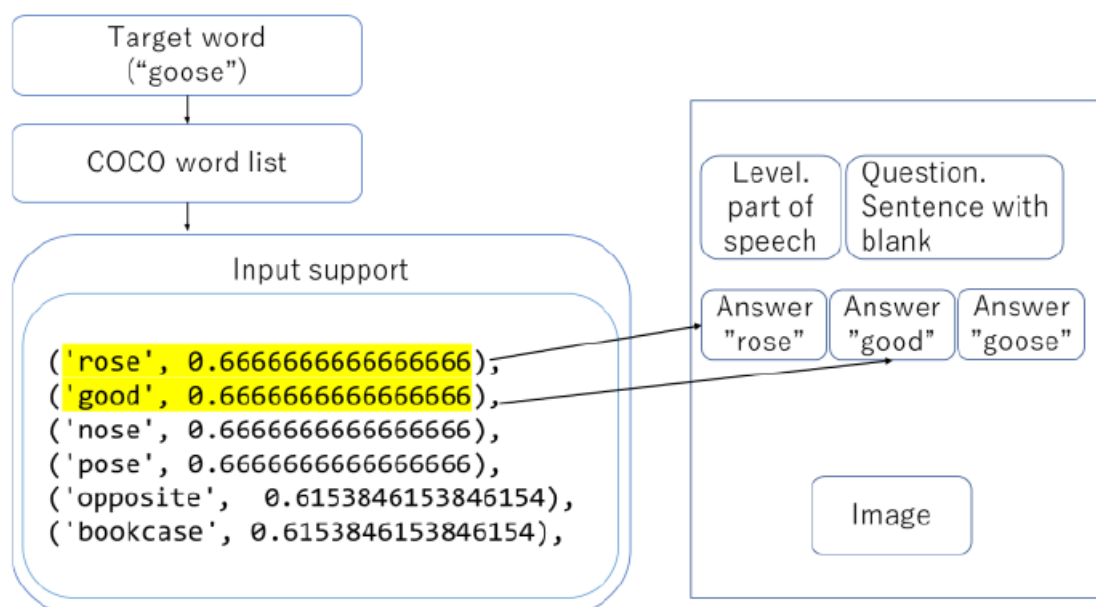


図 3-7 回答支援タスクモデル図

3.5 画像生成タスク

画像生成タスクでは、学習者の誤答に対して、対応した画像を生成し提示する。英語語彙学習に適した画像生成モデルを選定する上で、関連研究に於ける画像生成モデルを比較したうえで、本研究で用いた手法を紹介する。

空所補充問題を基本とし、学習者の誤答に応じた画像を表示させるうえで、生成される画像を局所的に変化させる手法、すべての画像を変化させる手法が考えられる。Nazeri らは、画像上のマスクされた領域を再構成する技術である、“Image Inpainting”の研究に於いて、Edge Connect を提案し、活用方法の一つとして不要な対象を削除するタスクが存在する[20]。ここで、学習者が誤った回答を選択した際に、その誤答を画像表現することを実現させるうえで、一度空所に該当する正答を表す対象物を消去し、誤答となる物体をイメージを重ねることで、学習者の誤答を表現させる方法が考えられる。図 3-8 では、MS-COCO データセットのイメージから Edge Connect により対象となる物体を消去し、事前に用意された画像を対象座標に重ねている。ここでは、学習用問題の正答に対応するのは“ライダー”であり、学習者の入力した誤答は“犬”を想定している。一方、この手法に於いて、学習対象となりうる語が物体のみであり、動きや状態を対象とすることは困難である。

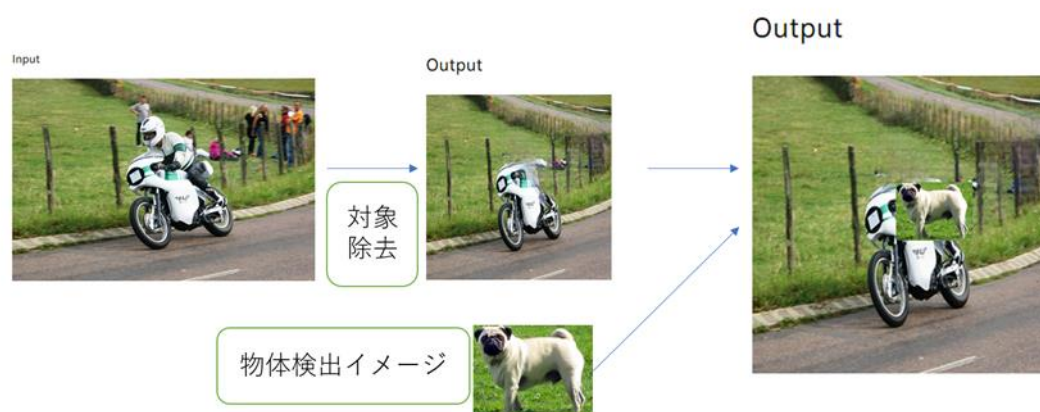


図 3-8 Edge Connect による対象物体の消去、および画像合成のイメージ [20]

本研究では、学習者の誤答を含む文章全体から画像を生成させることにより、対象を物体に限定しないことを目指す。文全体から画像を生成するモデルでは、現在様々なモデルが存在する。OpenAI により開発され DALL-E [21]は大規模言語モデルである GPT-3 をベースに構築された生成モデルであり、Boris らは DALL-E を踏襲し、GPT-3 より規模の小さい BERT をモデルとして適用した DALL-E mini [22]を提案した。DALL-E mini は DALL-E と比較し、軽量である一方、生成される画像の精度が低い問題が生じる。本研究では、DALL-E の後継であり、より多くのデータセットより学習された大規模なモデルである DALL-E2 を適用する。DALL-E2 を API としてシステムに実装することにより、学習者の誤答に応じた画像生成を実現させる語彙学習支援システムを構築した。図 3-9 では、DALL-E2 が入力文に応じた画像生成が可能であることを示している。上段の入力文“the women is (taking a photo of) the white bear next to the river”に対して、女性が白熊を川沿いで撮影している様子が再現できている。一方括弧内の語を変化させた下段の文に対する生成画像では、変更点“dancing with”を表現した画像生成が可能となっており、学習者の誤答に応じた画像生成に適していると捉え、提案手法のモデルとして選定した。

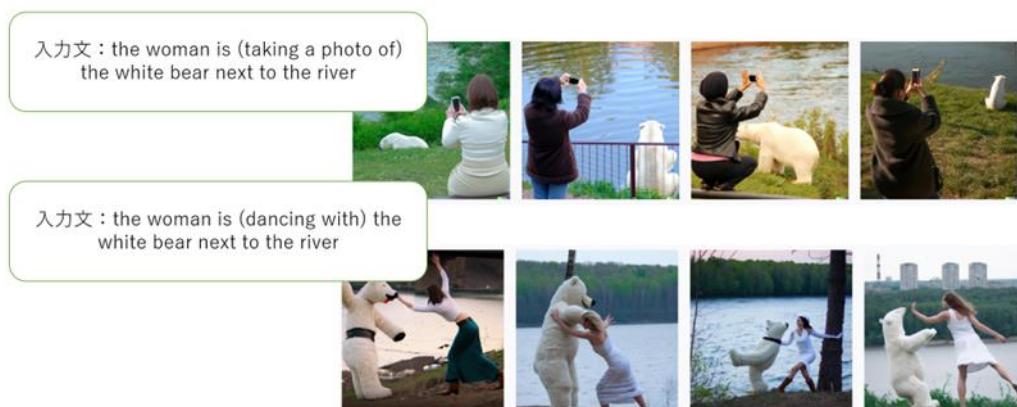


図 3-9 DALL-E による画像生成例 [21]

第4章 プロトタイプの実装

4.1 概要

本研究では、英語語彙学習に於ける解答選択式の空所補充問題を基本とし、学習者の誤答時に選択した誤答を含む文に対応した画像生成を行う。誤答のイメージでの可視化が学習者に効果的に印象を与え学習者の誤答への内省を促し、一度誤った問題に対して繰り返しの誤りを防ぐことを目的とする。第3章では、本研究のシステムの構成要素について言及したが、本章では、それら構成要素に含む課題に言及し、被験者実験用のプロトタイプの実装について述べる。

4.2 開発環境

プログラムの開発環境には、コードエディタである、Visual Studio Code に、Docker コンテナを用いた開発環境を運用可能な拡張機能である Remote-Containers を利用した。今回の実装に於いて用いた API が構築環境に依存した場合に於いても、他 PC に於ける環境の転用を可能とした。又、後述する画像生成済み学習用問題集を定義し API による呼び出しを排除しシンプルなコードで動く「システムを Google colab 上で構築することにより、どの PC からでも Google へのログインのみでシステムを稼働することを実現した。

4.3 システムの3つの構成要素に於ける課題

第3章に於いて、紹介したシステムでは、学習者が学習を行う際に、学習者の難易度に応じて自動で学習用コンテンツが生成されることを想定している。一方で、システムの構成要素の一つである質問生成タスクでは、学習者の対象語に対して、使用データ 21,463 キャプションに対して探索を行い、合致したものを空所とする為、普段あまり使われない難易度が高い語彙に関しては、データの増大に伴い検索時間が増大する可能性がある。提案手法で定義した COCO Word list に於ける、B2 レベルの語彙に関しては構築した環境に於いて、一つの問題生成に於いて 2 分程度時間を要するケースが存在した。一方で探索に時間がかからない比較的良好に見受けられる語に於いても品詞同定などのプロセスを要する為、20 秒程度は要する。学習時に適宜問題を生成する上では語彙学習を行う上で問題がある。一方、それ以外のタスクである回答支援タスクは計算コストを要さないが、画像生成タスクに於ける DALL-E2 による生成では、API を利用

し 5 秒程度の生成時間を要する。これらの問題に対し、プロトタイプの実装では、事前に画像および質問が含まれるデータセットを画像生成済み学習用問題集として事前に構築することで読み込み時間の問題を解消した。以下にデータセット構築の流れについて述べる。

画像生成済み学習用問題集の構築の流れは、提案手法で述べた COCO Word list の語彙を対象に、空所付き問題生成タスクにより空所付き問題を自動生成し、入力支援タスクである解答候補を付与することにより図 4-1 で示す学習用問題集を構築する。図 4-2 では、実際に生成された空所付き問題のデータのうちのひとつを抜粋したものである(1)では学習の初期画面に表示される MS-COCO データセットの画像の ID を示している。(2)は、先述した空所付き問題生成タスクにより自動で生成された空所付き問題を示している。(3)は、入力支援タスクに於ける Levenshtein 距離を活用した近似したスペルの抽出によって得られた解答候補となる単語を示している。ここでは、三択の選択肢に対して、3つの候補となる単語が表示されているが、これは後述する被験者実験に於いて、全て誤った回答から構築される問題の為のものである。

