IEICE TRANS. INF. & SYST., VOL.E100–D, NO.9 SEPTEMBER 2017

PAPER

An Algorithm to Evaluate Appropriateness of Still Images for Learning Concrete Nouns of a New Foreign Language*

Mohammad Nehal HASNINE†, ***Nonmember***, **Masatoshi ISHIKAWA**††, ***Member***, **Yuki HIRAI**†††, **Haruko MIYAKODA**††††, ***Nonmembers, and*** **Keiichi KANEKO**†a), ***Member***

SUMMARY Vocabulary acquisition based on the traditional pen-and-paper approach is outdated, and has been superseded by the multimedia-supported approach. In a multimedia-supported foreign language learning environment, a learning material comprised of a still-image, a text, and the corresponding sound data is considered to be the most effective way to memorize a noun. However, extraction of an appropriate still image for a noun has always been a challenging and time-consuming process for learners. Learners' burden would be reduced if a system could extract an appropriate image for representing a noun. Therefore, the present study purposed to extract an appropriate image for each noun in order to assist foreign language learners in acquisition of foreign vocabulary. This study presumed that, a learning material created with the help of an appropriate image would be more effective in recalling memory compared to the one created with an inappropriate image. As the first step to finding appropriate images for nouns, concrete nouns have been considered as the subject of investigation. Therefore, this study, at first proposed a definition of an appropriate image for a concrete noun. After that, an image re-ranking algorithm has been designed and implemented that is able to extract an appropriate image from a finite set of corresponding images for each concrete noun. Finally, immediate-after, short- and long-term learning effects of those images with regard to learners' memory retention rates have been examined by conducting immediate-after, delayed and extended delayed posttests. The experimental result revealed that participants in the experimental group significantly outperformed the control group in their long-term memory retention, while no significant differences have been observed in immediate-after and in short-term memory retention. This result indicates that our algorithm could extract images that have a higher learning effect. Furthermore, this paper briefly discusses an on-demand learning system that has been developed to assist foreign language learners in creation of vocabulary learning materials.

key words: *appropriate image, concrete noun, on-demand learning system, AIVAS, AIVAS-IRA*

1. Introduction

Duppenthaler [3] indicates that a vast majority of instructors and learners reported that vocabulary acquisition is an essential part of first and second language learning. Language learning, whether first or second, starts with learning words [4], [5]. Richards and Renandya [6] asserted that the vocabulary is a core component of the language proficiency and provides much of the basis for how well learners speak, listen, read and write. Wilkins [7] posited that without a grammar very little can be conveyed, but without a vocabulary nothing can be conveyed. In addition, Brown [8] stated that the vocabulary is the heart of any language.

Although vocabulary acquisition refers to memorizing words from different parts of a sentence, noun is considered to be a key component of it. In markedness theory [9], the concept of asymmetry is presupposed. That is, some language structures and concepts are considered to be more general compared to others. For instance, children acquire nouns before verbs. This is because nouns, particularly concrete nouns, correspond to specific "objects" and are easier to acquire than verbs, which represent "events", whose meaning is more difficult to comprehend. This asymmetric relationship between nouns and verbs can be rephrased as follows: the more general, unmarked structures or concepts such as nouns are easier for children to acquire, while the less general, more complex items such as verbs are acquired later. In this sense, since the acquisition of verbs implies the acquisition of nouns, noun acquisition is considered to be the starting point in grammatical categorization both in L1 and L2. It has been noticed that, foreign language learners also start acquiring a new language by memorizing nouns first. Therefore, acquiring a significant number of nominal lexical items is considered to be important to enrich a learner's knowledge of foreign language.

Nevertheless, vocabulary acquisition in the classroom has always been less prioritized because the common belief between instructors and learners is that vocabulary acquisition is to be attained by self-study. In this regard, Allen [10] agreed that vocabulary learning in classroom has been neglected during the past decades because i) instructors felt that the grammar should be emphasized more than the vocabulary, ii) specialists in methodology believed that students would make mistakes in sentence construction if they learn too many words before they have mastered the basic grammar, and iii) language instructors believed that word meaning can be acquired only through experience and cannot be taught in the classrooms. Therefore, foreign language learners around the globe are struggling to acquire significant vocabularies in the classroom.

Manuscript received November 17, 2016.
Manuscript revised May 16, 2017.
Manuscript publicized June 21, 2017.

†The authors are with Graduate School of Engineering, Tokyo University of Agriculture and Technology, Koganei-shi, 184–8588 Japan.
††The author is with Faculty of Business Administration, Tokyo Seitoku University, Tokyo, 114–0033 Japan.
†††The author is with Admissions Center, Shinshu University, Matsumoto-shi, 390–8621 Japan.
††††The author is with Faculty of Liberal Arts, Tsuda University, Kodaira-shi, 187–8577 Japan.

*This work is based on enhancement of the papers presented in conferences [1], [2].

a) E-mail: k1kaneko@cc.tuat.ac.jp (Corresponding author)
DOI: 10.1587/transinf.2016EDP7463

With the advent of computer-assisted language learning technologies, foreign language learning has become more interesting and exciting. Web-based learning has provided new opportunities for language learners and instructors to create and share learning contents and interact among others. Learners are able to study foreign languages with the help of the learning materials without direction or immediate supervision from their instructors. Hayati et al. [11], Cavus and Ibrahim [12], Lu [13], and Thornton and Houser [14] have reported their observations on how second language learning through mobile devices is becoming more and more popular than traditional approaches in Asian countries such as Iran, Turkey, Taiwan, and Japan, respectively. However, foreign language acquisition with the multimedia-enriched environment is sometimes problematic because of students' inadequate knowledge of technologies.

