Master ’s Thesis

# Integrating Object Recognition and WordNet for Japanese Thesaurus Acquisition

XUE MEIHUA

Supervisor Hasegawa Shinobu

Graduate School of Advanced Science and Technology
Japan Advanced Institute of Science and Technology
(Information Science)

March,2024

# Abstract

In the current global landscape, where interconnectivity between nations and cultures is at its peak, the skill of being multilingual has taken on a new level of significance. Within this multilingual spectrum, the Japanese language has become a particularly vital language to acquire. This holds especially true for international students residing in Japan, where fluency in Japanese goes far beyond academic achievements, transforming into an essential component for effective daily life and deeper cultural assimilation. For these students, learning Japanese is not merely a pursuit of linguistic proficiency for educational purposes. Instead, it represents a key to unlocking a fuller, more enriched experience within Japan. Proficiency in Japanese allows them to navigate the complexities of everyday interactions, from the simplicity of market transactions to the intricacies of social customs and traditions. It also plays a pivotal role in establishing meaningful connections with the local community, enabling a richer understanding of the cultural nuances and historical contexts that define Japanese society. In essence, Japanese language proficiency is more than a mere academic endeavor for international students in Japan. In an increasingly interconnected world, it is indispensable for day-to-day living, cultural understanding, and personal growth. In response to these challenges, this study introduces the PICSU system, whose pioneering approach integrates object recognition technology with the extensive WordNet lexical database, setting a new precedent in Japanese vocabulary learning.

The research explored the innovative use of images smartphones captured as a vocabulary learning tool. This approach integrates the learning of synonyms and antonyms through a thesaurus-based system, seamlessly embedding it within the framework of image-based learning. The system, named PICSU, represents a substantial shift from traditional language learning methodologies. It significantly enhances learner engagement by providing a contextually rich environment crucial for effective language acquisition. Unlike the conventional rote memorization techniques often associated with flashcards, PICSU leverages the visual stimuli from everyday life captured in photographs. This method not only aids in retaining new vocabulary but also helps in understanding the practical application of these words in real-world scenarios. By doing so, it offers a more holistic and immersive learning experience. Furthermore, the research delves into the comparative effectiveness of this innovative approach against the traditional flashcard methods, highlighting the benefits of integrating visual elements in language education.

The research employed a novel methodology, combining the YOLO object recognition

model with the WordNet database to enrich learners' vocabulary through a photo-based learning environment. An extensive experimental study involving 20 students from the Japan Advanced Institute of Science and Technology (JAIST) was conducted. This study assessed the PICSU system's efficacy compared to traditional flashcard methods. Participants were divided into four groups, each engaging in learning Japanese nouns through both the PICSU system and the flashcard method. The experimental design was meticulous, ensuring a comprehensive evaluation of the system's effectiveness in enhancing vocabulary learning.

The study's results clearly demonstrated the superior effectiveness and learner engagement of the PICSU system compared to traditional methods. Participants using PICSU showed significantly higher success in memory assessments, indicating better retention and understanding of vocabulary. They also experienced less memory loss and made fewer errors in tests, suggesting a deeper and more lasting grasp of the learned words. This success is attributed to PICSU's innovative approach, combining visual stimuli with contextual learning. The system's use of smartphone-captured images and thesaurus integration creates a more immersive and relatable learning environment. Additionally, learners reported higher levels of motivation and enjoyment with PICSU, underscoring its potential as a modern, effective tool for language learning.

These findings support the hypothesis that visual aids, a core component of the PICSU system, considerably enhance memory retention and facilitate more effective learning. The empirical data underscored the advantages of integrating visual aids into language learning methodologies, particularly in the context of complex languages like Japanese.

The study represents a significant advancement in language learning technology. By synergizing advanced object recognition technology with a comprehensive language database, the PICSU system has effectively demonstrated its potential to revolutionize language acquisition, especially focusing on thesauruses. The system's ability to integrate seamlessly into learners' daily lives, providing an engaging and interactive learning environment, sets it apart from traditional language learning methods.

The study opens the door for further exploration in key areas such as long-term retention effects of the PICSU system, understanding the cognitive mechanisms behind its learning process, and examining its adaptability across various learner demographics. Future enhancements of PICSU are planned to include gamification elements to boost engagement, as well as auditory components and voice functionality, aiming to create a more immersive and comprehensive learning experience. These developments are targeted not just towards facilitating effective Japanese language acquisition, but also catering to a broader range of learners, thereby enriching their journey towards fluency.

The ultimate goal is to evolve PICSU into a tool that transcends traditional vocabulary acquisition, fostering not only language learning but also promoting deeper cultural understanding and integration, making it a pivotal tool for cultural exchange and global communication.

In summary, the PICSU system marks an important advancement in the field of language education, signaling the start of a promising new chapter in this area. By ingeniously integrating cutting-edge technology with highly effective learning strategies, PICSU stands out as an avant-garde system. It offers a unique, engaging, and efficient approach to language acquisition, perfectly aligning with the needs and preferences of today's technology-oriented generation. The system's innovative use of smartphone-captured images and integration of thesaurus-based learning provide a contextually rich and visually stimulating educational experience. This not only aids in faster vocabulary acquisition but also ensures a deeper understanding and retention of the language. Furthermore, the potential of PICSU to revolutionize the landscape of language education is immense. Enhancing the learning experience for current students significantly contributes to increased motivation, engagement, and, ultimately, better learning outcomes. Its adaptability to incorporate future advancements in technology and pedagogy positions it as a dynamic and evolving tool. This adaptability ensures that PICSU will continue to set new benchmarks in language education, meeting the evolving needs of learners. Additionally, as demonstrated by the research, its proven effectiveness over traditional methods underscore its potential to become a standard in language learning, paving the way for a more interactive, immersive, and effective educational experience. In this way, PICSU is not just a tool for the current generation of learners but also fosters the future of language education, creating a legacy that will benefit future generations by providing a more engaging and technologically advanced learning environment.

Keywords: Japanese vocabulary learning, thesauruses learning, object detection, object recognition, learning system, YOLO, WordNet.

# Contents

# List of Figures

# List of Table

# Chapter 1 Introduction

## 1.1 Background

In recent years, the global trend in language learning has seen a significant upsurge, reflecting the growing importance of multilingual abilities in a globalized society. Each year, an increasing number of people show interest in learning new languages, driven by various motivations ranging from personal development to professional needs. However, despite initial enthusiasm, many learners encounter challenges that hinder their progress and motivation. Learning a new language can be daunting, with obstacles in mastering unfamiliar grammar, vocabulary, and pronunciation.

In language learning, reading book knowledge is boring and difficult to remember. Pictures can be an effective tool for language learning and promote spontaneous communication. Using pictures to learn languages has gained attention because of its fun and unique nature [1]. Visual cues help learners recall and learn language in a different form from the written words and can increase speaking flexibility or spontaneity. Using pictures for language learning can also make learning more interesting and unique.

In the context of Japan, these challenges take on specific dimensions for international students. For those residing in Japan, the necessity of learning Japanese is paramount, impacting virtually all facets of daily life [2]. The experience of international students in Japan underscores the importance of developing effective language learning strategies that cater to their unique needs. For instance, language learning is divided into many key parts, including basic pronunciation, grammar and sentence patterns, vocabulary building, Kanji learning, listening, reading, speaking practice, etc. Some researchers have pointed out that the most crucial factor in language learning is expanding vocabulary space [3]. Suppose international students strengthen their vocabulary learning. In that case, it will be easier to use in daily life, and if they learn the thesauruses of vocabulary, it will also make it more flexible and easier to use vocabulary to understand [4]. Their journey in navigating a new linguistic landscape highlights the need for innovative approaches in language education, particularly ones that are tailored to the complexities of the Japanese language.

Most existing Japanese learning software can only memorize and combine fixed words. A lot of Japanese language learning is only based on learning simple vocabulary, which

is not convenient for learners to understand and use. Inherently, learning content that already exists in textbooks will be difficult to apply in real life. Moreover, it is not easy to achieve continuous or regular learning. Learners will be unwilling to use learning software frequently due to tedious and inconvenient conditions, or it will be challenging to maintain self-discipline continuously, and it will not be easy to memorize words and vocabulary correctly. Technically, unstable connection environments, etc., may cause inconvenience in use.

In modern life, everyone uses mobile phones frequently, and the mobile phone's camera function is used more frequently than the phone function. This also shows that pictures significantly influence people's daily lives. If learning vocabulary through pictures can be more interesting and effective in daily life, it can help international students learn Japanese anytime and anywhere [5]. Therefore, using pictures to learn thesauruses of Japanese vocabulary has become an important research topic so that international students can learn Japanese more conveniently and flexibly in their daily lives in Japan [6]. However, in traditional language learning using pictures, the focus is on the learner associating pictures with words, and it is impossible to determine whether a picture is correct or incorrect or to create problems.

## 1.2 Research Objectives and Contributions

### 1.2.1 Research Objectives and Questions

This study aims to develop a vocabulary learning support system, called PICSU, for Japanese thesaurus using photographic object recognition and WordNet.

The specific research questions are as follows:

RQ1: How can photos of familiar objects that exist or are taken with smartphones be used in vocabulary learning?

RQ 2: How can thesaurus learning be integrated into vocabulary learning?

RQ 3: What are the differences in learning outcomes in a picture-based Japanese vocabulary and thesaurus learning application compared to traditional flashcards?

### 1.2.2 Research Contributions

By systematically addressing these research questions, the study aims to design and implement a Japanese vocabulary learning application that not only enhances learners' interest, long-term memory, and retention of the language but also accelerates the acquisition of Japanese through the utilization of smartphone cameras, enabling efficient learning anytime and anywhere by capturing images of surrounding objects or utilizing existing photos. The goal is to provide a tool that not only supports international students in their Japanese language learning journey but also integrates seamlessly into their daily lives, fostering a more widespread and practical use of the language.

## 1.3 Structure of the Thesis

The structure of this thesis is as follows:

- **Chapter 1: Introduction**

The first chapter introduces the importance of Japanese language learning to international students living in Japan and the fact that vocabulary learning occupies the central part of language learning. It also introduces the advantages of using pictures for learning and points out the shortcomings of traditional language learning methods in research. Finally, it describes this research's objective, questions, and contributions.

- **Chapter 2: Related Works**

Chapter 2 introduces some case studies on language learning so far, as well as research on language learning using technologies such as images or virtual reality. Then it introduces the improvements of this study based on existing research. What is different from previous research is that it not only uses image processing technology to learn language vocabulary, but also enriches learners' vocabulary by learning thesauruses of vocabulary to help international students use more vocabulary in their lives in Japan**.**

- **Chapter 3: Research Design and Methodology**

Chapter 3 proposes a photo-based thesaurus learning support environment combining YOLO and Wordnet. This research mainly uses Japanese Wordnet to convert thesauruses of known vocabulary and then enrich the vocabulary by learning of original words and thesauruses. By adding the YOLO model, the photo learning function can be realized, and image processing technology can be used to make Japanese learning more interesting and intuitive.

- **Chapter 4: Experimentation**

Chapter 4 describes the comparative analysis of the learning and testing results of the language learning support system PICSU developed in this study and the learning and testing results of the traditional flash card method, thereby obtaining the impact of the new thesaurus learning method using images on learners.

- **Chapter 5: Conclusion and Future Works**

Chapter 5 summarizes the research results and future work of this article.

# Chapter 2 Related Works

## 2.1 The Importance and Problems of Learning Japanese Vocabulary

The importance and challenges of learning Japanese vocabulary are immense. First, vocabulary is the cornerstone of any language communication, and this is especially true for Japanese. A rich vocabulary enables more nuanced understanding and expression, which is critical in everyday conversations as well as in professional or academic settings. Additionally, learning vocabulary is crucial to reading and writing in Japanese, given its complex writing system, which includes kanji (characters borrowed from Chinese), hiragana, and katakana. These items seem easy to learn at first glance because they are frequently used frequently and are used regularly in daily life. In vocabulary teaching, it is essential to understand the semantic boundaries between semantically related thesauruses and polysemous extensions [7]. This was considered to indicate the need for guidance that also focuses on the conversion process.

However, learning Japanese vocabulary comes with unique challenges. The existence of Chinese characters, each character may have multiple readings and meanings, adds a layer of complexity. For international students, especially native Chinese speakers, it is more difficult to learn due to the interference of their mother tongue [8].

Learning a language can be roughly divided into four aspects: listening, speaking, reading, and writing. Word nuances and context-specific usage [9] can also be complicated for learners, especially those whose native language has a very different structure and cultural background. Additionally, the vocabulary required for fluency can be daunting. These challenges require consistent practice and immersion in the language to learn effectively.

Figure 2.1 illustrates the interplay between the four language skills—listening, speaking, reading, and writing—and the role of paraphrasing. Solid arrows depict activities that bridge spoken and written forms, such as "writing what was heard" or "speaking what was written," necessitating paraphrasing due to stylistic differences. Dotted arrows indicate the need for concise paraphrasing when "speaking what was heard" or "writing

what was read." These visuals underscore paraphrasing as a key component in versatile language use.

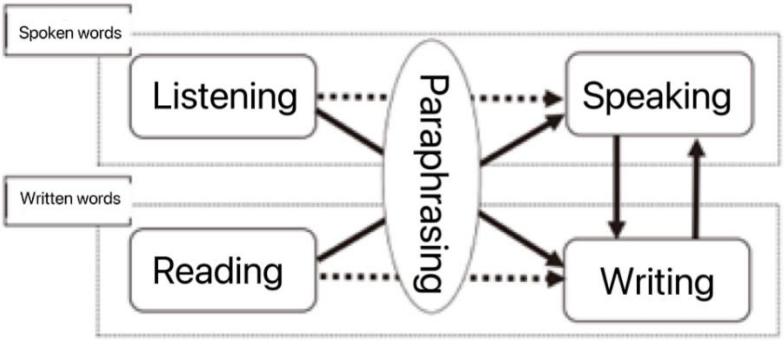


**Figure2. 1** Four skills and paraphrasing [9]
(Translated into English from the original version)

The need for paraphrasing depends on the means of communication, context, purpose, and audience, making contextually linked learning essential. This figure encapsulates the need for paraphrasing across different language skills, highlighting its importance in language learning and context-based education.

## 2.2 W-DIARY: Enhancing Language Learning with Photos

There are diverse prior studies on vocabulary learning using images, such as this study on an application called W-DIARY (W: Word D: Diary I: Image A: Addict RY: memoRY) developed by Kikuchi et al. using diary-style word learning using past photos [10].The study found that associating English words with existing photos can deepen the connection between events and words, making memory more effective, and proposes a new method for learning to search for English words. W-DIARY was equipped with a function that allows learners to write diaries using photos and mark them, study vocabulary books for review, and evaluate the final learning content.

Figure 2.1 shows the process of the learner writing a diary. When using W-DIARY, learners need to mark English words they want to remember onto objects or events in photos they take. These activities allow them to deepen the relationship between objects and words and learn words more effectively.


**Figure2. 2** Creating a learning vocabulary list [10].

Kikuchi's approach to language learning involves using photographs as mnemonic tools, where learners can associate any word with a photograph to aid memory without employing object recognition. This method focuses on personal associations between words and images but does not facilitate the recognition of specific words from the photographs. In contrast, PICSU introduces object recognition, enabling the identification and learning of specific words directly from images. Additionally, while Kikuchi's method does not incorporate thesaurus learning, PICSU is explicitly designed to allow learners to expand their vocabulary by understanding and memorizing thesauruses through visual cues. This significant difference lies in PICSU's use of technology for direct word recognition and its unique focus on thesaurus learning, offering a more structured and technologically integrated approach to language acquisition.