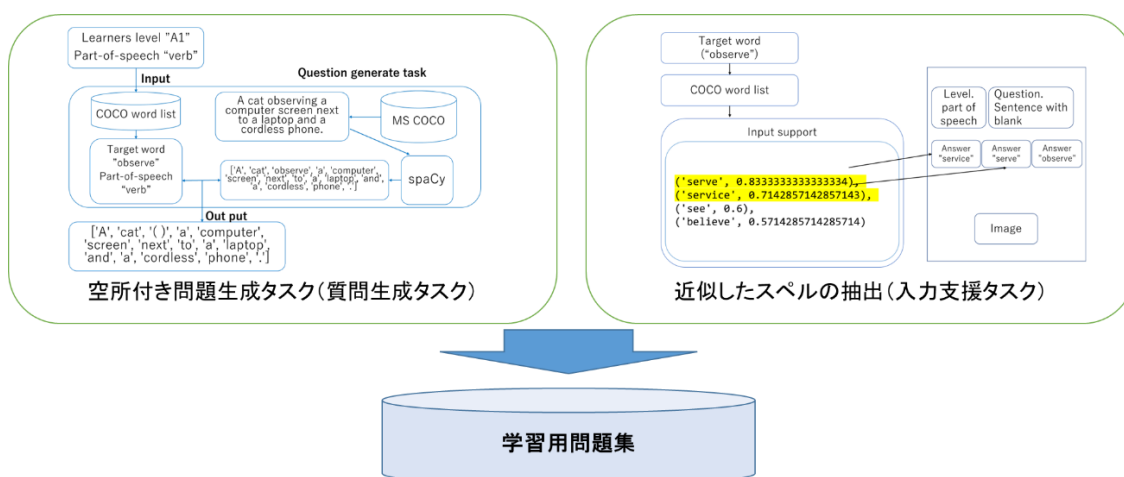


図 4-1 学習用問題集構築フロー

学習用問題集

- (1) 'image_id': '000000279887', 'id': 812973, 'caption': 'Many motorcycles are displayed in the store with descriptive signs posted near them. ',
(2) 'blankquestion': ['Many', 'motorcycles', 'are', 'displayed', 'in', 'the', 'store', 'with', '()', 'signs', 'posted', 'near', 'them', '.'], 'answer': 'descriptive'
(3) 'candidate': ['decorative', 'festive', 'primitive'],

図 4-2 学習用問題集データ

これら、学習用問題集に対して、学習者の誤答に応じて生成される画像情報を付与したものが、図 4-3 である。図 4-3、(a)では、学習者の解答候補を当該空所付き問題に補充した際によって得られる分を画像生成モデルである DALL-E2 に入力した際に得られる画像情報をそれぞれの選択した候補単語に対して付与した。(b)は先述した全て誤答によって構成される問題か否かの判定を“fake”:1 or 0 とすることで、条件分岐に依る誤答のみで構成された問題の提示を実現した。

```
{'image_id': '000000279887', 'id': 812973, 'caption': 'Many motorcycles are displayed in the store with descriptive signs posted near them. ', 'blankquestion': ['Many', 'motorcycles', 'are', 'displayed', 'in', 'the', 'store', 'with', '()', 'signs', 'posted', 'near', 'them', '.'], 'answer': 'descriptive', 'candidate': ['decorative', 'festive', 'primitive'],  
(a) 'learnerans1': 'decorative', 'genimage1': 'experimentimageqb2cocoadjective/812973_select_decorative.png',  
'learnerans2': 'festive', 'genimage2': 'experimentimageqb2cocoadjective/812973_select_festive.png',  
'learnerans3': 'primitive', 'genimage3': 'experimentimageqb2cocoadjective/812973_select_primitive.png',  
(b) 'fake': 0, 'wordlist_a_listid': 0, 'qid': 6, 'pos': 'adjective'}
```

図 4-3 画像生成済み学習用問題集データ

又、図 4-4 では、学習者の学習に応じ記録されるラーニングログを示している。Ⓐで示す“wordlist_a_listid”は、Word list ごとに振られた識別番号であり、被験者の問題がランダムで提示された場合に於いてもラーニングログより当該デー

タにアクセスすることを可能にした。⑤の“select_ans”は学習者が学習時に選択した回答を示している。このデータの参照により学習者が繰り返し誤りを行ったか否かの判定を可能にした。⑥ラーニングログは、学習者の解答ごとに蓄積され、システムエラーに依るデータの損失を防いでいる。又、ラーニングログよりシステムエラーにより中断された問題を同定し途中の問題から実験を直ぐに再開できる仕様とした。

```
•{'user_id': 8, 'qid': a, 'wordlist_a_listid': 45, 'initial_image':  
'000000093261', 'question': 'A young man on the beach () a  
kite', b, 'select_ans': 'resemble', 'pos': 'verb', 'trial_num': 1,  
'fake': c, 'condition': '2'}
```

図 4-4 ラーニングログ

図 4-5 は、上述した画像生成済み学習用問題集データを用いることにより実現した被験者実験用のプロトタイプユースケース図を表す。第3章に於いて述べたシステムとの主要な違いは、学習用問題集を構築することにより、質問生成、回答支援タスクに於ける解答候補の選定、画像生成の時間が解消し、軽量なアプリケーションの構築を実現させたことである。

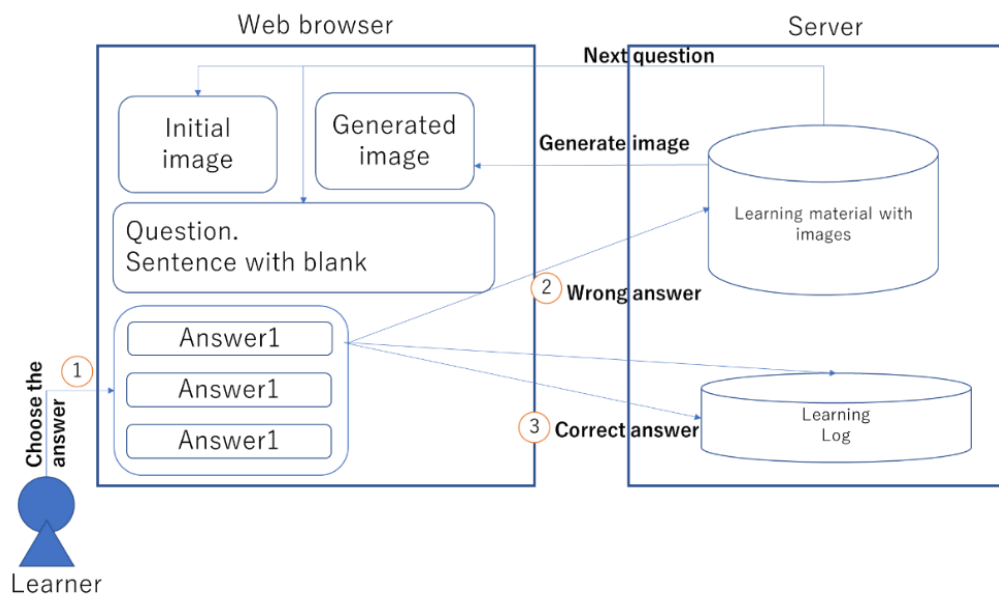


図 4-5 実験用プロトタイプユースケース図

4.4 システムの機能

提案システムに於ける操作を図 4-6,4-7 に示す. 図 4-6 の例では, 学習者が空所に誤答を選択した際に, 学習者の誤答に応じた画像が生成されていることを示している. 図 4-7 では学習者が正答した場合を示し, 誤答に対応する画像は表示されない.

LVIG_Learning Vocabulary with Image Generation



Question No.2

Two sheep and a () stand next to a fence in the yard

Question No.2

roam
 tram
 ram

Go / Next

LVIG_Learning Vocabulary with Image Generation



不正解 Wrong answer

Two sheep and a (tram) stand next to a fence in the yard

Question No.2

Two sheep and a () stand next to a fence in the yard

Question No.2

roam
 tram
 ram

Go / Next

図 4-6 システム操作例 1 (誤答を選択した場合)

LVIG_Learning Vocabulary with Image Generation



Question No.3

Motorcyclist on chromed motorcycle () a curve roadway .

Question No.3

ring
 rounding
 ruin

Go / Next

__Do not push,maintain__

LVIG_Learning Vocabulary with Image Generation



正解です Correct answer

正解

Motorcyclist on chromed motorcycle (rounding) a curve roadway .

Question No.3

Motorcyclist on chromed motorcycle () a curve roadway .

Question No.3

ring
 rounding
 ruin

Go / Next

図 4-7 システム操作例 2 (正答を選択した場合)

第5章 ケーススタディ

5.1 目的

本章では、被験者に対して提案した英単語学習方法がどのような影響を与えるか、被験者実験を通して調査を行った。英語語彙学習に於いて、誤答の画像による可視化により得られる、印象、示唆性が学習者の内省を促し、繰り返しの誤答を防ぐ事に繋がるか、効果的に記憶に定着するかについて評価を行った。ここでは、画像生成ありのモデル、画像生成なしのモデルを基に、対照実験による検証を行った。学習者が両モデルに依る学習を2回繰り返し、その前後に於いて繰り返しの誤りが回避できたかどうかを調査し、3日後に誤答をした問題に対するアンケートを行う事により、提案手法に於ける誤答に対する効果的な印象付け、イメージに依る理解並びに記憶の定着に寄与したか否か調査を行った。

5.2 実験方法

被験者の対象者は、大学院に所属する英語を第二言語とする L2 学習者を対象とした。国籍は、インド、インドネシア、タイ、中国、ミャンマー、日本、バングラデシュ、ベトナムの8か国から成り、英語習熟度は、TOEIC（500点～900点程度）を有する中から上級の英語力を有する25名の被験者を対象に行った。尚、システムエラー等に依るデータ欠損を除いた22名のデータを分析の対象とした。被験者が有する英語習熟度の内訳を図5-1に記す。

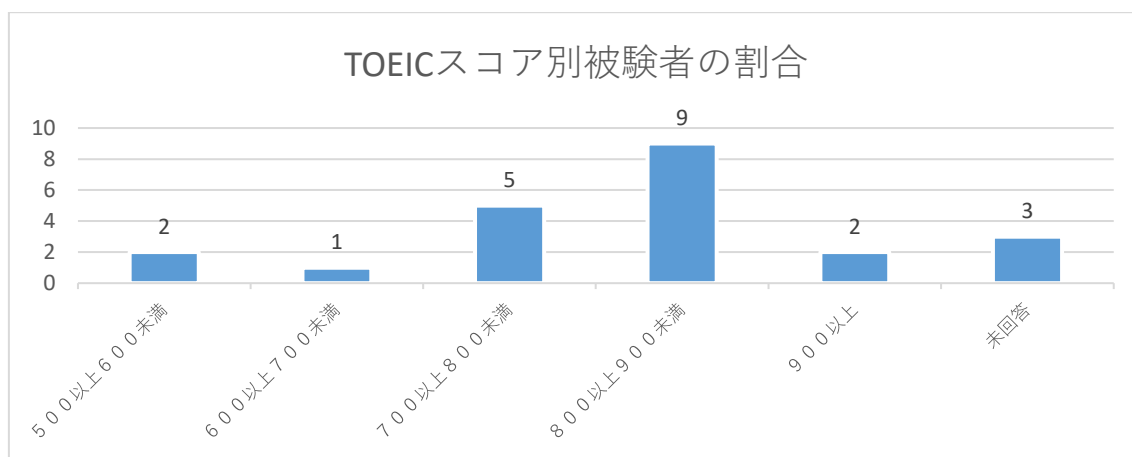


図 5-1 実験参加者の英語習熟度別内訳

5.2.1 実験条件

被験者は、パソコンのウェブブラウザ上で操作を行った。①誤答時の画像生成が含まれる Condition1, ②誤答時の画像生成が含まれない Condition2 の 2 種類の条件を設定した。学習対象語は、合計 100 個の単語からなる語彙リストを Word list A, Word list B に分割した。図 5-2 は、Word list A, Word list B の語彙リストの品詞に於ける内訳である。それぞれのワードリストは、合計 100 語から成るワードリストよりランダムかつ、各品詞の割合が一定になるように分割されている。これらワードリストは、提案手法により定義した COCO-Word list より選定されたもので、対象学習者を大学院レベルの高水準と定めている為、一部単語数が少ない品詞を除き、B2 レベルのものから中心に選定した。又、学習対象者が繰り返しの誤答を避けられたか判定する為、意図的に 20%程度すべての選択肢が誤りである問題を混在させた。高い習熟度の学習者が自らが正当と思っていた回答を繰り返し誤るプロセスは、Fossilization によって自然言語化した言語を修正することが困難である過程を疑似的に再現したものである。