The representation of multimedia applications and the presentation method of materials in various forms (audio, graphics, animation, video, text, and/or images) have become the subject of much debate among researchers. Chun and Plass [15] revealed that words associated with actual objects or corresponding images are learned more easily than those without. In addition, Carpenter and Olson [16] reported that one of the most common findings in cognitive research is that image-based learning is easier for learners in long-term memory recalling. Moreover, the phenomenon called picture superiority effect [17], [18] and the theory called dual coding theory [19] proved that pictures often have advantages over words in recalling of memory. However, it still remains a challenge for CALL (Computer Assisted Language Learning) researchers to find appropriate pictures for actual objects.

The purpose of this study is to extract an appropriate still image for each concrete noun in order to aid foreign language learners in acquisition of foreign vocabulary. The term 'appropriate' image has been the key focus in this study, by which we tried to bring out only one image for each concrete noun presuming that an appropriate image will boost learners' memories in recalling an acquired noun. This study proposes a definition of an appropriate image that can be represented for a concrete noun. To actualize the definition, an image re-ranking algorithm has been implemented. Later on, the effects of the appropriate images recommended by the algorithm have been assessed. Although this paper reports on our approach to find appropriate images for concrete nouns, the utmost aim of this ongoing investigation is to extract appropriate images for all kinds of nouns.

The rest of this paper is structured as follows. First, in Sect. 2, we survey the state-of-the-art computer-assisted vocabulary learning (CAVL) activities. Next, we introduce the findings of our preliminary experiments, the definition of an appropriate image to represent a concrete noun, the design and implementation of an image re-ranking algorithm, and the development of the on-demand vocabulary learning system in Sect. 3. Then, we evaluate the appropriate images by conducting a learning effect investigation in Sect. 4, and discuss our study related observations and the limitations of the current study in Sect. 5. Finally, we conclude this study by discussing the experimental results and future considerations in Sect. 6.

2. CAVL: State of the Art

Research related to technology-based vocabulary acquisition has been well investigated and documented in the literature. This section discusses the surveyed literature with respect to learning material representation, existing platforms to create self-learning materials, and the philosophies behind image selection.

Al-Seghayer [20] scrutinized the efficiency of multimedia annotations for learning unknown words, and yielded the conclusion that words presented under texts coupled with video clips were more effective in teaching unknown words than those using still images and texts. On the contrary, Yeh and Wang [21] investigated the effects of multimedia annotations and learning styles on vocabulary learning. They compared the effects of vocabulary acquisition by text annotation only versus text + image versus text + image + audio. The posttest result demonstrated that annotations using text + image were most effective for vocabulary acquisition. This study also investigated the reasons that lie behind the ineffectiveness of vocabulary learning via text + image + audio approach, and concluded that i) learners' learning styles and ii) the speech rate of the recorded annotation were involved. Furthermore, Chun and Plass [15] investigated the effectiveness of annotations with different multimedia types while acquiring a new vocabulary. The study result with regard to the effectiveness of different types of annotations concluded with significantly higher scores for words supported by text + picture than words with text-only and video + text. Consequently, this study considered still images along with texts and pronunciation data as the effective form of learning material representation.

PhotoStudy [22] creates learning materials based on the collaborative use of images that were captured by camera phones. The system's learning materials are automatically generated by sending images captured by learners via email and are stored in the database for collaborative learning. However, the collaborative policy of the learning materials was victimized by email spam, which led to failing image links. Also, a huge volume of incorrectly marked up images were found which are problematic for learners. Subsequently, Hasegawa et al. [23] developed the PSI (Personal Super Imposing) system. It is a web-based vocabulary learning system that has a function of creating English vocabulary learning materials based on short movies. Later on, SIGMA, an autonomous and collaborative vocabulary-learning environment, was developed by Kaneko et al. [24], [25] to reinforce PSI functionalities. Though the authors investigated the learning effects by comparing it with the pen-and-paper based learning method, this study did not demonstrate any significant difference. Moreover, the article does not give a proper indication of the video clips used to generate learning materials. Huang et al. [26] introduced a framework named Ubiquitous English Vocabulary Learning (UEVL) that assists learners in experiencing systematic vocabulary learning process. The UEVL system is able to detect learners' locations through radio-frequency identification and global positioning system technologies. The system uses video clips to generate learning materials. This study does not report on the learning effects of the system in terms of foreign vocabulary acquisition. Afterward, Anonthansap et al. [27] developed a system equipped with a dynamic and interactive interface that allows learners to study the vocabulary by using a mnemonic technique. However, no significant difference was observed when they compared this approach with a traditional dictionary-based learning approach and a static visualization approach where images are displayed statically without any special interaction. In addition to the systems mentioned above, there exist several web-based mobile-learning tools that support second language learners in vocabulary acquisition.