## 2.3 Image-to-Text Recognition in Language Learning

The research was conducted by Rustam Shadiev, Ting-Ting Wu, and Yueh-Min Huang. They conducted an experimental study to investigate the effectiveness of using image-to-text recognition (ITR) technology to facilitate vocabulary acquisition in authentic contexts [11]. The study involved native Russian speakers learning English as a foreign language and aimed to assess the learners' perceptions of the learning system and its impact on vocabulary acquisition. The research employed a pre-test–post-test/delayed

post-test design to evaluate the effectiveness of the proposed approach. The study also explored the learners' experiences with the technology and identified strategies for improving the accuracy rates of the image-to-text recognition system.

The experimental results showed that the learners in the experimental group expressed high agreement toward ease of system use, positively perceived the usefulness of the system for learning, had high behavioral intention to use the system for learning in the future, and had high levels of learning satisfaction. The study also identified valuable strategies for achieving better accuracy rates when using ITR, such as using unambiguous images for queries and setting the camera to take pictures at lower resolutions using low Internet bandwidth. Overall, the study suggests that image-to-text recognition technology has the potential to support vocabulary acquisition in authentic contexts.

Figure 2.2 shows a screenshot of the learning system with image-to-text recognition technology used in the study. The system included several primary functions: a camera, ITR, notes, and textbook. The ITR function used the Google Images service to generate English labels for objects of interest captured by the learners.

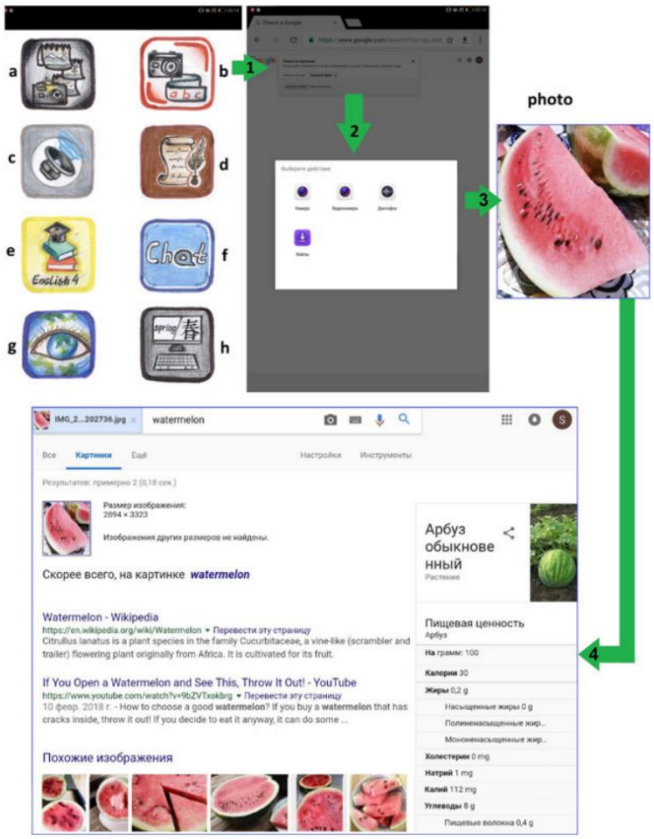


**Figure2. 3** system and image-to-text recognition process [11]

Shadiev's research focuses on the effectiveness of using real-world images in vocabulary learning through a system that supports learning via image search of photographed photographs. This approach emphasizes the impact of real-world imagery in enhancing vocabulary acquisition, underscoring the value of visual context in learning. However, Shadiev's study does not address the challenge of learning multiple expressions for the same concept, particularly relevant in languages like Japanese, with varied ways to express a single idea. In contrast, our work with PICSU incorporates image recognition technology and emphasizes learning thesauruses, providing a comprehensive understanding of multiple expressions for the same object.

## 2.4 Augmented Reality in Japanese Vocabulary Learning

The research was conducted by Riri Safitri, Resnia Trya Muslima, and Sandra Herlina from the Informatics Department at the University of Al Azhar Indonesia [12]. The research involved the development of an educational media using Augmented Reality and a mnemonic approach to teach Japanese vocabulary and hiragana letters. The study used the Multimedia Development Life Cycle method. It included materials evaluation, user acceptance testing, and direct testing on children to evaluate the impact of the application on their understanding of Japanese vocabulary. The study aimed to provide a new user experience for children through animation, 3D, and interaction, making learning fun and effective.

Figure 2.3 shows the use case diagram for the AR Hiragana application, which includes the user, camera, and book as participants. The application has four main menu options: Vocabulary, Hiragana, Quiz, and Exit. The Vocabulary menu displays 3D objects from five categories with Japanese pronunciation. The Hiragana menu displays the Hiragana letters and their pronunciation. The Quiz menu has six categories for evaluation. The Exit menu allows the user to exit the application.

The experimental results showed that the AR Hiragana application effectively aided children's learning of Japanese vocabulary and hiragana letters. The user acceptance test had 70 respondents, consisting of 21 children aged 6-12 years, 44 students aged 20-24 years, and 4 parents aged 29-32 years. The respondents rated the application as "very good" in all categories. Direct testing on children over 5 years showed an increase in the average quiz score before and after using the application, from 11 to 80.

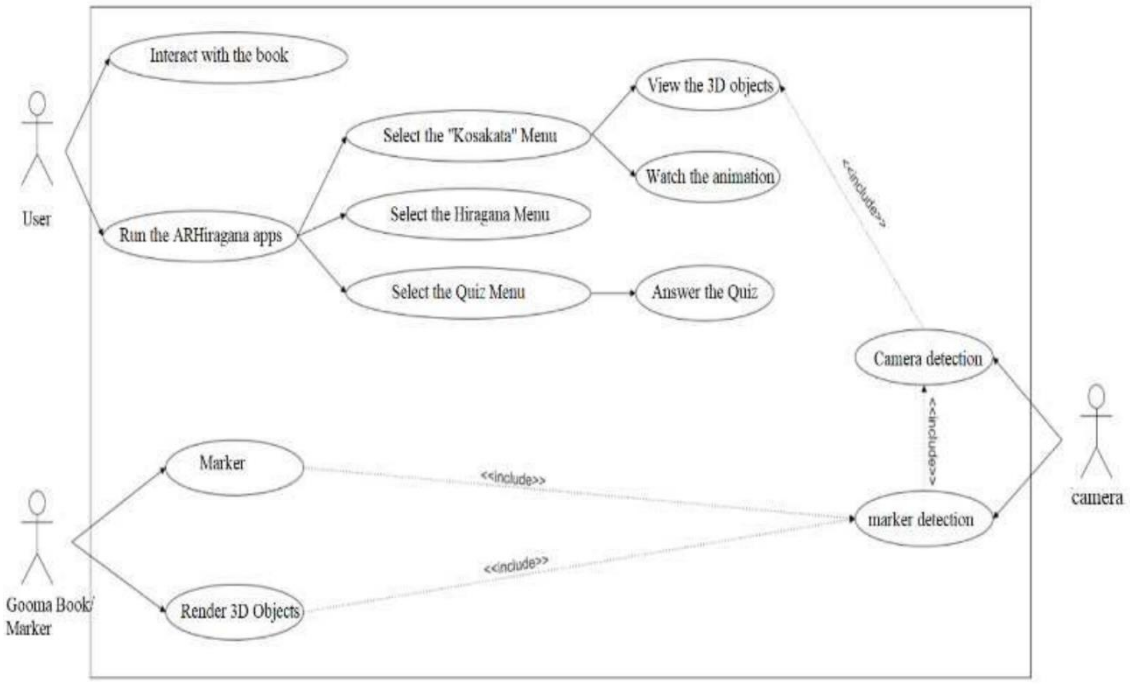


**Figure2. 4** AR Hiragana's Use Case Diagram [12]

## 2.5 Summary

Our research is considered novel because it learns vocabulary for various names with the same meaning from real-world photos. While the previous work implemented "vocabulary learning support using real-world images," it did not mention what are considered challenges or important paraphrases in Japanese vocabulary learning. The novelty of our study is the integration of photographic object recognition and WordNet to enhance the acquisition of Japanese thesauruses.

**Table 2. 1** Comparison between PICSU and related works

| System | Picture | Object Recognition | Japanese | Thesaurus |
|---|---|---|---|---|
| [9] | √ | × | × | × |
| [10] | √ | √ | × | × |
| [11] | √ | √ | √ | × |
| PICSU | √ | √ | √ | √ |

# Chapter 3 Research Design and Methodology

## 3.1 Proposed method

### 3.1.1 Overview

The proposed method consists of an interactive picture-based language learning application developed to help users expand their vocabulary through direct interaction with their surroundings. The system captures images of objects by smartphones, identifies them using a state-of-the-art object recognition algorithm (YOLO), and then provides thesauruses of the identified words by referencing a comprehensive lexical database (WordNet). This method introduces new vocabulary and reinforces learning by associating words with visual elements.

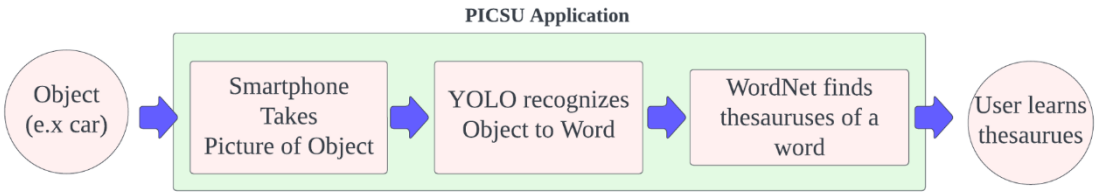


**Figure 3. 1** Workflow of the system

**Object (e.g., car):** This is where the journey begins, with the user selecting an object in their environment to learn.

**Smartphone Takes Picture of Object**: The user then uses a smartphone to photograph the chosen object, which the application will analyze.

**YOLO Recognizes Object to Word**: Our application employs the YOLO algorithm to process the image and accurately identify the object, transforming the visual input into a textual representation.

**WordNet Finds Thesaurus of a Word**: Once the object is recognized, WordNet finds and lists thesauruses for the identified word, offering a rich array of language learning possibilities.

**User Learns Thesaurus**: The final step in the process is where the user is presented with the thesauruses, thus engagingly and interactively facilitating the acquisition of new vocabulary.

In order to encapsulate the complexity of the system's architecture and facilitate a comprehensive understanding, a mathematical model has been contrived, encapsulating the workflow from the initial object recognition to the final review stage:

$$\boldsymbol{F(O) = TR(L(WN(Y(P(O)))))}$$

Specifically:

- $O$ be an object in the real world.
- $P(O)$ be the process of taking a picture of object $O$ using a smartphone.
- $Y(P(O))$ be the process of YOLO recognizing the object in picture $P(O)$ and converting it to a corresponding word W
- $WN(W)$ be the process of WordNet finding the thesauruses $T$ of word $W$.
- $L(T)$ be the learning process where the user learns thesauruses T.
- $TR(L(T))$ be the test and review process based on the learning outcome $L(T)$.

## 3.1.2 YOLO

### 3.1.2.1 Rational for The Selection of YOLO

YOLO, an acronym for "You Only Look Once [13]," is a state-of-the-art, real-time object detection system that stands out in the field of computer vision for its speed and accuracy. It utilizes a single convolutional neural network (CNN) to simultaneously predict multiple bounding boxes and class probabilities for those boxes.

YOLO is chosen for this research due to its remarkable efficiency and effectiveness in processing images. This algorithm can detect objects in real-time, which is crucial for a mobile application where quick and responsive interaction is desired. Additionally, YOLO offers a favorable balance between speed and accuracy, ensuring that the object detection process is not only swift but also precise. Its ability to generalize well from natural images to new domains makes it ideal for an educational tool where users may present a wide variety of objects for recognition.

Some of the reasons why YOLO is ahead of the competition include: speed, detection accuracy, good generalization, and open source. First, YOLO is extremely fast because it does not handle complex pipelines. It can process images at 45 frames per second (FPS). Additionally, YOLO achieves more than twice the average precision (mAP) compared to other real-time systems, making it an excellent candidate for real-time processing. Secondly, YOLO far exceeds other state-of-the-art models in terms of accuracy and has very few background errors. This is especially true for newer versions of YOLO. With these advances, YOLO takes a step forward by providing better generalization to new

domains, making it ideal for applications that rely on fast, powerful object detection.

For example, the paper "Automatic detection of melanoma using Yolo deep convolutional neural network" [14] shows that the first version YOLOv1 has the lowest average accuracy in automatically detecting melanoma disease compared to YOLOv2 and YOLOv3. YOLO's open source also drives the community to continuously improve the model. This is one of the reasons why YOLO has made so many improvements in such a limited time. Table below shows the average accuracy (MAP) comparison of the YOLO version.

**Table 3. 1** Mean Average Precision (MAP) comparison of YOLO versions [14]

| Framework | Benign | Malignant | mAP |
|---|---|---|---|
| **YoloV1** | 0.41 | 0.33 | 0.37 |
| **YoloV2** | 0.85 | 0.82 | 0.83 |
| **YoloV3** | 0.79 | 0.75 | 0.77 |

Released in July 2022 in the paper-Trained bag-of-freebies sets new state-of-the-art for real-time object detectors [15], YOLOv7 sets a new state-of-the-art for real-time object detectors. This version makes significant progress in object detection, surpassing all previous models in accuracy and speed.

YOLOv7 reforms its architecture by integrating the Extended Efficient Layer Aggregation Network (E-ELAN), enabling the model to learn more diverse features for better learning. Additionally, YOLOv7 extends its architecture by connecting the architectures of its derived models, such as YOLOv6 [16], YOLOv8 [17], and YOLO-R [18]. This enables the model to meet the needs of different inference speeds. In addition, bag-of-freebies refers to improving the accuracy of the model without increasing the training cost, which is why YOLOv7 improves the inference speed and detection accuracy.

Table 3.2 compares baseline for YOLO models. In which, each version of YOLO (denoted by prefixes like v4, R-u5, v4-CSP, v6, v7, …) is brought onto the table, along with its evaluation metrics to illustrate the performance in a clearer way.

The metrics shown on the table are:

- #Param.: The number of parameters in the model. A higher number means the model is more complex.
- FLOPs: Floating-point operations per second, an indication of the computational complexity of the model.
- Size: The input resolution size that the model uses.

- AP: Average Precision is a metric used to evaluate the performance of object detection models. It measures the precision (the ratio of true positives to the sum of true positives and false positives) across different levels of recall (the ratio of true positives to the sum of true positives and false negatives). In simpler terms, it assesses how accurate the model is at detecting objects across various thresholds of certainty.
- $Ap^{val}$: Average Precision on the validation set, a common metric to evaluate the accuracy of object detection models. The higher the better.
- $AP^{val}_{50}$, $AP^{val}_{75}$, $AP^{val}_{S}$, $AP^{val}_{M}$, $AP^{val}_{L}$: These are different Average Precision metrics at various IoU (Intersection over Union) thresholds (50% and 75%) and object scales (small, medium, and large). The IoU is a measure of overlap between the predicted bounding box and the ground truth box. $AP^{val}_{50}$ is generally easier to achieve than $AP^{val}_{50}$.