POS	number
Adjective	10
Adverb	10
Conjunction	4
Determiner	6
Noun	30
Preposition	10
Pronoun	10
Verb	20

POS	number
Adjective	5
Adverb	5
Conjunction	2
Determiner	3
Noun	15
Preposition	5
Pronoun	5
Verb	10

Word list A

POS	number
Adjective	5
Adverb	5
Conjunction	2
Determiner	3
Noun	15
Preposition	5
Pronoun	5
Verb	10

Word list B

図 5-2 学習対象語品詞別内訳

学習順序に依る効果を低減する為、被験者ごとにそれぞれの語彙リストをランダムで出題した。それぞれの条件に於ける順序効果、学習対象語に於ける順序効果を解消するカウンターバランスを取るため、被験者は4つのグループのいずれかに分類された。24名の被験者は、各グループ6名ずつ分類された。ここでは、これらの順序は、図5-3に示す順序となる。

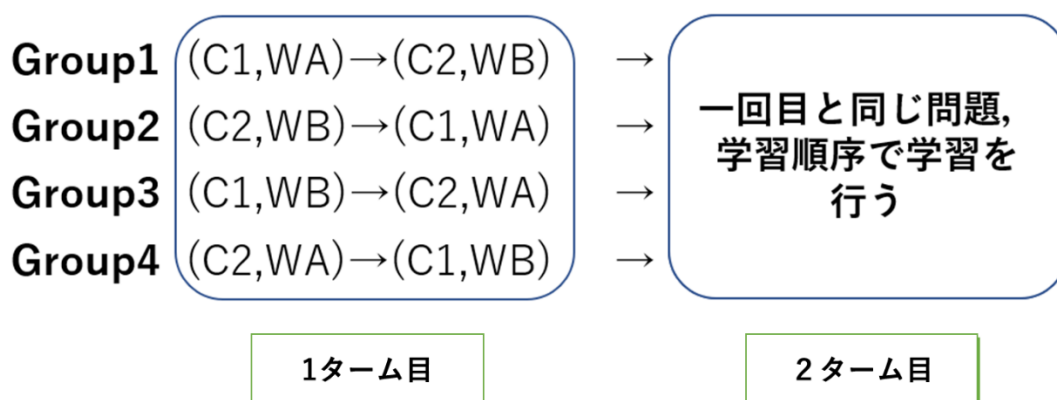


図 5-3 被験者実験に於ける条件, 順序

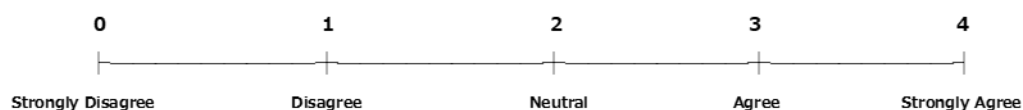
例えば、Group1 に於ける被験者は、初めに画像生成が含まれる Condition1 で、Word list A を学習し、Condition2 で、Word list B を学習する。以上を1タームとし、2ターム目では、1ターム目の問題と同じ問題を、同じ学習順序で学習を行う。図5-4に示す被験者によるリッカートスケールを用いたアンケートにより、1. MSCOCO データセットに含まれる画像に対する想起性、2. 生成画像に対する想起性、3. 生成画像から受けた印象の3点について情報を収集した。アンケートの内容はそれぞれ、

1. 初めに提示される画像は、文の内容を連想させるものか。
2. 誤答時の生成画像は、文の内容を連想させるか。
3. 誤答時の生成画像は、印象に残るか。

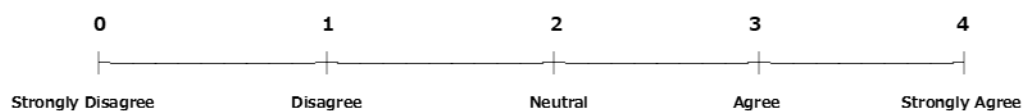
である。Condition1 に於ける学習で、かつ学習者が誤答を行い、画像が生成された場合に、1, 2, 3 の設問に対して回答を行う。Condition1 に於いて正答

した場合、画像生成が含まれない Condition2 の時は、1 の設問のみに対して回答を行う。当該アンケートについては付録に掲載する。被験者はアンケート用紙に問題解答ごとの記入を行う。アンケートは、Condition1, Condition2 でそれぞれ個別に用意した。

No1. The image displayed on the initial screen suggests the content of the sentence



No2. The images generated by wrong answers suggest the content of the sentence



No3. The images generated by wrong answers are impressive.

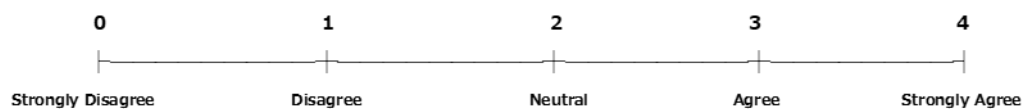


図 5-4 被験者によるリッカートスケールを用いたアンケート

5.2.2 実験手順

被験者に依る実験を以下の(1)～(5)の手順で行った。【】内の時間は所要時間を表す。

(1). 実験の流れとシステム操作説明【10分】

被験者に実験の目的等を含めた実験の全容の説明を行い、実際のデモを見せることで、システムの操作に慣れる時間を用意した。

(2). 学習1ターム目【50分】

(3). 各グループに応じ、それぞれの Condition, Word list の組み合わせから実験を開始する。1タームには50語の問題が含まれており、それぞれの Condition の合計100語に対し、回答、問題ごとのアンケートの記入を行う。1ターム50分を想定する。

(4). 休憩【5分】

(5). 学習2ターム目

被験者は、ターム1で学習した内容と同じ内容の学習を再び行う。

(6). 3日後アンケート

学習者が学習システムに於いて誤答をした単語について、確認のアンケートを行う。ここでは、condition1, 2それぞれに対して誤答した語を対象とし、当該単語について、既知のものであったか、若しくは知らなかったがイメージとして連想できるか、全く連想できないかの選択式のアンケートを行った。学習時の誤答に依る画像生成が、当該単語に於けるつながりとして記憶に定着したかどうかを判断するためのものである。図5-5は実験の実施環境を示す。実験環境は、図外部モニターに接続しスクロールが少なく済むよう配慮した。

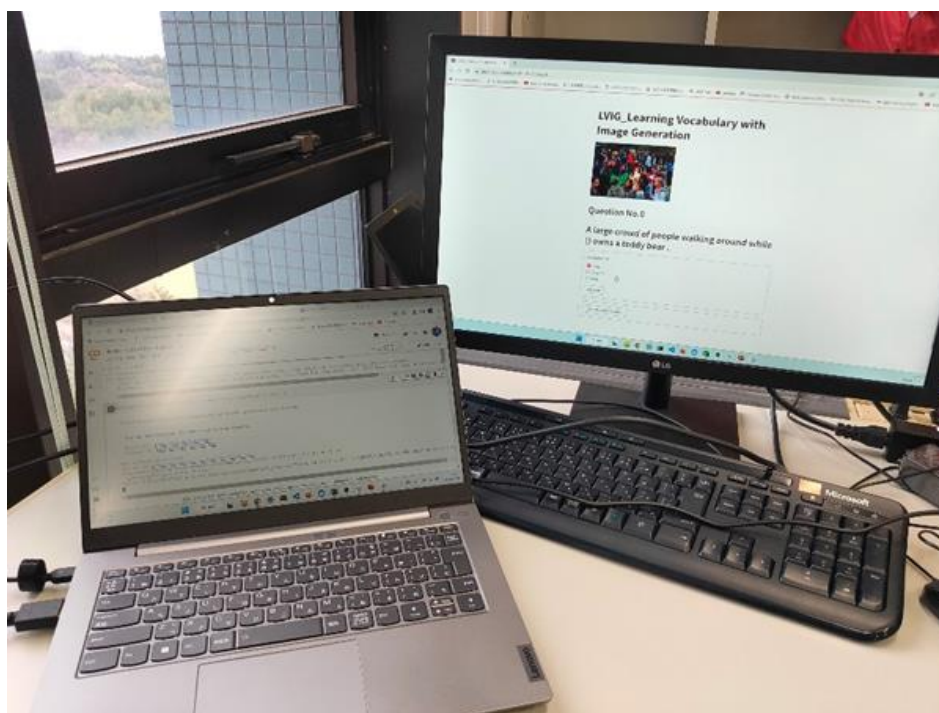


図 5-5 実験実施環境

5.3 実験の結果

本実験は、英語語彙学習に於ける学習効果、モデルの精度としての評価の二つの軸に依る評価を行う。

5.4 学習効果に対する評価

5.4.1 繰り返しの誤りの回避に対する評価

提案手法である画像生成有のモデルによる Condition1, 従来手法である画像生成が無いモデルによる Condition2 の両者に於ける学習効果を評価する。提案手法である画像生成ありのモデルが、従来の手法と比較し、繰り返しの誤りを防ぐ事ができたかどうか評価を行う。ここでは、目的変数に対応する要因が一つである、1 要因 2 水準の単変量解析となる。又、同一被験者が 2 つの水準を行う被験者内計画により行う。ここでは、Condition1 と Condition2 のペアが確定している“対応あり”となる。今回は、研究の目的である繰り返しの誤答を避けるという目的の元、繰り返し誤答が行われるかについて、1 ターム目で誤った語彙に対

し 2 ターム目で繰り返しの誤りを避けられたか否かで評価を行う。表 5-1 では、それぞれの被験者に於ける、Condition1、Condition2 でのそれぞれの繰り返しの誤答の頻度を表す。

表 5-1 各 Condition に於ける被験者の繰り返し誤り回数

被験者ID	Condition1	Condition2
3	2	3
4	4	23
5	3	14
6	5	20
7	6	2
9	3	4
10	1	1
11	18	12
12	1	7
13	12	8
14	0	4
15	1	8
16	1	1
17	13	16
18	1	8
19	11	10
20	14	14
21	4	7
22	1	8
23	14	16
24	17	2
25	1	2

検定手法として、図 5-6 の度数分布表より、データの正規性が認められない為、連続変数を扱うノンパラメトリックな手法として適した、Wilcoxon の符号順位検定を適用した。提案手法では 22 人の対象に対して、Condition1、Condition2 に対する観察を行い、2 群の差の検定を行った。Wilcoxon の符号順

