

Ph.D. Thesis

---

Recommendation of Appropriate Images for Vocabulary Learning

---

Author

**Mohammad Nehal Hasnine**

Supervisor

Professor Keiichi Kaneko

Graduate School of Engineering  
Tokyo University of Agriculture and Technology

March, 2018

## COMMITTEE MEMBERS

Masaki Nakagawa, Ph.D.

Professor of Department of Computer and Information Sciences  
Tokyo University of Agriculture and Technology

Keiichi Kaneko, Ph.D. (Supervisor)

Professor of Department of Computer and Information Sciences  
Tokyo University of Agriculture and Technology

Hironori Nakajo, Ph.D.

Associate Professor of Department of Computer and Information Sciences  
Tokyo University of Agriculture and Technology

Seiji Hotta, Ph.D.

Associate Professor of Department of Computer and Information Sciences  
Tokyo University of Agriculture and Technology

Kaori Fujinami, Ph.D.

Associate Professor of Department of Computer and Information Sciences  
Tokyo University of Agriculture and Technology

# **Recommendation of Appropriate Images for Vocabulary Learning**

Mohammad Nehal Hasnine

Department of Electronics and Information Engineering

Graduate School of Engineering

Tokyo University of Agriculture and Technology

## **ABSTRACT**

This thesis presents a scientific study on vocabulary learning as an application domain of computer-science and language pedagogy. We investigated a problem in the domain of image-based vocabulary learning/teaching, namely how to recommend appropriate images to represent words to be learned. I propose approaches for extracting appropriate images to represent concrete and abstract nouns. First, a definition of an appropriate image to represent a concrete noun is proposed. Then, an algorithm is designed to judge still images and determine the most appropriate one. For this, we employed FFT-based feature extraction method in power spectrum to analyze image features. Then a definition of an appropriate image to represent an abstract noun is proposed and evaluated. The evaluation failed to perform well in choosing only one appropriate image for representing an abstract noun. Therefore, we adapted feature-based categorical approach for recommending appropriate images to represent abstract nouns. We employed deep CNN-based unsupervised learning feature extraction method on our image set to analyze features. An appropriate image recommendation system has been developed based on our algorithms. The system is able to extract an appropriate image for representing a concrete noun together with the categorical recommendation for representing an abstract noun. Evaluations with subsets of both concrete and abstract nouns have been carried out to assess the vocabulary-learning effect of the appropriate images. This thesis concludes that our proposed image recommendation system is able to extract and recommend appropriate images for representing nouns that aid in quick memorization and long-term memory retention.

## THESIS ORGANIZATION

Chapter 1 starts with a general introduction on the importance of foreign language in linking to the roles of vocabulary in foreign language development. In this chapter, we emphasize the roles of noun acquisition in learning a new language. We correlated nouns with images and discussed noun imageability from the perspective of pedagogy. We also highlight our major contributions.

Chapter 2 contains a summary of the literature that have been reviewed for understanding the contemporary research and development in the field of foreign vocabulary acquisition.

Chapter 3 presents the theoretical background behind our motivation to put appropriate image recommendation as the central focus of this study. We also discuss the approaches that we have followed to define appropriate images for representing concrete nouns and abstract nouns.

Chapter 4 describes our key technological developments. Vocabulary learning material creation system, image sets preparation, algorithm design, appropriate image recommendation system development etc. are discussed in this chapter. Furthermore, technical specifications are also articulated in this chapter.

Chapter 5 details all the experiments we have conducted to support this study. This chapter discusses the experimental procedures, results, and outcome of each experiment together with the information about participants, used materials, feedback, and so on.

Chapter 6 reflects on the related aspects of the suitability of images in foreign vocabulary learning for memorizing vocabulary from the other parts of speech like verbs, adjectives etc.

Chapter 7 concludes this thesis by summarizing the key outcomes of this doctoral study.

Chapter 8 points out some limitations of this thesis is written, and discusses approaches to overcome them in future studies.

# Table of Contents

ABSTRACT .....	III
THESIS ORGANIZATION .....	IV
TABLE OF CONTENTS .....	V
<b>1. INTRODUCTION.....</b>	<b>12</b>
1.1 PROBLEM IDENTIFICATION IN TRADITIONAL LEARNING APPROACHES .....	13
1.2 TECHNOLOGY-ENHANCED VOCABULARY LEARNING .....	14
1.3 NOUNS BEFORE VERBS .....	14
1.4 THE POWER OF IMAGES IN MEMORIZATION .....	15
1.5 CONTRIBUTIONS .....	16
1.6 SUMMARY.....	18
1.7 OUTLINE .....	18
1.8 TERMINOLOGY.....	20
<b>2. STATE-OF-THE-ART .....</b>	<b>23</b>
2.1 TECHNOLOGY-DRIVEN RESEARCH ON LEARNING SYSTEM DEVELOPMENT.....	23
2.1.1 <i>CALL-mediated Spaced Retention.....</i>	<i>24</i>
2.1.2 <i>Web-based Learning.....</i>	<i>24</i>
2.1.3 <i>Variants .....</i>	<i>25</i>
2.2 REPRESENTATION OF MULTIMEDIA ANNOTATIONS.....	27
2.3 EMERGING TECHNOLOGIES FOR SUPPORTING FOREIGN VOCABULARY LEARNING .....	28
2.3.1 <i>Multimedia Annotation-based Learning Material Creation Systems .....</i>	<i>28</i>
2.3.2 <i>Augmented Reality-based Situated Learning Systems .....</i>	<i>43</i>
2.3.3 <i>Game-based Learning Systems.....</i>	<i>43</i>
2.3.4 <i>SMS-based Learning Systems .....</i>	<i>44</i>
2.4 SUMMARY.....	46
<b>3. APPROPRIATE IMAGES FOR NOUNS .....</b>	<b>48</b>
3.1 MOTIVATION.....	48
3.2 AN APPROPRIATE IMAGE FOR REPRESENTING A CONCRETE NOUN.....	50
3.2.1 <i>Pedagogical Investigation I.....</i>	<i>50</i>
3.2.2 <i>Pedagogical Investigation II .....</i>	<i>53</i>
3.2.3 <i>The Proposed Definition of an Appropriate Image.....</i>	<i>54</i>
3.3 APPROPRIATE IMAGES FOR REPRESENTING ABSTRACT NOUNS.....	55

3.3.1	<i>Defining an Abstract Noun</i> .....	55
3.3.2	<i>Approach 1</i> .....	57
3.3.3	<i>Approach 2</i> .....	61
3.4	SUMMARY.....	62
<b>4.</b>	<b>AIVAS SYSTEM</b> .....	<b>63</b>
4.1	OVERVIEW .....	63
4.2	APPROPRIATE IMAGE RECOMMENDATION SYSTEM.....	66
4.2.1	<i>Algorithm Design</i> .....	67
4.2.2	<i>AIVAS Image Sets</i> .....	68
4.2.3	<i>Determination of Feature Extraction Methods</i> .....	76
4.2.4	<i>Algorithms Implementation &amp; Output Demonstration</i> .....	80
4.3	LEARNING MATERIAL CREATOR .....	87
4.4	EXPERIMENTAL ENVIRONMENT .....	91
4.5	TECHNICAL SPECIFICATIONS .....	92
4.6	SUMMARY.....	94
<b>5.</b>	<b>EXPERIMENTS</b> .....	<b>95</b>
5.1	IMAGE EVALUATION EXPERIMENT I.....	95
5.1.1	<i>Approach</i> .....	95
5.1.2	<i>Result</i> .....	97
5.1.3	<i>Discussion</i> .....	98
5.2	LEARNING EFFECT INVESTIGATION I.....	99
5.2.1	<i>Approach</i> .....	99
5.2.2	<i>Result</i> .....	102
5.2.3	<i>Discussion</i> .....	103
5.3	IMAGE EVALUATION EXPERIMENT II.....	105
5.3.1	<i>Approach</i> .....	105
5.3.2	<i>Result</i> .....	107
5.3.3	<i>Discussion</i> .....	108
5.4	LEARNING EFFECT INVESTIGATION II.....	110
5.4.1	<i>Approach</i> .....	110
5.4.2	<i>Result</i> .....	115
5.4.3	<i>Discussion</i> .....	116
5.5	SUMMARY.....	117
<b>6.</b>	<b>ASPECTS OF IMAGE APPROPRIATENESS IN VOCABULARY LEARNING</b> .....	<b>118</b>
<b>7.</b>	<b>CONCLUSION</b> .....	<b>121</b>

8. LIMITATIONS AND FUTURE DIRECTIONS.....	124
APPENDICES.....	126
BIBLIOGRAPHY.....	129
RELATED PUBLICATIONS.....	138
ACKNOWLEDGEMENT.....	139
INDEX.....	140

# List of Tables

<i>Table 1-1 Glossary of Terms</i> .....	20
<i>Table 2-1: Features of the Experimental Systems that Use Multimedia Annotations</i> .....	40
<i>Table 3-1 Word List Used in Pedagogical Investigation I</i> .....	51
<i>Table 3-2 Distribution of the Participants</i> .....	51
<i>Table 3-3 Result of the Pedagogical Investigation I</i> .....	53
<i>Table 3-4 Participants Detail</i> .....	53
<i>Table 3-5 Examples of Abstract Nouns</i> .....	56
<i>Table 3-7 The Subset of the Abstract Nouns</i> .....	57
<i>Table 3-8 Examples of Targeted and Non-targeted Abstract Nouns</i> .....	58
<i>Table 3-9 Types of Inappropriate Images</i> .....	58
<i>Table 4-1 The Pseudo Code</i> .....	67
<i>Table 4-2 Overview of AIVAS Image Sets</i> .....	69
<i>Table 4-3 List of the English Words Used in Preparing the AIVAS-CNCRT59 Image Set</i> .....	70
<i>Table 4-4 Word List Used for Preparing the AIVAS-ABST-LS68 Image Set</i> .....	71
<i>Table 4-5 Participants Details</i> .....	71
<i>Table 4-6 Details on Image Collection</i> .....	72
<i>Table 4-7 Participant Details</i> .....	73
<i>Table 4-8 Development Tools</i> .....	92
<i>Table 5-1 Words and Their Corresponding Category</i> .....	96
<i>Table 5-2 Result of Image Evaluation Experiment I</i> .....	98
<i>Table 5-3 List of Russian-English Word Pairs</i> .....	99
<i>Table 5-4 Distribution of Participant Nationalities</i> .....	100
<i>Table 5-5 Result of the Learning Effect Investigation I</i> .....	102
<i>Table 5-6 Result of the Analysis</i> .....	102
<i>Table 5-7 Male -vs-Female Participants in Experimental Group</i> .....	103
<i>Table 5-8 Male -vs-Female Participants in Control Group</i> .....	103
<i>Table 5-9 Male -vs- Female Participants</i> .....	103
<i>Table 5-10 Feedbacks</i> .....	104
<i>Table 5-11 List of the Words</i> .....	105
<i>Table 5-12 Participant Details</i> .....	106
<i>Table 5-13 Result of the Image Evaluation Experiment II</i> .....	107
<i>Table 5-14 List of English-Polish Word Pairs</i> .....	110
<i>Table 5-15 Details on Google Top-ranked-vs-Appropriate Images</i> .....	112
<i>Table 5-16 Distribution of Participant Nationalities</i> .....	114



<i>Table 5-17 Result of the Learning Effect Investigation II.....</i>	<i>115</i>
<i>Table 5-18 Result of the Analysis.....</i>	<i>115</i>
<i>Table 5-19 Feedbacks.....</i>	<i>116</i>

# List of Figures

<i>Figure 2-1 Design of WBLL Material (Hamamrad, 2016)</i> .....	25
<i>Figure 2-2: Learning Environment (Joseph, 2005)</i> .....	29
<i>Figure 2-3 Interfaces and Study Environment in Vidioms System (Thornton P. &amp;, 2005)</i> .....	30
<i>Figure 2-4 The Backbone of PSI System (on left) and a Learning Material (on right) (Hasegawa K., 2007) ....</i>	32
<i>Figure 2-5 The Architecture of the MultiPod System (Hasegawa K., 2007) .....</i>	33
<i>Figure 2-6 Overview of the PHI System (Kaneko, 2007) .....</i>	33
<i>Figure 2-7 Learning Materials for Preposition Learning (Wong, 2010)</i> .....	34
<i>Figure 2-8 The Framework of the UEVL System (Huang, 2012)</i> .....	35
<i>Figure 2-9 A Sample Learning Material Created by UEVL system (Huang, 2012)</i> .....	35
<i>Figure 2-10 Learning Environment (Agca, 2013)</i> .....	36
<i>Figure 2-11 A Learning Material (Anonathanasap, 2014)</i> .....	37
<i>Figure 2-12 User Interfaces of Word Learning-CET6 App (Wu, 2015)</i> .....	38
<i>Figure 2-13 SCROLL-created Learning Material (Ogata H. L.-B., 2011)</i> .....	39
<i>Figure 2-14 Situated Vocabulary Learning Contents Using AR Technology (Santos, 2016) .....</i>	43
<i>Figure 2-15 Snapshot of the Learning Environment (Sahrir, 2012)</i> .....	44
<i>Figure 2-16 The Architecture of the SMS-based Vocabulary Learning System (Hayati, 2013)</i> .....	45
<i>Figure 3-1 Sample Learning Materials</i> .....	52
<i>Figure 3-2 Evaluation Questionnaire .....</i>	52
<i>Figure 3-3 An Example of the Variations in a Search Query for An Abstract Noun</i> .....	57
<i>Figure 3-4 Examples of Inappropriate Image Type 1 .....</i>	58
<i>Figure 3-5 Examples of Inappropriate Image Type 2 .....</i>	59
<i>Figure 3-6 Examples of Inappropriate Image Type 3 .....</i>	59
<i>Figure 4-1 Architecture of the AIVAS</i> .....	64
<i>Figure 4-2 AIVAS Subsystems</i> .....	65
<i>Figure 4-3 Survey Questionnaire .....</i>	74
<i>Figure 4-4 Frequency Domain-based Features (www.cs.auckland.ac.nz, n.d.)</i> .....	76
<i>Figure 4-5 CNN-based Feature Extraction and Classification (MathWorks, n.d.)</i> .....	78
<i>Figure 4-6 The Architecture of the Pre-trained AlexNet Neural Network (Krizhevsky, 2012)</i> .....	79
<i>Figure 4-7 Output Demonstration for Concrete Nouns</i> .....	81
<i>Figure 4-8 Identification of the Elbow Point</i> .....	83
<i>Figure 4-9 Distribution of the Images in 6 Clusters</i> .....	84
<i>Figure 4-10 Categorical Recommendation for an Abstract Noun</i> .....	86
<i>Figure 4-11 The Operation of the AIVAS-LMC</i> .....	88
<i>Figure 4-12 AIVAS-LMC Interface</i> .....	89

<i>Figure 4-13 The Format of a Learning Material in the AIVAS System.....</i>	<i>90</i>
<i>Figure 4-14 Learning Material Samples.....</i>	<i>90</i>
<i>Figure 4-15 An Interface of the Experimental Environment.....</i>	<i>91</i>
<i>Figure 4-16 AIVAS-AIRS Interface .....</i>	<i>93</i>
<i>Figure 5-1 An ABC Book Gallery.....</i>	<i>96</i>
<i>Figure 5-2 Comparison of Learning Materials.....</i>	<i>100</i>
<i>Figure 5-3 Flow of Learning Effect Investigation I.....</i>	<i>101</i>
<i>Figure 5-4 Example of a Satisfactory Output.....</i>	<i>109</i>
<i>Figure 5-5 Example of an Unsatisfactory Output.....</i>	<i>109</i>
<i>Figure 5-6 Determine an Appropriate Image .....</i>	<i>111</i>
<i>Figure 5-7 Examples of Learning Material .....</i>	<i>113</i>
<i>Figure 5-8 Flow of Learning Effect Investigation II.....</i>	<i>114</i>

# 1. Introduction

Second language learning is required in many countries, and therefore students must take courses like English as the Second Language (ESL), French as the Foreign Language (FFL), Spanish as the Additional Language (SAL) or English for the Speakers of Other Languages (ESOL) etc. together with their general education. There can be different reasons for acquiring a second or a third language, however, the benefits of being multilingual are enormous. Whether viewed from an educational point of view or from a social point of view, a foreign language helps to make real connections with people. Second language learning, along with general education, helps one's personal growth by enhancing global perspective and cross-cultural communication skills, and by sharpening cognitive skills and critical thinking. Furthermore, multilingualism improves employment potential, travel opportunity, and understanding of international literature, music, and films etc. For all these reasons, there has been a noticeable increase in the second language learning among students and adult learners over the last decades.