The highlighted percentages indicate the improvement or decline of the YOLOv7 models compared to their counterparts in each metric. Green indicates an improvement, and red indicates a decrease in performance. For example, YOLOv7 has improved AP^val by +1.3 points compared to YOLOv4. The improvements in AP values indicate that YOLOv7 could be more precise and reliable in object detection tasks compared to the other YOLO versions, such as YOLOv4 and its variations. Additionally, despite its high accuracy, YOLOv7 has fewer parameters and requires fewer FLOPs than some of the other models, suggesting that it could also be more efficient to run, which is an important consideration for real-time applications.

**Table 3. 2** Comparison of baseline object detectors [15].

| Model | #Param. | FLOPs | Size | $AP^{val}$ | $AP^{val}_{50}$ | $AP^{val}_{75}$ | $AP^{val}_{S}$ | $AP^{val}_{M}$ | $AP^{val}_{L}$ |
|---|---|---|---|---|---|---|---|---|---|
| **YOLOv4 [3]** | 64.4M | 142.8G | 640 | 49.7% | 68.2% | 54.3% | 32.9% | 54.8% | 63.7% |
| **YOLOR-u5 (r6.1) [81]** | 46.5M | 109.1G | 640 | 50.2% | 68.7% | 54.6% | 33.2% | 55.5% | 63.7% |
| **YOLOv4-CSP [79]** | 52.9M | 120.4G | 640 | 50.3% | 68.6% | 54.9% | 34.2% | 55.6% | 65.1% |
| **YOLOR-CSP [81]** | 52.9M | 120.4G | 640 | 50.8% | 69.5% | 55.3% | 33.7% | 56.0% | 65.4% |
| **YOLOv7** | 36.9M | 104.7G | 640 | **51.2%** | **69.7%** | **55.5%** | **35.2%** | **56.0%** | **66.7%** |
| improvement | -43% | -15% | - | +0.4 | +0.2 | +0.2 | +1.5 | = | +1.3 |
| **YOLOR-CSP-X [81]** | 96.9M | 226.8G | 640 | 52.7% | **71.3%** | 57.4% | 36.3% | 57.5% | 68.3% |
| **YOLOv7-X** | 71.3M | 189.9G | 640 | **52.9%** | 71.1% | **57.5%** | **36.9%** | **57.7%** | **68.6%** |
| improvement | -36% | -19% | - | +0.2 | -0.2 | +0.1 | +0.6 | +0.2 | +0.3 |
| **YOLOv4-tiny [79]** | 6.1 | 6.9 | 416 | 24.9% | 42.1% | 25.7% | 8.7% | 28.4% | 39.2% |
| **YOLOv7-tiny** | 6.2 | 5.8 | 416 | **35.2%** | **52.8%** | **37.3%** | **15.7%** | **38.0%** | **53.4%** |
| improvement | +2% | -19% | - | +10.3 | +10.7 | +11.6 | +7.0 | +9.6 | +14.2 |
| **YOLOv4-tiny-3l [79]** | 8.7 | 5.2 | 320 | 30.8% | 47.3% | 32.2% | **10.9%** | 31.9% | 51.5% |
| **YOLOv7-tiny** | 6.2 | 3.5 | 320 | **30.8%** | **47.3%** | **32.2%** | 10.0% | **31.9%** | **52.2%** |
| improvement | -39% | -49% | - | = | = | = | -0.9 | = | +0.7 |
| **YOLOR-E6 [81]** | 115.8M | 683.2G | 1280 | 55.7% | 73.2% | 60.7% | 40.1% | **60.4%** | 69.2% |
| **YOLOv7-E6** | 97.2M | 515.2G | 1280 | **55.9%** | **73.5%** | **61.1%** | **40.6%** | 60.3% | **70.0%** |
| improvement | -19% | -33% | - | +0.2 | +0.3 | +0.4 | +0.5 | -0.1 | +0.8 |
| **YOLOR-D6 [81]** | 151.7M | 935.6G | 1280 | 56.1% | 73.9% | 61.2% | **42.4%** | 60.5% | 69.9% |
| **YOLOv7-D6** | 154.7M | 806.8G | 1280 | 56.3% | 73.8% | 61.4% | 41.3% | 60.6% | 70.1% |
| **YOLOv7-E6E** | 151.7M | 843.2G | 1280 | **56.8%** | **74.4%** | **62.1%** | 40.8% | **62.1%** | **70.6%** |
| improvement | = | -11% | - | +0.7 | +0.5 | +0.9 | -1.6 | +1.6 | +0.7 |

Studies posit that utilizing substantial datasets for training enhances the precision and reliability of object recognition, which, in turn, contributes to the accuracy of subsequent learning processes [19]. Based on those results, we will employ the YOLOv7 algorithm for capturing and detecting images within our research [20]. This technique will be instrumental in advancing the development of 'PICSU', our Japanese learning support system designed to function as an application program.

## 3.1.2.2 Architecture of selected YOLO Model

Figure3.2 shows the YOLOv7 architecture used in this research. It has 24 convolutional layers, four max pooling, and two fully connected layers. The architecture works as follows:

- Resize the input image to 448x448 before passing it through the convolutional network.
- A 1x1 convolution is applied first to reduce the number of channels, and then a 3x3 convolution is applied to produce a cubic output.
- Except for the last layer, which uses a linear activation function, the underlying activation function is ReLU.
- Some additional techniques, such as batch normalization and dropout, regularize the

model and prevent it from overfitting.

Figure 3.2 is the specific architecture of YOLO.

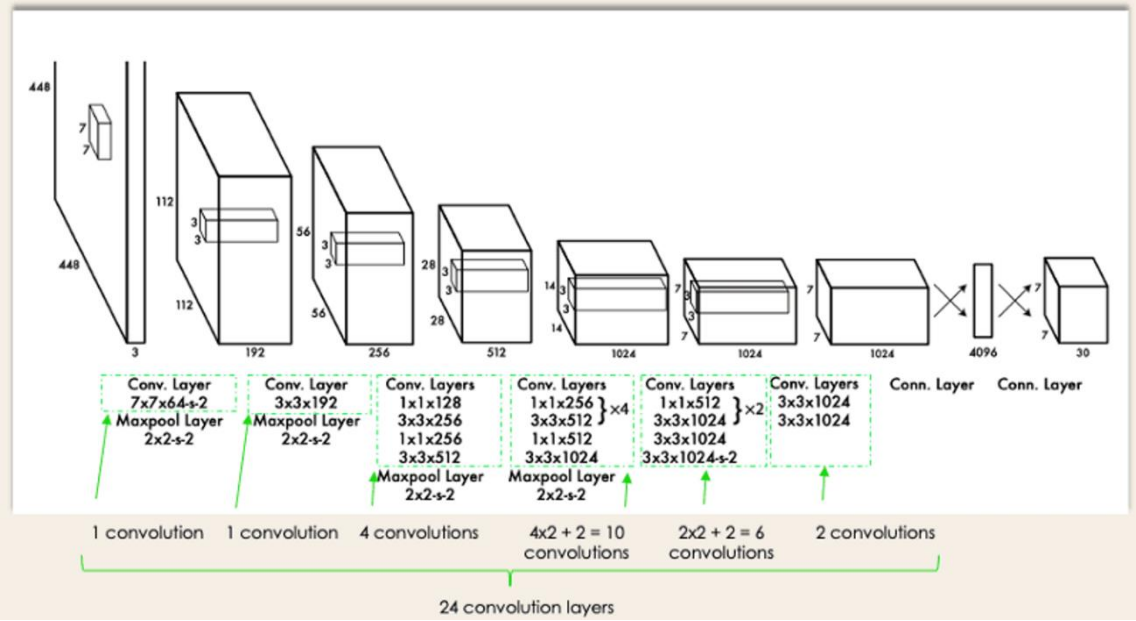


**Figure 3. 2** YOLO Architecture[21]

## 3.1.2.3 Integration of YOLO

The integration of the YOLOv7 model into the PICSU System is a testament to the system's advanced object recognition capabilities within a language learning context. Through the strategic use of the smartphone's SDK, YOLOv7 connects and communicates efficiently with the system's other modules, allowing for seamless interaction and data exchange. This symbiotic integration ensures that the object detection process is not only swift but also accurate and user-friendly, enhancing the overall effectiveness of the PICSU application.

**Image Capture and Input:** Users capture images via their smartphone, which are then inputted into the PICSU System using the platform-specific Android/iOS SDK.

**YOLOv7 Object Detection:** The YOLOv7 neural network receives the image for processing, employing its sophisticated algorithms to detect and identify the object within.

**Textual Output Generation:** Once the object is detected, YOLOv7 outputs the corresponding object name as text.

**SDK Mediated Communication:** The SDK facilitates the transfer of the identified word from the YOLOv7 model to other PICSU System modules.

**WordNet Thesaurus Retrieval:** The system then utilizes the identified word to query

WordNet, retrieving relevant thesauruses that enhance the user's learning experience.

**Cross-Platform Compatibility:** The integration is designed to be compatible with multiple operating systems, namely Android, iOS, and Windows, ensuring a broad user base can access the PICSU System.

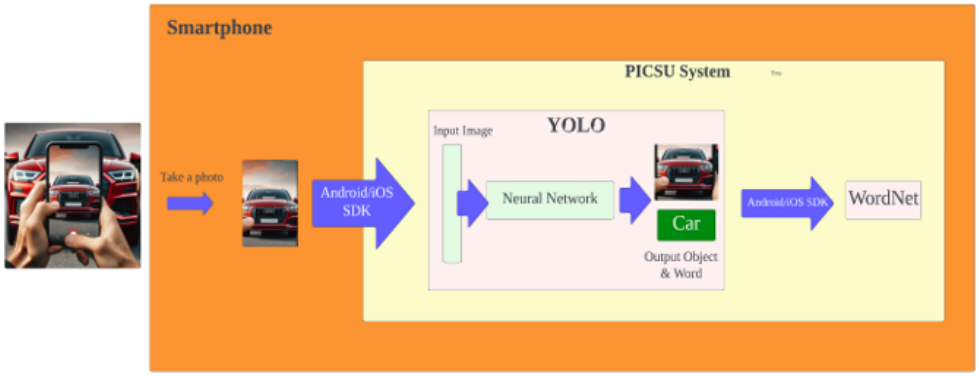


**Figure 3. 3** Integration of YOLO into PICSU system.

## 3.1.3 Japanese Wordnet

### 3.1.3.1 Introduction

Japanese Wordnet is a comprehensive vocabulary database developed by the National Institute of Information and Communications Technology of Japan in 2006 [22]. It was first released in 2008 and includes a Japanese version of WordNet that corresponds to the Princeton Thesaurus [23] (a set of thesauruses with a common meaning).

Figure 3.4 shows some results when experimenting with WordNet. In which I try to print all thesauruses of kuruma. The process can be described as:

1. Import WordNet module.
2. Get the WordNet instance of Japanese.
3. Initialize synsets object of Japanese word “車”.
4. Print out all the words in each set which will be thesauruses of the initialized word.
5. Other thesauruses are shown in the figure.

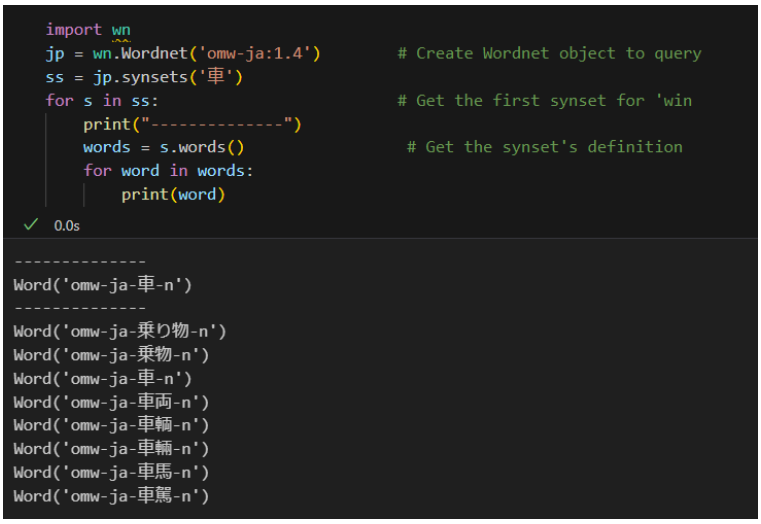


**Figure 3. 4** Integration of YOLO into PICSU system.

## 3.1.3.2 Integration of WordNet

The mechanism of WordNet in Figure 3.5 can be explained as following:

1.**Input:** The process begins when the system receives an input, in this case, the English word "Car".

2.**WordNet Module:**

•The input word is first processed through an English Japanese Common Dictionary within the WordNet module. This dictionary translates "Car" into its standard Japanese equivalent, providing an intermediate output.

•This intermediate output, which is the Japanese word for "car" (自動車 or 車), is then utilized within the module to look up thesauruses.

•A Thesauruses Lookup Table within the WordNet module is then queried with the Japanese word to find associated thesauruses.

3.**Output:** The final output of the process is a list of Japanese thesauruses for "car," such as 自動車 (jidosha), かー (ka), 車 (kuruma), and 乗り物 (norimono), which are then presented to the user.

Figure below illustrates the workflow of the mechanism:

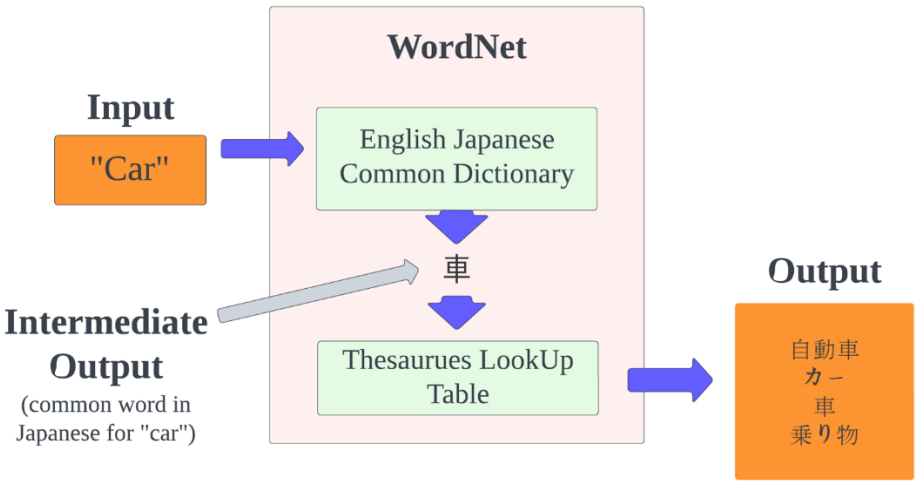


**Figure 3. 5** Mechanism of WordNet module

## 3.2 User Interaction Design

This section describes the preliminary design of the functions of the PICSU system screen before the implementation, including functions such as learning, review, and score viewing. In addition, there are also system screens used in the testing phase, which were all implemented when the system was finally implemented and merged.

Figure 3.6 is the initial design of the learning screen. The first image is a smartphone app mockup showcasing a red sports car image, the labeled word "Sports car," and its thesauruses. The second image outlines the app's workflow, displaying photos, vocabulary, and thesauruses.