位検定では、 $(x_{11}, x_{21}), (x_{21}, x_{22}) \dots (x_{1n}, x_{2n})$ としたとき、 $D_i = |x_{1i} - x_{2i}|$ とする 22 個の D_i に対して、小さい順に順位をつける。表 5-1 では、被験者 3 と被験者 9 などのように観測値の中に同一の値があり、この場合それぞれに対し平均順位を割り当てる。又、被験者 10, 16, 20 のように、2 要因間での差がない場合は、サンプルから除く。ここでの検定統計量は、 D_i が正であるものの順位の合計 W^+ 、 D_i が負であるものの順位の合計 W^- となる検定統計量 W^+ の値が、棄却限界値より大きいとき、若しくは、 W^- の値が棄却限界値よりも小さいとき、帰無仮説を棄却する。ここでは、提案手法に於ける学習システムと従来の学習システムに於ける違いが認められないとする、帰無仮説 $H_0: \theta = 0$ について検定を行った。対立仮説は、片側検定に於ける $H_1: \theta < 0$ もしくは、 $H_1: \theta > 0$ となる。ここでは、Wilcoxon の符号順位検定では、 n が 10 より大きい場合に、 W の平均を $E[W] = n(n+1)/4$ 、 W の分散を $V[T] = n(n+1)(2n+1)/24$ と表し、正規分布に近似的に従うことが知られている。これらを利用し検定を行った。表 5-2 は Wilcoxon の符号順位検定で必要となる値を表す。下部に記された、大、小の値は、正負それぞれの符号を有する数値の合算値となり、先述した W^+ と W^- がそれぞれ対応する。ここでは、小さい数値を W と選定する為、 $W = 45$ となる。サンプルサイズは、表 5-2 黄色に示したように差が 0 つまり中心点と完全に一致する点数は除外する為、 $n = 19$ となる。表 5-2 により、導出した Z 値は、 -2.0121 となり、5 パーセント有意水準に於いて、 Z 値が -1.654 以下、 W^- の値が棄却限界値よりも小さい為、帰無仮説を棄却する。又、Wilcoxon の符号順位検定に於いて n が小さいときに用いられるサインランク検定表を用いた検定に於いては、対象数 $n = 19$ 、 $\alpha 0.05$ の時の値 53 を W と比較する。 $W < 53$ となり、帰無仮説を棄却する。提案手法である画像生成モデルを適用したモデルが、繰り返しの誤りの回避に効果的であったことを示した。

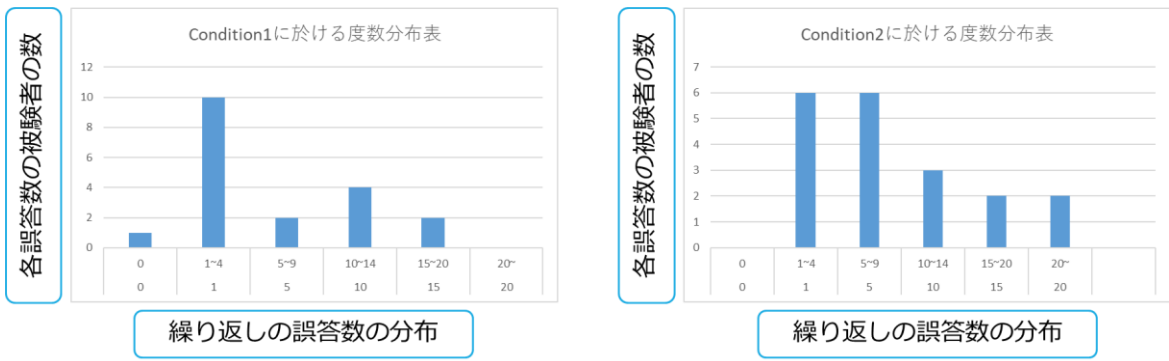


図 5-6 Condition1,2 に於ける繰り返しの誤答数に対する被験者数

表 5-2 被験者の繰り返しの誤りの回数に於ける検定表

被験者ID	Condition1	Condition2	差 (Con2-Con1)	絶対値	順位	符号	+	-
3	2	3	1	1	2	2	2	
4	4	23	19	19	18	18	18	
5	3	14	11	11	15	15	15	
6	5	20	15	15	16.5	16.5	16.5	
7	6	2	-4	4	8	-8		-8
9	3	4	1	1	2	2	2	
10	1	1	0					
11	18	12	-6	6	10.5	-10.5		-10.5
12	1	7	6	6	10.5	10.5	10.5	
13	12	8	-4	4	8	-8		-8
14	0	4	4	4	8	8	8	
15	1	8	7	7	13	13	13	
16	1	1	0					
17	13	16	3	3	5.5	5.5	5.5	
18	1	8	7	7	13	13	13	
19	11	10	-1	1	2	-2		-2
20	14	14	0					
21	4	7	3	3	5.5	5.5	5.5	
22	1	8	7	7	13	13	13	
23	14	16	2	2	4	4	4	
24	17	2	-15	15	16.5	-16.5		-16.5
25	1	2	1	1	2	23	23	
合計							149	-45
大								149
小								45
統計量Z								-2.0121

5.4.2 定性評価による生成画像の特性と繰り返しの誤りの回避に対する評価

これまでの説明により、画像生成モデルを適用したモデルが繰り返しの誤答を避けることができたことについて明らかにしたが、生成された画像のうちどのような画像が学習効果に寄与したかについて、被験者の定性評価であるリッカートスケールを用いて収集したデータを基に分析を行った。ここでは、繰り返し誤答を行った語彙、繰り返しの誤答を避けられた語彙に対して得られた情報を基に、生成された画像が文を連想させるものかどうかの指標とする画像の想起性、印象、それぞれの学習効果に対する影響を調査する。ここでは、被験者 22 名が繰り返し行った誤答 72 語、繰り返しの誤答が避けられた 129 語を対象とする。調査のために、得られたアンケートをラーニングログにタグ付けを行い、誤答回数、想起性スコアの平均、印象スコアの平均をそれらの単語に対しタグ付けを行った、ノンパラメトリックな指標である Spearman の順位相関係数による分析を行った。

初めに、画像に於ける想起性と繰り返しの誤答についての調査を行う。図 5-7 図では、y 軸を画像の想起性、x 軸を誤答の繰り返し回数を表している。右記の数値は、想起性と繰り返しの誤答数に於ける相関係数を表す。ここでは、繰り返しの誤答の数と想起性に於いて、相関は得られなかった。

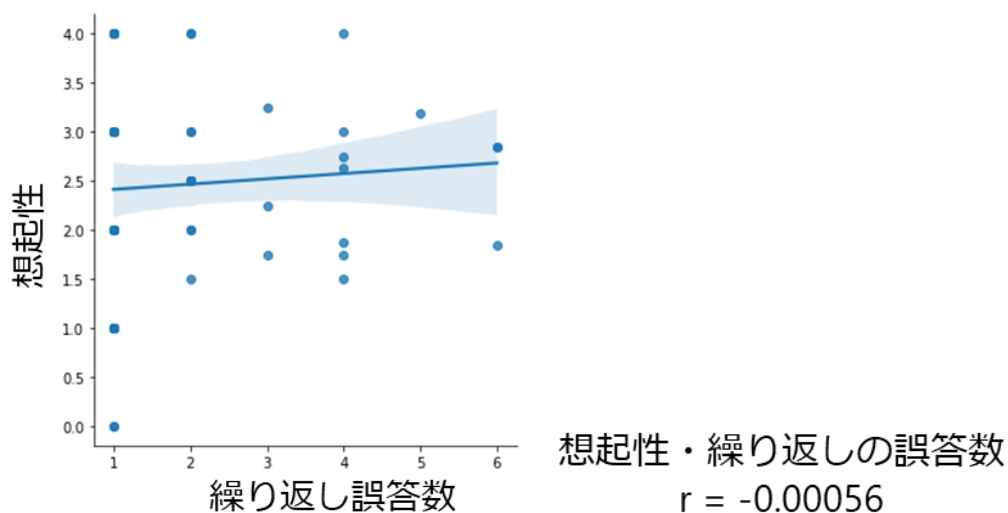


図 5-7 画像の想起性と繰り返しの誤答が発生した語彙の相関関係

一方、繰り返しの誤答数と印象に対する相関関係は、図 5-8 のとおり、弱い負の相関が得られた。これらの両者の比較では、生成画像の想起性より、印象の度合いがより繰り返しの誤答に影響を与えていることが伺える。生成画像による繰り返しの誤答への抑止は画像に含まれる印象によって得られることが伺える。

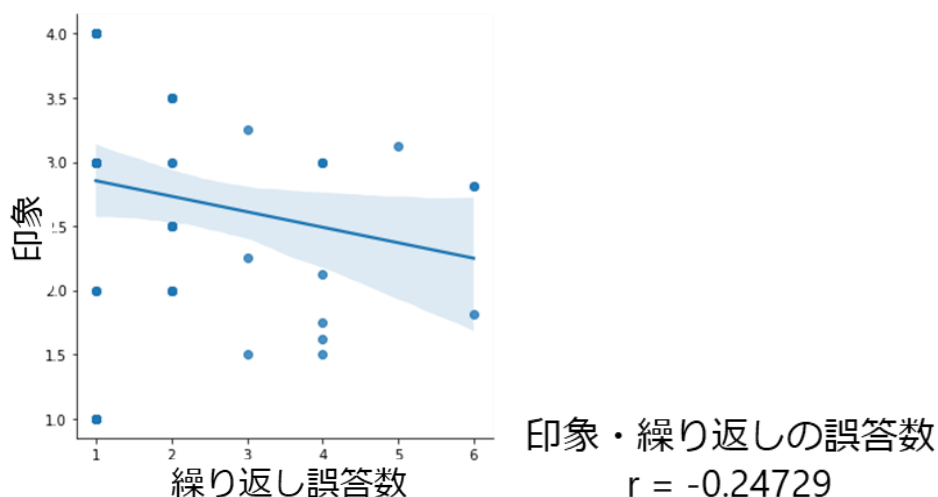
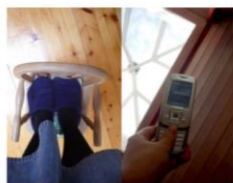


図 5-8 画像の印象と繰り返しの誤答が発生した語彙の相関関係

これらの結果より、画像生成に依る誤答の示唆性を活用した語彙学習に於いて、画像が有する想起性より、画像が持つ印象が繰り返しの誤答の回避に有効であることが示された。

次に、特に顕著に誤答が繰り返された画像を一部抜粋し分析を行う。ここでは、繰り返された誤答“tool”, “someone”に対して、例示する。図 5-9 では、“a person lifting a little() with their feet and a person holding a cell phone”という文の空所補充に対して、“tool”を選択したことによって画像生成が行われた例を示している。ここでの生成画像による想起性の評価の平均は、2 で印象は 2 であった。一方で、初期画像の想起性は 3 であり、ほかの選択肢である“school”, “pool”と比較し初期画面の妥当性の観点からも tool を再び選んでしまったものと思われる。

LVIG_Learning Vocabulary with Image Generation



Question No.30

a person lifting a little () with their feet and a person holding a cell phone

Question No.30

tool

school

pool

Go / Next

LVIG_Learning Vocabulary with Image Generation



不正解 Wrong answer

a person lifting a little (tool) with their feet and a person holding a cell phone

図 5-9 繰り返し誤答が発生した語彙に対する生成画像例①

一方で，“someone”に対して生成された画像を見ると，図 5-11 では，初期画面の想起性は，2.6，生成画像の想起性は2であり，印象も2であった。

LVIG_Learning Vocabulary with Image Generation



Question No.31

A large crowd of people walking around while () owns a teddy bear .