Kaneko et al. [28] introduced the concept of universal images in vocabulary acquisition. They also discussed some cultural aspects and benefits of universal images in terms of foreign language learning. However, a concrete indication of universal image selection was not discussed. ImageNet [29], [30], which was built upon the backbone of WordNet [31], [32], contains a large number of images. However, these studies do not draw a clear definition of educational images and clarify the significance of ImageNet images on learners' long-term memory retention. Consequently, a definition of an appropriate image for vocabulary acquisition is still ambiguous in the available literature.

We could not observe any recognized study that clearly indicates what an appropriate image is to begin with, and how it can be extracted in the creation of learning materials. Also, currently available vocabulary learning environments have limitations to provide on-demand creation facility to the learners. A system developed based on our algorithm will underpin other existing systems because it provides an automatic function to extract an appropriate image in the creation of a learning material for each concrete noun. Additionally, in the process of a learning material creation, translation data and voice data are extracted automatically from the web. Therefore, the current system facilitates the creating process because learners can create learning materials by themselves without looking for other external resources.

3. Design and Implementation of an Image Reranking Algorithm

Concrete nouns are easier to create images compared to abstract ones because concrete objects actually exist. However, how to define an "appropriate" image is a challenge. Most of the existing images for the image-based vocabulary learning systems are prepared beforehand by the instructors themselves. A system that automatically extracts the "appropriate" image for the corresponding word would facilitate the material creating process. Our first step is to focus on concrete nouns. Thus, in this section we propose a definition of an 'appropriate' image for representing a concrete noun. Additionally, we discuss on the preliminary experiments, algorithm design and implementation, system architecture, learning materials' format and representation etc.

3.1 Background and Approach

Current image search engines on the web rely on the keywords around the images and file-names, which produce much garbage in the search result [33]. Ben-Haim et al. [33] also pointed out that web-based image search engines are blind to the actual content of images. To explain the limitations of keyword-based image search, Jain and Varma [34] articulated that, "there is no straightforward, fully automated, way of going from textual queries to visual features. Thus, image search engines primarily rely on static and textual features for ranking". Since image search engines primarily rely solely on static and textual features for ranking, the standard images recommended by these search engines may not necessarily be suitable for foreign language learners in acquisition of new vocabulary. Resulting therefrom, in the field of still image-based vocabulary learning, extracting the appropriate image from the web may be considered to be challenging and may cause a significant amount of time waste for learners. To overcome this problem, our research trial is to understand the characteristics of still images and define the 'appropriate' images for acquiring concrete nouns.

Preliminary investigations looked into the characteristics of commonly found still images in standard image search engines. Preliminary investigation 1 examined the learning effect between an image that contains single object without having any background object (that is, mono-color background) -versus- an image that contains multiple objects in the background. Ten Bengali words were randomly selected and their corresponding images were used in creation of learning material. We extracted two images given by Google image search API for each word. One of them was a top-ranked image without background object (that is, mono-color background), and the other one was a top-ranked image with multiple objects as the background. Each vocabulary learning material was created with an image and the word meaning. Note that, to observe the effects of images in recalling memory, pronunciation data were not included in the learning materials. 20 participants were randomly divided into two groups and each was asked to participate in a 10-minutes study session. In the pretest, participants were orally asked whether they have any background of Bengali language and found that none of the participants had previous knowledge of the language. Therefore, the pretest scores were set to zero. Then, posttest 1 was conducted immediately after the study session while posttest 2 was conducted a week after posttest 1. Mann-Whitney's U-test indicated that no significant difference was noted ($U = 35$, $p = 0.07$). Consequently, we came to the conclusion that, backgrounds in still images do not play significant roles in vocabulary acquisition.

Preliminary investigation 2 investigated the object representation in an image frame. We have conducted a questionnaire with 26 participants. The questionnaire had been prepared containing 10 different representations (images) of the same object. Proportion and position of the object in the image frame had been changed; however, the shape of the object was left unchanged. The participants were asked to rank those images on a scale of 1 to 10 with the instruction of "If you were the instructor, which of the following images would you like to use to teach your students? Please rate each image on a scale of 1 to 10 (and reasons, if any)". The statistical data compared by ANOVA ($F_{9,250} = 170.1$, $p < 0.01$) and the past Steel-Dwass multiple comparison tests revealed that the image that was represented with 'the highest proportion of the actual object highlighted in the center position of the image frame' significantly outperformed others.

After analyzing the characteristics of still images, this study assumed that appropriate images for concrete nouns are those having the actual object(s) located in the middleground, and the highest proportion of the actual object(s) in the image frame. We hypothesized that images containing those features would play significant roles in learners' memory retention rates in comparison to the inappropriate ones. Thus, this study emphasizes on extracting appropriate images that will help learners in both short-term and long-term memory retention. To extract appropriate images, the current study designed and implemented an algorithm named AIVAS-IRA (Appropriate Image-based Vocabulary Acquisition System - Image Re-ranking Algorithm). Given a set of corresponding images related to a concrete noun, AIVAS-IRA finds the most appropriate image among them to create a learning material for the noun by re-ranking the images.

3.2 Design of Algorithm

The design of AIVAS-IRA involves three phases. They are initial phase, intermediate phase and re-ranking phase. The initial phase is non-repetitive. On the other hand, the intermediate and the re-ranking phases are iterative. This section briefly discusses the algorithm design and implementation of the AIVAS-IRA.