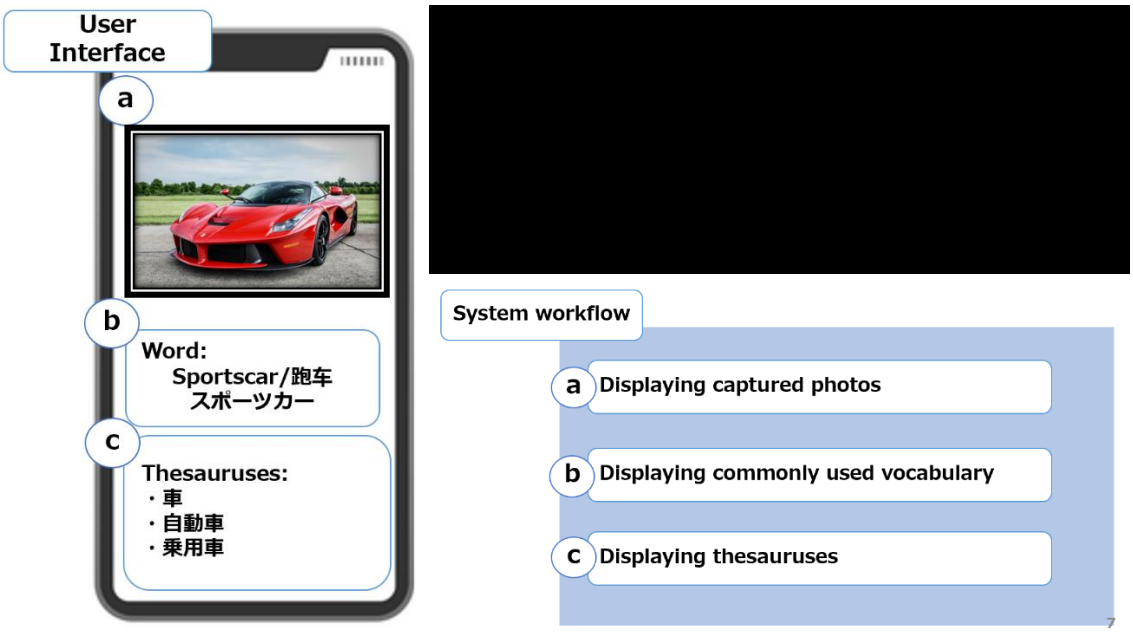


**Figure 3. 6** Initial design of learning screen

The three screens provided in Figure 3.7 illustrate the user interface flow for the PICSU application, guiding the user from the welcome screen to the main functional area of the app. Here's the workflow explained:

**1.Welcome Screen:**

•The user is greeted with a "Welcome to PICSU" message. This is the initial screen when opening the application.

•There is a "Start" button, which presumably the user would press to proceed to the next stage of the application, which is logging in or signing up.

**2.Login/Sign Up Screen:**

•After pressing "Start," the user is taken to a screen where they can enter their ID (username) and Password.

•This screen provides two options: "Login" for returning users to access their account and "Sign Up" for new users to create an account.

•Once the user enters their credentials and selects either to log in or sign up, they are taken to the main page of the application.

**3.Main Page Screen:**

•Upon successful login or sign-up, the user is welcomed with a personalized greeting, "Hello, [ID]!" Where [ID] is the user's account name.

•The main page presents three options: "Learning," "Review," and "Score," which the user can navigate through to access different features of the app.

•"Learning" would start the language learning activity, "Review" allows revisiting previously learned materials, and "Score" shows the user's progress or results.

•There are also navigation buttons "Back" and "Next," allowing the user to move to previous screens or proceed to the next part of the app, respectively.

•This user interface design allows for a straightforward and user-friendly experience, guiding the user intuitively through the process of starting the app, logging in, and accessing the main learning features.

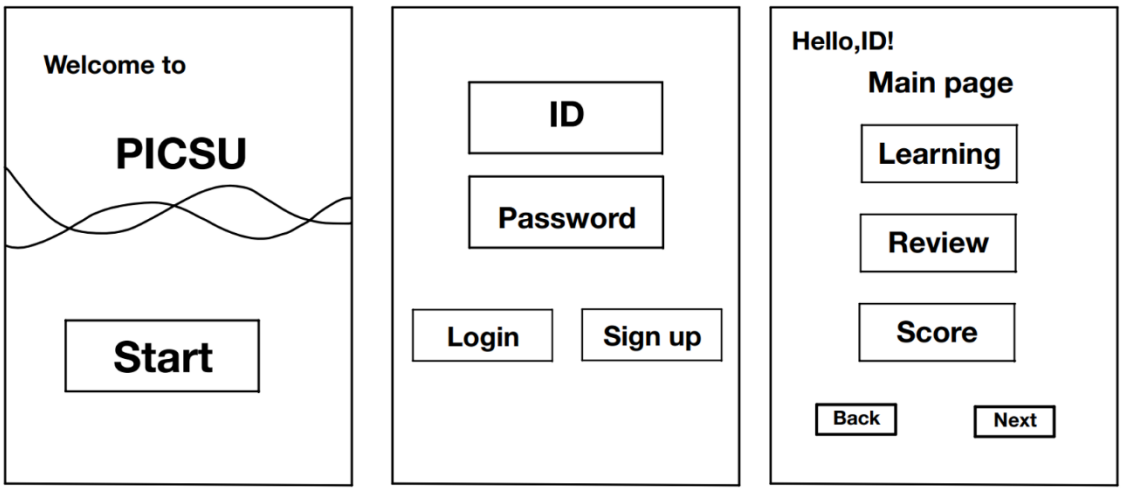


**Figure 3. 7** Initial design of PICSU main interface

These 3 screens in Figure 3.8 represent sequential steps in the user interaction in PICSU:

1.**First Learning Screen:** The user begins the learning activity, presented with a visual (picture) and the corresponding word in the source language. Three potential thesauruses are listed, with options to select 'Yes' or 'No' to indicate familiarity. Navigating the session is facilitated by 'Back' and 'Next' buttons.

2.**Final Learning Screen:** This continues the learning process with another visual and associated word and options to affirm knowledge of the listed thesauruses. The 'Finish' button suggests this is the final step in the learning activity sequence.

3.**Main Page Screen:** After completing the learning activities, the user is brought to a main page, which greets them by ID and offers options to continue 'Learning', 'Review' previous sessions, or check 'Score' for progress. Users can navigate to the previous screen or forward with the 'Back' and 'Next' buttons.

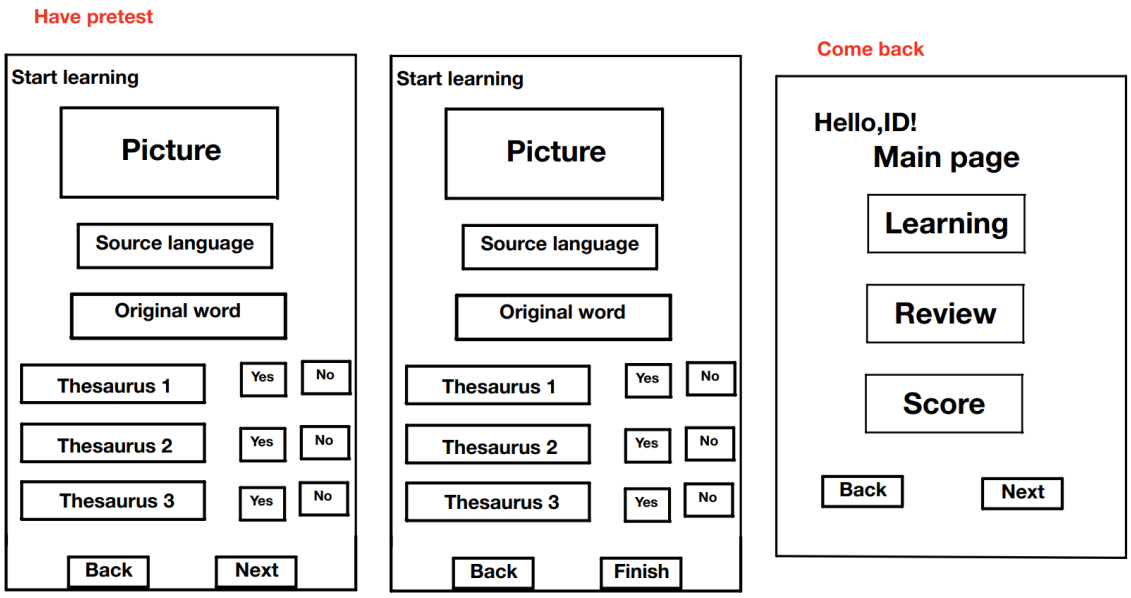


**Figure 3. 8** PICSU initial pre-test interface

The last five screens in Figure 3.9 illustrate the user interface flow for a testing group within a language learning application, focusing on vocabulary testing and learning methods.

1.**Test Start Screen:** The user is presented with a "Test" title and a "Start" button, initiating the testing phase.

2.**Main Page Screen:** After starting, the user is greeted with "Hello, [ID]!" indicating a personalized experience. The main page offers two options: "Learning method" to choose the study approach and "Score" to view the user's testing results.

3.**Learning Method Selection Screen:** The user can select between two different learning methods, "PICSU" and "Fc" (Flashcard), each with two sets to choose from. The user can navigate back or proceed with the "Next" button.

4.**Testing Screens:** These screens appear during the testing phase, where words are presented alongside their explanations in the source language. The user can navigate through different words using the "Back" and "Next" buttons, with the final screen providing a "Finish" button to conclude the test.

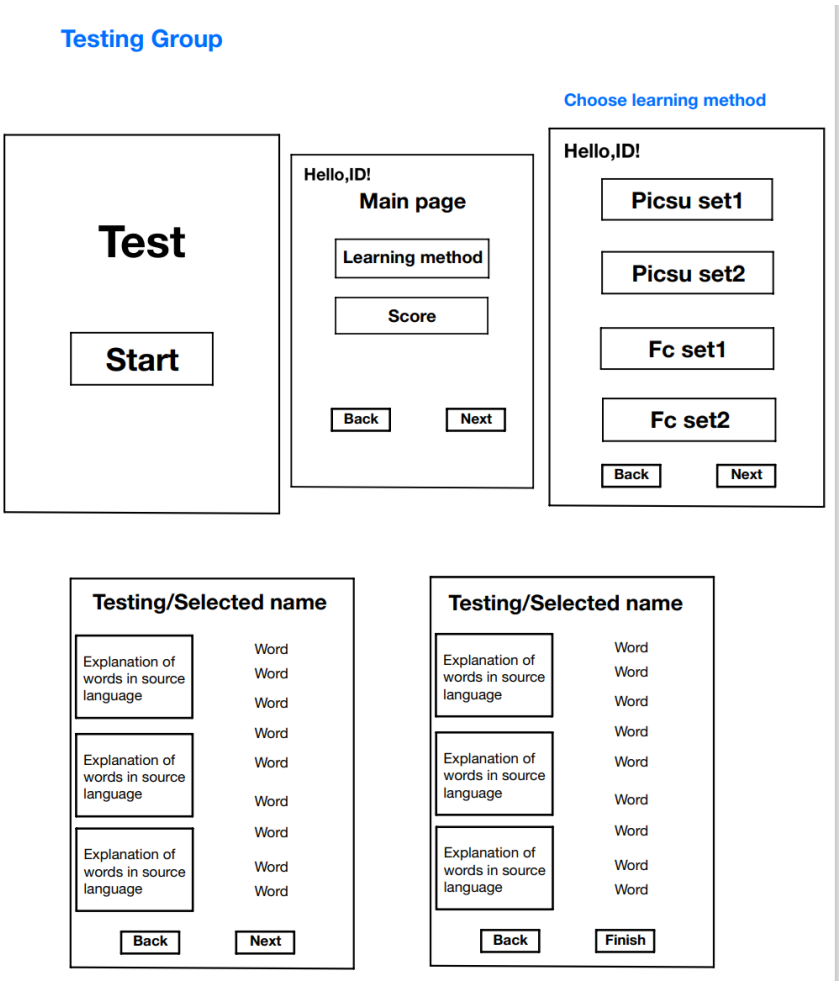


**Figure 3. 9** Initial test interface design

Based on the design of the initial user interface and test interface, many modifications were made in the process of completing the implementation of the PICSU learning support system. Based on the original design, two experimental methods were designed: PICSU's picture learning method and the traditional flashcard learning method for comparative experiments, and after learning, the learned word parts were tested respectively. result. The specific experimental PICSU design interface is introduced in the next section.

## 3.3 PICSU System

### 3.3.1 Overview

PICSU is a sophisticated web application designed to facilitate the learning and evaluating Japanese vocabulary. Developed as a comprehensive and user-friendly platform, PICSU aims to bridge the gap between conventional learning methods and the evolving needs of modern learners. The application incorporates an intuitive interface, allowing users to navigate various learning modules and assessment tools seamlessly.

### 3.3.2 Technical Specifications

**Platform Type**

- Web Application: Accessible through standard web browsers, ensuring cross-platform compatibility and ease access without downloading additional software.

**Programming Languages**

- Python: The back-end logic and server-side functionalities are primarily developed in Python, known for their readability and efficiency.
- HTML/CSS/JavaScript: Front-end development is executed using HTML, CSS, and JavaScript, providing an engaging and responsive user interface.

**Framework**

- Django Web Framework: Utilizes Django, a high-level Python web framework that encourages rapid development and clean, pragmatic design. Django's robust and scalable architecture is ideal for handling the extensive database interactions and user management required by PICSU.

**Database**

- PostgreSQL: Incorporates PostgreSQL as the relational database management system. Chosen for its reliability, flexibility, and compatibility with Django, PostgreSQL effectively manages the extensive data sets involved in vocabulary learning, including user progress tracking, vocabulary data, and assessment records.

**Cloud Hosting**

- HEROKU by Salesforce: Hosted on HEROKU, a cloud platform service by Salesforce. This hosting solution ensures high availability, easy scalability, and seamless deployment, contributing to an uninterrupted and consistent user experience.

### 3.3.3 User interface of PICSU

The user interface of PICSU used in the experiment is introduced here. According to this design, participants can complete vocabulary learning in PICSU and traditional flash card vocabulary learning, respectively ,and take the test after learning and the delayed test two days later.

Figure 3.10 is the PICSU login interface, with two options: login and register.

Figure 3.11 is the PICSU registration interface. Learners need to set the password twice. If they forget the password, they can also retrieve it.

Figure 3.12 is the PICSU selection login interface. If learners already have an account, they can log in directly.

Figure 3.13 is the main interface of PICSU, where user information and learning vocabulary library are displayed, and learning or testing steps can be selected.

Figure 3.14 shows the PICSU pre-test interface. Here is the first step of the experiment: selecting a known word and learning thesauruses in the subsequent learning process. The steps are to keep the known words and choose the unknown words. The interface is a scrolling answer method. Click the “Do not Know” button, and the word will disappear automatically.

Figure 3.15 is the PICSU learning interface, which provides pictures in Chinese and English to help users learn thesauruses of vocabulary.

Figure 3.16 is the flashcard learning interface. In this learning, learners click on the card to flip it. The front of the card shows the original word and its Chinese and English translations, while the back has three thesauruses to learn.

Figure 3.17 shows user test interfaces. There is a time limit during the test. It will automatically end in 15 minutes, as shown in the picture. Learners can choose from multiple options of thesauruses they have learned based on the Chinese or English translation of the prompts, type them all within the time, and then submit them smoothly.