Question No.31

someone

own

anyone

Go / Next

LVIG_Learning Vocabulary with Image Generation



不正解 Wrong answer

A large crowd of people walking around while (someone) owns a teddy bear .

図 5-10 繰り返し誤答が発生した語彙に対する生成画像例②

これらの単語に共通して言えるのは、普段使い慣れた簡易な単語である点である。被験者による実験の結果から、高い習熟度の学生に於いても、繰り返しの誤答が生じることが伺える。提案手法では、意図的にすべてが誤答となる設問が20%含まれており、英語の習熟度が高いからこそ、自らの回答に過信してしまい、間違いを見過ごしてしまう可能性が伺える。一部の英語の高い習熟度を有する被験者には、従来手法に於いて繰り返しの誤り回数が顕著であった。これは実験の設定で設問に解くスピードが求められることから自らの間違いに対する（今回の問題設定では実際には間違えていない可能性を秘める）内省の機会を得られなかったと捉えられる。一方提案手法では、それらの傾向を有する学習者の繰り返しの誤答が防げられていたことが伺え、これは、誤答に対する内省の機会を十分に与えられたことを示した。Fossilization では学習者が自らの誤りに気が付かずに、繰り返し使うことにより自然言語化してしまう。それ故に内省の機会を得られないことが、一度自然言語化した語彙に対する誤りの是正を困難にしている。一方、提案手法では、疑似的に学習者に Fossilization を体験させ、習得済みの言語に対する内省の効果がみられた。

5.4.3 遅延テストに於ける誤った語の理解度に対する評価

遅延テストでは、先述した繰り返しの誤りの回避に対する評価同様、提案手法である画像生成を有する Condition1、従来手法である画像生成を含まない Condition2 の両者に於ける、誤った語の理解度に対する評価を行う。ここでの条件は、先述した評価同様、目的変数に対応する要因が一つである、1 要因 2 水準の単変量解析となる。又、Condition1 と Condition2 のペアが確定している“対応あり”となる。今回は、被験者がそれぞれのコンディションに於ける誤った語についての理解度を問うアンケートにより評価を行う。アンケートの内容は図 5-11 に示す通り、被験者ごとに誤った語を提示し、それらの語彙に対して、“既に知っている”、“想像できる”、“想像できない”のいずれかを回答する方式で行った。ここでは、誤った語はどちらのコンディションで回答したものか被験者から判断ができないようリストの作成を行った。

Thank you very much for participating in my experiment the other day.

I would appreciate your cooperation in the following tests.

Please use the following example as a guide.

If you already know the word, write "1" in the "**already know**" field. If you did not know the word before, but can imagine it, please enter "1" in the "**imaginable**" field. If you do not remember it at all, please enter "1" in the "**unimaginable**" field.

Q.No	word	already know	imaginable	unimaginable
Example	like	1		
1	decorative		1	
2	messy	1		
3	ordinary	1		
4	condition	1		
5	beside	1		
6	so	1		
7	living	1		
8	square	1		
9	while	1		
10	the	1		
11	those	1		
12	swing	1		
13	like	1		

図 5-11 遅延テストのアンケートの一部抜粋

被験者に於ける各コンディションの理解度の評価として、“想像できる”と回答した語彙数から“想像できない”と回答した語彙数の差を誤った語の理解度の評価値とした。表 5-3 では、それぞれの被験者に於ける、Condition1, Condition2 それぞれについての誤った語の理解度の評価値を表す。

表 5-3 各 Condition に於ける誤った語の理解度の評価値

被験者ID	Condition 1	Condition 2
3	3	-1
4	0	0
5	3	4
6	3	-1
7	0	0
9	2	-1
10	2	-3
11	0	1
12	1	0
13	-1	0
14	0	1
15	1	4
16	0	-1
17	1	0
18	2	2
19	2	-1
20	0	0
21	0	0
22	0	-1
23	0	-1
24	-1	0
25	1	0

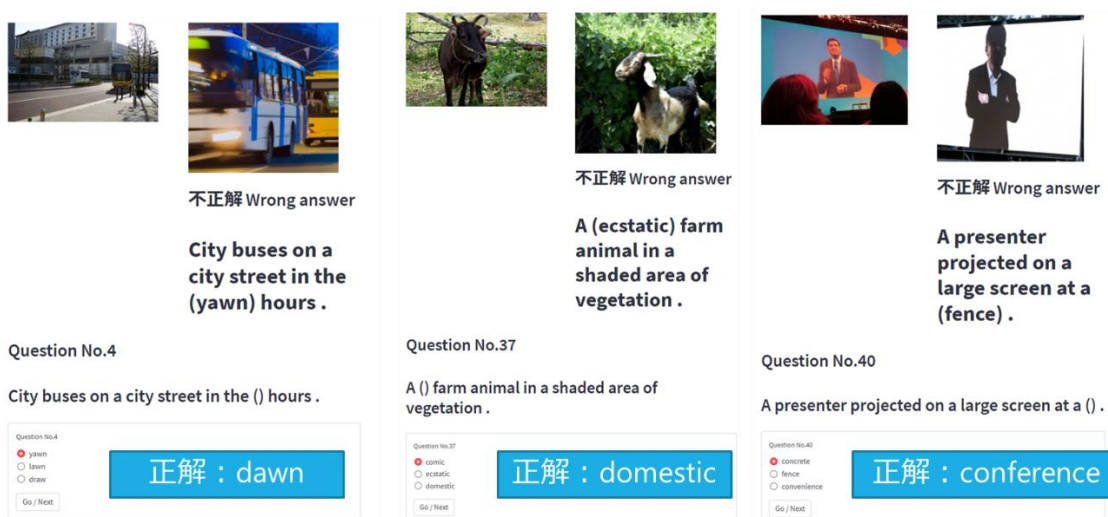
ここでは、先述した繰り返しの誤りの回避に対する評価手法と一部共通する点がある為、一部説明を割愛する。検定手法としてデータの正規性が認められない為、連続変数を扱うノンパラメトリックな手法として適した、Wilcoxon の符号順位検定を適用した。22 人の対象に対して、Condition1, Condition2 に於ける、誤った語の理解度の評価値を比較し、2 群の差の検定を行った。表 5-4 黄色に示したように差が 0 である点数を除外する為、対象数 $n=17$ となる。 n が 10 以上より大きい為、正規分布に近似的に従わせる検定手法をとる。ここでは、正負それぞれの符号が有する値の合算値に於いて、小さい値である $W=49$ を採用する。導出した Z 値は -1.3018 となり、5 パーセント有意水準に於いて、 Z 値が

-1.654 以上、 W^- の値が棄却限界値よりも大きい為、帰無仮説を採択する。又、Wilcoxon の符号順位検定に於いて n が小さいときに用いられるサインランク検定表を用いた検定に於いては、対象数 $n=17$ 、 $\alpha 0.05$ 、の時の値 41 を W と比較する。 $W > 41$ となり、帰無仮説を採択する。検定の結果、Condition1, Condition2 に於いて、誤答した語彙の理解度について、有意な差は見られなかった。繰り返しの誤りの回避に対する評価の結果、今回の結果より誤答への効果的な印象を与え、繰り返しの誤答を防げたが、誤った語を理解するという点に於いては、不十分である可能性を有していることが伺える。

表 5-4 誤った語の理解度の評価値に於ける検定表

被験者ID	Condition 1	Condition 2	差	絶対値	順位	符号	+	-
3	3	-1	4	4	15.5	16.5	16.5	
4	0	0	0					
5	3	4	-1	1	6	-7		-7
6	3	-1	4	4	15.5	16.5	16.5	
7	0	0	0					
9	2	-1	3	3	13	14	14	
10	2	-3	5	5	17	18	18	
11	0	1	-1	1	6	-7		-7
12	1	0	1	1	6	7	7	
13	-1	0	-1	1	6	-7		-7
14	0	1	-1	1	6	-7		-7
15	1	4	-3	3	13	-14		-14
16	0	-1	1	1	6	7	7	
17	1	0	1	1	6	7	7	
18	2	2	0					
19	2	-1	3	3	13	14	14	
20	0	0	0					
21	0	0	0					
22	0	-1	1	1	6	7	7	
23	0	-1	1	1	6	7	7	
24	-1	0	-1	1	6	-7		-7
25	1	0	1	1	6	7	7	
合計							121	-49
大								121
小								49
統計量Z								-1.3018

次に、誤った語に対する被験者の回答で、“想像できなかった”という回答が 2 回以上選ばれた語彙について生成画像より分析を行う。ここでは、“yawn”, “ecstatic”, “fence”を例示する。図 5-12 では、それら誤った語を入力とした際に生成される画像、及び問題を示している。青色で記された語彙は、その時の正答を示す。一部回答候補に回答がないものは、その問題が 5.2.1 実験条件で言及した間違いのみを回答候補とする問題を表している。図 5-12 下部の値については、それらの語の誤りの“繰り返し回数”, “想起性”, “印象”の値を示す。ここでは、これらの生成画像に於いて、画像とセンテンスの対応を表す, “想起性”に於いて, “どちらももいえない”である 2 以下の低い数値を示した。ここでの“想像できなかった”とされる画像に共通している点として, 図 5-12 より空所として補充した語彙が画像に反映されない特徴消失が発生していることが伺える。



[対象語彙：繰り返し回数, 想起性, 印象]

'21_yawn': 2, 2.0, 3.0

'50_ecstatic': 0, 2.0, 1.5

'66_fence': 1, 1, 1

図 5-12 特徴消失が発生した生成画像例

以上の結果より、誤った語に対する理解度の評価から、誤った語である空所補充語彙が画像内に表出しない、特徴消失問題によって、それら語彙への理解が困難となっていることが伺える。英語語彙学習に於いて、誤った語の習得は重要な要素の一つであり、これら特徴消失問題を解消することが、適用語彙の拡張を考えた上で不可欠となる。

5.5 モデルの精度に対する評価

5.5.1 MS-COCO データセットに含まれる画像、及び生成画像に対する評価

ここでは、今回の提案手法で用いた MS-COCO データセットより抽出した画像に於ける精度、DALL-E2 によって生成された画像に対する評価を基に、画像と文に於ける想起性について調査する。ここでは、先述した 0 から 4 までの 5 段階のリッカートスケールによって調査した各語彙に於ける被験者から得られた評価値の平均を算出し、0~1 未満,1~2 未満,2~3 未満,3~4 の 4 段階による分類を行った。2 をどちらでもないとし、数値が低いほど想起性は低く、数値が高いほど良いとする。5-4-2 で紹介した定性評価に依る評価では、アンケートの集計方法について深く言及しなかったため、ここでは使用したデータを例示し、集計方法を紹介する。図 5-13 では、図 4-4 で紹介したラーニングログに、学習者のアンケートの結果をタグ付けしたものを表す。ここでは、黄色で記した 'wordlist_a_listid' は、学習用コンテンツ固有の数字を示し、一つの問題に対し一意に与えられる。緑色で示す 'select_ans' は、問題に対し学習者が選択した語彙を示す。水色で示す 'userinitial' は、MS-COCO データセットより抽出した画像である初期画像の想起性を示す。橙色で示す 'usergene_recall' は、生成画像に対する想起性を示している。

```
[{'user_id': 6, 'qid': 19, 'wordlist_a_listid': 18, 'initial_image': '000000155154', 'question': '', 'select_ans': 'condition', 'pos': 'noun', 'trial_num': 1, 'fake': 1, 'condition': '1', 'result': 'wrong', 'userinitial': 0, 'usergene_recall': 3, 'usergene_impre': 4} ... {'user_id': 6, 'qid': 49, 'wordlist_a_listid': 11, 'initial_image': '000000024144', 'question': '', 'select_ans': 'either', 'pos': 'conjunction', 'trial_num': 3, 'fake': 1, 'condition': '1', 'result': 'wrong', 'userinitial': 1, 'usergene_recall': 3, 'usergene_impre': 4}]
```