In the initial phase, we apply Fast Fourier Transform (FFT) to sample 'appropriate' images to calculate their center. That is, the power spectrum in the spatial frequency domain, which is obtained by applying FFT to each image, is regarded as a vector, and the center of these vectors for all sample images is calculated. In Step (1): At first, a list of 59 English words had been prepared. These English words were randomly chosen from Alan Beale's English vocabulary word list that is used by English as Second Language (ESL) learners. After that, 59 images associated with 59 English words in total have been accumulated that we have recognized as sample 'appropriate' images. The sample appropriate images were accumulated based on the following criteria: i) actual object(s) highlighted in the center position, and ii) higher proportion of actual object(s) in the image. In accumulation of the sample appropriate images, we have not paid attention to the background of the images. In Step (2): Power spectrums in the spatial frequency domain of the accumulated images have been determined by FFT. To determine the power spectrums of images, all 59 images have been gray-scaled (256-gradations) and resized to 32×32. In Step (3): Generated vectors have been converted into unit vectors by vector normalization. In Step (4): Average of the generated vectors found in Step (3) has been calculated. In Step (5): The derived average from Step (4) has been converted to a unit vector by vector normalization, and labeled as the 'center point' of sample appropriate images. In this study, the center point of the accumulated appropriate images has been considered as the scale of appropriateness of images in the re-ranking phase. That is, by using this center point, we judge appropriateness of the images for all words. Therefore, it is not necessary to prepare any new sample image when the user inputs a new word.

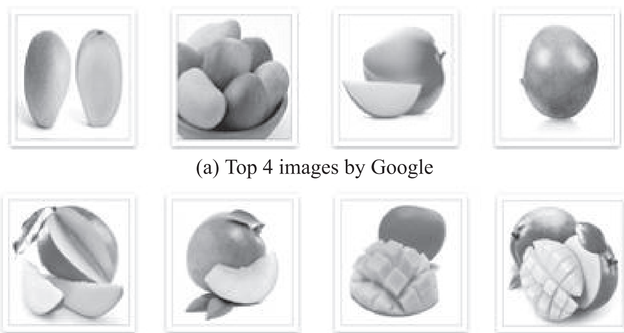
In the intermediate phase, firstly, a set of corresponding images for each noun is downloaded from an image search engine. Secondly, power spectrums in the spatial frequency domain of all the downloaded images are determined. In determination of power spectrums, images are gray-scaled (256-gradations) and resized to 32×32. Thirdly, derived vectors are converted into unit vectors. And finally, the Euclidean distance of each vector from the center point of sample appropriate images (that was determined previously in initial phase) was calculated.

In the re-ranking phase, firstly, the Euclidean distances found in intermediate phase were compared. And finally the most appropriate image was extracted.

AIVAS-IRA was coded with PHP and Java. To collect corresponding images for each noun, we have used Google image search API because it is widely used and offers over ten billion images. Furthermore, Google's tweaks image search algorithm and SafeSearch feature filters pornographic contents in its search return, and therefore we considered that Google image search API may be the most suitable to test our algorithm. Because the images provided by the Google image search API depend on several factors, the outputs may vary based on the time of the query. We report that, AIVAS-IRA requires an average of 40 to 60 seconds to generate the output when it ranks a set of 24 corresponding images for each noun. Figure 1 (a) shows the top four images among 24 images given by the Google image search API for the word 'mango'. Figure 1 (b) shows the top four images that are obtained by AIVAS-IRA by re-ranking the 24 images by Google image search API.

3.3 Implementation of Algorithm

AIVAS generates five-second-long learning materials, each of which includes the spelling, the meaning, the pronunciation data, and an appropriate image of the word to be learned. Initially, a learner is required to input a word that



(a) Top 4 images by Google

(b) Top 4 images by AIVAS-IRA

Fig. 1 Output comparison.

(a) Top-ranked by Google (b) Top-ranked by AIVAS-IRA

Fig. 2 Comparison of learning materials.

he/she intends to acquire. AIVAS accepts the word to be inputted in either a language that he/she is familiar with or in an unknown language. Thereafter, AIVAS generates a learning material in four steps: Firstly, it finds the meaning of the input word from a web translation engine. Secondly, it searches the corresponding images from Google image search API, and extracts an appropriate image by using AIVAS-IRA. The image searching operation is conducted based on the English translation if any non-English word was inputted by the learner. It can be mentioned that a polysemic word is dealt with its default meaning provided by the web translation engine used to implement the system. Thirdly, AIVAS extracts the pronunciation data from a text-to-speech engine. Finally, it embeds the spelling and the meaning as subtitles and generates a learning material by superimposing the subtitles, the pronunciation, and the image together. After that, once a learning material is created, AIVAS displays the created material to the learner so that he/she can acquire the vocabulary.

Figure 2 (a) displays the learning material for the word 'tomato' created by AIVAS with an image that is top-ranked by Google image search API. In contrast, Fig. 2 (b) displays the learning material created by AIVAS with an appropriate image that is recommended by AIVAS-IRA.