Vocabulary learning refers to memorizing unfamiliar words either in one's native language or in a new language. Whether one is learning a native or a new language, one starts by memorizing unfamiliar words (Thornbury, 2006) (Mashhadi, 2015), which can be distressing (Yang W. &, 2011). Wilkins, emphasized the need to acquire significant vocabulary for communication, "without grammar very little can be conveyed but without vocabulary, nothing can be conveyed" (Wilkins, 1972). Hence, vocabulary is regarded as a vital element of the language capability and delivers much of the foundation for how well novice learners communicate (Lam, 2002). No language learning, whether native or foreign, can take place without the memorization of significant vocabulary. Acquiring significant vocabulary is also considered to be a pillar of learners' ability that accelerates studying any kind of language-related tasks (Gorjian, 2012). Insufficient vocabulary skills can be the reason for miscommunication or poor communication. Moreover, significant vocabulary can arguably differentiate between nervous speakers and expert speakers. Unarguably, vocabulary learning is a compulsory part of being skillful in a new language. Consequently, vocabulary learning is critically important to typical language learners in learning a new language (Coady, 1997).

## 1.1 Problem Identification in Traditional Learning Approaches

The natural ways of memorizing vocabulary in a native language are from the surroundings, through real-life contexts, relating a picture with a word, social experiences or in the classroom. However, vocabulary memorization in a new language through natural ways can be daunting. Therefore, the conventional way of foreign vocabulary learning is educator-centered, where an educator is responsible to teach foreign vocabulary in a formal educational setting such as a classroom or one-to-one lesson. Nevertheless, vocabulary learning in formal educational settings has been given little attention. Vocabulary learning is often considered to be a time-consuming process in the formal educational settings. Therefore, the mutual understanding between the teachers and the students is that the vocabulary has to be acquired by self-study. In the formal educational settings, the theoretical discussions, writings, the teaching of syntactic structures and pronunciation, especially the functional knowledge of the former is regarded as the key to foreign language acquisition (Marton, 1977). Vocabulary learning is often treated as a problem marginal to other language learning activities, as it has been a matter of common belief that the acquisition of foreign lexicon is a by-product of having the learner participate in these other activities (Marton, 1977). Furthermore, vocabulary teaching in the formal educational settings has been given less priority during the last decades for the following reasons. Firstly, teachers believed that in the classroom, more emphasis should be given on the grammar than on the vocabulary. Secondly, experts in language pedagogy believed that students often make mistakes in sentence creation if they have memorized too many words before mastering the basic grammar. Thirdly, linguists believed that word meaning should be acquired through daily experiences and cannot be taught effectively in the classroom (Allen, 1983). As a result, vocabulary is considered to be one of the neglected areas in the formal educational settings.

Forgetting newly memorized words is inevitable, because what is hard to memorize is often easy to forget. Newly acquired vocabulary, whether learned in the classroom or through self-study, requires continuous practice. As a result, efforts need to be made outside the classroom to recall newly learned words. Because of this, intentional vocabulary learning in informal settings is gaining more popularity among motivated foreign language learners. Supporting research also concluded that the intentional vocabulary learning in informal settings is the key for most of the English as the Foreign Language (EFL) learners' vocabulary expansion, because new words in a foreign language are often problematic and slow to be memorized without ambiguity out of context (Wu, 2015). As a result, research attention to developing experimental and commercial systems to support foreign vocabulary learners in informal

settings is gaining popularity. A number of applications on both web and mobile platforms have already been introduced.

## **1.2 Technology-enhanced Vocabulary Learning**

The emergence of computer technologies accelerated the dramatic implementation of new technology-enhanced vocabulary learning systems to support both formal and informal learning. The affordability of computers and the pervasion of internet technologies have enabled learners to engage themselves in vocabulary learning in their free time. Moreover, the convenience of smart phones and mobile phones have raised the learners' interest greatly to learn new words interactively and collaboratively. We have witnessed a release of several experimental systems from the early 2000s to until now. Also, there are many commercially available free software like duolingo, rosetta stones etc., and paid application packages like Vocab1, WordPal, SpeedStudy etc. which are available to the English as the Second Language (ESL) learners. With the help of these systems, learners can learn whenever they want without direct supervision of a language teacher. An advantage commonly claimed by these systems is that they offer personalized intelligent tutoring, though this is disputed. Regardless of the debate on how intelligent the existing vocabulary learning systems are, we cannot ignore the fact that these systems have been implemented using the cutting-edge learning/teaching methodologies. Also, multimedia technologies (collective forms of media and contents together with each other at the same time) have been adapted in a scientific manner so that the learners can learn effectively. We cannot overlook the importance of technology-enhanced vocabulary learning tools in an informal setting (when an instructor is not available, or outside the classroom).

## **1.3 Nouns before Verbs**

In a general sense, vocabulary learning refers to memorizing various meanings that can be used in constructing sentences. However, either in native or a foreign language, nouns are considered to be the key component in language development. Several studies established the logic behind memorizing nouns before other components including verbs and other predicate terms. It is often said that the first words that children acquire are nouns. This has been documented as evidence that the concepts referred to by nouns are particularly attainable to infants: they are non-identical from, and conceptually more fundamental than the concepts referred to by verbs or prepositions (Gentner, 1982). Moreover, the concept of asymmetry is presupposed in the markedness theory (Andrews, 1990), according to which some language

formations and notions are considered to be more conventional compared to others. For example, children often memorize nouns before verbs or prepositions. This happens because nouns, especially concrete nouns, correlate with specific “objects” and are easier to memorize than verbs, which represent “events”, whose meaning is much more complicated to comprehend. This relationship between nouns and verbs is known to be an asymmetric relationship. This means that for novice learners of any language, and particularly for children, the more general and unmarked structures (or concepts) are much easier to memorize. Whereas the less general and more complex items such as verbs are difficult to memorize, and are acquired later. In this sense, the memorization of verbs implies the memorization of nouns. Hence, noun memorization is considered to be a starting point in grammatical categorization for both native and foreign languages. Although there are some arguments to the contrary (Tardif, 1996), we assume that foreign language learners, in learning a new language, often start by memorizing nouns.

#### 1.4 The Power of Images in Memorization

The English idiom “*a picture is worth of thousand words*” has motivated many researchers to find the right image to represent a word. We cannot ignore the power of visuals in human brain. Images have many cognitive benefits in the human brain while experiencing a new thing or recalling an old memory. Some cognitive benefits of images in learning new things can be articulated as:

- Images stick in our long-term memory. Quoting from a book titled *Visual Literacy: Learn To See, See To Learn* (Burmark, 2002) “*...unless our words, concepts, ideas are hooked onto an image, they will go in one ear, sail through the brain, and go out the other ear. Words are processed by our short-term memory where we can only retain about seven bits of information. Images, on the other hand, go directly into long-term memory where they are indelibly etched*”
- Images help to transmit messages faster to our brains. Visuals are processed 60000 times faster in the brain than text (Alliance, 2014).
- Images improve comprehension by affecting learners on a cognitive level and stimulating imagination, thereby enabling learners to process information faster (Gutierrez, 2014). A 1997 study found that images can improve memorization by up to 400 times (Machine, 1997).
- Images trigger emotions by engaging with the content, and such emotional reactions influence information retention (Gutierrez, 2014).

- Images motivates learners to respond better to visual information than text alone (Gutierrez, 2014).
- Images play a role in improving memory and stimulate the growth of cerebral cortex (Dewar, 2014).

Memorizing words with the help of images is not a new thing to practice, especially for nouns because they are considered to be more imageable than verbs, adjectives, preposition etc. Noun imageability (discussed in more detail in Sec. 3.1) had been an active area of research interest by experimental psychologists and linguists since the 1960s because of its many cognitive benefits. However, no significant research outcomes have been observed because of the challenges involved in creating the right image to represent a noun. Finding the right image to represent a noun is quite problematic for both humans and computers. As a result, in the old days, noun imageability was not an active research area among computer scientists. However, with dramatic developments in the computer technology, contemporary researchers focus on many areas of image processing for specific purpose such as medical, educational, commercial etc. Currently available image search engines are much more powerful compared with what was available two decades ago. Nowadays, image search engines are considered to be the best source for finding images. Many systems have been implemented that primarily recommend images for commercial products. However, no significant research contributions have been observed to find the right images to create image-based vocabulary learning material. This study focuses on this area to provide a solution. Thus, the focus of research in this study is on the extraction and recommendation of appropriate images to represent nouns.

## 1.5 Contributions

The focus of this study is to assist motivated foreign language learners in the acquisition of foreign language vocabulary in an informal learning with the help of appropriate images. At present, image-based vocabulary learning in either formal or informal setting is a research area in the field of educational technology, and is getting attention around the globe (Joseph, 2005) (Hasegawa K., 2007) (Ishikawa, 2007) (Kaneko, 2007) (Wong, 2010) (Huang, 2012) (Agca, 2013) (Kalyuga, 2013) (Anonathanasap, 2014) (Wu, 2015) (Santos, 2016) (Uosaki N. O., 2017).

A key limitation of existing approaches is the recommendation of appropriate images for learning foreign vocabulary. Although both experimental and commercially available systems use images in preparing their learning materials, none of them have clarified what types of



images have cognitive benefits for learning. Also, existing studies (details will be provided in Section 2.3.1) failed to reveal mechanism to extract right images for representing words that can be considered as appropriate educational resources. As a result, instructor-suggested images are used in most of these systems for preparing the learning material.

In this study, in order to provide a solution to this problem, we worked on the idea to recommend appropriate images for representing words that can be utilized for foreign vocabulary learning. By implementing this idea, we aim to contribute to solving the problem that foreign language learners face in determining appropriate images for words to be learned. This study emphasizes not just building systems but also analyzing the pedagogical aspects involved in foreign vocabulary learning. Here is a summary of our contributions:

1. The concept of an appropriate image for vocabulary learning is introduced. As the first step to establish the concept, we carried out two pedagogical investigation, and based on their results a definition for an appropriate image for representing a concrete noun is proposed.
2. Next, an algorithm that is able to evaluate still images and extract only one appropriate image for representing a concrete noun is proposed. This algorithm is able to rank images that meet our definition.
3. Then, we worked on two hypotheses for determining appropriate images to represent abstract nouns. The first hypothesis yields a definition for an appropriate image to represent an abstract noun belonging to a particular subset of abstract nouns. The second hypothesis worked on a category-based recommendation of appropriate images. Algorithms that can achieve these tasks are implemented too.
4. Without relying on existing image sets, we prepared our own image sets to test the algorithms. Image sets are prepared based on learners-recommended and authors-selected sample appropriate images.
5. Next, an image recommendation system was implemented based on these algorithms. Our system is able to recommend appropriate images for both concrete and abstract noun. For a concrete noun, the system can recommend the most appropriate image to represent it. On the other hand, appropriate images are recommended for abstract noun in categories, instead of individual single image.
6. A web-based vocabulary learning system that supports multiple language learning is developed. The system allows a learner to create his/her learning material on demand. Moreover, while creating the learning material in this system, all the learning resources (an appropriate image, text, pronunciation, and meaning) are extracted automatically from the cloud services. As a result, unlike other existing

systems, learners do not have to accumulate learning resources by themselves in our system. This is expected to save a considerable amount of time for the learner in preparing learning material.

7. Finally, this study reports the feedback and learning data from 204 participants, which was collected and analyzed between 2014 and 2017. We also report on data collection and methods of analysis.

We believe that the concept of appropriate image will bring a new dimension to the field of technology-enhanced vocabulary learning, particularly for image-based vocabulary acquisition, which is still a young subdomain in educational technology development.

## 1.6 Summary

In this chapter, we have focused on foreign vocabulary learning and discussed related issues. We briefly reflected on the necessity of foreign language learning together with general education; and then discussed the role of vocabulary in language development. After this we identified the problems faced by learners in traditional learning approaches, followed by a discussion of the merits of technology-enhanced learning. Then we discussed why nouns come before verbs when memorizing new words, thereby choosing to focus on nouns for this research. After this we reflected on how the linguists' research on the imageability of nouns has become a topic of interest for computer scientists. Finally, we pointed out some key limitations of the existing studies and listed our key contributions to solve these limitations.

This chapter also describes the outline of this dissertation and the terminology used in writing it.

## 1.7 Outline

This thesis is organized into seven chapters. A brief outline of the remaining chapters is as follows:

Chapter 2 contains a summary of the literature we have reviewed for understanding the contemporary research and development in the field of foreign vocabulary acquisition. The review emphasizes three areas to understand the research on technology-enhanced vocabulary

learning: first, technology-driven research on vocabulary learning system development, second, the representation of multimedia annotations, and third, the emerging technologies for supporting an informal learning of foreign vocabulary.

Chapter 3 presents the theoretical background for our approach to the extraction of the appropriate images for representing nouns (concrete and abstract nouns only). We discuss our motivation by articulating the importance of suitable images for memorizing nouns and some limitations of the currently available image search engines. After that, in two sections we discuss our approach to define appropriate images for representing concrete nouns and abstract nouns, respectively.

Chapter 4 presents our key technological developments. Here we briefly introduce our system AIVAS and describe its subsystems. The creation of vocabulary learning material, image sets, AIVAS-IRA algorithm design, appropriate image recommendation system etc. are all discussed in this chapter. Furthermore, technical specifications for AIVAS system are articulated in this chapter as well.

Chapter 5 describes the evaluation experiments we carried out to assess our ideas. We describe the experimental procedures, results, and outcome of each experiment together with the information about the participants, material used, feedbacks, and so on.

Chapter 6 discusses some aspects on the appropriateness of images in foreign vocabulary learning for other parts of speech, such as, verbs, pronoun, adjectives etc. This chapter explains why we limited our research focus on nouns instead of considering other parts of speech as well.

Chapter 7 concludes this thesis by summarizing the key results of this doctoral study.

Finally, Chapter 8 points out some current limitations of this study, and suggestions to overcome them in future studies.