図 5-13 アンケートの結果をタグ付けしたラーニングログ

図 5-14 では、すべての被験者から得られた、図 5-13 で示したデータを基に、算出した初期画像に対する想起性の平均値を表している。黄色で示した数値は、先述した学習用コンテンツ固有の数字を示し、水色で示す数値は、初期画像の想起性の全被験者の平均値を表している。この例では、先述した 4 つの分類のうち、2~3 未満に分類される。

```
{'78': 2.49393367767334, '64': 3.9822349548339844, '52': 3.528841972351074, '77':  
3.6940927505493164, '71': 0.4799175262451172, '53': 2.008944511413574, . . . . '23':  
2.7107443809509277}
```

図 5-14 各初期画像の想起性の平均値

図 5-15 では、学習用コンテンツのすべての初期画像の想起性の平均値を算出し、先述した 4 段階の分類を行い、割合を示したものである。このグラフから、MS-COCO データセットより抽出した画像である初期画像に於いて、高い想起性を有していることが伺える。

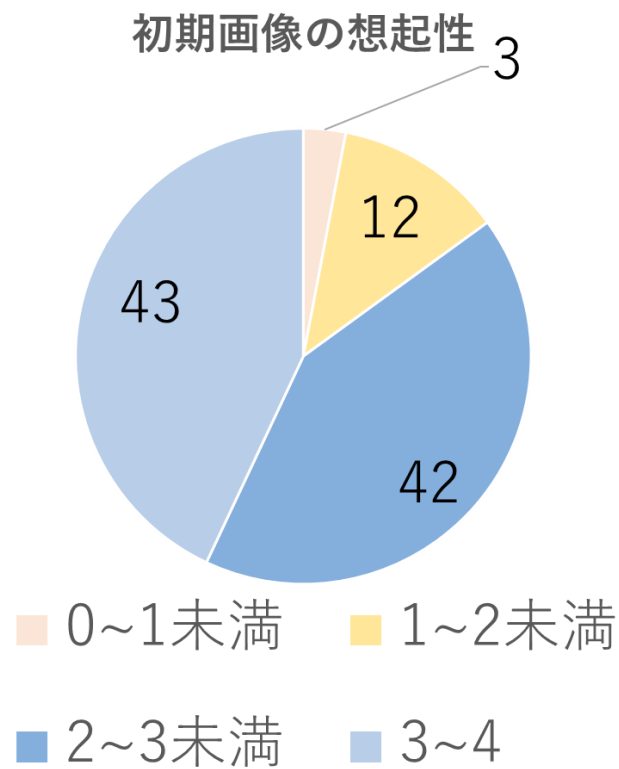


図 5-15 MS-COCO データセットから抽出した初期画像の想起性

又，図 5-16 では，DALL-E による画像の想起性を示しており，文のみから生成を行う画像生成モデルに於いても，高い想起性を有していることが伺える。

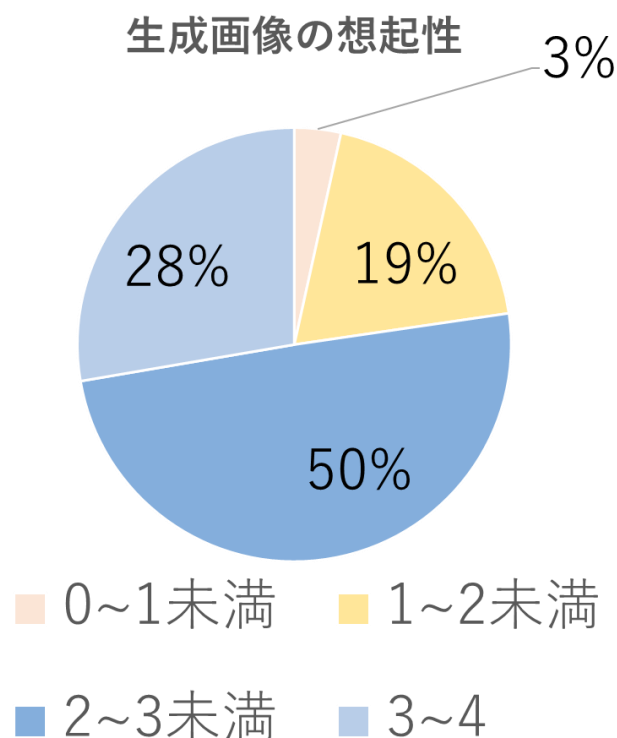


図 5-16 画像生成モデルによって生成された画像の想起性

5.5.2 画像生成モデルに依る生成画像に対する考察

モデルの精度に於ける結果は、MS-COCO データセットに含まれる画像、DALL-E2 で生成された画像、双方に於いて十分な想起性を有していることが明らかとなり、学習用コンテンツとして十分に機能する事を示した。一方で、画像の想起性は高いが図 5-17 のように学習者の選択した回答が正しいにも関わらず誤答としてしまう例がある。この例では、“a very big fridge with () inside it.” という文に対し、正答は“nothing”であるが、実際には、冷蔵庫の中には、ものが入っていることが伺え、“anything”が適している。実際の生成画像に於いても、“anything”が選択されたことによって生成された画像は、初期の画像と近似している。これらの誤りは、モデルの再現度が高いが故に語彙に対する誤った認識を与えてしまう可能性を示している。実際に被験者からは、モデルが高く表現できるが故に、高い習熟度の学習者に於いても、自身の有している語彙の知識に懐疑的になる場面が生じたという意見を得た。

LVIG_Learning Vocabulary with Image Generation



Question No.9

a very big fridge with () inside it

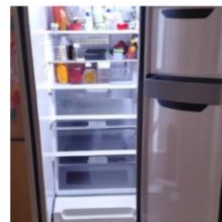
Question No.9

nothing

anything

something

Go / Next



不正解 Wrong answer

a very big fridge
with (anything)
inside it

Question No.9

a very big fridge with () inside it

Question No.9

nothing

anything

something

Go / Next

図 5-17 学習者に誤った認識を与える生成画像例

その他の問題として、学習対象語が画像として表出しない、特徴消失問題がある。5.4.3 遅延テストに於ける誤った語の理解度の評価の項目でも言及したが、画像生成に於いて特徴消失が発生する場合、その学習対象語に対する理解の手がかりを得ることができない問題がある。ここでは、モデルの定性評価にて高い想起性を有している場合に於いても、学習対象となる語が生成画像に特徴として表出しない場合、学習コンテンツとしては不適切である。ここでは、空所補充する学習対象語に応じて、その語彙の特徴を有した画像を生成されることが望ましく、特徴の有無を定量的に評価する指標が求められる。今回、定性的評価によって、画像が有する想起性、画像が有する印象、という二つの軸に対する効果を検証したが、それらを定量的に評価することが可能となると、画像生成を用いた語彙学習研究に於ける、生成画像の指標となることが期待される。

第6章 まとめと今後の課題

本研究では、英語語彙学習に於ける空所補充問題を基本とし、画像生成モデルを適用することにより、学習者の誤答への内省を促し、繰り返しの誤りを防ぐ効果を示した。①質問生成タスクでは、大規模データセットである MS-COCO データセットを利用し自然言語処理の技術を活用することで、画像が含まれた空所補充問題を自動で生成することを実現した。②回答支援タスクでは、定義した COCO-Word list より、自動で候補単語を選定することにより、学習者が誤答をおこしやすい環境の構築を実現した。③画像生成タスクでは、DALL-E2 をモデルとして適用することで、学習者の誤答に応じた画像生成を実現させた。提案手法では、既存のデータセットを活用することで、難易度に応じた問題を自動生成することを実現し、学習者が誤答を起こしやすい環境を構築することにより、誤答に対する学習環境を構築した。被験者に依る実験では、提案手法が、画像生成を含まない画像と文の対となった従来の手法との比較を行い、提案手法が繰り返しの誤答を防ぐことが可能であることを示した。又、既存のデータセットに含まれる画像、および画像生成モデルによって生成された画像は質的に高いという結果を得られた。一方、提案モデルでは、その高い想起性により、学習コンテンツに誤りが含まれた場合に、学習者に誤ったまま語彙を学習させてしまう可能性を有していることが明らかとなった。又、遅延テストに依る誤った語の理解度の評価の結果より、誤った語の概念の理解として、提案手法では特徴消失問題により、語彙への理解の手がかりを得られず、誤った語への理解が困難であることが明らかとなった。

今後の課題としては、英語語彙学習に於いて誤った語の習得は重要な要素の一つであるという認識の中、入力された語彙に対する特徴消失問題への対処が求められる。ここでは、文と語彙の空所補充によって生成された画像に於いて、補充された語彙が画像に特徴として有しているか判断する定量的な評価が求められる。ここでの評価軸は、提案手法で定めた想起性に該当するが、今後は、定量的な評価に依る厳密な特徴の比較としての再現性の評価が求められる、本研究により、画像の印象の値が想起性より、語彙の繰り返しの誤りを防ぐうえで効果的であったことが明らかとなった。今後は、印象に対する定量的手法も模索していきたい。これらを実現したとき、英語語彙学習に於ける、学習に効果的な生成画像を評価することが可能となる。又、英語語彙学習に於いて、有効である画

像と文のペアのデータに対する基準値を設けることにより、近年様々な画像生成モデルが台頭する中、英語語彙学習に於ける良質なモデルを選定することが可能となり、画像生成モデルを活用した英語語彙習得研究に於いて、多大なる貢献を果たすと考えられる。今後は、生成画像に対する指標の策定、特徴消失問題を解消したモデルの選定、構築により、多様な語彙学習への展開が可能であると期待できる。画像生成を活用した語彙学習研究の発展が、世界共通言語である英語の効果的な習得を実現し、国際共創が実現することを切に願う。

对外発表

・ Kazuki Sugita, Wen Gu and Shinobu Hasegawa: Error-Driven Learning: Development of Vocabulary Learning Support System that Utilizes Suggestibility of Error by Image Generation, IIAI Letter on Informatics and Interdisciplinary Research, Vol. 003, LIIR078, (2022).
<https://doi.org/10.52731/liir.v003.078>

謝辞

本研究を始めるにあたり、研究の本質から全てに於いてご指導を頂いた、主指導教員の長谷川忍教授に心より感謝致します。

又、副テーマ研究に於いてご指導、ご助言を頂いた白井清昭准教授、貴重なご助言を下された小谷一孔教授、岡田将吾准教授に心よりお礼申し上げます。又、国際交流では、川西俊吾教授、元山琴菜講師から多くの留学生との交流の機会を頂戴し感謝致します。そして、忙しい時期に、実験にご協力を下された皆様方に深く感謝致します。最後になりますが、国際会議の発表に於いて多大なる指導を頂いた Gu Wen 助教、研究室 MTG で助言を下された太田光一助教、同研究室の皆様、GLTS、DGGL の皆様に本当に感謝致します。