In the representation [23], [28] of a learning material, the corresponding image and the spelling of the word are displayed throughout the five seconds. The pronunciation is repeated twice: one and three seconds after, respectively. The meaning is displayed on the screen after two seconds from the beginning. Intention of the delay is to give the

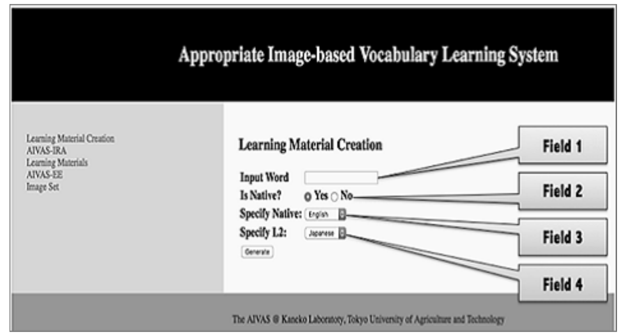


Fig. 3 User Interface of AIVAS

learners an interval to see the spelling of the word and judge whether they already know the word or not.

AIVAS supports eleven widely spoken languages: Chinese, English, French, German, Italian, Japanese, Korean, Polish, Russian, Spanish, and Swedish. Currently, AIVAS does not support other languages. Hence, a non-native learner of the above-mentioned languages is expected to be familiar with at least one of those languages in order to acquire vocabulary through AIVAS. The interface also has been designed with a consideration of learners with inadequate IT skills (Fig. 3). The interface contains four fields. The four fields are Field 1) Input a word to be learned, Field 2) Specify whether the inputted word is in the learner's spoken language or not. Field 3) Specify the spoken language, and Field 4) Specify the target language. In order to generate a learning material, a learner is required to specify all four fields, followed by clicking on the 'Generate' button. Field 1 allows a learner to type/insert a word into the text box. The input should be a single word or a compound word with or without a hyphen. In Field 2, by simply checking into the radio button, a learner is required to specify if the input word is in his/her spoken language or not. AIVAS is programmed in a way that it can generate learning materials for words in both spoken and target languages. Based on the instruction received in Field 2, the system determines the type of learning material to be created. For readers' clarity, if a learner specifies the input as a word of the spoken language (in Field 2), then he/she needs to specify the spoken language (in Field 3) and the target language to be learned (in Field 4). Upon clicking on the 'Generate' button, a learning material will be created to acquire the target language. On the other hand, if an input word is specified as not a word of the spoken language, then the system automatically counts it as the target language which the learner is supposed to acquire. Thereafter, upon specifying the spoken language (in Field 3) and target language (in Field 4), a learning material will be generated to acquire that unknown word. Since AIVAS is not functioned with an automatic detection system of languages, learners are required to select the input and target languages in both Fields 3 and 4 manually.

AIVAS runs on most web browsers, and it was implemented by using HTML5, PHP, and JavaScript technologies. CreateJS, which is a suite of modular libraries, together with tools to create interactive content incorporating with HTML5 was used in the material creation process. Microsoft Translator, the VoiceRSS Text-to-Speech engine, and Google image search API were used as the translation tool, the pronunciation data tool, and the image search tool, respectively. PostgreSQL has been used as the database technology.

4. Learning Effect Investigation

Learning effect investigation has been conducted with 52 participants from 21 different nationalities to assess whether there exist any significant difference between the two different types of learning materials, i.e. those created by the images recommended by AIVAS-IRA and those created by images suggested by Google image search API as top-ranked. Table 1 shows the distribution of the participants' nationalities. The experiment examined the learners' immediate-after, short- and long-term memory retention rates. The experiment includes a pretest immediately before the study session, followed by a posttest 1 (PT1), a delayed posttest (PT2), and an extended delayed posttest (PT3).

A subset of concrete nouns had been set prior to the experiment. In this paper, the subset had been limited to concrete nouns, those representing animal names, fruit names, vegetables names, compound nouns, and object names. It came to our observation that, novice (K1) EFL (English as Foreign Language) learners in some Asian countries often start memorizing words with an ABC book (or English wordbook). An ABC book often provides a gallery of words that are commonly found and frequently used in their native language with the representation of alphabet-word-image, often belonging to the subset. Hence, we conducted experiments using those words in the subset.

We asked a native Russian speaker to choose 20 Russian-English word pairs that would be suitable for Russian learners at the introductory level. Table 2 shows the list of Russian-English word pairs that have been used in the generation of learning materials.

Participants in this experiment were high-school, college, undergraduate and graduate students without having any Russian language background. We prepared a pretest questionnaire consisting of 20 preselected Russian words and provided it to the participants. We requested all the participants to mark the words that they have previously acquired and are familiar with. The pretest result indicated that none of the participants was familiar with any word in the list. Therefore, we have set the pretest scores of all participants to be zero.

Participants were randomly grouped into Experimental Group (EG) and Control Group (CG) so that each group contained 26 participants. The procedure of the experiment was explained to the participants beforehand. Participants in the EG studied Russian words with learning materials created by using the appropriate images recommended by AIVAS-IRA. On the other hand, participants in the CG studied the words with learning materials created by using the top-ranked images suggested by Google image search API. In the study session, each participant was equipped with a laptop with high-speed internet connection, a headphone, and a mouse. A pen and a sheet of white paper were also provided to assist their study session. We allowed the participants to take notes during the study session but we did not allow them to use them during the posttests. We asked the participants to study these 20 Russian words by using learning materials. We set the study session to 10 minutes.