## 1.8 Terminology

**Table 1-1** Glossary of Terms

Acronyms and Terms	Definitions
Appropriate Image	An educational image that will have cognitive benefits over to an inappropriate (i.e. any standard) image (The term appropriate image was the key in this thesis)
AIVAS	Appropriate Image-based Vocabulary Learning System
AIVAS-IRA	AIVAS-Image Reranking Algorithm
AIVAS-AIRS	AIVAS-Appropriate Image Recommendation System
AIVAS-EE	AIVAS-Experimental Environment
AIVAS-LMC	AIVAS-Learning Material Creator
API	Application Program Interface
ADDIE	Analysis, Design, Development, Implementation, Evaluation
B4A	A platform to develop android applications
CALL	Computer Assisted Language Learning
ESL	English as the Second Language
EFL	English as the Foreign Language
ESOL	English for the Speakers of Other Languages
e-Learning	Electronic Learning, learning with the help of electronic media
FFT	Fast Fourier Transformation
FSL	French as Second Language
GPS	Global Positioning System
HodgePodge	HodgePodge is a system that automatically adds aural information and subtitles to the movies/still images
Intelligent Tutoring	Presents individualized inputs based on a student' s needs and learning behavior (Malpani, 2011)

InIT	Inappropriate Image Type
L1	First language generally one's mother tongue
L2	Refers to a second language an individual acquires
Multimedia Technologies	In developing a computer-based application, it refers to applying interactive multimedia annotations (such as text, image, video, sound, animation, graphical drawing etc.) to convey a message
Multimedia Annotations	Examples are, text annotation, image annotation, video annotation, sound annotation etc.
MultiPod	MultiPod is a vocabulary learning system that runs in mobile platform
MP3	An audio coding format
m-Learning	Mobile Learning, learning with the technologies that are portable
MALL	Mobile Assisted Language Learning
MEM	Memrise, a language learning platform
MyCLOUD	My Chinese Language ubiquitous learningDays
NASA-TLX test	The NASA-Task Load Index is a widely used multidimensional subjective technique to measure workload (Colligan, 2015)
PSI	Personal Super Imposing
PDA	Personal Digital Assistant
RFID	Radio Frequency Identification
Rosetta Stone	An online language-learning software
SMALL	Seamless Mobile-Assisted Language Learning Support System
SVL	Systematic Vocabulary Learning
SCROLL	System for Capturing and Reminding Of Learning Log
SAL	Spanish as the Additional Language
SLA	Second Language Acquisition
TMM	Tell Me More, a language learning software
u-Learning	Ubiquitous Learning, learning with the help of ubiquitous technologies with the intention of anytime-anywhere convenience

UEVL	Ubiquitous English Vocabulary Learning
Vidioms	A Web-based English Idioms Learning Site
WBLL	Web-based Language Learning
WOW	Writing Online Workshop, a ESL writing tool
WiFi	Wireless Network
3G	Third Generation

## 2. State-of-the-Art

Technology-enhanced vocabulary learning either in web or in mobile platform aims to provide the learners with mobility and portability. Numerous cognitive benefits of multimedia annotations stimulated the interest of researchers and motivated them in developing socially beneficial systems that can replace conventional learning/teaching methods. This chapter reports different scientific articles published so far in the area of technology-enhanced vocabulary learning, specifically in the context of informal learning.

The reviewed articles are classified into three categories: firstly, widely adapted vocabulary learning methods; secondly, representation of multimedia annotations in different educational settings; and thirdly, the emerging technologies developed till today. Each category can be divided into multiple subcategories. This chapter first summarizes the research methods that are adapted in the domain of vocabulary learning system development (in Section 2.1). Next, we report on the representation of multimedia annotations in regard to vocabulary learning (in Section 2.2). In Section 2.3, we present our survey on the experimental systems that intend to support foreign vocabulary learning between the years 2005 to 2017. Finally, we summarize this chapter by articulating the differences between our work and the existing studies in Section 2.4.

### 2.1 Technology-driven Research on Learning System Development

The adaptation of technology-enhanced vocabulary learning over conventional learning/teaching methods has increased dramatically in recent years. Many Asian countries have moved forward to engage their students in language learning with the assistance of computer technologies. P. Thornton & C. Houser (Thornton P., 2005), M. Lu (Lu, 2008), N. Cavus & D. Ibrahim (Cavus N., 2009), and A. Hayati et al., (Hayati, 2013) have published their findings on how foreign language learning with the help of mobile devices is gaining popularity over the conventional methods in Asian countries such as Japan, Taiwan, Turkey, and Iran. Many non-Asian countries, especially Europe and the North American regions, have adapted technology-supported learning/teaching since the 1990s. Keeping technology in the central position, several vocabulary learning methods have been proposed by researchers and implemented by computer scientists.

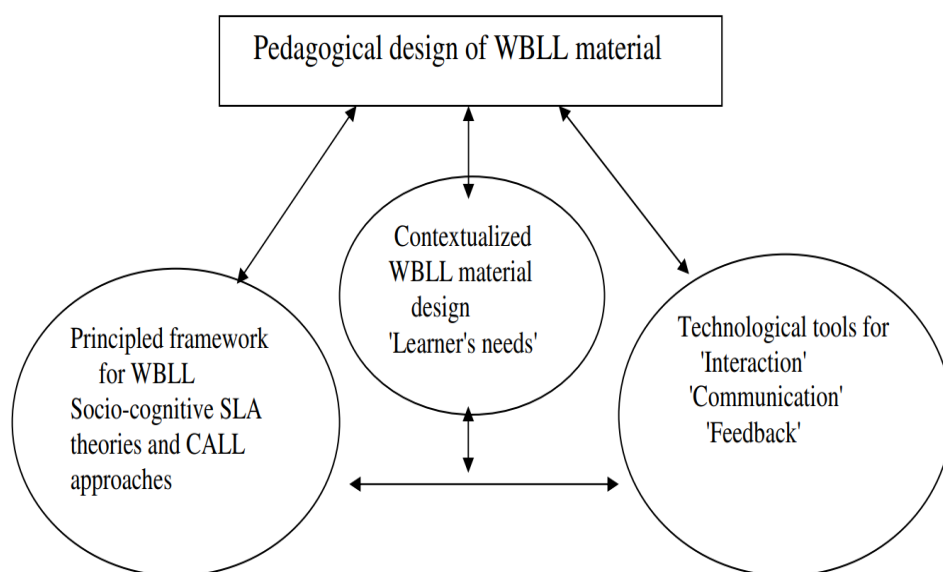
### **2.1.1 CALL-mediated Spaced Retention**

A key advantage of vocabulary memorization using Computer Assisted Language Learning (CALL) programs is that it provides systematic repetition of words, ensuring that the learned words are not forgotten immediately. CALL-mediated spaced retention method (Baddeley, 1997) is often adopted by computer scientists in the implementation of computer programs, because it provides a time interval between subsequent reviews of previously learned words. Hence, a newly learned word stays in the memory longer than usual. Few studies have tested the cognitive effects of this spaced repetition method. T. Nakata (Nakata, 2008) investigated the effectiveness of English vocabulary learning by Japanese high school students using either word lists, word cards (also known as flash cards) or CALL-mediated spaced repetition approach. Another study (Hirschel, 2013) revealed the short- and long-term learning effects between learning vocabulary through vocabulary notebooks and a CALL program with spaced repetition. Both studies conclude that CALL-mediated spaced retention method has a significant impact on the learners' memory compared to the conventional methods.

### **2.1.2 Web-based Learning**

A commonly adopted method is the Web-based Language Learning (WBLL) technique. A WBLL-based presentation method can be described as an audio and an external arrangement of the reading comprehension in addition to a pictorial representation of the comprehension. WBLL technique provides autonomy by creating a learning framework that provides a more productive and richer content, and a more adaptable learning speed (Hamamorad, 2016). This method is noticeably popular among foreign vocabulary learners, and therefore many computer scientists applied it in developing computer programs. Furthermore, this technique is able to provide a real-life context and a favorable background for autonomous learning (Ward, 1998). Web-based learning is also able to enhance the students' meta-knowledge and stimulate their competence for organizing, monitoring and assessing their own learning progress (Hamamorad, 2016). This, in turn, boosts autonomous learning (O'malley, 1990). Figure 2-1 shows a recent 2016 pedagogical framework proposed by (Hamamorad, 2016) indicating the relationship among learner needs, second language acquisition (SLA) and computer-assisted language learning (CALL).





**Figure 2-1** Design of WBL Material (Hamamorad, 2016)

The significance of this method has also been verified. An example is a study conducted by Bahman G (Gorjian, 2012), which used [www.cnn.com](http://www.cnn.com) as the source of expository passages (more precisely, 12 passages for the experimental purpose). Results indicate that WBL-based presentation helped the learners in short-term recall over the conventional method. Hamamorad (Hamamorad, 2016) also remarked that WBL is a lucrative method of CALL that can be embraced in teaching/learning the target language. It provides the students with a customized approach to acquiring a new language which fits their learning styles (some learners may prefer ‘listening’ or ‘reading’, while others may prefer ‘visualized information or visual aids’). Moreover, this strategy provides a consistent access to a variety of different educational contents.

### 2.1.3 Variants

Commonly observed variants of technology-driven research that are often practiced by both linguistics and computer scientists for vocabulary learning curriculums and developing programs are:

- Cognitive web-based vocabulary learning (Kritikou, 2014): Cognitive strategies are repetition and use of mechanical means to study vocabulary by identifying a student’s progress by taking an individualized course and a set of educational activities. The vocabulary level of a student is mainly taken into consideration along with visual and auditory information. From this, the most appropriate learning environment is determined, which facilitates the learning process.

- Intentional (Bereiter, 1989) and incidental (Huckin, 1999) learning: Intentional learning refers to the cognitive learning processes that have a learning goal rather than an incidental outcome (Bereiter, 1989). The interest in intentional learning is increasing dramatically. An intentional learner often engages in learning during their free time without direct supervision of an instructor. On the other hand, for incidental learners, vocabulary is acquired incidentally through engaging in extensive reading (Huckin, 1999).
- Explicit and implicit learning (Choo, 2012): Implicit learning in knowledge psychology is defined as acquiring knowledge by a process that takes place spontaneously without any conscious operation. On the other hand, explicit learning is characterized by a more conscious operation where an individual makes a hypothesis and tests it in search for that particular structure (Choo, 2012) (Ellis, 1994).

Supporting studies have revealed the efficacies of these variants. Merits of cognitive web-based vocabulary learning method for the Greek language as the second language are tested by (Kritikou, 2014). The effectiveness of intentional and explicit/implicit learning in vocabulary acquisition has been discussed by (Barcroft, 2009) and (Choo, 2012), respectively. Further research findings, from the perspective of vocabulary learning approaches among EFL Iranian medical science students, have been reported by (Hashemi, 2015)

Adaptation of the right method is important in the implementation of a computer program. To adapt multiple methods in a single program has also been attempted. Determining a suitable method (or multiple methods) depends on several factors including the targeted users, types of vocabularies to offer, educational settings, the configuration of a computer program, resources to develop a program, and empirical pedagogical study.

## 2.2 Representation of Multimedia Annotations

Many researchers have examined the impact of multimedia annotations on language learning. Foreign vocabulary acquisition with the corresponding image(s) and/or sound is effective to repeat the memorization activities (Schmitt, 1997) (Wright, 2005). Foreign vocabulary associated with the right objects (known as imagery techniques) are memorized more swiftly than texts (Kellogg, 1971). Noticeable pedagogical studies between the years 1960s to 1990s including ‘the pictorial superiority effect’ and ‘the dual coding theory’ have suggested that the participants often have an advantage in recalling memory if images are used to represent foreign words (Paivio A. R., 1968; Paivio A., 1973) (Webber, 1978) (Omaggio, 1979) (Hudson, 1982) (Herron, 1995) (Chun, 1996). Technology boom (often known as the dot-com boom) between 1995 and 2001 encouraged computer scientists to test other annotations, such as text, image, video, animation, sound, in a multimedia-enriched learning system. As a result, several studies have been carried out during the late 1990s and early 2000s for comparing different multimedia annotations in foreign language learning. For example, a 1996 study suggests that vocabulary by text plus picture annotation has a significant impact over words with text-only and text plus video annotations in incidental learning (Chun, 1996). Later, a 2001 study showed that lexical items prepared by text coupled with video clips were more beneficial in teaching unknown lexical items than those prepared using pictures and texts (Al Seghayer, 2001). Subsequently, a 2003 study revealed that annotations using text plus image were most advantageous for vocabulary memorization compared to text annotation only and text plus image plus audio (Yeh, 2003). Finally, a recent 2012 study concluded that recalling newly acquired Swahili words is easier when paired with pictures rather than the translation data (Carpenter, 2012). Many popular software products found in the market use images to represent their contents. One reason for this is that are easier to find or create images compared with other visual aids. Also, images have cognitive benefits such as recalling old memories or events.

## 2.3 Emerging Technologies for Supporting Foreign Vocabulary Learning

A significant interest in developing technology-enhanced vocabulary learning systems for supporting foreign vocabulary learning have been noticed between the years 2005 to 2017. A recent study reviews the impact of popular language learning software (Rosetta Stone, TMM, MEM and ESL WOW) in English (Nicholes, 2016). Although this study focused on the progress of Chinese students' learning in writing, listening, speaking, reading, and grammar; vocabulary can be acquired in each of these areas. Along with popular systems, several experimental systems have been proposed for assisting learners in acquiring vocabulary by self-study. We classify these experimental systems into four types:

- Multimedia annotation-based learning material creation systems (Sec. 2.3.1)
- Augmented reality-based situated learning system (Sec. 2.3.2)
- Game-based learning systems (Sec. 2.3.3)
- SMS-based learning systems (Sec. 2.3.4)

After a brief description of each type, a comparative discussion of different systems will be presented. Furthermore, a comparison of key features to understand the system capability will be discussed. After that, a table is created to describe the results and limitations of these studies. Observing the short- and long-term memory retention is an important area to investigate. Therefore, learning effect investigation is taken into consideration and articulated.

### 2.3.1 Multimedia Annotation-based Learning Material Creation Systems

An intentional learner is highly motivated to learn new things, takes responsibility, and actively engages in techniques that facilitate learning. Given that the vocabulary has to be acquired during free time by self-study, language instructors often spend less time to instruct vocabulary in the classroom. Therefore, over the past decades, the popularity of intentional vocabulary learning systems has increased noticeably. Use of visual aids as multimedia annotations is common. Therefore, generation of video-clip/still-image/animation-based learning material is commonly observed. Several systems provide learning material that are embedded with pronunciation (voice) data, word meaning, and definitions.

In this subsection, we first discuss the roles and limitations of recently developed (between the years 2005 and 2017) experimental systems that use video-clip/image/animation/other-aid

along with pronunciation data and word meaning to generate vocabulary learning material. Finally, a comprehensive list of these systems and their features is shown in Table 2-1.

A 2005 study conducted by S. Joseph et al., (Joseph, 2005) introduced PhotoStudy, a vocabulary learning system that runs on both web and mobile platforms. This study purposed to assist EFL students to enhance their vocabulary skills by memorizing words associated with images. The system generates vocabulary learning material based on the collaborative utilization of images taken by camera-phones. However, this study reported that the collaborative policing was victimized by email spams, which led to failing the links to images?. Figure 2-2 shows the study environment for a learner while using this system. A questionnaire survey to support this study has been conducted, which revealed that the appearance of the word seems to play a significant role in memorizing vocabulary, while the spelling of the word was not constantly chosen as a memorization strategy. Therefore, this study needed further development in their proposed system. It can be reported that numerous incorrectly marked up images were detected in the shared database, which proved troublesome for learners. Also, this study did not assess the learning efficacy of their learning materials.



a) Main menu, b) Content overview, c) Example image, d) Multiple choice, e) Failure, f) Success feedback



a) Content Selection, b) Success feedback, c) Summary

Figure 2-2: Learning Environment (Joseph, 2005)

Another 2005 study conducted by P. Thornton & C. Houser (Thornton P., 2005) reported on three case studies on English education in Japan. As a part of this study, a web-based learning site for English idioms named Vidioms has been developed. In this system, student-recorded animations were used to show each idiom's literal meaning, and a video clip shows the idiomatic meaning. Additionally, the text-based material includes an explanation, a script, and a quiz. The first case study inquired about Japanese students' (N=333) usage of mobile devices. The survey result indicated that 99% of the students send emails on their mobile devices exchanging over 200 emails per week. The second case study examined the students' (N=44) English vocabulary learning ability for identical materials using mobile devices in comparison with paper or the web. The result revealed that the students learned using mobile devices significantly more than the other approaches. The third case study evaluated (N=31) the site Vidioms, which consisted of English-idiom learning material generated by student-recorded animations. The system sent video and web material through cellular phones and PDAs to the learners. Figure 2-3 shows the interface of the Vidioms system. The survey results (21 questions) found that the assessment of its hardware, web pages, videos, and audios was significantly positive. The results also suggested that the learners using the PDAs gave a significantly higher rating to its video quality and learning idioms compared with students using mobile devices. Nevertheless, this study did not investigate the learners' short- and long-term memory retention rates. Also, this study did not clearly indicate the nature of animations and video clips used to represent English idioms in Vidioms system.