参考文献

- [1] Rao, P.S. The role of English as a global language. *Res. J. Engl.* 2019, 4, 65–79. Available online: [https://www.rjoe.org.in/Files/vol4issue1/new/OK%20RJOE-Srinu%20sir\(65-79\).pdf](https://www.rjoe.org.in/Files/vol4issue1/new/OK%20RJOE-Srinu%20sir(65-79).pdf) (accessed on 1 December 2022)
- [2] Laufer, Batia. (1989). A factor of difficulty in vocabulary learning: Deceptive transparency. *AILA Review*. 6. 10-20.
- [3] C. Kobayashi (2016) “Lexical Errors in the English Compositions of Japanese Learners,” *Foreign language education: Theory and practice*, vol42, pp39-51.
- [4] Nagy, W. (1995). *On the Role of Context in First- and Second-Language Vocabulary Learning*. Technical Report No. 627.
- [5] Hasnine, M.N.; Flanagan, B.; Akcapinar, G.; Ogata, H.; Mouri, K.; Uosaki, N. (2019) Vo-cabulary 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
- [6] Laufer, Batia. (1988). The concept of “synforms” (similar lexical forms) in L2 vocabulary acquisition. *Language and Education*. 2. 113-132. 10.1080/09500788809541228.
- [7] Takao IMAI. (2017) The Effects of Explicit Instruction of “Image English Grammar for Communication” on Tertiary English Classes, *ARELE: Annual Review of English Language Education in Japan*, 2016, Volume 27, Pages 137-152
- [8] 田中茂範 (2013) 『愛蔵版 表現英文法』, コスモピア, 433-434 頁
- [9] 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. 113-132. 10.1080/09500788809541228.
- [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] 菊池開, 長谷川忍: “過去の写真を利用する日記スタイル単語学習アプリケーション W-DIARY の開発”, 信学技法, ET2015-61, pp. 15-19, (2015).
- [12] Hasnine, M.N.; Flanagan, B.; Akcapinar, G.; Ogata, H.; Mouri, K.; Uosaki, N. (2019) Vo-cabulary learning support system based on automatic image captioning technology. In Pro-ceedings of the International Conference on Human-Computer Interaction, Orlando, FL, USA, 26–31 July 2019; Springer: Cham, Switzerland, pp. 346–358
- [13] Lin, Tsung-Yi & Maire, Michael & Belongie, Serge & Hays, James & Perona, Pietro & Ramanan, Deva & Dollár, Piotr & Zitnick, C. (2014). Microsoft COCO: Common Objects in Context.
- [14] Kaplan-Rakowski, R., Lin L., & Wojdynski, T. (2021). Learning vocabulary using 2D pic-tures is more effective than using immersive 3D stereoscopic pictures. *International Journal of Human-Computer Interaction*, 37(12)
- [15] 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.
- [16] 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
- [17] [Chujo, K., Oghigian, K. and Akasegawa, S. “A corpus and grammatical browsing system for remedial EFL learners,” In A. Leńko-Szymańska and A. Boulton (eds.), *Multiple Affordances of Language Corpora for Data-driven Learning*. Amsterdam: John Benjamins, 2015, 109-128.
- [18] Honnibal, M., & Montani, I. (2017). spaCy 2: Natural language understanding with Bloom embeddings, convolutional neural networks and incremental parsing
- [19] 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.
- [20] Nazeri, Kamyar and Ng, Eric and Joseph, Tony and Qureshi, Faisal and Ebrahimi, Mehran[2019]
- [21] Ramesh, A., Dhariwal, P., Nichol, A., Chu, C., and Chen, M.(2022),

“Hierarchical Text-Conditional Image Generation with CLIP Latents”,
arXiv e-prints.

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

付録

ここでは，本研究で使用した資料を付録として収録する．

1. 被験者実験で用いたマニュアル
2. 被験者実験で用いたアンケート
3. 実験後テスト



Manual for experiments

[Objective]

Prior to the experiment, we would like you to participate in the experiment with peace of mind by confirming and agreeing on the following points between you and us, the researchers. The following points are explained below. You will learn English vocabulary using two different English vocabulary learning support systems. For each system, one in which the contents of wrong answers are generated as images when the learner makes a mistake. And the other in which no images are generated. Using each system, it will be evaluated the experiment result based on your impressions and learning results.

[Methods of protecting personal information and matters relating to informed consent]

Personal information to be collected includes name, contact information, and age and gender for statistical information. All personal information except for experimental results will be discarded after publication. Information obtained from the experiments will be used only for research purposes and will not be used for any other purpose. Your personal

information (name, contact information, etc.) and experimental results will be kept strictly confidential in the laboratory of your research supervisor. The results of experiments other than personal information will eventually be published in academic conferences, papers, books, etc., but the results will not be disclosed in a form that allows individuals to be identified. Personal information, such as age and gender, may be disclosed as statistical information, but names and contact information will not be disclosed.

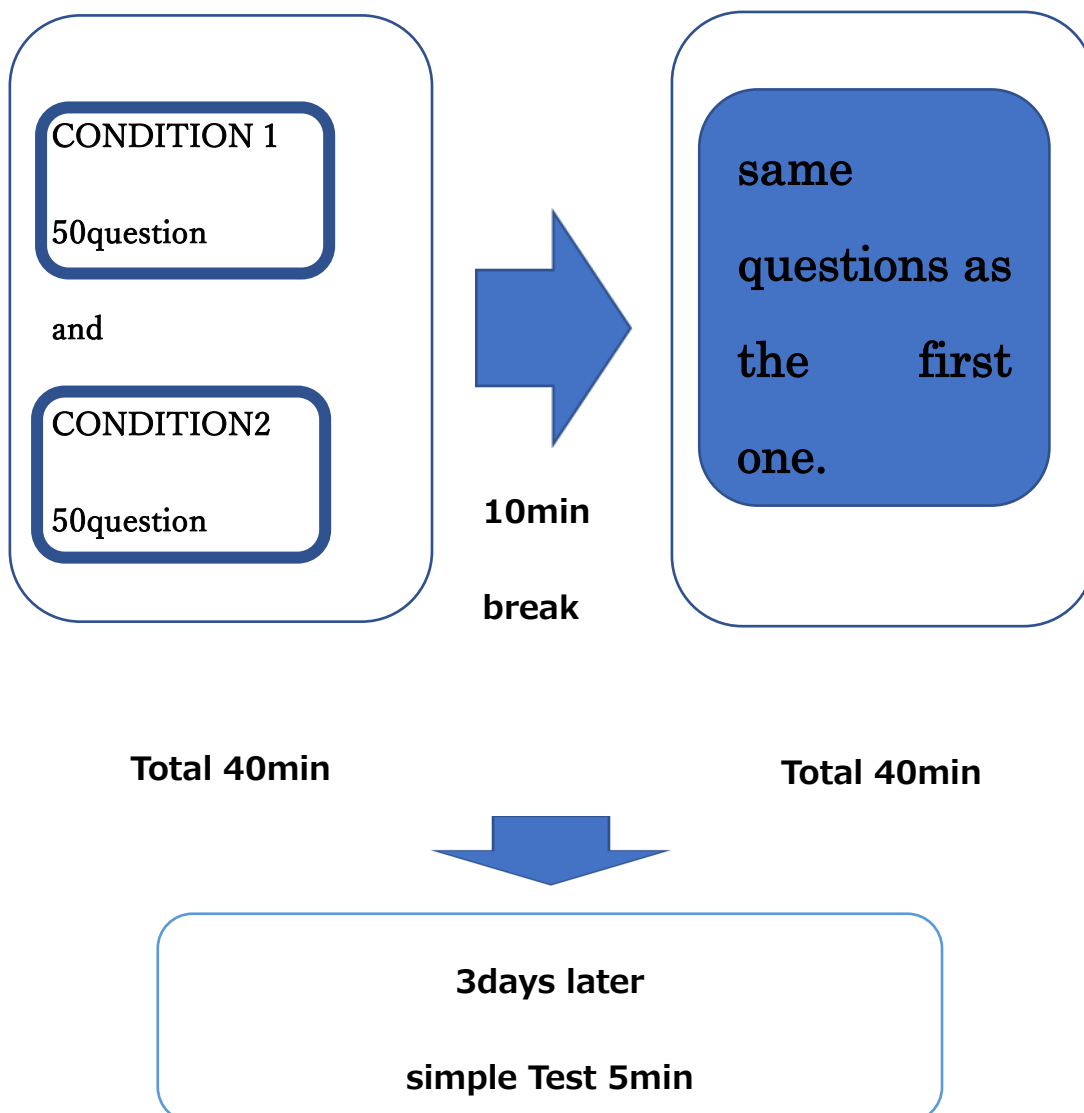
[Reward]

For participating in this experiment, we will pay 1,000 yen per hour.

[Learning flow]

The learner is required to follow the steps shown in the figure below. The difference between Condition1 and Condition2 is whether or not image generation is included. You will learn a total of four conditions, and these sets will be run twice. Each condition has 50 questions. Please allow approximately **10 minutes per condition** for each answer. Learners will

be asked to fill out a **experimental evaluation sheet** for each question answered. There are two types of this sheet. Each condition requires a different sheet to be filled out. After each condition, the person in charge will **change the settings**, so please let us know when you see the page signaling the end of the session, which will be explained later. We will send you a **brief 5minutes test three days after the experiment** is over. we ask that **you respond within the day after the announcement.**



In the study, the same level of questions will be provided at random, regardless of the learner's proficiency level. The number of questions in the study will be 200 questions per session, and the learner will be asked to solve **the same question a second time.**

We will include questions that do not include a correct answer in the question.

In the figure on the right, the correct answer is "blank," whereas the possible answers are "black," "back," and "bag," respectively.

Please choose the one that you think best fits the context for such questions.

[System Description]

This section describes the operation methods of the systems, one with image generation and one without image generation, in this order. The

LVIG_Learning Vocabulary with Image Generation



正解

A red traffic stop sign with a (blank) white sign below it .

Question No.13

A red traffic stop sign with a () white sign below it .

不正解

A red traffic stop sign with a (back) white sign below it .

- black
- back
- bag

explanation of the system without image generation is brief since the operating procedures are identical.

LVIG_Learning Vocabulary with Image Generation



The images displayed correspond to the learner's questions shown below.

Question No. represents the number of the question. It corresponds to the number on **the experimental evaluation sheet** to

Question No.13

A red traffic stop sign with a () white sign below it.

Question No.13

back

black

blank

Go / Next

1

The learner selects the appropriate word for the vacant space in the English sentence.

LVIG_Learning Vocabulary with Image Generation



Question No.13

A red traffic stop sign with a () white sign below it .

Question No.13

back

black

blank

Go / Next

2

62 Select the option that you believe is the correct answer and press the Go/Next button once.

LVIG_Learning Vocabulary with Image Generation



正解です

Correct Answer

Here we show that the learner's response was correct.

A red traffic stop sign with a (blank) white sign below it.

Sentences containing the entered word will be displayed.

Question No.13

A red traffic stop sign with a () white sign below it.