We asked the participants to answer PT1 questions immediately after the study session. We provided print-outs of PT1 questions to the participants and asked them to state their answers on them within 10 minutes. In the PT1, PT2, and PT3 questionnaires, we listed all of the 20 Russian words and asked the participants to write down the meanings of them in Japanese or English. We asked the participants not to study the Russian words by themselves during the intervals between the posttests.

Three weeks after the study session and PT1, we requested all of the 52 participants to sit in for PT2. The purpose of PT2 was to observe the participants' short-term memory retention rates of the newly acquired words. The purpose of the extended delayed posttest (PT3) was to investigate participants' long-term memory retention rates. We requested all of the participants to sit in for PT3 three weeks after PT2. Table 3 shows the results of the experiment.

The t-tests revealed that there was no significant difference in the average scores by the participants in EG and in

Table 1 Distribution of Participant Nationalities.

Group	Nationality (Number of Participants)
EG	Japan (10), Thailand (4), Bangladesh (2), Cambodia (2), China (1), Iran (1), Indonesia (1), India (1), Kenya (1), Pakistan (1), Philippines (1), USA (1)
CG	Japan (8), Thailand (3), Bangladesh (2), Afghanistan (1), Czech Republic (1), Honduras (1), Indonesia (1), India (1), Laos (1), Nigeria (1), Pakistan (1), Qatar (1), South Korea (1), Uyghur (1), USA (1), Vietnam (1)

Table 2 List of Russian-English word pairs.

Russian	English	Russian	English
собака	dog	кошка	cat
тигр	tiger	голубь	pigeon
манго	mango	киви	kiwi
дыни	melon	виноград	grape
огурец	cucumber	картофель	potato
морковь	carrot	капуста	cabbage
самолеты	aircraft	спальня	bedroom
книжнаяполка	bookselft	наушники	earphone
тул	chair	будильник	clock
дерево	tree	камеры	camera

Table 3 Result of the learning effect investigation.

	Average score of PT1 (S.D.)	Average score of PT2 (S.D.)	Average score of PT3 (S.D.)
EG (N = 26)	13.19 (4.24)	5.30 (2.74)	4.65 (2.73)
CG (N = 26)	11.34 (3.94)	4.03 (2.04)	3.00 (2.09)
p-value (t-test)	0.117	0.069	0.020

CG in the immediate-after (p = 0.117) and in the short-term sessions (p = 0.069). However, the average score by the participants in EG significantly outperformed that by the participants in CG in the long-term memory retention rates (p = 0.020). This result indicates that our algorithm could extract images that have a higher learning effect. Actually, the average distance from the center point and the standard deviation of the 20 images top-ranked by AIVAS-IRA for the 20 Russian words were 0.200 and 0.116, respectively. On the other hand, those of the 20 images top-ranked by Google image search API were 0.478 and 0.148, respectively. A t-test on these average distances indicated that there is a significant difference between them ($p = 1.54 \times 10^{-7}$). Thus, we can conclude that our algorithm recommends better images than the Google image search engine by our measurement.

5. Discussion

Nouns can be acquired through still images, graphical images, illustrated images, sketched photos or even video clips. Concrete nouns represented with still images often give learners the advantage to recall old memories. However, inappropriate or irrelevant images may often cause confusion in the learners' mind, leading to delay in learning, or hindering memorization. Determining an appropriate image for each concrete noun is still a matter of debate amongst researchers. A simple text-based query into an image search engine (Yahoo, Google, Flickr etc.) on the web returns thousands of images for a concrete noun. The characteristics of the still images that are commonly found in search return often vary massively. Therefore, both manual and automatic extraction of only one appropriate image among thousands is difficult for both human and system, respectively. Furthermore, learners are often unaware about the types of the images that are considered as appropriate learning resources. This problem is still unsolved in existing literature, although few thoughts regarding this matter have been articulated. This study acquaints its readers with our definition of an appropriate image for representing a concrete noun. To transform the proposed definition into reality, an algorithm has been introduced in this study that is able to re-rank a finite set of corresponding images for each concrete noun, and extract the most appropriate one from it. We believe that, this study will help foreign language learners and instructors in creation of effective learning materials.

However, the current study has several limitations, which will certainly be investigated in future studies. Firstly, this paper revealed the learning effects of the appropriate images for a subset of concrete nouns. However, learning effect investigations with respect to other types of concrete nouns have not been tested. Secondly, AIVAS has limitations in dealing with the words with polysemy, and dealing with long articulated words (long-sound). Thirdly, AIVAS purely depends on web servers; therefore failure of any of the external servers may cause an interruption in learning material creation. Fourthly, in the process of an appropriate image selection (i.e. image re-ranking), AIVAS-IRA is unable to detect irrelevant images. Finally, familiar representations of an object may vary depending on the cultural backgrounds or nationalities. Therefore, it may affect the result of the experiment that included relatively many Japanese participants. Actually, we have divided the 52 participants into a Japanese group and a group of others, and compared their average scores of PT1, PT2, and PT3 by a t-test. As a result, the average score by the group of others is significantly higher than that by the Japanese group at PT2 (p = 0.035). However, there was not a significant difference at PT1 or PT3. Hence, we think that further investigation is necessary.