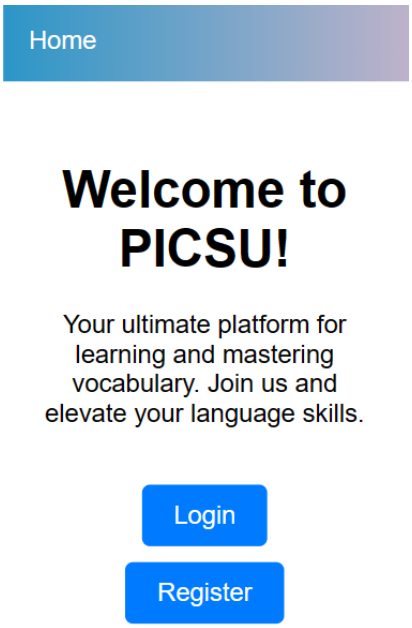


**Figure 3. 10** PICSU login interface

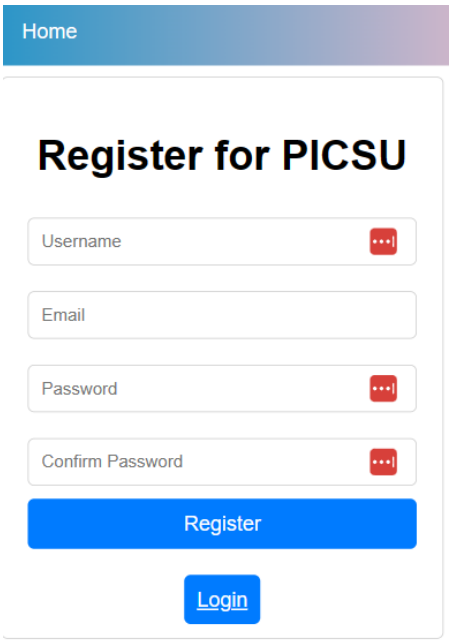


**Figure 3. 11** PICSU registration interface

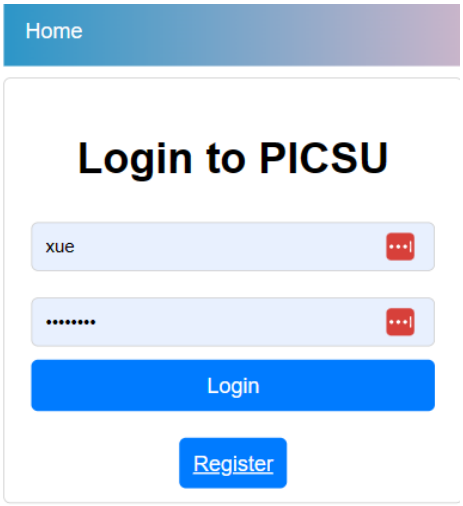


**Figure 3. 12** PICSU choose login interface.

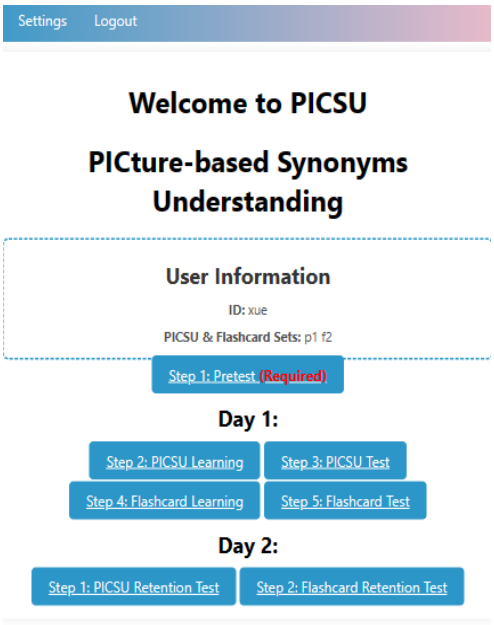


**Figure 3. 13** PICSU main interface


**Figure 3. 14** PICSU pre-test interface


**Figure 3. 15** PICSU learning interface.


**Figure 3. 16** Flashcard learning interface.

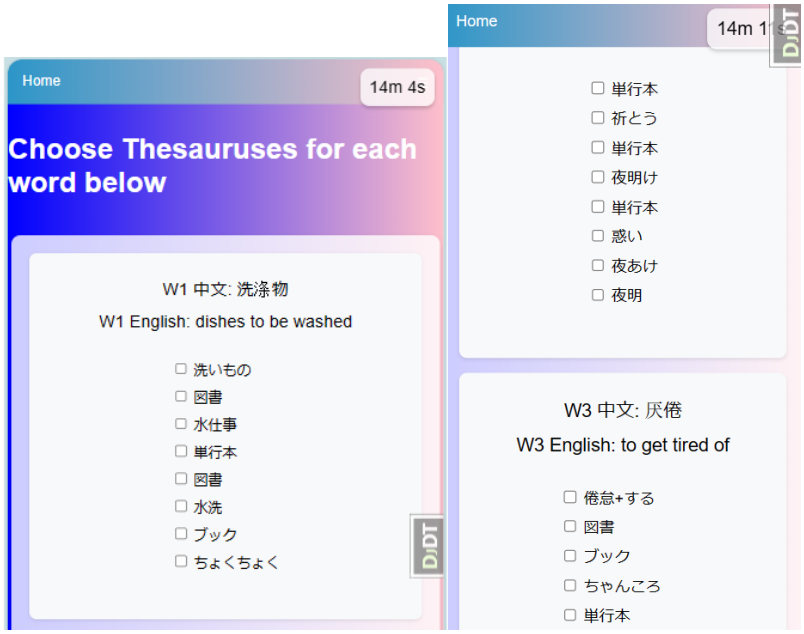


**Figure 3. 17** Test interface

## 3. 4 Target users

The subjects of this study are international students who are not native speakers of Japanese and mainly use Chinese and English language assistance. If the experiment is the goal, the Japanese language level of the participants must be at least JLPT N3 or above.

Among these learners, instead of using Japanese grammar and meaning to learn Japanese, they can achieve the effect of using Japanese vocabulary correctly by filtering known words and learning their thesauruses. At the same time, Chinese and English are also provided during the learning process, and thesauruses are learned concerning the original words.

# Chapter 4 Experimentation

## 4.1 Experiment introduction

### 4.1.1 Experimental methods and purposes

This experiment evaluates the efficacy of two Japanese vocabulary learning methods: a picture-based learning system (PICSU) and a traditional flashcard method. The purpose of this study is to investigate which methods, such as visual image integration and repetition confirmation, more effectively increase learning efficiency and memory retention in Japanese language acquisition. It involved 20 students from JAIST, including 19 Chinese and 1 Vietnamese participants, they consist of 10 males and 10 females, with diverse Japanese language proficiencies: 8 held JLPT N1 certification, 8 had N2, 4 were at N3 level. Their ages ranged from 21 to 30, with an average age of 26. The study involved learning Japanese noun thesauruses using both the PICSU system and flashcards, followed by tests to evaluate word recall and retention. The 20 students were divided into four groups to minimize the impact of learning content and order differences:
Group 1: PICSU set1, fc set2: 5 people
Group 2: PICSU set2, fc set1: 5 people
Group 3: fc set1, PICSU set2: 5 people
Group 4: fc set2, PICSU set1: 5 people
Total: 20 people

### 4.1.2 Research Design

The research will implement a within-subjects design to compare the efficacy of two different methods for studying Japanese thesaurus. All participants will be exposed to the PICSU, a picture-based learning system, and a traditional flashcard method. Two distinct sets of non-overlapping vocabulary will be assigned, one for each learning method, to ensure the comparison is valid.

Participants will first engage with the PICSU method to study the first set of vocabulary. Following this, they will utilize the flashcard method to learn the second set of vocabulary. This approach will mitigate any potential order effects, with half of the participants

starting with the PICSU method and the other half beginning with the flashcard method. Upon completing each learning session, participants will take a test designed to assess their acquisition of the vocabulary relevant to the method just used.

By comparing test results, the study aims to determine which method is more effective for learning Japanese thesauruses. The within-subjects design allows each participant to serve as their own control, enhancing the study's internal validity by controlling individual differences in learning ability and prior knowledge. The design of this study is divided into the following parts:

1.Developing a PICSU System that offers:

(1)A picture-based thesauruses learning method powered by object recognition technology and a large vocabulary corpus.

(2)A traditional flashcard learning method for thesaurus acquisition.

2.Creating vocabulary sets respective to the two learning methods above and designing and setting up the experiment.

3.Designing metrics to evaluate the efficiency of the vocabulary learning system.

4.Analyzing data from the experiment to compare the effectiveness of the picture-based method (PICSU) and the traditional flashcard method of learning Japanese thesauruses.

## 4.2 Flow of the experiment

Flow of the Experiment: Comparative Study on Vocabulary Learning Methods

Total Duration: 3 hours

Learning and first test 2 hours ＋ ②delay test 1 hour(After 2 days or more (within 1 week))

①Learning and first test 2 hours：

0) Introduction and Setup (10 minutes):
   - Welcome participants and explain the purpose of the study.
   - Provide instructions about the two different learning methods and the overall experiment.

1) Pre-test (10 minutes)
   - Check if participants already know the word for which they want to learn thesauruses.

2) Learning Japanese Vocabulary/ Thesauruses with PICSU System (25 minutes):
   - Participants engage in learning Japanese vocabulary and thesauruses using the

PICSU system.

- Ensure that participants have adequate resources and support during this phase.

3) First Rest Period (5 minutes):

- Participants will take a short break to relax and refresh.

4) First Recall Test (15 minutes):

- Conduct a test to assess how many words participants remember from the PICSU system session.

5) Second Rest Period (10 minutes):

- Another break for participants to prepare for the next learning method.

6) Learning Japanese Vocabulary with Flashcard Method (25 minutes):

- Participants switch to learning vocabulary using flashcards.
- This session will have the same level of difficulty and duration as the first session.

7) Third Rest Period (5 minutes):

- A final rest period before the last test.

8) Second Recall Test (15 minutes):

- A test to evaluate how many words participants remember from the flashcard session.
- Like the first test, this should measure the effectiveness of the flashcard method in vocabulary retention.

②delay test 1 hour(After 2 days or more (within 1 week))：

1) PICSU memory Test (20 minutes)

- A test will be conducted to assess the words participants remember from the PICSU system session three days later.

2) Break Time (10 minutes)

- Participants will take a short break to relax and refresh.

3) Flashcard Memory Test (20 minutes)

- This test evaluates the words participants remember from the flashcard session three days later.

4) Questionnaire (10 minutes):

- Participants will be asked to respond to a post-experiment questionnaire.
- Offer an opportunity for participants to give feedback or ask questions about the study.

Figure 4.1 shows the experiment flow, which lasted 3 hours in total and was divided into two days. The pre-test, study, and testing phases dominate the first day. The second day

is a delayed test conducted two days later to determine whether the learner's memory results are valid.

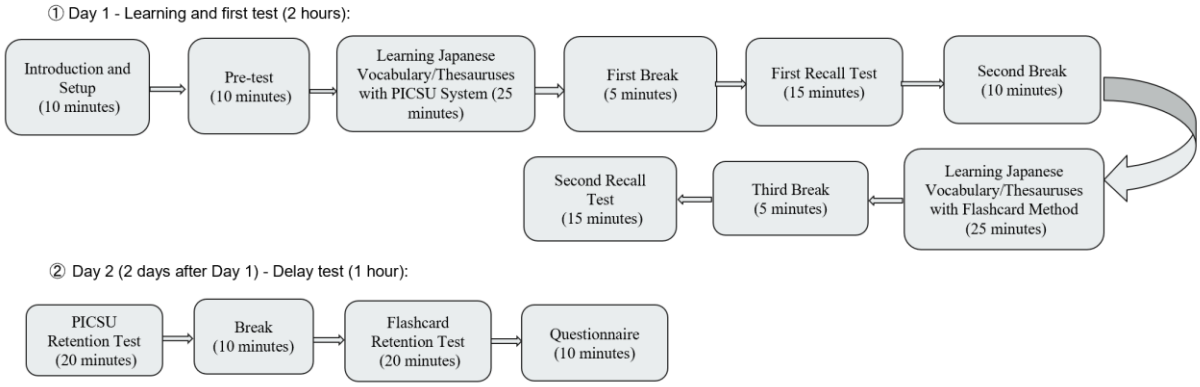


**Figure 4. 1** Flow of experiment

The questionnaire has three questions in total, which are as follows:

1.Which learning method did you prefer and why?

2.What improvements, if any, would you suggest for each learning method?

3.Please feel free to write down any observations or insights you have regarding this experiment.

## 4.3 Dataset Preparation

To standardize experimental conditions as well as control the learning environment, this study used the Japanese Language Education Vocabulary List [24] (http://jhlee.sakura.ne.jp/JEV/). The Japanese word database used in this study is mainly the vocabulary introduced in the Japanese education vocabulary list published on the Internet. After screening to a certain degree of difficulty, it is roughly at the level of Japanese JLPT N3 or above, and the names are used as the learning. The center then uses Wordnet to convert thesauruses and sets it so that each original word has three corresponding learning thesauruses. In addition, in this study's dataset, there are Chinese and English translations corresponding to the original words, which helps learners understand and remember when learning thesauruses.

The data of this experiment is divided into two groups: the PICSU vocabulary learning library with pictures and the traditional flashcard library without pictures. Divide the two

sets of data into two sets of the same amount of data again and set different vocabulary libraries for different learning objects to obtain stable data results.

In total, 320 words are chosen to form the dataset. In which, PICSU words set contains 160 words, and Flashcard words set contains 160 other words. Each learning method's dataset is then divided into 2 non-overlapping subsets with equal 80 words: subset P1 and P2 as for PICSU, subset F1 and F2 for Flashcard. The 4 subsets are permuted together into 4 pairs: (P1, F1), (P1, F2), (P2, F1), (P2, F2). Each pair later will be randomly assigned to an arbitrary participant, and the order of 2 subsets in the pair can be changed.