Figure 2-3 Interfaces and Study Environment in Vidioms System (Thornton P., 2005)

A system to create English vocabulary learning material based on short video clips was introduced by Hasegawa et al. (Hasegawa K., 2007). This system, Personal Super Imposing (PSI), was able to create sets of 5-second vocabulary learning material, which consisted of the spelling, the meaning, and a short video clip together with the pronunciation data of the word to be learned. In creating the learning material, the spelling and the meaning is embedded in the video clip as subtitles, and the pronunciation data is automatically extracted from a database called the Pronunciation Input (PI) system. The interactive PI system allows a user to record the pronunciation of a word by using a microphone and displays a list of English words with their meanings. The system is able to eliminate computer-generated noise while the user is recording the pronunciation data of a word using a microphone. Besides, the system is also able to create learning material when a video clip is sent directly via email to a specific address. For this, the subject line of the email needs to contain the spelling and the meaning of the word to learn. Both PSI and PI systems have been developed in Microsoft Visual Basic.NET 2003 and Microsoft Visual Basic.NET 2005, respectively along with several pieces of free software. Figure 2-4 displays the backbone of PSI system and an example of generated learning material. A learning effect investigation comparing the PSI method and the conventional pen-and-paper-based learning approach has been conducted. The result revealed no significant learning effect between the two learning approaches. However, the memory retention rate two months after the experiment was significantly higher in the PSI method than the conventional method. This study has few limitations as follows. Firstly, the PSI method was not found to be significantly effective compared with the conventional pen-and-paper based memorization approach. Secondly, the article did not articulate a proper guideline for the video clips used to generate the learning material. Thirdly, no description was given on how computer-generated noises were omitted while recording a pronunciation. Fourthly, gathering (/recording) a video clip for each word can be a troublesome and a time-consuming process. It can be mentioned that determining a suitable scene that represents a word can be debatable, and a question may arise on the acceptability of one's recorded video clip by other users. Fifthly, video clip-based learning material may be suitable for learning verbs but may not be an effective way for learning nouns and adjectives. Finally, the process of creating the learning material in the system may be difficult for users with inadequate computer skills.

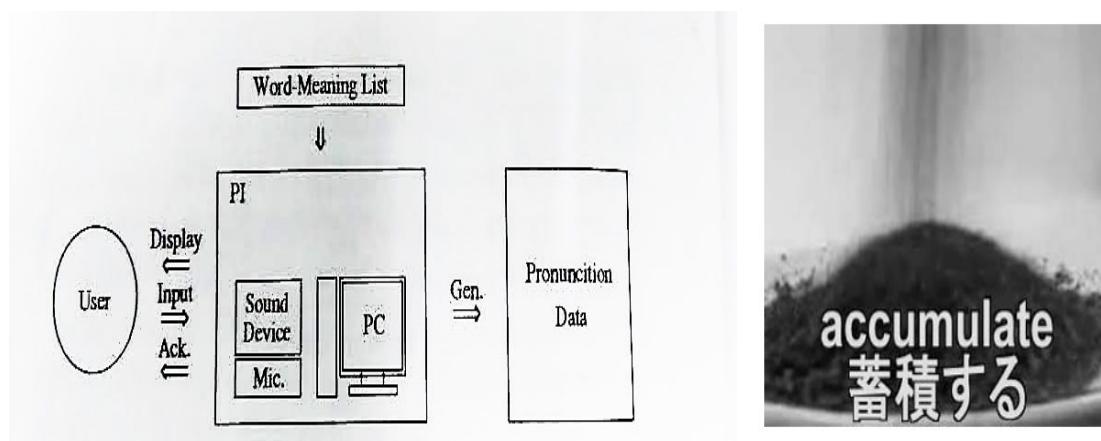


Figure 2-4 The Backbone of PSI System (on left) and a Learning Material (on right)  
(Hasegawa K., 2007)

MultiPod (Hasegawa K., 2007), a vocabulary learning system in the mobile platform has been introduced for the user's mobility and portability. The system could be used on the users' iPods and enabled them to move data from a database to their own devices. The learning material in MultiPod system consists of the pronunciation of the foreign word together with a 5-second movie to represent its meaning. Two subsystems, HodgePodge and PodBase were developed to support MultiPod's operation (Ishikawa, 2007). HodgePodge was designed to automatically add aural information and subtitles to the movies/still images in the creation of learning material. On the other hand, PodBase was designed to manage the learning material produced by HodgePodge. Figure 2-5 shows the system architecture of MultiPod. A comparison of MultiPod-based learning with pen-and-paper-based learning was carried out with ten participants to verify the efficacy of this system. No significant differences were observed in the participants' memory retention during post-test 1 (just after) and post-test 2 (two weeks after). Later on, in post-test 3 (two months after), a significant difference was noted (Amemiya, 2007). Few limitations have been observed in these studies. Firstly, the system creates learning material from a given set of selected words that are chosen by instructors. Therefore, MultiPod does not provide on-demand creation facilities. Finally, this study did not clearly indicate the types of movie clips and/or still images that were used, and how to accumulate them.



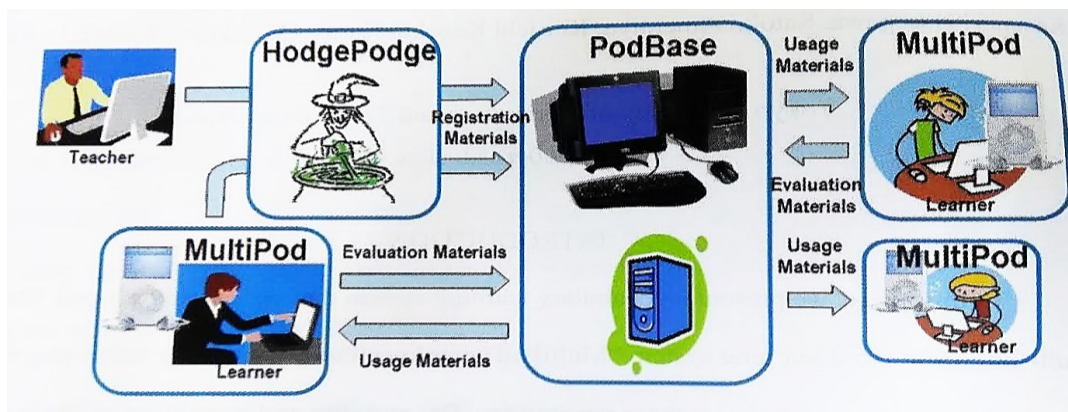


Figure 2-5 The Architecture of the MultiPod System (Hasegawa K., 2007)

In later research, a multi-lingual vocabulary learning system was proposed by K. Kaneko et al., (Kaneko, 2007) that generates learning material based on universal images. To create the learning material, PSI and PI (Amemiya, 2007) (Ishikawa, 2007) systems were used. A system called Personal Handy Instructor (PHI) was developed by which the learners can learn the learning material created by PSI system. Figure 2-6 shows an overview of the PHI system. The format of the learning material in both the studies is the same. Images common to multiple languages are named as universal images and are used to create the learning material. To test the suitability of universal images in foreign vocabulary acquisition, a leaning effect investigation with ten participants was conducted. Twenty unfamiliar words were selected and the participants were asked to acquire them during the study session. The control group participants followed the conventional pen-and-paper-based approach. The results showed that the experimental group participants had better results with respect to memory recall, but the difference was not significant. A key limitation of this study is that although the concept of universal images in foreign vocabulary acquisition has been introduced, it is not clear how to extract them. Also, it was not discussed what kind of parameters an instructor or a learner has to follow in the selection of universal images. Finally, the study did not investigate long-term memory retention of learners.

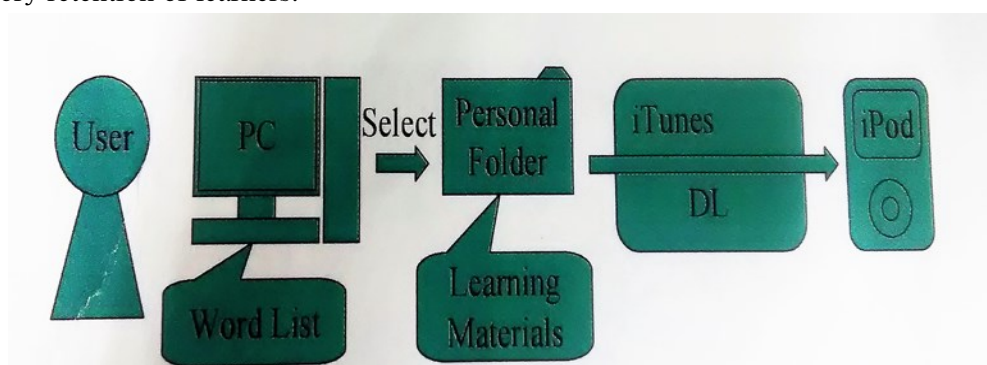
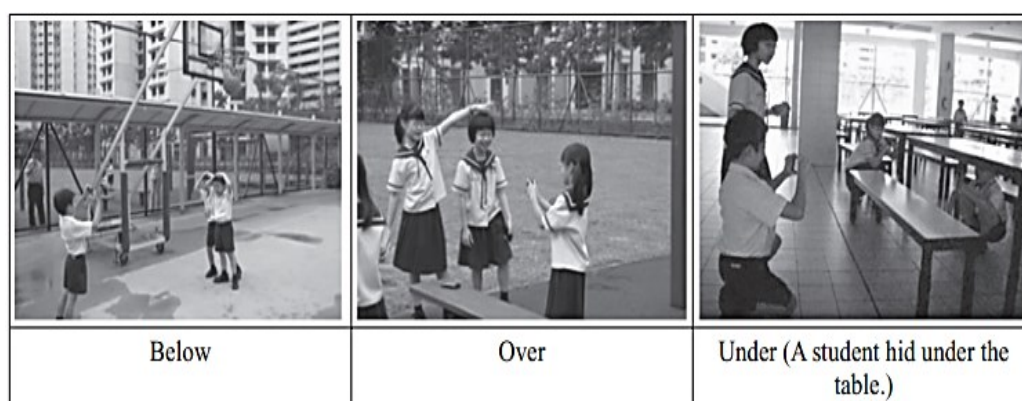


Figure 2-6 Overview of the PHI System (Kaneko, 2007)

A Mobile Assisted Language Learning (MALL) framework for learners to create vocabulary learning content by themselves is implemented by (Wong, 2010). To acquire English prepositions and Chinese idioms, the system assists an elementary school student to create a vocabulary learning content (as shown in Figure 2-7) on a PocketPC by capturing a photo based on his/her real-life contexts. The system also lets the students engage in online discussion using the Sketchy™ tool. This research also contains two case studies. The first case study reports on the outcome of a lesson on six English prepositions, and yields the result that most students felt positive about the lesson, especially with respect to sharing/showing their created content among the classmates. The second case study investigated the students' learning activities for Chinese idiom acquisition. Four activity processes were investigated: i) in-class contextual idiom learning, ii) out-of-class, contextual, individualistic sentence making, iii) out-of-class, online peer learning, and iv) in-class consolidation. The results of this second case study indicated that after a 9-week period the students' contribution was huge and the participation levels were more stable in rendering sentence reappraisal. Moreover, in-class consolidation showed a significant potential and provided encouragement to the students. This study did not investigate the efficacy of student-captured photos in content creation. The learning effect with respect to the learners' memory retention was also not tested.



**Figure 2-7** Learning Materials for Preposition Learning (Wong, 2010)

Y. M. Huang et al. developed a system called UEVL (Ubiquitous English Vocabulary Learning) to assist students with SVL (Systematic Vocabulary Learning) using ubiquitous technologies such as RFID tags, QR codes and GPS (Huang, 2012). An SVL system to learn vocabulary can be implemented by facing, obtaining, comprehending, consolidating and using RFID and GPS technologies to locate a student's geographical location (Hatch, 1995). On detecting the student's location, the system displays guided information to notify them about any learning opportunity, and provides them with learning material. Video clips were used in the UEVL system to create the learning material as follow. First, the student inputs a query sentence (e.g. the subtitle of a video clip). Second, the system breaks the query sentence into

multiple separate words. In the meantime, the student is allowed to choose a query word in order to get a clear picture of its form and meaning. Third, the system represents both examples and definitions of the query word. Simultaneously, the student must match the examples with the definitions in order to improve her/his awareness of the query word. Finally, when the students finish the matching exercise, the system displays the correct answers to the students. Thus, the students can obtain a clear image of the form of the query word and can deepen their understanding of its meaning. The system is able to provide necessary hints and learning material to the students according to their situation. This study also compared performances of active and passive students. The results show that, first, both the system functionalities and the material attributes of the UEVL system emphatically and remarkably influenced on the system, and the active and passive students get more interested in perceived usefulness and perceived ease of use, respectively. This study did not report on the learning effects of the UEVL system for new vocabulary acquisition. Moreover, the system does not support on-demand creation facility. Figure 2-8 and Figure 2-9 show the UEVL architecture and a sample learning material created by the system, respectively.

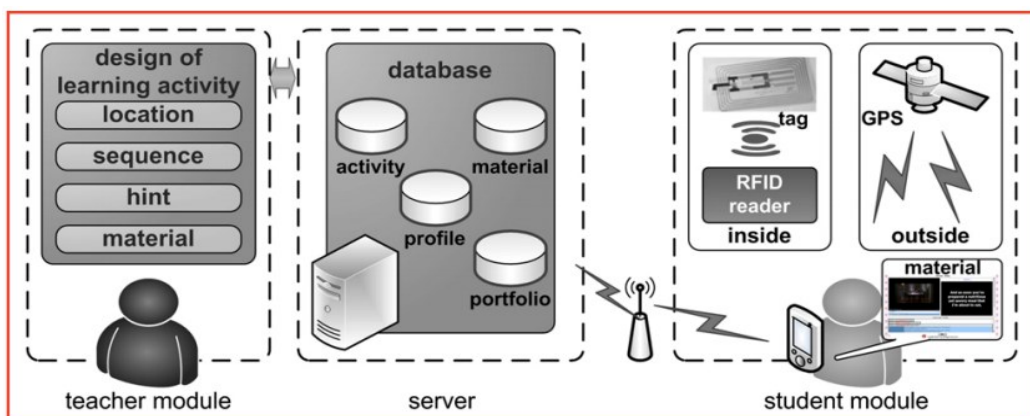


Figure 2-8 The Framework of the UEVL System (Huang, 2012)



Figure 2-9 A Sample Learning Material Created by UEVL system (Huang, 2012)

A MALL environment to build L2 vocabulary with the Microsoft Tag technology has been introduced by R. K. Agca & S. Ozdemir (Agca, 2013). In this system, online learning material and Microsoft Tag technology are used together in the development of a mobile learning environment. This study investigated the effectiveness of the mobile leaning environment compared with learning by printed course books. In the experiment, the participants scanned the Microsoft Tags on pages with mobile devices, and displayed the word definitions and images related to the words in the course books. Only cell-phones were used in the experiment. Forty students in each group participated in the experiment. The results indicated a significant difference between the post-tests in favor of the students who studied with mobile-supported foreign language learning environment. The efficacy of the Microsoft Tag technology was also investigated with a questionnaire survey. Most of the students agreed that Microsoft Tag technology removes the distance between the printed material and the digital learning environment, and also provides a faster access to the online environment. Due to a lack of information on the Microsoft Tag technology, students may face technical difficulties while using this environment. Additionally, this study does not clearly indicate the word definitions and the pictures that have been used in creating the learning material. Figure 2-10 shows the learning environments in this MALL system.