6. Conclusion and Future Considerations

This study worked on the possibilities of a system to extract an appropriate image for a concrete noun. On that account, we proposed a definition of an appropriate image that can represent a concrete noun in order to assist foreign language learners in learning of new vocabulary. An image re-ranking algorithm called AIVAS-IRA has been designed and implemented. AIVAS-IRA is able to determine an appropriate image for each concrete noun and recommend it to a learner in the process of learning material creation. This study specifically examined the efficacy of the appropriate images in concrete nouns acquisition by measuring participants' memory retention rates after considerable interval of time. 52 participants bearing different cultural backgrounds have participated in the experiment. The experiments measured participants' memory retention rates in immediate-after the study session, short-term (3-week), and long-term (6-week) by conducting a posttest, a delayed posttest, and an extended delayed posttest, respectively. The t-tests revealed that no remarkable differences have been observed in learners' immediate-after (p = 0.117) and in short-term (p = 0.069) memory retention rates. However, a significant difference on learners' memory retention rates has been noted in extended delayed posttest (p = 0.020). Therefore, this study yields the conclusion that, our proposed system AIVAS-IRA is able to recommend images for concrete nouns which can be addressed as appropriate images. Hence, those images can be considered as appropriate learning resources to acquire concrete nouns of a new language.

Further improvements of the AIVAS-IRA will be a key focus in our future studies. We should investigate other parameters such as image color distributions and object features to re-rank images effectively. Also, additional functions in AIVAS should be implemented to deal with polysemic words.

Acknowledgments

This study was partly supported by a Grant-in-Aid for Scientific Research (C) of the Japan Society for the Promotion of Science under Grant No.17K01118.

References

[1] M.N. Hasnine, Y. Hirai, M. Ishikawa, H. Miyakoda, and K. Kaneko, "A vocabulary learning system by on-demand creation of multilinguistic materials based on appropriate images," Proceedings of the 2014 e-Case & e-Tech 2014, pp.343–356, Nov. 2014.

[2] M.N. Hasnine, Y. Hirai, M. Ishikawa, H. Miyakoda, and K. Kaneko, "Learning effects investigation of an on-demand vocabulary learning materials creation system based on appropriate images," Proceedings of the 4th ICT International Student Project Conference, May 2015.

[3] P. Duppenthaler, "Vocabulary acquisition: the research and its pedagogical implications," Tezukayama Gakuin University Research Journal, vol.42, pp.1–14, 2007.

[4] S. Thornbury, How to teach vocabulary, Pearson Education Limited, 2002.

[5] F. Mashhadi and G. Jamalifar, "Second language vocabulary learning through visual and textual representation," Procedia - Social and Behavioral Sciences, vol.192, pp.298–307, 2015.

[6] J.C. Richards and W.A. Renandya, eds., Methodology in language teaching: An anthology of current practice, Cambridge University Press, Cambridge, UK, 2002.

[7] D.A. Wilkins, Linguistics in language teaching, Edward Arnold, 1972.

[8] H.D. Brown, Teaching by Principles: An Interactive Approach to Language Pedagogy, 2 ed., Longman, 2001.

[9] E. Andrews, Markedness Theory, Duke University Press, 1990.

[10] V.F. Allen, Techniques in teaching vocabulary, Oxford University Press, New York, USA, 1983.

[11] A. Hayati, A. Jalilifar, and A. Mashhadi, "Using short message service (sms) to teach english idioms to efl students," British Journal of Educational Technology, vol.44, no.1, pp.66–81, Jan. 2013.

[12] N. Cavus and D. Ibrahim, "m-learning: An experiment in using sms to support learning new english language words," British Journal of Educational Technology, vol.40, no.1, pp.78–91, Jan. 2009.

[13] M. Lu, "Effectiveness of vocabulary learning via mobile phone," Journal of Computer Assisted Learning, vol.24, no.6, pp.515–525, Dec. 2008.

[14] P. Thornton and C. Houser, "Using mobile phones in english education in japan," Journal of Computer Assisted Learning, vol.21, no.3, pp.217–228, 2005.

[15] D.M. Chun and J.L. Plass, "Effects of multimedia annotations on vocabulary acquisition," The modern language journal, vol.80, no.2, pp.183–198, 1996.

[16] S.K. Carpenter and K.M. Olson, "Are pictures good for learning new vocabulary in a foreign language? only if you think they are not.," Journal of Experimental Psychology: Learning, Memory, and Cognition, vol.38, no.1, pp.92–101, 2012.

[17] A. Paivio and K. Csapo, "Picture superiority in free recall: Imagery or dual coding?," Cognitive Psychology, vol.5, no.2, pp.176–206, Sept. 1973.

[18] A. Paivio, T.B. Rogers, and P.C. Smythe, "Why are pictures easier to recall than words?," Psychonomic Science, vol.11, no.4, pp.137–138, April 1968.

[19] J.M. Clark and A. Paivio, "Dual coding theory and education," Educational Psychology Review, vol.3, no.3, pp.149–210, Sept. 1991.

[20] K. Al-Seghayer, "The effect of multimedia annotation modes on L2 vocabulary acquisition: A comparative study," Language Learning & Technology, vol.5, no.1, pp.202–232, Jan. 2001.