Figure 4.2 is the vocabulary library of PICSU, which mainly includes the original word, three thesauruses of the original word to be learned, Chinese and English explanations, and pictures for reference learning.

| Number | Word | Thesaurus1 | Thesaurus2 | Thesaurus3 | Chinese | English | Image_link |
|---|---|---|---|---|---|---|---|
| 1 | 相手方 | 対戦相手 | 対手 | 対抗馬 | 对方 | Opponent | https://previews.123rf.com/images/zcreamz11/zc |
| 2 | 洗い物 | 水仕事 | 水洗 | 洗いもの | 洗涤物 | dishes to be washed | https://www.photolibrary.jp/mhd6/img182/450-2( |
| 3 | 挨拶 | 会釈 | 式礼 | 御挨拶 | 问候 | greeting | https://4.bp.blogspot.com/-_Lx0MQ57Irw/VXOUJI |
| 4 | 青色 | ブルー | 青み | 青 | 蓝色 | blue | https://www.beiz.jp/images_P/blue/blue_00081.j |
| 5 | 青空 | 碧天 | 碧空 | 碧落 | 蓝天 | blue sky | https://www.beiz.jp/images_P/sky/sky_00198.jpg |
| 6 | 赤色 | レッド | さ丹 | 丹色 | 红色 | red | https://illustimage.com/photo/dl/1703.png |
| 7 | 赤ちゃん | ベイビー | ねんね | ベビー | 婴儿 | baby | https://1.bp.blogspot.com/-laOgJ0DLwb4/UUhH3 |
| 8 | 暁 | 夜あけ | 夜明け | 夜明 | 黎明 | dawn | https://pbs.twimg.com/media/EUJdi8SWAAET2vr? |
| 9 | 秋 | オータム | フォール | 商秋 | 秋天 | autumn | https://4.bp.blogspot.com/-8KI4d-zdUn0/VQF_IAL |
| 10 | 空き巣 | 偸盗 | 泥坊 | 泥棒 | 空巢 | burglary | https://1.bp.blogspot.com/-qeJHpS-EXtc/VoziF89 |
| 11 | 飽きる | うんざり+する | 倦む | 倦怠+する | 厌倦 | to get tired of | https://shori.link/wp-content/uploads/2018/07/9 |
| 12 | 開く | オープン+する | 打ち開く | おっ開く | 开 | to open | https://1.bp.blogspot.com/-UG7ci_zxYes/X4aVsG |
| 13 | 悪 | 不届き | 不当 | 不正 | 坏 | bad | https://4.bp.blogspot.com/--WcbDAiKCLM/U-8Fx |
| 14 | 悪質 | ちんけ | 下等 | 不良 | 劣质 | poor quality | https://3.bp.blogspot.com/-hG8weryELEw/WVdjz |
| 15 | 悪態 | ののしり | 下品な言葉 | 冒涜的な間投詞 | 恶言 | abuse | https://mainichigahakken.net/hobby/img/033-00 |
| 16 | 憧れ | アンビション | ドリーム | 冀望 | 憧憬 | aspiration | https://3.bp.blogspot.com/-jgfbyHohGZM/WASJM |
| 17 | 朝 | モーニング | 午前 | 昼まえ | 早晨 | morning | https://2.bp.blogspot.com/-d7wvDnB61sw/VRE4V |
| 18 | 欺く | いかさま+する | いんちき+する | ごまかす | 欺骗 | to deceive | https://stat.ameba.jp/user_images/20221011/05/ |
| 19 | 預ける | 任せる | 任す | 信託+する | 寄存 | to deposit | https://1.bp.blogspot.com/-NyaG5T2ZVgg/WIGp0 |
| 20 | 温かい | 厚い | 懇ろ | 手厚い | 温暖 | warm | https://1.bp.blogspot.com/-y48MZUI83IA/XNE-uA |
| 21 | 温める | 加温+する | 加熱+する | 暖める | 加热 | to warm up | https://1.bp.blogspot.com/-2FAGaqCb4qo/XobTW |
| 22 | 穴 | うろ | 孔 | 洞 | 洞 | hole | https://1.bp.blogspot.com/-HDQZ488nCTo/Xwkxb |
| 23 | 油絵 | オイルペインティング | 油彩 | 油画 | 油画 | oil painting | https://4.bp.blogspot.com/-rqHX3m-8A_8/WEOPH |
| 24 | アボカド | アヴォカード | アボカード | アヴォカド | 牛油果 | avocado | https://4.bp.blogspot.com/-Ms9kyuWO-II/UnyGD( |
| 25 | 尼 | 修道女 | 修道尼 | 女僧 | 尼姑 | nun | https://4.bp.blogspot.com/-TJwrNqSND1U/VxC3e |

**Figure 4. 2** PICSU's vocabulary library

Figure 4.3 is the vocabulary library of Flashcard. It contains vocabulary at the same difficulty level as the vocabulary library of PICSU. The difference is that there are no pictures for reference learning.

| Number | Word | Thesaurus1 | Thesaurus2 | Thesaurus3 | Chinese | English |
|---|---|---|---|---|---|---|
| 1 | 後々 | フューチャー | 今後 | 以後 | 以后 | in the future |
| 2 | 余り物 | あまり | 余りもの | 余り | 剩余物 | leftovers |
| 3 | あるいは | ひょっとしたら | ひょっとすると | ひょっとして | 或者 | or |
| 4 | 言い出す | おっしゃる | いう | 云う | 开始**说** | to start talking |
| 5 | 生き残り | サヴァイヴァル | サバイバル | 存命 | 幸存者 | survivor |
| 6 | 幾らか | 一掬 | 些少 | 僅 | 一些 | a little |
| 7 | 一時に | いっぺんに | いっきに | 一度に | 一**时** | temporarily |
| 8 | 一部 | セクション | パーツ | パート | 一部分 | a part |
| 9 | 一刻 | かたくな | いこじ | えこじ | 一刻 | a moment |
| 10 | 一体 | ひととおり | おおかた | たいてい | 究竟 | really |
| 11 | 一式 | ひとまとまり | ひとそろい | セット | 一套 | a full set |
| 12 | 一寸 | いくぶん | ずいぶん | ちょいと | 一点 | a little |
| 13 | 相変わらず | あいかわらず | 何時ものごとく | 何時も通り | 依旧 | as usual |
| 14 | 愛護 | プロテクション | 保護 | 守り | **爱护** | protection |
| 15 | 愛好 | 傾倒 | 尊崇 | 崇拝 | **爱**好 | hobby |
| 16 | 愛国心 | パトリオティズム | 愛国 | 愛国主義 | **爱**国心 | patriotism |
| 17 | 合い言葉 | キャッチワード | スローガン | モットー | 口号 | slogan |
| 18 | 愛する | いとおしがる | いとおしむ | 傾慕する | **爱** | to love |
| 19 | 間 | コネクション | コネ | 中らい | **间**隔 | space |

**Figure 4. 3** Flashcard's vocabulary library

Also, to ensure that the YOLOv7 can produce a good performance for the system, we tried to run the object detection with collected images used later in the experiment. The purpose is to observe the accuracy of the model. Table 4.1 shows the results of our dataset totally containing 160 prepared pictures for PICSU.

**Table 4. 1** Evaluation of YOLOv7 on our dataset of learning pictures.

| | mAP@50 | Precision | Recall |
|---|---|---|---|
| YOLOv7 | 86% | 91% | 88% |

Upon reviewing the evaluation metrics, it is evident that the model demonstrates satisfactory accuracy levels. However, it encounters difficulties in accurately recognizing specific labels, such as 'adjustment,' 'chemicals,' 'copyright,' 'main point,' 'research,' and 'blank space.' Despite their frequent usage, these terms represent abstract nouns that are challenging to depict visually. This limitation has prompted a strategic decision to select more effectively representable nouns through imagery for improved model performance.

## 4.4 Experimental results

### 4.4.1 Overall

The experimental test was structured into a pre-test, a post-learning test, and a retention test. In the pre-test, participants identified known words and excluded unknown words, then only learned thesauruses of the known words. The pre-test results were not included in the final analysis. Due to varying pre-test knowledge, the number of words learned differed among participants. Therefore, percentage-based metrics were employed for consistent comparison in subsequent analyses, accommodating individual differences in baseline vocabulary knowledge.

The table 4.2 shows the test scores (unit of percentage, calculated by division of participant's correct answers and total correct answers) of 20 participants in the experiment. Here FF, RF, FP, RP stands for First test Flashcard, Retention test Flashcard, First test PICSU, Retention test PICSU respectively.

**Table 4. 2** Test scores of all participants

| Participant | FF | RF | FP | RP |
|---|---|---|---|---|
| 1 | 66.67 | 63.33 | 90.91 | 87.88 |
| 2 | 87.08 | 81.25 | 93.71 | 94.34 |
| 3 | 71.05 | 68.86 | 91.11 | 82.96 |
| 4 | 65.42 | 72.08 | 83.89 | 92.22 |
| 5 | 71.21 | 67.68 | 96.49 | 83.33 |
| 6 | 90.3 | 87.08 | 95.56 | 87.08 |
| 7 | 90.83 | 96.11 | 96.38 | 98.55 |
| 8 | 88.75 | 93.75 | 93.33 | 97.18 |
| 9 | 93.33 | 92.5 | 99.42 | 97.66 |
| 10 | 80 | 87.08 | 94.12 | 87.18 |
| 11 | 87.92 | 85.83 | 93.33 | 90.32 |
| 12 | 92.92 | 91.25 | 91.67 | 94.44 |
| 13 | 88.61 | 84.39 | 83.33 | 89.44 |
| 14 | 98.33 | 98.33 | 98.85 | 99.4 |
| 15 | 89.87 | 85.23 | 94.44 | 94.44 |
| 16 | 69.48 | 65.26 | 70.62 | 68.9 |
| 17 | 81.94 | 81.02 | 93.33 | 87.88 |
| 18 | 67.08 | 72.5 | 88.33 | 91.67 |
| 19 | 94.58 | 98.33 | 98.33 | 99.44 |
| 20 | 60.89 | 60 | 56.79 | 67.9 |

After conducting a simple descriptive statistic calculation with Table 4.2, we have the following results:

**Table 4. 3** Descriptive statistics of test results

| | FF | RF | FP | RP |
|---|---|---|---|---|
| **Mean** | 81.813 | 81.593 | 90.197 | 89.6105 |
| **Standard Error** | 2.618743 | 2.727114 | 2.294774 | 1.971711 |
| **Standard Deviation** | 11.71137 | 12.19602 | 10.26254 | 8.817757 |
| **Minimum** | 60.89 | 60 | 56.79 | 67.9 |
| **Maximum** | 98.33 | 98.33 | 99.42 | 99.44 |
| **Count** | 20 | 20 | 20 | 20 |

Friedman Test Result:

- Critical value: 30.781
- P-value: 9.45e-07

Table 4.4 shows followed the Friedman Test which can detect the difference between sample groups, we conducted a post-hoc test named Friedman-Nemeyi Test in order to have a more specific view on the localization of significant difference:

**Table 4. 4** Friedman-Nemeyi Test Result of Correct Answers by Participants

| FRIEDMAN-NEMENYI TEST | | | alpha | 0.05 | |
|---|---|---|---|---|---|
| group | R sum | size | std err | q-crit | R-crit |
| FF | 38.5 | 20 | | | |
| RF | 30.5 | 20 | | | |
| FP | 63 | 20 | | | |
| RP | 68 | 20 | | | |
| | | 20 | 5.773503 | 3.633 | 20.97514 |
| Q TEST | | | | | |
| group 1 | group 2 | R sum | q-stat | p-value | |
| FF | RF | 8 | 1.385641 | 0.7610052 | |
| FF | FP | 24.5 | 4.243524 | 0.014467 | |
| FF | RP | 29.5 | 5.10955 | 0.0017567 | |
| RF | FP | 32.5 | 5.629165 | 0.0004121 | |
| RF | RP | 37.5 | 6.495191 | 2.732E-05 | |
| FP | RP | 5 | 0.866025 | 0.9281758 | |

From the Table 4.3, it is evident that the average scores obtained through the Flashcard approach in the initial and subsequent retention evaluations (FF, RF) are notably inferior to those achieved via PICSU (FP, RP). The Standard Error and Standard Deviation across all four assessments display a high degree of similarity. Although PICSU's first test has the lowest minimum score compared to the rest, the highest scores in all assessments are close to perfect. Overall, the most striking indication is that the PICSU method appears to enable learners to attain higher scores. Nonetheless, to gain a more detailed understanding of the assessment outcomes, the following two sections will delve deeper into the data. Also from Friedman Test result: this extremely low p-value, well below the

conventional threshold of 0.05, indicates that there are statistically significant differences in the scores across these four tests. This suggests that the learning methods and test timings had a significant impact on the vocabulary learning outcomes in this set of students as well.

Later, Table 4.4 shows by observing the Friedman-Nemeyi Post Hoc Test, we have:

1.Groups Tested: Four groups (FF, RF, FP, RP) were compared, likely representing scores from different test conditions or learning methods.

2.Rank Sums: Each group's rank sum is shown, with FP having the highest (63) and RF the lowest (30.5).

3.Sample Size: The size for each group is consistent at 20.

4.Standard Error: The standard error of the rank sums is 5.773503.

5.Critical Values: The critical Q value (q-crit) is 3.633, and the critical range (R-crit) is 20.97514.

The post-hoc analysis suggests that the FP and RP methods are significantly different from FF and RF, with FP and RP yielding better outcomes given their higher rank sums. No significant difference was found between FF and RF, or between FP and RP, indicating similar performances within these pairs. The significant p-values in the comparisons involving FP and RP suggest these methods are more effective than FF and RF under the conditions tested.

Similar to Table 4.2, we also have Table 4.5 that shows the percentage of incorrect (false positive) answers. They are all answers that participants thought to be correct, but indeed, they were incorrect answers:

**Table 4. 5** Percentage of Incorrect Answers by Participants

| Participant | FF | RF | FP | RP |
|---|---|---|---|---|
| 1 | 2.08 | 1.67 | 0 | 0 |
| 2 | 4.58 | 3.75 | 1.26 | 0.63 |
| 3 | 17.98 | 28.07 | 7.41 | 7.41 |
| 4 | 2.92 | 4.58 | 1.67 | 0 |
| 5 | 6.57 | 10.1 | 0.88 | 0.88 |
| 6 | 8.44 | 10.83 | 3.89 | 3.89 |
| 7 | 4.17 | 4.58 | 1.45 | 0.72 |
| 8 | 1.67 | 1.25 | 1.67 | 2.26 |
| 9 | 6.25 | 7.08 | 2.92 | 2.34 |
| 10 | 5.83 | 12.92 | 5.88 | 10.9 |
| 11 | 2.08 | 3.75 | 3.89 | 4.3 |
| 12 | 1.25 | 2.5 | 2.78 | 0 |
| 13 | 4.22 | 5.06 | 7.22 | 1.11 |
| 14 | 1.67 | 1.67 | 1.15 | 0.57 |
| 15 | 3.8 | 4.22 | 1.11 | 2.78 |
| 16 | 2.35 | 0.94 | 0 | 0.56 |
| 17 | 5.56 | 5.56 | 4.85 | 0.61 |
| 18 | 9.58 | 4.58 | 3.33 | 2.22 |
| 19 | 3.33 | 1.67 | 2.22 | 0.56 |
| 20 | 7.56 | 20.44 | 4.94 | 6.79 |

From Table 4.5, after conduction calculation for descriptive statistics and Friedman Test, yielded as:

**Table 4. 6** Percentage of incorrect answers by participants

| | FF | RF | FP | RP |
|---|---|---|---|---|
| **Mean** | 5.0945 | 6.761 | 2.926 | 2.4265 |
| **Standard Error** | 0.863042 | 1.547741 | 0.49588 | 0.655287 |
| **Standard Deviation** | 3.859642 | 6.921706 | 2.217641 | 2.930531 |
| **Sample Variance** | 14.89684 | 47.91002 | 4.917931 | 8.588013 |
| **Minimum** | 1.25 | 0.94 | 0 | 0 |
| **Maximum** | 17.98 | 28.07 | 7.41 | 10.9 |
| **Sum** | 101.89 | 135.22 | 58.52 | 48.53 |
| **Count** | 20 | 20 | 20 | 20 |

Friedman Test Result:

- Critical Value: 21.995
- p-value: 0.00654

Table 4.7 shows same as the part for correct answers analysis, we now also conduct the Friedman-Nemeyi Post Hoc Test which yielded the following results:

**Table 4. 7** Friedman-Nemeyi Post Hoc Test for Incorrect Answers of Participants

| FRIEDMAN-NEMENYI TEST | | | alpha | 0.05 | |
|---|---|---|---|---|---|
| group | R sum | size | std err | q-crit | R-crit |
| FF | 59.5 | 20 | | | |
| FP | 41.5 | 20 | | | |
| RF | 66 | 20 | | | |
| RP | 33 | 20 | | | |
| | | 20 | 5.773503 | 3.633 | 20.97514 |
| Q TEST | | | | | |
| group 1 | group 2 | R sum | q-stat | p-value | |
| FF | FP | 18 | 3.117691 | 0.122255 | |
| FF | RF | 6.5 | 1.125833 | 0.856243 | |
| FF | RP | 26.5 | 4.589935 | 0.006525 | |
| FP | RF | 24.5 | 4.243524 | 0.014467 | |
| FP | RP | 8.5 | 1.472243 | 0.725298 | |
| RF | RP | 33 | 5.715768 | 0.000319 | |

Analysis of descriptive statistics, Friedman Test and Friedman-Nemeyi Post Hoc Test are presented accordingly follows:

Table 4.6 shows statistical metrics for four score sets: FF, FP, RF, and RP. RF scores have the highest mean (6.761) and variability, indicated by the highest standard deviation (6.921706) and variance. FP and RP have lower means (2.926 and 2.4265, respectively), showing more consistent but lower performance. The ranges in minimum and maximum scores suggest diverse outcomes, especially in RF. Despite the varied sums, all groups have an equal count of 20 observations, highlighting differing effectiveness or difficulty levels across the methods. Also, from the Friedman Test result: since the p-value is significantly lower than the standard threshold of 0.05, we can conclude that there are statistically significant differences in the scores across the four tests. This suggests that the method of learning (PICSU versus Flashcard) and the timing of the test (pre-test versus post-test) had a significant impact on the students' vocabulary learning outcomes.