Figure 2-10 Learning Environment (Agca, 2013)

A study has been conducted by M. Kalyuga et al. to investigate the use of online activities in teaching foreign language vocabulary (Kalyuga, 2013). An online activities program was proposed to assist elementary level Russian language learners enrolled at Macquarie University, Australia. This system was based on matching foreign words with familiar words. The activity consisted of presenting L2 (Russian) lexical items with their L1 (English) equivalents together with other multimedia elements such as sound and images. This program allows the learners



to choose between three types of exercises: L2 word to L1 verbal information, L2 word to image, and L2 word to pronunciation. The students are allowed to switch among difficulty levels while acquiring vocabulary, and the program is designed to adjust their needs accordingly. The key limitations of this study are: firstly, the system has not been developed. Secondly, the effectiveness of utilizing images in the L2 vocabulary learning has not been examined.

To memorize words of an unfamiliar language, O. Anonthanasap et al. proposed an educational system that is equipped with dynamic and interactive interfaces (Anonthanasap, 2014). This system used a mnemonic-based interactive learning approach (Atkinson, 1975) (Paivio A. , 1969) for memorizing L2 words. The system allows the users to seamlessly browse a collection of foreign words while suggesting phonetically-related words of a known language for helping to memorize words in an unfamiliar language. To represent learning material for mnemonic words, 400 most frequently used nouns generated by the Google custom search API were used. D-Flip (Vi, 2013) was used to facilitate the system's interactivity. The system was based on the Soundex (USA Patent No. Soundex,US Patent 1, 1918) phonetic algorithm, and to measure the difference between two strings, the Levenshtein distance (Levenshtein, 1966) was used. Figure 2-11 shows an example of learning material generated by this system. To evaluate the system, this study compared traditional pen-and-paper, static visualization and mnemonic-based vocabulary (using the proposed system) learning approaches. However, the result (derived from a NASA-TLX test) could not find any significant difference between the proposed method and the other two approaches for acquiring mnemonic words. To improve the system's performance in suggesting mnemonics words to the learner, other phonetics algorithms need to be tested and compared with the Soundex algorithm.

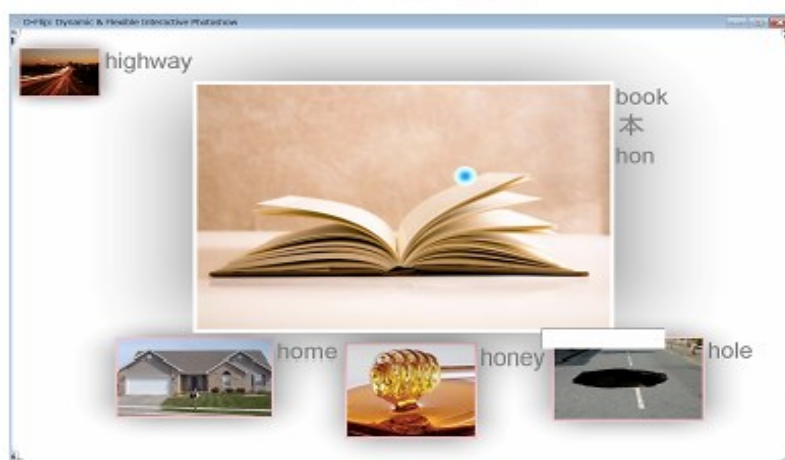


Figure 2-11 A Learning Material (Anonthanasap, 2014)

Word Learning-CET6 (Wu, 2015), a B4A smartphone application consisting of 1274 English words was developed to help EFL students in intentional vocabulary acquisition (Nation, 2001). In the development of Word Learning-CET6 app, 1274 words were programmed into a database with three features: the spelling, the pronunciation, and the Chinese definition. The beta version of this app executes three basic functions: i) displaying the words alphabetically, ii) selecting or deselecting words from the database and building a new word pool to increase the efficiency of learning, and iii) performing the sample test to evaluate a learner's learning efficiency.

Figure 2-12 displays the user interface of this system. The efficacy of utilizing smart phones as a tool for teaching and learning English vocabulary in a natural environment has been assessed by vocabulary knowledge tests. The post-test result found a significant difference between the participants in the experimental group (learned with Word Learning-CET6) and control group (learned without the app). One of the limitations of this system is that it lacks any function to communicate between users. Additionally, new functions to correct errors or to incorporate new contents need to be implemented. The app cannot be used in the classroom environment unless these modules are provided.



Figure 2-12 User Interfaces of Word Learning-CET6 App (Wu, 2015)

SCROLL (System for Capturing and Reminding Of Learning Logs) is the next generation e-learning environment designed to support informal learning of foreign vocabulary with the help of multimedia annotations (Ogata H. L.-B., 2011) (Uosaki N. O., 2012) (Uosaki N. O., 2017). The system primarily supports the learners in acquiring knowledge and skills by accumulating their daily experiences as learning logs. The system records the learners' personal and geographical data, contents used (image, translation data, context etc.) in learning material creation etc. as learning logs. Later, the system analyzes these learning logs and recommends learning content to the learners. The system incorporated a module implemented for assisting learners in foreign vocabulary learning with the help of an image, translation data, and voice data. It has been observed that SCROLL users have been reluctant to collect images in the creation of learning material. Therefore, most of the learning material in SCROLL systems has been created without using any image, which makes the system less interactive to the users. Figure 2-13 shows an example of SCROLL-created learning material.

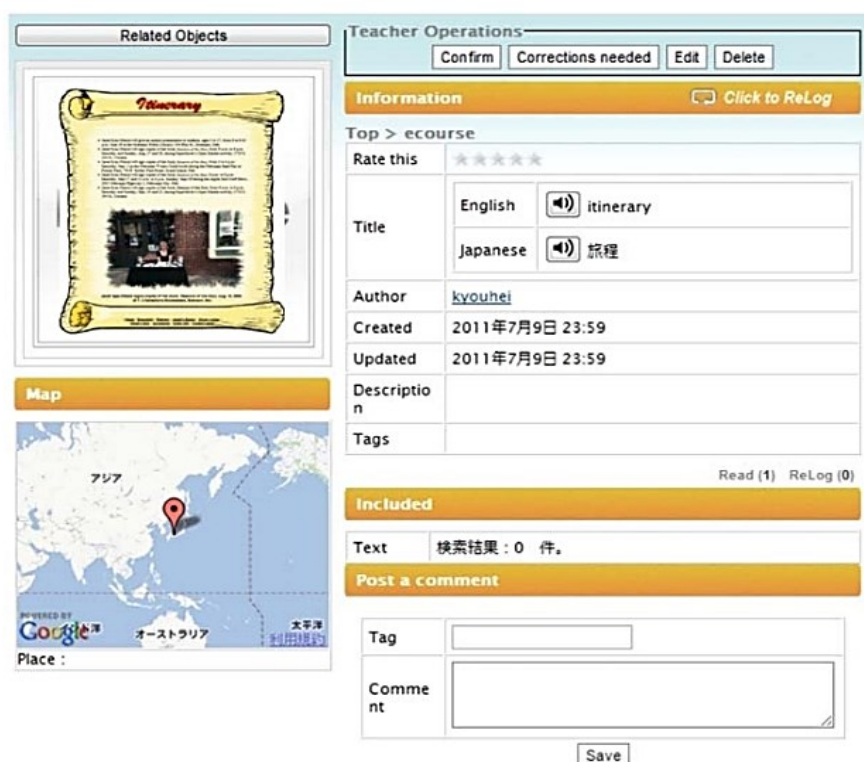


Figure 2-13 SCROLL-created Learning Material (Ogata H. L.-B., 2011)

In Table 2-1, we summarize the features of the above-mentioned experimental systems.









### 2.3.2 Augmented Reality-based Situated Learning Systems

Vocabulary can be acquired through situated learning settings. In this approach, instead of displaying names and directions, the system displays words and animations to teach new vocabulary that is relevant to the objects found in the environment. A 2016 study by M. E. C. Santos et al. showed the efficacy of Augmented Reality (AR) for situated vocabulary learning, and proposed a system based on AR technologies for educational settings (Santos, 2016). A handheld AR system was developed to display any combination of multimedia, including image, sound, animation, and text in a real environment. The main part of the system contained a controller, which had access to learning contents, sensors, and user inputs. The controller received the marker ID and camera view matrix from the tracker, and used this information to specify the behavior of on-screen display. ARToolKit and OpenGL ES 2.04 technologies were used to develop the tracker and the renderer, respectively. The system runs only on iPad tablets. Situated vocabulary learning contents created by the system for nouns (displaying using labels) and verbs (displaying using animations on real objects) are shown below in Figure 2-14.

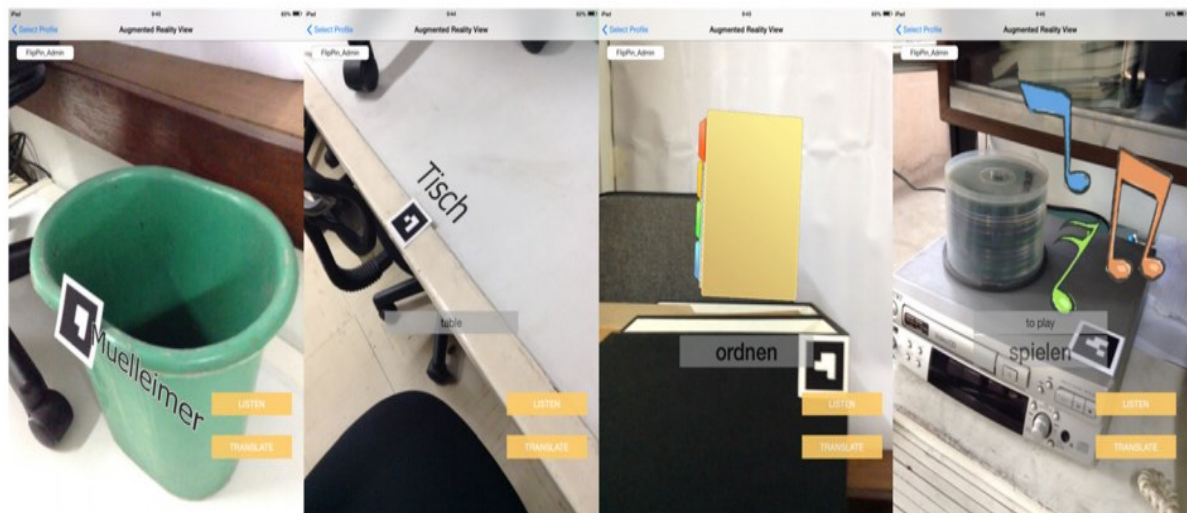


Figure 2-14 Situated Vocabulary Learning Contents Using AR Technology (Santos, 2016)

### 2.3.3 Game-based Learning Systems

In the early 2010s, research has focused on game-based vocabulary acquisition. Two key challenges for a game-based vocabulary learning system are (1) selecting an adequate instructional design model and (2) formative evaluation of a developed system (McGriff, 2000) (Kruse, 2005). A gaming environment often features storyboards, web pages, vocabulary games, publishing the game websites, formative evaluation, revisions, and data analysis. It also generates a report, which often engages learners for a longer time. A 2012 study by M.S. Sahrir & N. A. Alias implemented a game-based vocabulary learning platform to provide a new learning experience for Arabic learners at IIU, Malaysia (Sahrir, 2012). The study adapted the ADDIE model, a popular instrumental design model of an online game, in developing this system. Formative evaluation defined by M. Tessmer

(Tessmer, 1993) was carried out via the participation of subject matter experts, instructional design experts, and teachers. A game-based vocabulary learning system can be effective if it is developed with an adequate instrumentational design model, and a formative evaluation is ensured. However, the results of the formative evaluation were not discussed in the paper. Few other technical limitations have been observed, such as Arabic writing systems, displaying the overall scores for all the players during a competition, and the vowel sounds while pronouncing Arabic words. Figure 2-15 displays the learning environment.



Figure 2-15 Snapshot of the Learning Environment (Sahrir, 2012)

### 2.3.4 SMS-based Learning Systems

Vocabulary learning can be done through receiving short-message services often abbreviated as SMS services. Due to the convenience of mobile devices, adult learners have often engaged themselves in this type of study. SMS-based learning has led to enhanced learner motivation, learner curiosity, learner autonomy, learner's self-efficacy, learner's technological self-confidence when learning language, vocabulary, and concepts (Katz, 2013). Vocabulary learning via SMS services is often considered to be flexible, user-friendly, controlled and adaptive. ActiveX control package (Logiccode GSM SMS ActiveX DLL technology) was a popular technology during the 2000s. However, WiFi and Bluetooth technologies, developed in the 2010s, are often used now. Two recent studies (2009 and 2013) have implemented SMS-based vocabulary learning. One of them is an m-learning system developed to learn technical English words with the help of SMS text messaging (Cavus N., 2009). This system provides a single graphical user interface-based display to memorize technical English words. The other approach is an SMS-based material-supporting system for memorizing English idioms for native Persian speakers (Hayati, 2013). This system offers the

contextual vocabulary memorization. The system (Figure 2-16) offers the contextual vocabulary memorization. Both the systems are controlled by instructors.

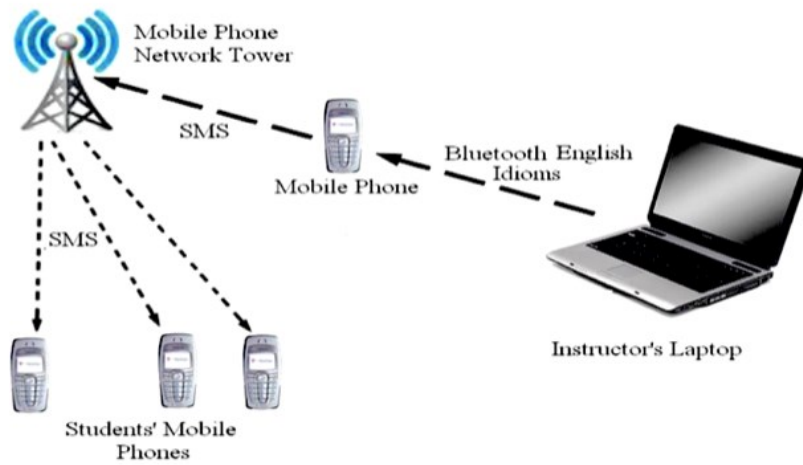


Figure 2-16 The Architecture of the SMS-based Vocabulary Learning System (Hayati, 2013)

## 2.4 Summary

This chapter is based on the scientific articles that have been reviewed to support the current study. For each of the articles, it mainly highlighted the research purpose, methodology, key findings, experimental procedures, and the technical specifications. Additionally, key limitations were articulated as a part of the review.

The review provided a rich picture of various researchers' perceptions of vocabulary learning either by self-study or in the classroom environment. We examined various aspects including the purpose, widely adapted vocabulary learning methods, multimedia annotations representations, experimental procedures, survey, feedback, data analysis, data collection from human participants and pedagogical viewpoints with examples.

The adaptation of an adequate method in designing a TEVL system depends on the educational setting, the learners demand, the instructors' way of teaching, and related observations. This may be a challenge for early-stage researchers because each method has advantages and pitfalls.

Representing a word with a single annotation or with a combination of two or more annotations is possible. In a single annotation, the common format is text-only or definition-only. On the other hand, text+image, text+video, text+animation, text+image+audio, and text+video+sound are identified as popular representations of a word with multiple annotations. Among them, text+image can be pointed out as the most effective form of multimedia representation.