[21] Y. Yeh and C.W. Wang, "Effects of multimedia vocabulary annotations and learning styles on vocabulary learning," CALICO Journal, vol.21, no.1, pp.131–144, 2003.

[22] S. Joseph, K. Binsted, and D. Suthers, "Photostudy: Vocabulary learning and collaboration on fixed & mobile devices," Proceedings of the 2005 IEEE International Workshop on Wireless and Mobile Technologies in Education, pp.206–210, 2005.

[23] K. Hasegawa, S. Amemiya, M. Ishikawa, K. Kaneko, H. Miyakoda, and W. Tsukahara, "PSI: A system for creating english vocabulary materials based on short movies," The Journal of Information and Systems in Education, vol.6, no.1, pp.26–33, Dec. 2007.

[24] K. Kaneko, K. Hasegawa, X. Jin, M. Ishikawa, N. Shinagawa, and H. Miyakoda, "Autonomous and collaborative learning environment for vocabulary acquirement," Proceedings of the 30th International Conference on Information Technology Interfaces, pp.147–152, June 2008.

[25] K. Kaneko and H. Miyakoda, "The history of hacking SIGMA: an autonomous and collaborative vocabulary learning system," Proceedings of the 2011 Eighth International Joint Conference on Computer Science and Software Engineering, pp.426–431, May 2011.

[26] Y.-M. Huang, Y.-M. Huang, S.-H. Huang, and Y.-T. Lin, "A ubiquitous english vocabulary learning system: Evidence of active/passive attitudes vs. usefulness/ease-of-use," Computers & Education, vol.58, no.1, pp.273–282, 2012.

[27] O. Anonthanasap, C. He, K. Takashima, T. Leelanupab, and Y. Kitamura, "Mnemonic-based interactive interface for second-language vocabulary learning," Proceedings of the Human Interface Society, vol.14, Sept. 2014.

[28] K. Kaneko, K. Hasegawa, S. Amemiya, M. Ishikawa, H. Miyakoda, and W. Tsukahara, "Multi-linguistic learning materials for vocabulary acquirement based on universal images," Proceedings of the Sixth International Internet Education Conference, Sept. 2007.

[29] J. Deng, W. Dong, R. Socher, L.J. Li, K. Li, and L. Fei-Fei, "Imagenet: A large-scale hierarchical image database," Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp.248–255, June 2009.

[30] J. Deng, A.C. Berg, K. Li, and L. Fei-Fei, "What does classifying more than 10,000 image categories tell us?," Proceedings of the 11th European Conference on Computer Vision: Part V, vol.6315, pp.71–84, Sept. 2010.

[31] G.A. Miller, R. Beckwith, C. Fellbaum, D. Gross, and K.J. Miller, "Introduction to wordnet: An on-line lexical database," International Journal of Lexicography, vol.3, no.4, pp.235–244, 1990.

[32] G.A. Miller, "Wordnet: A lexical database for english," Communications of the ACM, vol.38, no.11, pp.39–41, Nov. 1995.

[33] N. Ben-Haim, B. Babenko, and S. Belongie, "Improving web-based image search via content based clustering," Proceedings of the 2006 Conference on Computer Vision and Pattern Recognition Workshop, p.106, June 2006.

[34] V. Jain and M. Varma, "Learning to re-rank: Query-dependent image re-ranking using click data," Proceedings of the 20th International Conference on World Wide Web, pp.277–286, March 2011.

Mohammad Nehal Hasnine is a Ph.D. candidate in the Graduate School of Engineering at Tokyo University of Agriculture and Technology in Japan. His main research interest is pedagogical systems. He received the B.S. degree from Stamford University, Bangladesh in 2009 and the M.E. degree from Tokyo University of Agriculture and Technology in 2015.

Masatoshi Ishikawa is an Associate Professor at Tokyo Seitoku University in Japan. His main interests are in database systems, pedagogical systems and area informatics. He received the B.E. degree from Tokyo University of Agriculture and Technology in 1995. He received the M.E. and Ph.D. degrees from Nara Institute of Science and Technology in 1997 and 2004, respectively. He is a member of IEEE CS, ACM, IPSJ, and IEICE.

Yuki Hirai is a Lecturer at Shinshu University in Japan. His main research area is educational technologies. He received the B.L.A. and M.Ed. degrees from Tokyo Gakugei University in 2007 and 2009, respectively. He also received the Ph.D. degree from University of Tsukuba in 2012. He is a member of IPSJ and JSAI.

Haruko Miyakoda is a Professor at Tsuda University in Japan. Her main research areas include L1 language acquisition, phonology, phonetics, and second language acquisition in higher education. She received her B.A., M.A. and Ph.D. degrees from Sophia University. She is a member of ICPLA, IASCL, and JELS.

Keiichi Kaneko is a Professor at Tokyo University of Agriculture and Technology in Japan. His main research areas are functional programming, parallel and distributed computation, partial evaluation, fault-tolerant systems, and multimedia pedagogical systems. He received the B.E., M.E., and Ph.D. degrees from the University of Tokyo in 1985, 1987, and 1994, respectively. He is a member of ACM, IEEE CS, IEICE, IPSJ, and JSSST.