From the Table 4.7 provided Friedman-Nemenyi test results:

1.Groups Compared: The groups compared are FF, FP, RF, and RP, which could correspond to different testing or study conditions.

2.Rank Sums: The rank sums for each group are provided, with RF having the highest sum (66) and RP the lowest (33).

3.Sample Size: The sample size for each group is 20.

4.Standard Error: Given as 5.773503 for the rank sums.

5.Critical Values: The critical Q value is 3.633, and the critical range is 20.97514.

6.Q Test Results:

- FF vs. FP: No significant difference, with a q-stat of 3.117691 and a p-value of 0.122255.
- FF vs. RF: No significant difference, with a q-stat of 1.125833 and a p-value of 0.856243.
- FF vs. RP: Significant difference, with a q-stat of 4.589935 and a p-value of 0.006525.
- FP vs. RF: Significant difference, with a q-stat of 4.243524 and a p-value of 0.014467.
- FP vs. RP: No significant difference, with a q-stat of 1.472243 and a p-value of 0.725298.
- RF vs. RP: Significant difference, with a q-stat of 5.715768 and a p-value of 0.000319.

The results indicate that the RF method significantly differs from both the FF and RP methods, suggesting that it may be the most effective approach under the tested conditions. FF and FP, as well as FP and RP, do not differ significantly, implying similar performance. The RP method has significantly lower outcomes compared to both FF and RF. These findings highlight the varying effectiveness of the different methods used in the study.

### 4.4.2 Reduction of Correct Answers

Each experimenter learned the same number of PICSU vocabulary and the same number of Flashcard vocabulary respectively, and then tested the learned part after learning. Then, according to the test results, the correct quantity and incorrect quantity are obtained, and the PICSU and Flashcard are compared according to the obtained quantity. Based on the comparison results, it is judged whether the PICSU learning support system developed by the institute is effective.

The first thing recorded in this chart is comparing the learner's test results on that day

and the test results two days later concerning the amount of memory lost by the learner. Conclusions are drawn from comparing the number of memories of learned vocabulary words. The blue color in the chart represents the number of lost memories for PICSU vocabulary, and the red color represents the number of lost memories for Flashcard vocabulary. Loss of memory means the learner has not mastered the vocabulary well, so the higher the number in this table, the less ideal the result is.

Figure 4.4 is a comparison table showing the reduction of correct answers (in percentage) among 20 participants after learning in two different ways: PICSU and Flashcard.

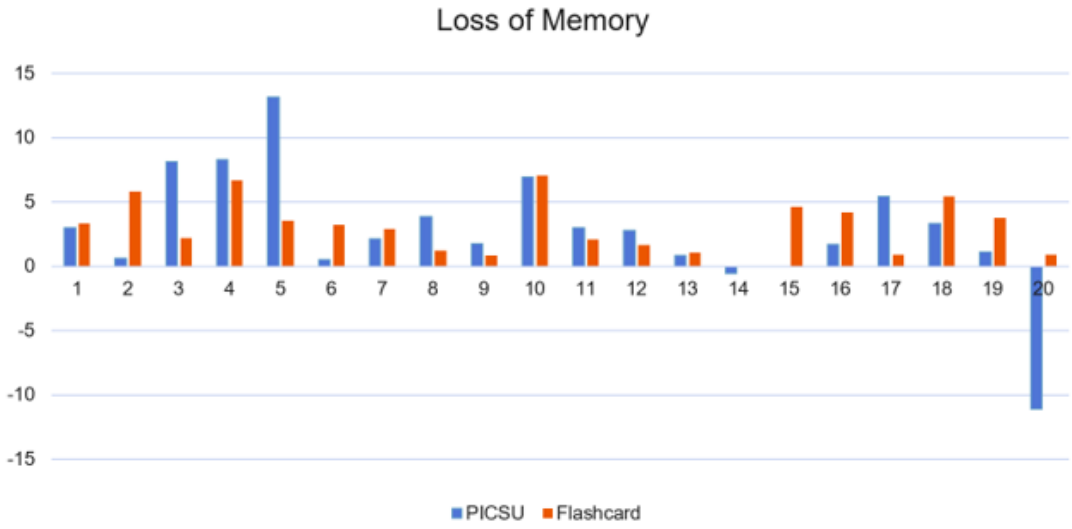


**Figure 4. 4** Loss of memory

It is obvious that the number of participants that maintain more correct answers of PICSU method are higher than Flashcard, which can imply that using PICSU method are producing reliable effectiveness. However, to have a clear view that the performance of 2 methods is significantly different, it is important to run another analysis for more concise evidence. Therefore, we conducted the Wilcoxon Signed-Rank Test [25] to verify if data of the 2 methods can tell the difference among their performance. Below is the result table of my Wilcoxon Signed-Rank Test:

**Table 4. 8** Wilcoxon Signed Rank Test for Reduction of Correct Answers

| **Loss of true positive answers** | | | | | | |
|---|---|---|---|---|---|---|
| PICSU | Flashcard | Difference | ABS Diff | Rank ABS Diff | Positive Rank | Negative Rank |
| 3.03 | 3.34 | -0.31 | 0.31 | 3 | | 3 |
| 0.63 | 5.83 | -5.2 | 5.2 | 17 | | 17 |
| 8.15 | 2.19 | 5.96 | 5.96 | 18 | 18 | |
| 8.33 | 6.66 | 1.67 | 1.67 | 9 | 9 | |
| 13.16 | 3.53 | 9.63 | 9.63 | 19 | 19 | |
| 0.55 | 3.22 | -2.67 | 2.67 | 14 | | 14 |
| 2.17 | 2.92 | -0.75 | 0.75 | 5 | | 5 |
| 3.85 | 1.25 | 2.6 | 2.6 | 12 | 12 | |
| 1.76 | 0.83 | 0.93 | 0.93 | 7 | 7 | |
| 6.94 | 7.08 | -0.14 | 0.14 | 1 | | 1 |
| 3.01 | 2.09 | 0.92 | 0.92 | 6 | 6 | |
| 2.77 | 1.67 | 1.1 | 1.1 | 8 | 8 | |
| 0.83 | 1.06 | -0.23 | 0.23 | 2 | | 2 |
| -0.58 | 0 | -0.58 | 0.58 | 4 | | 4 |
| 0 | 4.64 | -4.64 | 4.64 | 16 | | 16 |
| 1.69 | 4.22 | -2.53 | 2.53 | 11 | | 11 |
| 5.45 | 0.92 | 4.53 | 4.53 | 15 | 15 | |
| 3.34 | 5.42 | -2.08 | 2.08 | 10 | | 10 |
| 1.11 | 3.75 | -2.64 | 2.64 | 13 | | 13 |
| -11.11 | 0.89 | -12 | 12 | 20 | | 20 |
| SUM | | | | | 94 | 116 |
| Test Statistics (Smaller Sum) | | | | | 94 | |
| Sample Size (Number of non-zero Ranks) | | | | | 20 | |
| Critical Value (n=20, alpha=0.05) | | | | | 52 | |

The Table 4.8 shows that the Test Statistics (94) is higher than the Critical Value (52), which shows failure in the rejection of null hypothesis. This means there is not enough evidence to say that the data of 2 methods are significantly different.

### 4.4.3 Reduction of Incorrect Answers

Figure 4.5 represents the comparison of incorrectly answered questions representing the test results of learners using the two learning methods PICSU and Flashcard. Specifically, it is calculated based on the number of choice errors in the total number of learning using PICSU on the first day and the number of choice errors in the second delayed test. The resulting number is the reduction in the number of choice errors. The same is true for Flashcard. Comparing the number of test errors on the first day with the number of delayed test errors shows a reduction in the number of errors for Flashcard. Compare the amount of error reduction for two different learning styles. In the chart below, blue represents the number of errors reduced by PICSU, and red represents the number of errors reduced by Flashcard. In this chart, the larger the number, the better the effect.

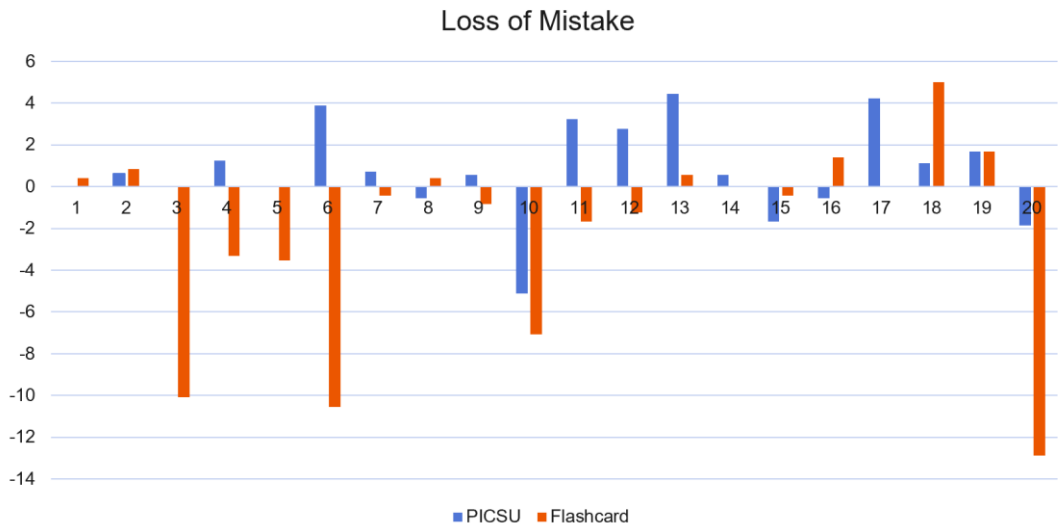


**Figure 4. 5** Loss of mistake

From the Table 4.9, the number of participants reducing more incorrect answers by learning PICSU is almost 3 times higher than ones learning Flashcard. However, it is better to conduct Wilcoxon Signed-Rank. Test to ensure the difference between 2 methods' results data. The test result is shown in the table below:

**Table 4. 9** Wilcoxon Signed-Rank Test for Reduction of Incorrect Answers

| **Loss of true negative answers (Unit: %)** | | | | | | |
|---|---|---|---|---|---|---|
| PICSU | Flashcard | Difference | ABS Diff | Rank ABS Diff | Positive Rank | Negative Rank |
| 0 | 0.42 | -0.42 | 0.42 | 2 | | 2 |
| 0.63 | 0.83 | -0.2 | 0.2 | 1 | | 1 |
| 0 | -10.09 | 10.09 | 10.09 | 17 | 17 | |
| 1.25 | -3.33 | 4.58 | 4.58 | 15 | 15 | |
| 0 | -3.54 | 3.54 | 3.54 | 10 | 10 | |
| 3.89 | -10.56 | 14.45 | 14.45 | 19 | 19 | |
| 0.72 | -0.42 | 1.14 | 1.14 | 5 | 5 | |
| -0.56 | 0.42 | -0.98 | 0.98 | 4 | | 4 |
| 0.58 | -0.83 | 1.41 | 1.41 | 7 | 7 | |
| -5.13 | -7.08 | 1.95 | 1.95 | 8 | 8 | |
| 3.23 | -1.67 | 4.9 | 4.9 | 16 | 16 | |
| 2.78 | -1.25 | 4.03 | 4.03 | 13 | 13 | |
| 4.44 | 0.56 | 3.88 | 3.88 | 11 | 11 | |
| 0.57 | 0 | 0.57 | 0.57 | 3 | 3 | |
| -1.67 | -0.42 | -1.25 | 1.25 | 6 | | 6 |
| -0.56 | 1.41 | -1.97 | 1.97 | 9 | | 9 |
| 4.24 | 0 | 4.24 | 4.24 | 14 | 14 | |
| 1.11 | 5 | -3.89 | 3.89 | 12 | | 12 |
| 1.67 | 1.67 | 0 | | | | |
| -1.85 | -12.89 | 11.04 | 11.04 | 18 | 18 | |
| SUM | | | | | 156 | 34 |
| Test Statistics (Smaller Sum) | | | | | 34 | |
| Sample Size (Number of non-zero Ranks) | | | | | 19 | |
| Critical Value (n=19, alpha=0.05) | | | | | 46 | |

(*Here, the critical value equals 19 because there is one participant with same reduction of correct and incorrect answer are 0, making the difference to be 0. Hence, we discard this one's entry from the number of sample)

From the Table 4.9, the Test Statistics (34) are lower than the Critical Value (46)[26], which means that we have enough confidence to say that the outcome data of the 2 learning methods are significantly different from each other.

According to the results of this chart, among the 20 people, 14 people from PICSU reduced the number of wrong answers, and 6 people from Flashcard reduced the number of wrong answers. From the comparison of the results, it can be seen that the error loss results of PICSU are better than those of Flashcard. In other words, in this experiment, PICSU performed better on learners' error loss and can help learners reduce errors.

## 4.5 Questionnaire results

In the context of this study, a post-experiment survey was meticulously administered. The primary objective of this questionnaire was to garner precise feedback and assessments from the participants who engaged in the experimental procedure. This feedback, elicited through a series of targeted inquiries, is poised to yield invaluable insights that are anticipated to drive the refinement and enhancement of future experimental iterations. The ensuing discourse delineates the specific feedback garnered from the participants regarding the content of the questionnaire employed in the study.

There are 19 of the 20 participants that filled out the questionnaire. The following content is an analysis based on the results of the 19 questionnaires. According to the participant's first answer and Figure 4.6 pie chart, more people like the PICSU learning method because the form of pictures makes it easier for everyone to remember, and the learning process is more interesting. The learning method of Flashcard is relatively boring, and it is easy to feel sleepy during the learning process.