The focus of contemporary developments on the web or mobile platforms have been on the intentional vocabulary learners, often called as motivated learners or intentional learners, who often engage in vocabulary learning during their free time without relying on a curriculum. TEVL systems mediated with cloud services are getting popular for several reasons: firstly, the convenience of smart applications; secondly, learners are interested in learning multiple foreign languages simultaneously; thirdly, interactive and collaborative learning experiences; fourthly, low cost (the expense of learning with a native instructor is increasing); fifthly, learners can create their own learning contents based on the situation or self-captured photo; and finally, learners can learn during their free time.

Several strategies have been adopted to motivate non-motivated learners. Video game-based learning is often used to increase motivation of vocabulary learners. As the meanings of many words are created and defined by a series of experiences, several cognitive approaches, behaviorist methods, and SMS-based learning approaches have been widely adopted to motivate foreign vocabulary learners.

Commonly observed limitations in the reviewed articles can be summarized as,

- Studies suggested that images play a significant role in vocabulary memorization. However, the nature of the images, how to determine an educational image, and how to collect them have not been discussed clearly. One possible reason for this might be that teachers, linguists, and computer scientists often have different thoughts on the suitability of images

for representing words. Many of the existing systems consider instructor-suggested images for their learning material.

- Although many authors explicitly stated the importance of recalling newly learned words, only a few articles investigated the short-term/long-term/extended-long-term memory retention rates.
- On-demand creation of learning material has not been addressed properly. These days, third-party APIs (search, text-to-speech, translation etc.) are very powerful, and therefore incorporating on-demand creation facility can be very useful.
- Constraints regarding the data collection from human participants have not been discussed adequately.
- Not many researchers have focused on the appropriate length of the learning material.

To provide a solution to these problems, our approach was first to introduce the concept of an appropriate image. We focused on nouns, and proposed a definition of an appropriate image for representing a concrete noun. Then we proposed a definition of an appropriate image for representing an abstract noun. After that, we considered a category-based appropriate image based on image features to represent an abstract noun. We propose an algorithm to evaluate still images and extract only one appropriate image for representing a concrete noun. The algorithm is modified so that it can recommend an appropriate image based on a category. We implemented an image recommendation system based on these algorithms.

Our next approach was to propose a web-based TEVL system that aids the learners in creating on-demand learning material. Here, our main focus was on automatic extraction of learning resources (image, text, pronunciation, and meaning) while creating the learning material. We assumed that this will help the learners by saving a considerable amount of time.

A key emphasis of this study is to investigate the short-term/mid-term/ long-term memory retention rates of newly learned vocabulary. We followed a scientific approach in designing experiments, deciding participant distribution, data analysis etc., and collected learning data from global participants. We also report on the limitations of our experiments.

### 3. Appropriate Images for Nouns

This chapter describes the theoretical background and pedagogical approaches that we have followed in determining appropriate images for representing nouns (concrete and abstract only). To begin with, we discuss our motivation in Section 3.1. Then in Section 3.2, we discuss the approach that we followed to define an appropriate image to represent a concrete noun. Next, Section 3.3 discusses our approaches for abstract nouns. Finally, we summarize the chapter in Section 3.4.

#### 3.1 Motivation

Noun imageability has been an area of investigation among linguists and computer scientists. Imageability (synonymously picturability or imagery) is defined as the ease with which a word gives rise to a mental image (Paivio A. Y., 1968). Imageability may influence the processing of words in the mental lexicon (Marianne Lind, 2012). Imageability of nouns is an area of interest among experimental psychologist because of the indispensable role that imagery may play in recalling old memories (Kintsch, 1972). Generally, nouns are considered as more imageable than verbs. It is because nouns denote entities, while verbs denote relations between entities. Entities, comparing with relations are inherently more imageable, which may explain the general and cross-linguistic difference in imageability found between the word classes (Marianne Lind, 2012). However, this is often shown to be incorrect while dealing with abstract nouns because of their conceptual imageability and polysemic behavior. In addition, not every abstract word of a language is a bona fide lexical item, but some words are often decomposed into more basic terms. For instance, the word ‘wisdom’ may not be a regular lexical entry but must be decomposed further and its meaning inferred from that of ‘wise’ (Kintsch, 1972). As a result, learners may find difficulty in forming an appropriate image for the word wisdom because it is a transformation of the word wise. Imageability is highly correlated with concreteness. Nonetheless, a word like ‘emotion’ is very imageable but low on concreteness, while some nouns such as ‘armadillo’, perhaps because they refer to objects rarely experienced, are concrete but very hard to imagine (Paivio A. Y., 1968) (Bird, 2001). Summarizing the linguists’ research, finding appropriate images for representing abstract nouns is an exceedingly challenging area that needs to be investigated.

Although a simple query in a standard image search engine will result in thousands of corresponding images to represent a single noun, major limitations have been observed in returning images that can be used for vocabulary learning purposes. The reasons are: first, standard image search engines mainly rely on the keywords around the images and file names, because of which image search outputs produce many irrelevant images in the search result (Ben-Haim, 2006). The second reason, as pointed out by Ben-Haim et al., is that currently web-based image search engines are unable to detect actual objects, and hence search engines are blind to the right contents of images. The third reason, quoting from Jain and Varma’s article, is, “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” (Jain, 2011). Standard image search engine suggested images may



not necessarily be suitable for educational purposes because the search output primarily relies on static and textual features of the images. For all these reasons, finding appropriate images from standard image search engines has been problematic. Therefore, determining the most appropriate image to represent a noun is quite a challenging task, and most of the existing vocabulary learning systems where images have been used to prepare learning contents use instructor-suggested images. We identified this as a limitation in image-based vocabulary learning research and contributed to overcoming it.

To provide a solution to this problem, our research trial was to define the concept of an appropriate image for vocabulary learning, and find an approach to extract such images with the computer technologies. Our trial was to build an appropriate image recommendation system that can extract an appropriate image to represent a noun.

In this chapter, we discuss the theoretical background of the approaches that we have followed to determine appropriate images for representing nouns. We first discuss about concrete nouns followed by abstract nouns. The term appropriate image is the key concept in this chapter.

## 3.2 An Appropriate Image for Representing a Concrete Noun

Concrete nouns are easy to visualize in the human brain, because the term concrete refers to a tangible item. Although concrete nouns easily evoke mental imagery, there are hardly any studies on the characteristics of still images that can be used as educational resources for vocabulary learning.

Two pedagogical investigations examined the characteristics of images retrieved by image search engines. The Pedagogical Investigation I (Sec. 3.2.1) compared the memory retention rates between an image consisting of a solo object without any background objects (that is, a mono-color white background) with an image consisting of miscellaneous objects in the image-frame. The Pedagogical Investigation II (Sec. 3.2.2) looked into the representation of the same object in different forms to ascertain what kind of object representation is preferred.

### 3.2.1 Pedagogical Investigation I

The purpose of this investigation was to examine whether or not different types of images play a major role when a learner is acquiring a completely new language. Two types of commonly observed images to represent a concrete noun were selected: 1) images with a solo object on a white background without having any other objects, and 2) images containing various other objects as background or foreground in the image frame.

Bengali was chosen to be the new language to be learned. Ten Bengali words were randomly selected in the creation of learning material: a list of these words and their English translations are shown Table 3-11.

Then two corresponding images to represent each word were chosen from Google image search engine. One image was of a solo object on a white background. The other image contained multiple objects in the background or the foreground of the image frame. These two images were chosen from the top-ranked images retrieved by the Google image search engine corresponding to the given word.

For each of the ten words, two sets of learning material were created corresponding to each type of image, resulting in twenty sets of learning material. Text-data and the translation data were also used in the creation of these sets of learning material.

Table 3-1 Word List Used in Pedagogical Investigation I

Bengali word (in Katakana)	Corresponding English Translation
বিড়াল (ビラル)	Cat
কুকুর (ククル)	Dog
গোরু (ゴル)	Cow
জুতা (ジユ)	Shoe
ফুল (フル)	Flower
বই(বই)	Book
হরিণ (ホリン)	Deer
পাখি (পাখী)	Bird
মুরগি (মুরগি)	Hen
আঙ্গুর (আঙ্গুর)	Grape

Twenty participants of different nationalities enrolled in the undergraduate and graduate programs participated in this experiment. The distribution of the participant nationalities is shown in Table 3-2.

Table 3-2 Distribution of the Participants

Group	Nationality(Number of Participants)
A	Japan(7), Vietnam(1), Iran(1) and India(1)
B	Japan(7), Vietnam(2) and China(1)

A pre-test was conducted to assess the participants' knowledge of Bengali. The pre-test scores suggested that none of the participants have any skills in the Bengali language. Therefore, the pre-test scores were assigned to zero.

The twenty participants were randomly divided into two groups based on their pre-test scores. Group A participants were supported with the learning material created with the help of single-object images. On the other hand, Group B participants were supported with the learning material created with the help of multiple-object images. Figure 3-1 shows an example of the learning material used in each of the two groups for a word. In this example, the Group A and the Group B participants used the learning material displayed on the left and on the right, respectively. To measure the memory retention rates these images may play over time, we did not include pronunciation data in the learning material.

A 10-minute study session was assigned to the participants. The study session was equipped with a paper-based representation of the learning material, a pen/pencil, and an eraser. Learners were allowed to take notes during the study session. However, those notes were not allowed to be used in post-tests.



ククル  
犬(Dog)



ククル  
犬(Dog)

Figure 3-1 Sample Learning Materials

The post-test 1 was carried out immediately after the 10-minute study session to assess their short-term learning effect. The post-test 2 was conducted a week after the post-test 1 to observe if the participants can recall the newly acquired words. Both post-test 1 and post-test 2 questionnaires were multiple-choice questions with four options in each question. Figure 3-2 shows the format of the evaluation questionnaire used in post-test 1. A similar structure was used in designing the post-test 2 questionnaires.

Post-test: ONE

Date: \_\_\_\_\_ Name: \_\_\_\_\_ Group: A / B

(Please choose the most appropriate answer from the choices given below and write in the )

Question 1: ボイ

1. 本(Book)

2. 女の人  
(woman)

3. 手(Hand)

4. 鳥(Bird)

Ans:

Question 2: ビラル

1. お茶(Tea)

2. 猫(Cat)

3. 花(Flower)

4. さくら  
(Cherry)

Ans:

Question 3: アングル

1. みかん  
(Orange)

2. 鹿(Deer)

3. 葡萄(Grape)

4. みどり(Green)

Ans:

Figure 3-2 Evaluation Questionnaire

Table 3-3 shows the results of post-test 1 and post-test 2. The average scores for both Group A and Group B were 10.00 in post-test 1. In post-test 2, the average scores of Group A and Group B were 10 and 9.5 respectively. However, the statistical analysis using Mann-Whitney's U-test revealed no significant difference between the two groups ( $U = 35, p = 0.07$ ). Therefore, we conclude that

background objects in the image frame of still images do not have any significant role in memorizing vocabulary of a new language.

**Table 3-3** Result of the Pedagogical Investigation I

	Average of Posttest 1	Average of Posttest 2
A	10	10
B	10	9.5
		U = 35, p = 0.07

### 3.2.2 Pedagogical Investigation II

If we observe closely, search outputs in an image search engine result vary widely with regard to the object representation in the image frame. Nobody can precisely determine what kind of object representation will be appropriate for representing a concrete noun. We were unable to find any recognized study on this topic. Hence, the Pedagogical Investigation II considered the object representation in an image frame. Our objective was to understand the learners' preferences of object representation in an image frame while acquiring new words with the help of these images.

At first, a survey questionnaire containing 10 different representation (image) patterns of the same object was prepared. We have chosen the image of an *apple* for this purpose. In the questionnaire, the proportion and the position of the object in the image frame was changed, but the shape of the object was not changed.

Twenty-six participants took part in this survey, who were all students in foreign language learning. The participants were Japanese or foreign students enrolled in full-time or in the exchange programs at Tokyo University of Agriculture and Technology. Table 3-4 shows the nationalities of the participants.

**Table 3-4** Participants Detail

Nationality(Number of Participants)
Japan(18), Vietnam(4), Thailand (2), South Korea(1) and Bangladesh (1)

In the survey, we asked the participants to rank each of the 10 representations on a scale of 1 to 10. The written instruction was,

*“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)”*

We analyzed the data with ANOVA. The statistical data compared by ANOVA ( $F_{9,250} = 170.1$ ,  $p < 0.01$ ) and the past Steel-Dwass multiple comparison tests indicated that the image that was represented with *the highest proportion of the actual object highlighted in the center position of the*

*image frame* outperformed significantly over other representations. The participants' second preferred representation was a *small proportion of the actual object positioned into the center of the image frame*.

### 3.2.3 The Proposed Definition of an Appropriate Image

Based on our pedagogical investigations, we propose the following definition of an appropriate image for representing a concrete noun:

*“An appropriate image for representing a concrete noun is the one having the actual object(s) located in the middle-ground, with the highest proportion of the actual object(s) in the image frame”.*

We presumed that an appropriate image will help the learners in quick acquisition and memory recall compared to random, inappropriate images. Hence, our trial was to build a system to extract the most appropriate image for visualizing a concrete noun. Moreover, we aimed to recommend other appropriate images (e.g. the 2<sup>nd</sup> most appropriate image, the 3<sup>rd</sup> most appropriate image, and so on) to a learner. In this way, a learner can choose his/her preferred images if the system-recommended image is not satisfactory.

### 3.3 Appropriate Images for Representing Abstract Nouns

The term *abstract* refers to intangible things. Linguists have defined the term in many ways. The Cambridge dictionary defines an abstract noun as *a noun that refers to a thing that does not exist as a material object*. A simple google query will result in the definition *a noun denoting an idea, quality, or state rather than a concrete object*. In essence, an abstract is a noun that cannot be seen, smelt, tasted, heard, or touched; rather it represents a quality, a concept, an idea, or maybe even an event. Therefore, the ambiguities related to the mental imagery of an abstract noun are hidden into its definition.

Despite numerous approaches to distinguish between abstract and concrete nouns; linguists have so far failed to come to a unanimous understanding of these categories (Khokhlova, 2014). Some words (such as pollution, air etc.) are too abstract and remain unclear to the human mind. Accordingly, finding an appropriate (educational or suitable) image for representing an abstract noun is extremely difficult, and is also a matter of debate.

In our research, we first propose a definition of an abstract noun. It is necessary to have a working definition in order to build a system for extracting the appropriate image(s) for representing an abstract noun.

#### 3.3.1 Defining an Abstract Noun

In defining an abstract noun, we considered a study conducted by Allen P et al. in 1968 as a reference (Paivio A. Y., 1968), where Abstractness-Concreteness (C) dimension was articulated. In this study, 925 nouns were scored on a 7-point Likert scale for Abstractness-Concreteness (C), Imagery (I), Meaningfulness (m), and Frequency (F) values. These four measures were articulated as follows:

Abstractness-Concreteness (C): C was defined in terms of directness of reference to sense experience;

Imagery (I): I was defined in terms of a word's capacity to arouse nonverbal images;

Meaningfulness (m): m represented the mean number of written associations in 30 seconds;

Frequency (F): F represented three broad categories (high, medium and low) based on the number of occurrences per million words.

In this study, nouns with Abstractness-Concreteness (C)'s mean (M) value less than or equal to 4 have been recognized as abstract nouns.

$$\text{Abstract Nouns} \approx C (M \leq 4)$$

Table 3-5 shows some examples of nouns that have (✓) and have not (✗) been categorized as abstract nouns. More specifically, nouns in ✓ meet the definition of an abstract noun and are taken into consideration in the extraction of appropriate images. On the other hand, nouns like *pollution* in ✗ are not considered abstract.