3

Question No.13

blank
 black
 back

Go / Next

Press the Go/Next button once to proceed to the next question. Please fill out **the evaluation sheet** and proceeding to the next

LVIG_Learning Vocabulary with Image Generation



Question No.13

A red traffic stop sign with a () white sign below it .

Question No.13

blank

black

back

Go / Next

Next, we will explain the screen when a learner gives an incorrect answer.

LVIG_Learning Vocabulary with Image Generation



不正解

Wrong Answer

A red traffic stop sign with a (back) white sign below it .

Question No.13

A red traffic stop sign with a () white it .

Depending on the learner's answer, an image representing the sentence containing the incorrect word and the sentence will be

Please fill out the evaluation sheet before moving on to the next question, whether you answered correctly or incorrectly.


Question No.13

- black
- blank
- back

Go / Next

**実験は終了です。お疲れ様でした。
次のセッションに移りますので、担
当者にお知らせください。**

**The experiment is over. Thank you for
your cooperation. Please inform the
person in charge so that we can move
on to the next session.**



When one condition is completed,
the above screen will appear. At
this stage, please inform the staff
member in charge of your request.

LVIG_Learning Vocabulary with Image Generation



不正解

Wrong Answer

**A red traffic stop sign with a
white sign below
it .**

In the system without images, no images are displayed when an incorrect answer is made.

Other than that, the operation method is the same as the system with images.

Question No.13

**A red traffic stop sign with a () white sign below
it .**

Question No.13

- black
- blank
- back

Go / Next

[Post-experiment flow]

After the experiment, participants will be asked to complete a simple test.

This test will be a simple 5-minute test conducted on the web.

An announcement will be made in the morning **three days after the completion of the experiment**. Because it is a simple test, we ask that **you respond within the day after the announcement**.

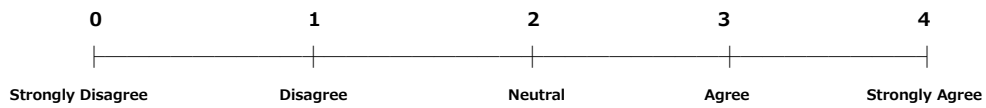
That is all for the explanation.

If you have any questions, please contact the person in charge.

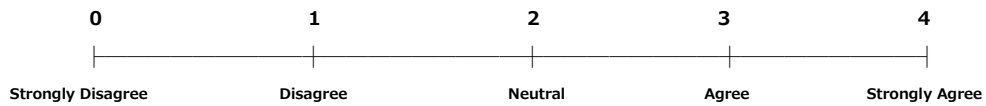
Evaluation sheet for condition1

Explanation : Referring to the example described at the bottom, please fill in the cells on the next page with a number for each of the following three elements for each Question. No.

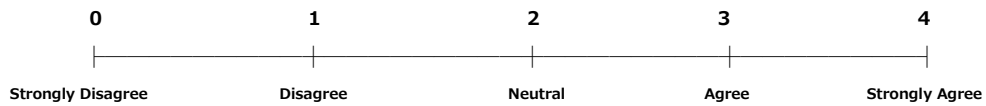
No1. The image displayed on the initial screen suggests the content of the sentence



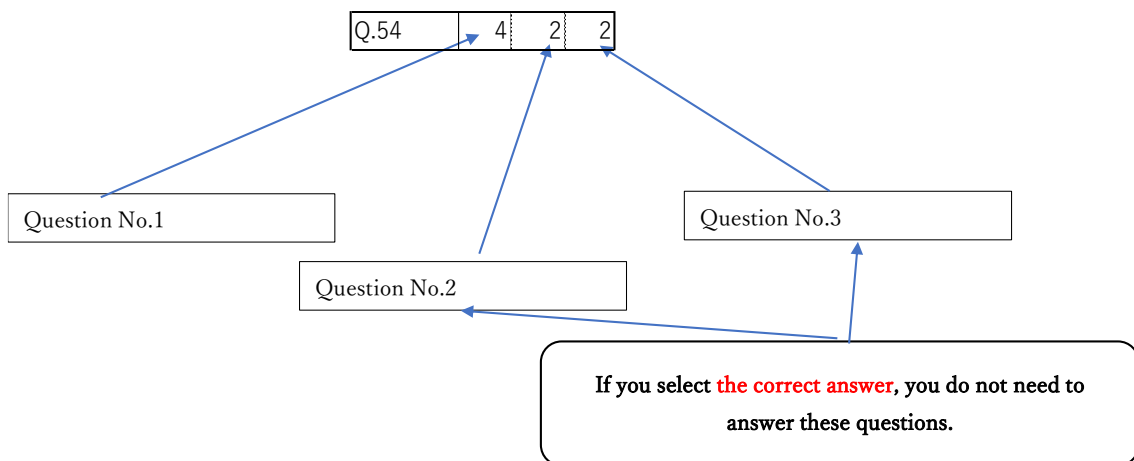
No2. The images generated by wrong answers suggest the content of the sentence



No3. The images generated by wrong answers are impressive.



Example



Condition 1

UserID

Trial Number 1

Age:

Gender:

Score (TOEIC etc):

Your native language

Q.Number	No.1	No.2	No.3						
Q.0			Q.17			Q34			
Q.1			Q.18			Q35			
Q.2			Q.19			Q36			
Q.3			Q.20			Q37			
Q.4			Q.21			Q38			
Q.5			Q.22			Q39			
Q.6			Q.23			Q40			
Q.7			Q.24			Q41			
Q.8			Q.25			Q42			
Q.9			Q.26			Q43			
Q.10			Q.27			Q44			
Q.11			Q.28			Q45			
Q.12			Q.29			Q46			
Q.13			Q.30			Q47			
Q.14			Q.31			Q48			
Q.15			Q.32			Q49			
Q.16			Q.33						

Trial Number 2

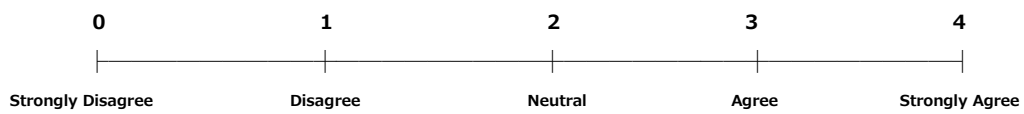
Q.Number	No.1	No.2	No.3						
Q.50			Q.67			Q84			
Q.51			Q.68			Q85			
Q.52			Q.69			Q86			
Q.53			Q.70			Q87			
Q.54			Q.71			Q88			
Q.55			Q.72			Q89			
Q.56			Q.73			Q90			
Q.57			Q.74			Q91			
Q.58			Q.75			Q92			
Q.59			Q.76			Q93			
Q.60			Q.77			Q94			
Q.61			Q.78			Q95			
Q.62			Q.79			Q96			
Q.63			Q.80			Q97			
Q.64			Q.81			Q98			
Q.65			Q.82			Q99			
Q.66			Q.83						

Evaluation sheet for condition2

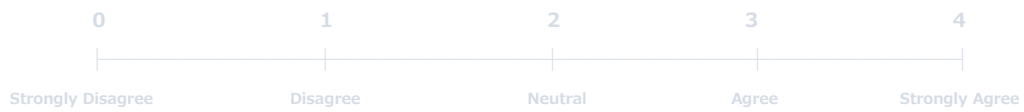
Explanation : Referring to the example described at the bottom, please fill in the cells on the next page with a number for each of the following three elements for each Question. No.

In this section, please fill in only the first cell.

No1. The image displayed on the initial screen suggests the content of the sentence



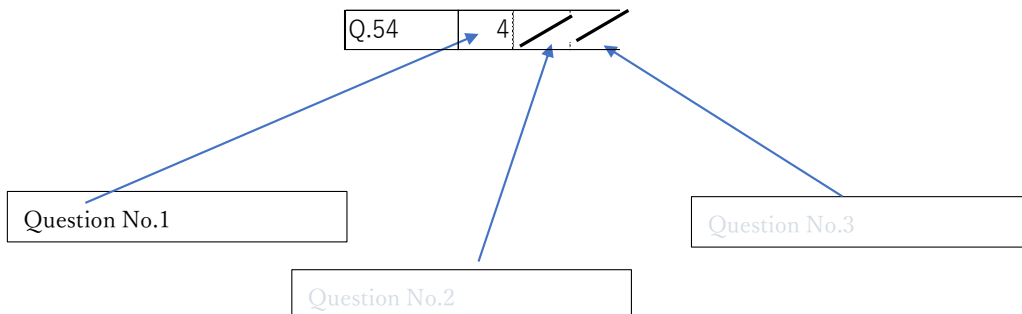
No2. The images generated by wrong answers suggest the content of the sentence



No3. The images generated by wrong answers are impressive.



Example



Condition 2

Trial Number 1

Age: _____

Gender: _____

Score (TOEIC etc): _____

Your native language _____

Q.Number	No.1						
Q.0			Q.17			Q34	
Q.1			Q.18			Q35	
Q.2			Q.19			Q36	
Q.3			Q.20			Q37	
Q.4			Q.21			Q38	
Q.5			Q.22			Q39	
Q.6			Q.23			Q40	
Q.7			Q.24			Q41	
Q.8			Q.25			Q42	
Q.9			Q.26			Q43	
Q.10			Q.27			Q44	
Q.11			Q.28			Q45	
Q.12			Q.29			Q46	
Q.13			Q.30			Q47	
Q.14			Q.31			Q48	
Q.15			Q.32			Q49	
Q.16			Q.33				

Trial Number 2

Q.Number	No.1						
Q.50			Q.67			Q84	
Q.51			Q.68			Q85	
Q.52			Q.69			Q86	
Q.53			Q.70			Q87	
Q.54			Q.71			Q88	
Q.55			Q.72			Q89	
Q.56			Q.73			Q90	
Q.57			Q.74			Q91	
Q.58			Q.75			Q92	
Q.59			Q.76			Q93	
Q.60			Q.77			Q94	
Q.61			Q.78			Q95	
Q.62			Q.79			Q96	
Q.63			Q.80			Q97	
Q.64			Q.81			Q98	
Q.65			Q.82			Q99	
Q.66			Q.83				

Thank you very much for participating in my experiment the other day.

I would appreciate your cooperation in the following tests.

Please use the following example as a guide.

If you already know the word, write "1" in the "**already know**" field. If you did not know the word before, but can imagine it, please enter "1" in the "**imaginable**" field. If you do not remember it at all, please enter "1" in the "**unimaginable**" field.

Q.No	word	already know	imaginable	unimaginable
Example	like	1		
1	enough			
2	material			
3	other			
4	ring			
5	those			
6	draw			
7	for			
8	between			
9	plum			
10	lie			
11	actively			
12	like			
13	slow			
14	minor			
15	beside			
16	inside			
17	list			
18	messy			
19	tap			
20	force			
21	disorganized			
22	which			
23	its			
24	festive			
25	liner			
26	return			
27	strain			
28	if			
29	location			
30	learn			
31	they			
32	yawn			1
33	at			
34	beyond			
35	soon			
36	ordinary			
37	such			
38	decorative			
39	living			
40	while			
41	direction			
42	special			
43	swing			
44	either			
45	sharp			
46	as			
47	the			
48	toot			
49	while			
50	someone			
51	lately			
52	something			
53	profession			
54	residence			
55	his			
56	alongside			
57	a			
58	land			
59	fence			
60	essential			
61	happily			
62	when			
63	anyone			
65	nearly			
66	heat			
67	load			
68	convenience			