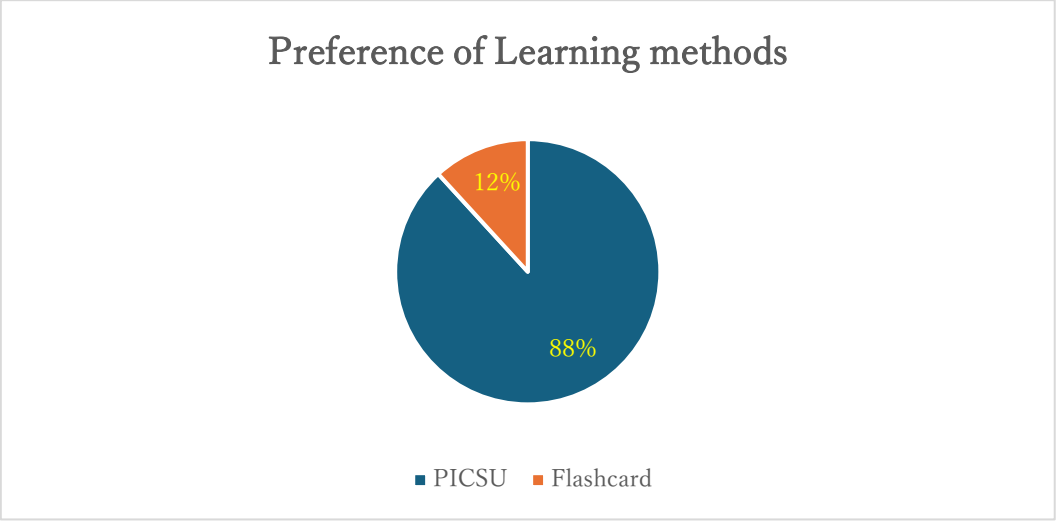


**Figure 4. 6** Preference of Learning methods

In the other two questions, most respondents indicated that including pronunciation guidance alongside vocabulary would significantly enhance the learning experience.

Additionally, there was a notable preference for the integration of comprehensive explanations or illustrative sentences, as well as the implementation of distractor options within assessments. Many participants also expressed that this was their inaugural experience with thesaurus-based vocabulary expansion, which they found innovative and engaging. These insights will be invaluable guiding principles for ongoing refinement and advancement in future research endeavors.

## 4.6 Discussion

In this study, we anticipated the PICSU system, which integrates visual aids in learning, outperforming the traditional Flashcard method in vocabulary acquisition. The premise is based on the cognitive principle that imagery, a key component of PICSU, enhances memory retention and facilitates learning. In contrast, Flashcards, though effective, tend to be monotonous and less stimulating for memory compared to the dynamic visual engagement offered by PICSU.

The experimental results supported our hypothesis: Participants using PICSU showed higher success rates in memory-based tests, indicating that visual aids in learning significantly bolster memory retention. Furthermore, learners employing PICSU exhibited less memory loss and fewer errors during tests, suggesting a deeper and more lasting understanding of vocabulary.

In examining two key evaluation metrics – the decrease in correct answers (indicating loss of memory) and the decrease in incorrect answers (indicating loss of mistake) – a notable observation emerged from the comparative study between PICSU and Flashcard learning methods in Japanese thesaurus acquisition.

Firstly, the data revealed a predominantly positive trend in reducing correct answers among both PICSU and Flashcard learners. This suggests that most participants experienced decreased memory loss, regardless of the learning method employed.

However, a divergent pattern was observed in the reduction of incorrect answers. Approximately 80% of the students utilizing the PICSU method exhibited a positive reduction in mistakes, indicating a decrease in the number of incorrect answers. In contrast, around 55% of the learners using Flashcards demonstrated a negative trend in this metric, implying increased errors post-learning. This distinction suggests a significant difference in the effectiveness of the two methods in reducing learning mistakes.

# Chapter 5 Conclusion and Future Works

## 5.1 Conclusion

In this study, we developed an innovative learning approach integrating object recognition technology, specifically YOLO (You Only Look Once), with a comprehensive Japanese language database, WordNet. Our underlying hypothesis posited that a picture-based learning strategy could significantly enhance the efficacy of acquiring thesauruses. To test this hypothesis, we compared our novel picture-based method and the traditional flashcard approach. The results of this evaluation demonstrated that our proposed method outperformed the conventional flashcard technique in reducing the frequency of errors made by learners. This finding underscores the potential advantages of incorporating visual aids and advanced technology in language learning methodologies.

Section 1.2 describes the four primary research questions. Through this research, we have these conclusions.

RQ1: How can photos of familiar objects that exist or are taken with smartphones be used in vocabulary learning?

A1: Photos of familiar objects taken with smartphones can be transformed into a dynamic vocabulary learning tool using a YOLO-based object recognition module within a language learning app. When a user snaps or uploads a picture, the module identifies the object and retrieves its Japanese word from a corpus, along with synonyms, enhancing vocabulary breadth. This method personalizes learning by linking words to the user's daily environment, ensuring active engagement, and promoting better retention through contextual and interactive learning experiences.

RQ2: How can thesaurus learning be integrated into vocabulary learning?

A2: Thesaurus learning can be seamlessly integrated into vocabulary learning by using a tool like WordNet for Japanese within a language learning application. When an object is recognized by the application's object recognition module, WordNet is queried to find all the thesauruses related to that word. These synonyms and related terms are then displayed on the smartphone screen, providing the user with a rich set of vocabulary associated with the initially recognized word. By presenting these terms together, learners can understand the nuances between different synonyms and how they can be used in various contexts, which is a critical skill in language proficiency.

RQ3: What are the differences in learning outcomes in a photo-based Japanese vocabulary

and thesaurus learning application compared to traditional flashcards?
A3: The learning outcomes in a photo-based Japanese vocabulary and thesaurus learning application show marked improvements when compared to traditional flashcard methods. In experimental settings, the proposed photo-based method demonstrated a slight increase in the number of learners retaining correct answers over time. However, the most significant difference lies in the reduction of incorrect answers. The data indicates that nearly three-quarters of participants using the photo-based application reduced their incorrect answers more effectively than those using flashcards. This suggests that the photo-based method not only aids in better retention of correct information but also more significantly helps learners correct and avoid their previous mistakes, leading to a more robust and effective learning process.

## 5.2 Future Works

Considering these conclusions, future research should explore the long-term retention effects of both methods, investigate the underlying cognitive mechanisms at play, and consider a broader range of subjects to determine the versatility and adaptability of the PICSU method across various disciplines. In the forthcoming research endeavors, the primary objective will be to refine and advance the architecture of the current system.

For example, combining YOLOs with other object recognition models can be considered. This orientation of MoE (Mixture of Experts) strategy may ensure the accuracy of the whole model. Expansion ideas are not merely confined to the realm of technical system optimization but are also envisioned to encompass a comprehensive vocabulary learning apparatus. The ambition is to transcend the existing framework by integrating additional mechanisms that augment the learning process's engagement and efficacy. One such enhancement under consideration is the incorporation of gamification elements. These are intended to stimulate sustained study habits and foster a deeper commitment to learning among users.

Moreover, there is an impetus to supplement the system with auditory components and elucidative example sentences, which can facilitate the assimilation of vocabulary in context. The addition of voice functionality is also anticipated, aiming to bolster oral proficiency and provide corrective feedback, thereby supporting the multifaceted development of language skills.

This research is driven by the aspiration to contribute significantly to the field of Japanese language education and to provide effective learning tools for non-native

speakers. The ultimate vision is to craft a system that not only fosters the progression of Japanese language acquisition but also empowers a broader demographic of learners to achieve fluency. The implication is that such advancements in language education technology will enable individuals to employ Japanese seamlessly in everyday interactions and cultural exchanges.

# Acknowledgement

First, I would like to express my gratitude to my professor Prof. Shinobu Hasegawa. During the two years at JAIST, he was always patient and guided and helped me with his professional knowledge. When I feel lost and confused, I can always come up with enlightening suggestions and point the way forward. I still remember that I felt very familiar with Prof. Hasegawa from the very beginning when I was an intern student. It was fate that brought me to this school to study, and I am honored to continue doing research with Prof. Hasegawa. I feel that the experience of these two years has been beneficial to me throughout my life and has even changed my path forward. I am very grateful to have met such a life mentor.

At the same time, I would also like to thank Prof. Shogo Okada and Prof. Kokolo Ikeda. I was very nervous during the mid-term presentation, but they encouraged me with a gentle attitude and gave me a lot of good suggestions to continue this research. full of motivation. I am also very grateful to my colleagues in the laboratory, including other classmates and friends at JAIST, who gave me a lot of constructive help and encouragement, allowing me to successfully persist until the end.

Finally, I would like to thank my family. My parents have been silently supporting and encouraging me. No matter when I feel confused or lost, they always cheer me up and recharge my batteries. In the future, I hope to make good use of the experience and knowledge I have learned in these two years to become an increasingly stronger person and never give up.

# Publications

Meihua Xue, Wen Gu, Koichi Ota, Shinobu Hasegawa:” Development of a vocabulary learning support system using photographic object recognition”, Student Research Presentation in Japanese Society for Information and Systems in Education, (2024 in press).

# References


[1]Y Helene -「マイ・ アルバム: 語学学習のための写真と文章を使った課題」, 作大論集, 2011

[2] 安暁旭, 吉野孝 -「留学生のためのメディア統合型モバイル日本語学習支援システムの構築」, 第 73 回全国大会講演論文集 2011 (1), 417-418, 2011-03-02

[3]得丸智子 -「アプリを活用した単語学習を中心とする日本語独習~TAE によるインタビュー分析」, 開智国際大学紀要, 2020

[4] Nation, I.S.P. (2001). Learning Vocabulary in Another Language. Cambridge University Press. (本)

[5]三室千草, 梶山朋子, 大内紀知 -「子どもの学習意欲を向上させる英単語アプリケーションの開発~ 自身が撮影した写真の活用による英単語帳の作成」電子情報通信学会技術研究報告; 信学技報, 2014

[6]マスデン, 眞理子, Mariko, Masden-「留学生の相談から見た日本語学習の必要性」, 熊本大学留学生センター紀要 12 65-70, 2008-12-26

[7]松田文子 -「日本語学習者による語彙習得: 差異化・一般化・典型化の観点から」, 世界の日本語教育. 日本語教育論集, 2000

[8] Fu Gaihua, 「中国語を母語とする日本語学習者の主語省略の習得について：主題予測性をめぐって」 , 『言語の普遍性と個別性 (11) 』, pp.91-100, (2020).

[9]鎌田美千子「パラフレーズに着目した日本語指導書開発のための一考察 質問紙調査から見えてきた課題――」、『宇都宮大学国際学部研究論集』― 49 号、51-59 頁、2020 年 2 月

[10]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).

[11] R. Shadiev, T.-T. Wu, and Y.-M. Huang, “Using image-to-text recognition technology to facilitate vocabulary acquisition in authentic contexts,” ReCALL, vol. 32, no. 2, pp. 195–212, May 2020, doi: 10.1017/S0958344020000038.

[12] R. Safitri, R. T. Muslima, and S. Herlina, “Mobile Augmented Reality for Japanese Vocabulary and Hiragana Letters Learning with Mnemonic Method,” in 2022 Seventh International Conference on Informatics and Computing (ICIC), Dec. 2022, pp. 1–7. doi: 10.1109/ICIC56845.2022.10006920.

[13] CY Wang, A Bochkovskiy, HYM Liao - 「YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors」, arXiv preprint arXiv:2207.02696, 2022

[14] Y Nie, P Sommella, M O'Nils, C Liguori, J Lundgren-「Automatic detection of melanoma with yolo deep convolutional neural networks」, 2019 E-Health and Bioengineering Conference (EHB), 2019

[15] Chien-Yao Wang, Alexey Bochkovskiy, Hong-Yuan Mark Liao-「YOLOv7: Trainable Bag-of-Freebies Sets New State-of-the-Art for Real-Time Object Detectors」, 2023, pp. 7464-7475

[16] 2209.02976.pdf (arxiv.org) YOLOv6
[17] 2305.09972.pdf (arxiv.org) YOLOv8
[18] arxiv.org/pdf/2105.04206.pdf YOLO-R
[19]西川 由理,佐藤 仁,小澤 順 -「一般物体検出 YOLO の分散深層学習による性能評価」, 研究報告ハイパフォーマンスコンピューティング（HPC）, 2018-HPC-166
[20] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, “You Only Look Once: Unified, Real-Time Object Detection.” arXiv, May 09, 2016. doi: 10.48550/arXiv.1506.02640.
[21] https://www.datacamp.com/blog/yolo-object-detection-explained
[22] H Isahara, F Bond, K Uchimoto, M Utiyama, K Kanzaki -「Development of the Japanese WordNet.」, 2008 - cs.brandeis.edu
[23] Princeton Thesaurus Christiane Fellbaum (1998, ed.) WordNet: An Electronic Lexical Database. Cambridge, MA: MIT Press.
[24] http://jhlee.sakura.ne.jp/JEV/
[25] Wilcoxon Signed-Rank Test Rey, D., Neuhäuser, M. (2011). Wilcoxon-Signed-Rank Test. In: Lovric, M. (eds) International Encyclopedia of Statistical Science. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-04898-2_616
[26] Pernet C. Null hypothesis significance testing: a short tutorial. F1000Res. 2015 Aug 25; 4:621. doi: 10.12688/f1000research.6963.3. PMID: 29067159; PMCID: PMC5635437.

# Appendix

# Agreement for cooperation in experiments

## Consent Form

Japan Advanced Institute of Science and Technology

Affiliation: Advanced science and technology Research Department　Advanced science and technology Major　Hasegawa Laboratory

Research Supervisor: Xue Meihua

I have received a written explanation of the research and experiments on the "Development of a Vocabulary Learning Support System Using Photographic Object Recognition" from Xue Meihua. I have understood the experiment's purpose, methodology, personal information protection measures, and safety management considerations. I willingly consent to provide the requested personal information, data, and any related information for the experiment.

Items explained and understood.

(Please indicate with a check mark (✓) on the left side of the items you have understood in the consent form, and indicate with a cross mark (×) on the left side of the items you have not understood.)

1 Outline of the experimental plan:

（ ）・Purpose and significance of the experiment

（ ）・Information and data to be provided

2 Personal Information Protection:

（ ）・Collection of personal information is necessary in accordance with the purpose and experimental plan.

（ ）・Methods for anonymizing the provided data, etc.

（ ）・Proper storage and management of data

3 Intrusion and security management:

（ ）・Expected discomfort, burden, etc.

4 Informed consent:

（ ）・Participation in the experimental plan is voluntary.

（ ）・There will be no adverse consequences if you won’t join this experiment.

（ ）・Even after agreeing to participate in the research plan, consent can be withdrawn in writing within three days.

（ ）・There will be no adverse consequences if you withdraw your consent.

（ ）・Provided data will be discarded upon withdrawal of consent.

（ ）・Collected data will not be shared with others without the individual's consent.

（ ）・Plans for presenting the experimental results include conference presentations and publication of papers.

（ ）・Payment of compensation for participating in the research plan.

Year Month Day

Name (Signature)

Contact (Email Address)

Post-experiment questionnaire

ID_____

Thank you very much for participating in the experiment.
We would greatly appreciate it if you could kindly take a moment to fill out the following questionnaire.

Gender_____, Age_____, Japanese level_____

Which learning method did you prefer and why?

_____

_____

What improvements, if any, would you suggest for each learning method?

_____

_____

Please feel free to write down any observations or insights you have regarding this experiment.

_____

_____