**Table 3-5** Examples of Abstract Nouns

Noun (Paivio A. Y., 1968)	✓	Advice Dream Love	C (M=2.08) C (M=3.03) C (M=1.80)
	✗	Offshoot Pollution Lord	C (M=4.20) C (M=4.14) C (M=4.18)

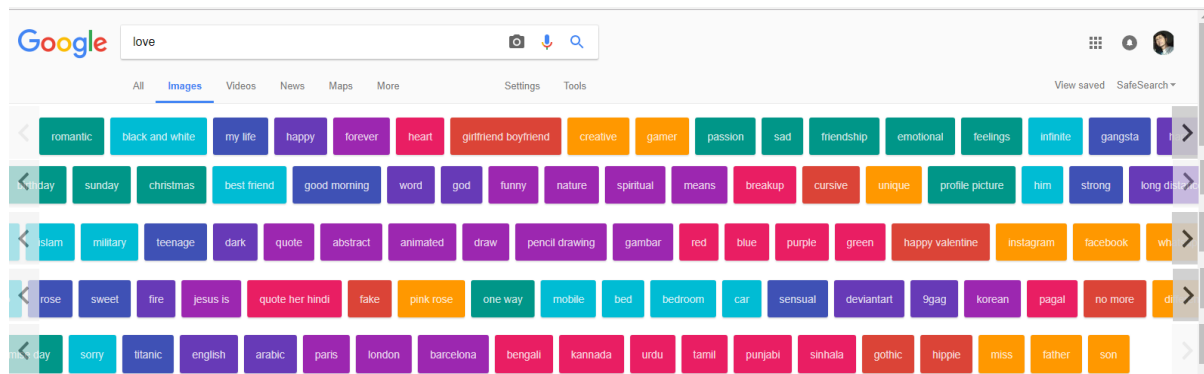
A total of 321 nouns meet our proposed definition [Abstract Nouns  $\approx$  C (M $\leq$ 4)] and recognized as abstract nouns. A list of these abstract nouns is given in Appendix A, alphabetically.



### 3.3.2 Approach 1

#### 3.3.2.1 Fixing a Subset of Abstract Nouns

Image recognition and classification is a complex task. The task gets more complex for a system when it is required to extract only appropriate images. For us humans, we learn the task of recognition from the moment we are born and we keep practicing it naturally and effortlessly as adults (Deshpande, 2016). Unlike human brain, computers read only numbers. Having patterns in numbers is important when an algorithm is trying to distinguish between images. Therefore, the generalization of entire abstract nouns is very difficult due to a wide variation in the corresponding images for each noun. It is also very problematic for an algorithm to establish common patterns between different images. Moreover, due to the heavy cultural influences, an individual’s view on appropriate educational images may vary. Polysemous meaning of words further limit the usefulness of images provided the image search engines in response to abstract noun queries. For instance, a query with the Google image search with the keyword ‘love’ results in thousands of images belonging to many different situations as shown in Figure 3-3.



**Figure 3-3** An Example of the Variations in a Search Query for An Abstract Noun

Therefore, it is essential to limit a subset of abstract nouns for our study. Our subset is limited to three types of abstract nouns as shown in Table 3-6. In approach 1, we consider to investigate about this subset.

**Table 3-6** The Subset of the Abstract Nouns

Type	Description
1	Basic abstract nouns that represent social contexts between humans
2	Abstract nouns that are related to feeling and/or emotion
3	Abstract nouns that represent human’s social and religious beliefs.

Other types of abstract nouns were not considered as the part of the approach 1. Table 3-7 shows some examples of the three types of abstract nouns that belong to the subset, and also some examples of abstract nouns that do not belong to the subset.

**Table 3-7** Examples of Targeted and Non-targeted Abstract Nouns

Type	Examples
1	Advice, Opinion, Joke, Idea, Recognition, Strength, Answer, and Knowledge etc.
2	Love, Hate, Victory, Freedom, Greed, Anger, Ego, Expression, and Betrayal etc.
3	Blessing, Dream, Heaven, Hell, Death, Ghost, and Devil etc.
X	Fly, Hot, Pleasure, Amount, Attitude, Citation, Code, Deduction, Poetry, Trouble, Illusion, Impact, Impulse, Array, Fact, Fault, Deed, Science, History, Profession, Capacity, and Equity etc.

As stated before, image search outputs given by the standard image search engines for text-based query corresponding to abstract nouns have been well observed. While analyzing the nature of still images, a variety of images have been observed. We assumed that in the field of TEVL, wide variations in the images may create confusion and distraction in the mind of the learner. Our observations on the nature of images outputted by the standard image search engines are that for abstract nouns, three types of inappropriate images are often found in the topmost position. We have identified those types of images as ambiguous and inappropriate learning resources for the purpose of image-based vocabulary learning. These three **Inappropriate Image Types** (hereafter addresses as InIT) are, as follows:

**Table 3-8** Types of Inappropriate Images

InIT	Characteristics of the Images
1	A still image that contains only textual data (texts/characters) in the image frame
2	Still images that are visually too abstract are may be inappropriate learning resources
3	Still image that contains multiple objects including the existence of human, object, animation, and textual data in the image frame.

InIT1 images often distract non-native English speakers and learners with poor English ability. In addition, many images contain irrelevant characters in the image frame, which makes the learning process slower and unattractive. Hence, we considered such images as inappropriate for learning resources. Figure 3-4 displays three examples (from left to right) for the words advice, idea, and ego. These images were ranked as the top images by Google image search engine on 2017-07-17.



**Figure 3-4** Examples of Inappropriate Image Type 1

InIT2 images are often observed in searching outputs. The term ‘visually too abstract’ refers to those

images with neither textual data nor concrete expression in their frames, as shown in Figure 3-5. Such images often creates confusion in the mind of the learner. Hence, InIT2 images may not be appropriate as learning resources. Figure 3-5 shows three visually abstract images for representing three abstract words, heaven, hell, and joke, which were used for image searching. These images were listed as the top images by the Google search engine on 2017-07-17.



**Figure 3-5** Examples of Inappropriate Image Type 2

InIT3 images for representing an abstract noun can be a reason of confusion and distraction in the mind of the learner because of their ambiguity. Some examples of such inappropriate images for representing hate, blessing, and knowledge are shown in Figure 3-6. All three images were included in the top ten images by the Google image search engine on 2017-07-17.



**Figure 3-6** Examples of Inappropriate Image Type 3

We recognize InIT1, InIT2 and InIT3 images as inappropriate, and therefore eliminated them from the recommended images. We assume that these images are inappropriate learning resources for memorizing foreign words.

To provide a solution to this problem, we propose a definition of an appropriate image for visualizing abstract nouns that belong to subset 1.

### *3.3.2.2 Definition of an Appropriate Image*

The definition of an appropriate image proposed for the abstract noun subset in approach 1 is,

*“An image having the physical or concrete existence possibly positioned in the central position in the image frame” .*

We propose this definition assuming that images containing human or object(s) may be more acceptable compared to those containing only textual data (texts/ characters) in the image frame, or that are visually too abstract. We also considered the results of the Pedagogical Investigation II in proposing this definition. Recall that in this investigation, we found that positioning the actual object in the central position in the image frame is the best way of representing objects. Moreover, in proposing this definition, we considered that it will be easier for an algorithm to find a pattern in images when extracting only one appropriate image for representing an abstract noun.

### 3.3.3 Approach 2

The approach 2 was proposed for frequently used abstract nouns. Here, without extracting a unique appropriate image for representing an abstract noun, we decided to recommend images in categories based on image features. With this goal in mind, we decided to treat frequently used abstract nouns as the first step. Hence, we first identified the subset of abstract nouns that can be considered as frequently used, which refer to the commonly used nouns in daily life.

#### *3.3.3.1 Identifying Frequently Used Abstract Nouns*

In identifying the subset of frequently used abstract nouns, we have used the study carried out by Paivio et al. (Paivio A. Y., 1968), who measured the frequency of 925 nouns based on the number of occurrences per million words. Paivio et al. categorized word frequency comprehensively into three categories based on Thorndike-Lorge frequencies: high-frequency words are denoted as A or AA, medium frequency words as 10-49, and low-frequency words as 1-9. We consider high-frequency words to be frequently used abstract nouns. Based on Paivio's measurements, 83 abstract nouns are found to meet our criteria of abstract noun (as defined in 3.3.1), which constitute the subset of frequently used abstract nouns in our study.

#### *3.3.3.2 Proposed Solution*

Our research goal was to design an algorithm for recommending appropriate images in categories based on the image features of these 83 abstract nouns. A list of these 83 frequently used abstract nouns is shown in Appendix B. We acknowledge that recommending only one image to represent an abstract noun can be debatable. Therefore, we propose a categorical recommendation of appropriate images so that a learner can elect his/her own appropriate image.

### 3.4 Summary

In this chapter, we discussed the theoretical background of our research. We emphasized the concept of appropriate images for noun memorization. The research questions that we have identified are:

- What kind of image can be called an appropriate image?
- How can we define an appropriate image?
- What would be the properties of an appropriate image?
- Can there be only one appropriate image to represent an abstract noun?
- What would be an approach to extract those appropriate images?

A definition of an appropriate image for visualizing a concrete noun is proposed. This definition was based on two pedagogical investigations. In one study, we compared the memory retention rates of an image containing a solo object on a white background with an image containing multiple objects in the background or in the foreground. The second study investigated the position of the object representation in an image frame so that it is effective while acquiring foreign vocabulary with the help of an image-based vocabulary learning system.

Then, in approach 1, a definition of an appropriate image for visualizing an abstract noun belonging to particular subset of abstract nouns is presented. Basic abstract nouns that represent social contexts between humans (such as advice, opinion etc.); that are related to feeling and/or emotion (such as, love, hate); and that represent our social and religious beliefs (such as, heaven, hell) belong to this subset. We supposed that perhaps only one appropriate image can be extracted to visualize an abstract noun that belongs to this subset, and we proposed a definition accordingly. Prior to introducing this definition, we defined abstract nouns, and identified three types of images that are inappropriate learning resources for foreign language noun memorization.

In the final step, we prepared a second subset of abstract nouns. This subset identified 83-frequently used abstract nouns in the English language with the intention to employ feature-based image recommendation approach to suggest multiple appropriate images for each abstract noun. Without relying on the extraction of only one appropriate image, our approach was to recommend appropriate images in categories so that a learner can choose one appropriate image by himself/herself. The next chapter describes the technologies developed to recommend appropriate images for representing concrete and abstract nouns (defined in this chapter) for foreign vocabulary acquisition.

## 4. AIVAS System

In this chapter, we describe the key details of our system that is aimed at assisting foreign language learners in memorizing foreign vocabulary in an informal setting utilizing appropriate images.

This chapter is organized as follows. Section 4.1 gives an overview of our proposed system to support informal learning. Section 4.2 describes our proposed recommendation system for extracting an appropriate image for representing a concrete noun, as well as feature-based categorical image recommendation for abstract nouns. Next, in Section 4.3 we describe our web-based system to create vocabulary learning material on demand with the help of a standard still image, the text data, the translation data and the pronunciation data. Section 4.4 discusses the experimental environment to help us in conducting evaluation experiments. Section 4.5 describes the major technical specifications for performing tasks in our system. Finally, we summarize this chapter in Section 4.6.

### 4.1 Overview

AIVAS stands for **A**ppropriate **I**mage-based **V**ocabulary **L**earning **S**ystem, which is a web-based application designed to run in web browsers but currently supports only PC-based vocabulary learning. As the name indicates, the foremost purpose of this system is to aid learners in acquiring foreign language vocabulary using appropriate images. Figure 4-1 displays the architecture of the AIVAS system.

The main characteristics of the AIVAS are as follows. Firstly, the system lets the learners create their own learning materials on demand. The learners are able to create a five-second-long learning material for a word that include its spelling, meaning, pronunciation, and a corresponding image. The on-demand architecture allows a learner to create his/her own learning material without relying on instructors, and learn at their best times. The entire process of creating learning material happens automatically, so the learner does not need to gather learning resources such as image, text data, translation data or pronunciation data by himself/herself. Unlike existing systems where learning resources need to be gathered by a learner, the AIVAS collects them from web services automatically. As a result, the system saves a considerable amount of time in the process of creating learning material.

Secondly, AIVAS supports multiple foreign languages: the prototype supports 11 widely used languages. Therefore, learners are able to acquire vocabulary in many languages using just one system. Thirdly, the system allows learners to generate their own learning materials, save it in the system database and reuse it. Fourthly, the system allows a learner to rate previously created learning material, and also to refer to the ratings given by other learners. Fifthly, the system is able to decide the most appropriate image for a concrete noun, and recommend images based on image features for abstract nouns. Finally, the system lets the learners select their own appropriate images if the system-recommended images are not considered satisfactory.

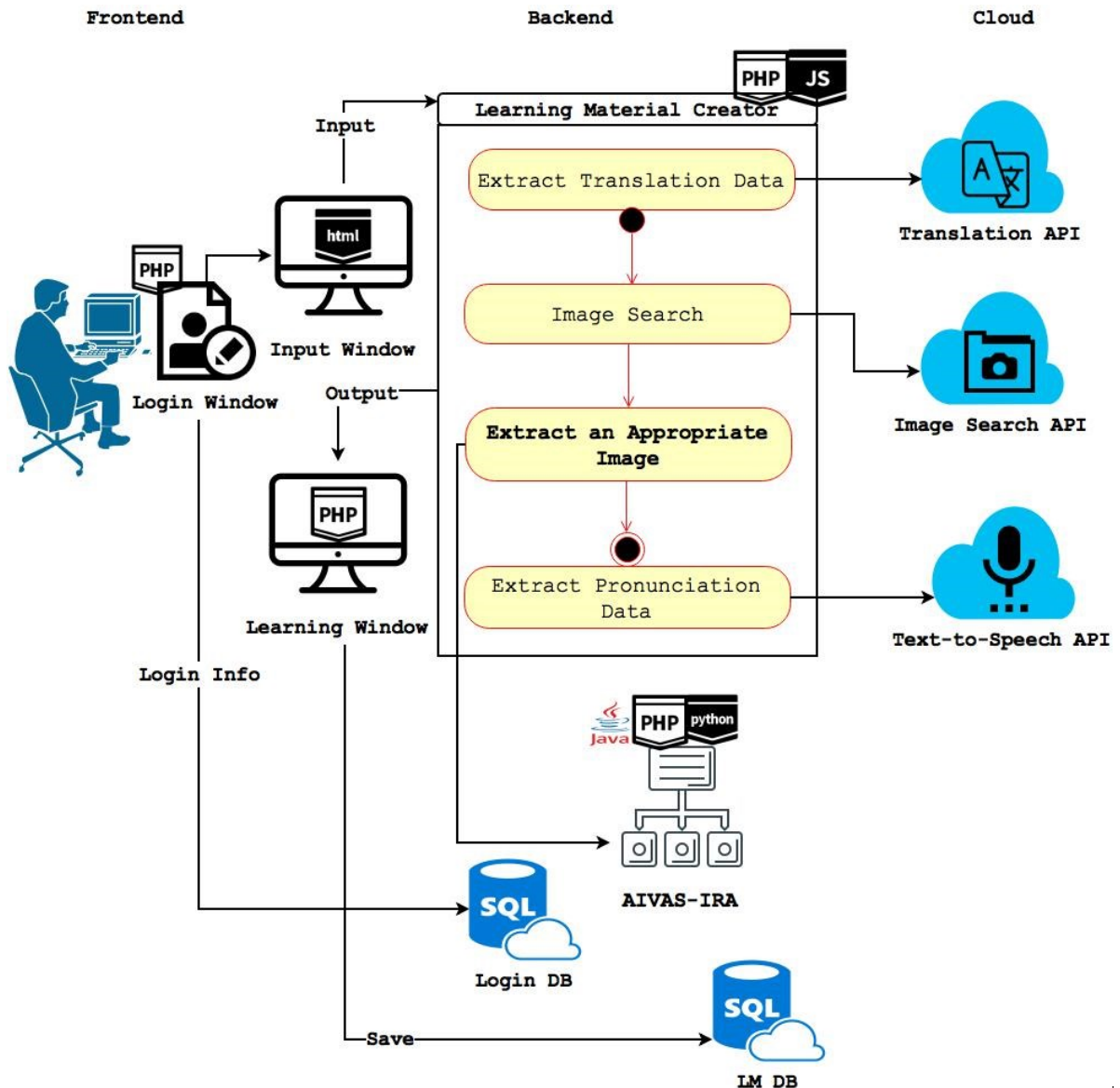


Figure 4-1 Architecture of the AIVAS

AIVAS system consists of five main subsystems:

- 1) Appropriate Image Recommendation System (hereafter, AIVAS-AIRS)
- 2) Learning Material Creator (hereafter, AIVAS-LMC)
- 3) Experimental Environment (hereafter, AIVAS-EE)
- 4) Archive
- 5) Image sets

The first subsystem, AIVAS-AIRS, recommends the most appropriate image for learning a word. The second subsystem, AIVAS-LMC, assists a learner in creating learning material. This subsystem also lets a learner rate a created learning material and save it into the databases. The third subsystem, AIVAS-EE, lets us (developers/authors) handle the learning data collection. The fourth subsystem



keeps records of the learning material created by a learner. Also, it lets learners browse through the collection of learning materials created by other learners. Finally, the fifth subsystem handles the image sets including visualizing the images and updating image sets etc. Figure 4-2 shows the subsystems of the AIVAS system. An additional subsystem named ‘multimedia annotations’ allows the learners to download multimedia annotations such as text-to-speech data, images and translation data for a word to be learned.

<h1>AIVAS</h1>					
<b>AIRS</b> <ul style="list-style-type: none"> <li>• Able to determine most appropriate image</li> <li>• Recommend appropriate images</li> </ul>	<b>LMC</b> <ul style="list-style-type: none"> <li>• On-demand ceation</li> <li>• Save learning material</li> <li>• Rate a learning material</li> <li>• Reuse a learning material</li> </ul>	<b>EE</b> <ul style="list-style-type: none"> <li>• Learning data accumulation</li> <li>• Assist Surveys</li> </ul>	<b>Imagesets</b> <ul style="list-style-type: none"> <li>• AIVAS-CNCRT59</li> <li>• AIVAS-ABST-LS68</li> <li>• AIVAS-ABST-LS795</li> <li>• AIVAS-ABST-8300</li> </ul>	<b>Archive</b> <ul style="list-style-type: none"> <li>• Manage learners' information</li> <li>• Manage learning material database</li> </ul>	<b>Multimedia Annotations</b> <ul style="list-style-type: none"> <li>• Accumulation of learning resources</li> </ul>

Figure 4-2 AIVAS Subsystems

## 4.2 Appropriate Image Recommendation System

AIVAS-Appropriate Image Recommendation System (AIVAS-AIRS) is an experimental system designed to extract appropriate images for representing nouns. This system, in recommending appropriate images, extracts those images that satisfy our proposed definition (discussed in the earlier chapter). We have noted that extracting an appropriate still image for representing a noun is a problematic and time-consuming process for learners. Our system is designed to provide a solution to this problem.

Although image search engines are effective at recommending relevant images for a noun, the adequateness of these images for memorizing the noun has rarely been discussed. One reason might be that the web-based image search engines mostly depend on the keywords surrounding an image and its file name, which leads to much irrelevancy in the search result (Ben-Haim, 2006). In addition, there is no straightforward and fully automated approach to advancing from textual queries to visual features of an image (Jain, 2011). Consequently, image search engines fundamentally depend on the static and textual features for ranking images, and are blind to the actual contents of an image (Ben-Haim, 2006) (Jain, 2011). Due to these reasons, we assumed that a standard image search engine recommended images may not be the most suitable learning resources for noun memorization. In the existing vocabulary learning systems where images are used, these images are mostly gathered from the instructors beforehand. Also, we observed that the image search engine suggested images for a noun do not satisfy the requirement of our proposed definition. Hence, we decided to develop this system.

### 4.2.1 Algorithm Design

AIVAS-AIRS is based on the AIVAS-IRA algorithms. The acronym AIVAS-IRA spells out AIVAS-Image Reranking Algorithm. The initial prototype of the AIVAS-IRA was designed for extracting an appropriate image to represent a concrete noun. However, the initial design of the AIVAS-IRA has been modified several times when we dealt with abstract nouns. We kept the name AIVAS-IRA, however in this thesis, and we use the plural form of it (that is, AIVAS-IRA algorithms). AIVAS-IRA algorithms refer to the multiple modifications of the initial AIVAS-IRA algorithm.

Four phases are involved in the design of our proposed algorithm, AIVAS-IRA: the initial phase, the intermediate phase, the re-ranking phase and the re-determining centroid(s) phase. The initial phase and the recalculate initial phase are non-repetitive. On the contrary, the intermediate and the re-ranking phases are iterative.

Table 4-1 is the pseudo code of the algorithm.

**Table 4-1** The Pseudo Code

<p><b>[Initial Phase]</b></p> <p><b>Step 1</b> Image set preparation</p> <p><b>Step 2</b> Apply an adequate image feature extraction method *Input: Images gathered in Step 1 *Output: Feature vector/feature map</p> <p><b>Step 3</b> Determine the appropriate number of the cluster(s) and their centroid(s) * Input: Features data derived from Step 2 *Output: Cluster centroid(s)</p> <hr/>
<p><b>[Intermediate Phase]</b></p> <p><b>Step 4</b> Download a finite set of corresponding images from a standard web-based image search engine into a folder *Input: Text-based query (i.e., any word that a learner intends to learn) *Output: A set of corresponding images</p> <p><b>Step 5</b> Feature extraction of all images that were downloaded in Step 4</p> <p><b>Step 6</b> Calculate the Euclidian distance(s) of each image from the centroid(s) that is the output of Step 3 *Input: Image feature of each image from Step 5 *Output: Distance(s) from the centroid(s)</p> <p><b>Step 7</b> Assign each image to its appropriate cluster (if the number of clusters is more than one)</p>
<p><b>[Re-ranking Phase]</b></p> <p><b>Step 7</b> Compare the Euclidean distance(s) of the images that belong to the same cluster(if the number of clusters is more than one)</p> <p><b>Step 8</b> Rank images in each cluster based on the measured distances (nearness to farness) to that particular cluster centroid (if the number of clusters are more than one)</p> <hr/>
<p><b>[Re-determining Centroid(s) Phase]</b></p> <p><b>Step 9</b> Re-determine the centroid(s) Step 1: Save an appropriate image into the relevant image set Step 2: Repeat Step 2 and Step 3</p>

In the initial phase, our intention was to prepare image sets containing sample appropriate images. We accumulated a set of sample appropriate images and analyzed them so that the center point of the accumulated appropriate images can be considered as the scale of appropriateness. With this goal in mind, several image sets were prepared for testing the performance of the AIVAS-IRA algorithms. This thesis reports on only four major image sets that we used for experiment (algorithm evaluation and learning effect investigation) and data collection. These three image sets are: AIVAS-CNCRT59, AIVAS-ABST-LS68, AIVAS-ABST-LS795 and AIVAS-ABST8300. Details on these image sets will be given in the next section.

In addition to preparing the image sets, we had to determine which feature extraction methods to apply. We employed both handcrafted FFT-based feature extraction method and deep CNN-based feature extraction methods on our image sets. Precisely, we used handcrafted FFT feature extraction in power spectrum to extract image features from AIVAS-CNCRT59 and AIVAS-ABST-LS68 image sets. Also, we used the pre-trained AlexNet convolutional neural network as the feature extractor for AIVAS-ABST-LS795 image set. We will discuss the reasons behind employing these two methods in 4.2.3.

#### 4.2.2 AIVAS Image Sets

Four main image sets have been prepared for the implementation of the AIVAS-IRA algorithms. They are AIVAS-CNCRT59, AIVAS-ABST-LS68, AIVAS-ABST-LS795 and AIVAS-ABST8300. We have followed *[(Project Name)-(Nature-of-the-Nouns)-(Other-information)-(Number-of-instances)]* naming format to name our image sets. The acronyms CNCRT, ABST and LS spell out concrete, abstract and learner-suggested, respectively. The number in each image set represents the number of images that were considered as sample appropriate images. Table 4-2 shows an overview of these image sets followed by a brief description of each image set.

**Table 4-2** Overview of AIVAS Image Sets

Image Set Name	Brief Description	Instances
		Format
		Preprocessing
		Creator
AIVAS-CNCRT59	59 sample appropriate images for representing 59-English words accumulated by the authors. Two measures were taken into consideration: actual object(s) highlighted in the center position, and the higher proportion of actual object(s) in the image.	59 images
		Still images (.jpg)
		None
		(Hasnine M. N., 2014) (Hasnine M. N., 2015) (Hasnine M. N., 2017)
AIVAS-ABST-LS68	68 sample appropriate images representing 14-English abstract nouns suggested by the learners of foreign languages accumulated based on a survey.	72 images (5 discarded)
		Still images, text (.jpg)
		Yes
		(Images with .jpg only) (Hasnine M. N., 2016)
AIVAS-ABST-LS795	795 sample appropriate images for representing 83 frequently used English abstract nouns that are recommended by the learners of foreign languages were accumulated.	795 images
		Still images, text (.jpg)
		None N/A
AIVAS-ABST8300	8300 sample appropriate images representing 83 frequently used English abstract nouns that were gathered by the authors. Precisely speaking, 100 images for representing each of those 83 abstract nouns. Noun relevancies were taken into account.	19759 images (11459 discarded) Still images, text (.jpg) Yes (Duplicate copy check, RGB value check) N/A

### AIVAS-CNCRT59 Image Set

AIVAS-CNCRT59 Image set contains 59 images related to 59 English words. At first, 59 English words were randomly picked from Alan Beale’s English vocabulary word list, which is a compilation of three small ESL (English as Second Language) dictionaries consisting of 21877 words. The list, often used by non-native English speakers to memorize English words, is available here (ManyThings.Org, n.d.). Table 4-3 shows the list of English words used in the accumulation of images in AIVAS-CNCRT59 image set. Then 59 images related to these 59 English words were accumulated, which we consider as sample appropriate images.

In accumulating these sample appropriate images, the following two criteria were followed:

- i) Actual object(s) highlighted in the center position, and
- ii) A higher proportion of actual object(s) in the image.

No special attention was paid to the background or foreground objects of the images. Therefore, images with/without background objects were considered. The 59 sample appropriate images cover a wide range of nouns including animals, fruits, vegetables, flowers, objects, and compound nouns.

**Table 4-3** List of the English Words Used in Preparing the AIVAS-CNCRT59 Image Set

<i>Airplane</i>	<i>Bicycle</i>	<i>Crown</i>	<i>Face</i>	<i>Newspaper</i>	<i>Soil</i>
<i>Apple</i>	<i>Balloon</i>	<i>Duck</i>	<i>Girl</i>	<i>Owl</i>	<i>Shirt</i>
<i>Ant</i>	<i>Bath</i>	<i>Deer</i>	<i>Gun</i>	<i>Pumpkin</i>	<i>Sunglass</i>
<i>Bathtub</i>	<i>Cup</i>	<i>Dolphin</i>	<i>Grape</i>	<i>Pear</i>	<i>Table-clock</i>
<i>Bath</i>	<i>Cap</i>	<i>Elephant</i>	<i>House</i>	<i>Popcorn</i>	<i>Table-fan</i>
<i>Banana</i>	<i>Cake</i>	<i>Fan</i>	<i>Hand</i>	<i>Rat</i>	<i>Tea</i>
<i>Bat (animal)</i>	<i>Clover</i>	<i>Football</i>	<i>Hen</i>	<i>Sugar</i>	<i>Tiger</i>
<i>Bus</i>	<i>Crow</i>	<i>Frog</i>	<i>Kite</i>	<i>Swim</i>	<i>Van</i>
<i>Bed</i>	<i>Car</i>	<i>Fort</i>	<i>Leg</i>	<i>Shoe</i>	<i>Vegetables</i>
<i>Book</i>	<i>Cherry</i>	<i>Fruits</i>	<i>Music</i>	<i>Sun</i>	

The AIVAS-CNCRT59 image set consists of still full-color images. The images vary in size. Accumulated images are roughly 400\*300.

### AIVAS-ABST-LS68 Image Set

AIVAS-ABST-LS68 image set is a collection of 68 images that represent 14 English abstract nouns. This set contains images suggested by learners engaged in foreign language acquisition. We considered the fact that one's cultural background may heavily influence the images for abstract nouns. Therefore, a survey was carried out to prepare AIVAS-ABST-LS68 image set.

To prepare the AIVAS-ABST-LS68 image set, at first a list of 14 English abstract nouns was chosen from our target abstract nouns list (Table 3-7). These 14 English abstract nouns were considered as the representative nouns, and were randomly selected by the authors based on the participants' familiarity with these nouns. All but one of the participants who took part in this survey were non-native speakers of English. Therefore, we chose commonly used words. The list of words used in preparing this image set is shown below in Table 4-4.

**Table 4-4** Word List Used for Preparing the AIVAS-ABST-LS68 Image Set

Anxiety	Devil	Love
Victory	Strength	Freedom
Death	Opinion	Dream
Heaven	Ability	Effort
Welfare	Silence	

Six participants joined in this experiment. All the participants reported that they are actively engaged in foreign language learning. Table 4-5 shows the details of the participants who took part in this experiment.

**Table 4-5** Participants Details

Id	Nationality/L1	L2s	Affiliation
1	Laos/Lao	English and Japanese	Waseda University
2	Cambodia/Khmer	English and Japanese	Tokyo Zokei University
3	Afghanistan/Pashto	English, Japanese, and Russian	Waseda University
4	Japan/Japanese	English and German	--
5	Yemen/Arabic	English	University of Brighton
6	British/English	Italian	University of Brighton

A survey questionnaire was prepared for preparing the images. Top 24 images for representing each abstract word listed in Table 4-4 were downloaded from Google image search API. During the experiment, printed handouts were used. The participants were asked to circle their top 10 images (from the given set of 24 images) for representing each word that they can consider as appropriate for learning that word.

The written instruction was

“Please circle your top 10 image preferences for the corresponding word”

And the oral instructions were

“Please circle only those images that you think can be considered as appropriate images for representing the word”

“Also, your image preference can be none (0) for any word”

A total of 73 sample appropriate images for representing 14 English abstract nouns were collected. Table 4-6 below shows further details on the number of images selected by the participants. Words were randomly distributed among the participants.

**Table 4-6** Details on Image Collection

Participant Id	Words Provided	No of Images Gathered
1	Anxiety	5
	Devil	2
2	Love	10
	Victory	4
	Strength	5
3	Freedom	10
	Death	2
	Opinion	3
4	Dream	5
	Heaven	5
5	Ability	6
	Effort	6
6	Welfare	3
	Silence	7

Five images were discarded due to unacceptable formats. Hence, a total of 68 images were considered as sample appropriate images and were used in testing the system.

We were expecting a total 140 images but the participants felt that the number of irrelevancies in the downloaded images is high. As a result, participants could not provide us the expected number of images. A total of 336 images were used in the experiment, but only 73 images (21.7%) were considered as sample appropriate images by the participants. It clearly indicates that the number of inappropriate images is high in Google image search outputs. Finally, our image set contains a total of 68 sample appropriate images. In other words, we only accumulated 48.6% images from our expected 140 images. As a result, this image set remained rather small in size.



### **AIVAS-ABST-LS795 Image Set**

AIVAS-ABST-LS795 image set contains 795 appropriate images for representing 83 frequently used English abstract nouns. To collect these 795 sample appropriate images, we carried out an experiment with 24 participants from nine nationalities. All participants were actively learning a foreign language.

As stated earlier, this image set was prepared based on learner-suggested appropriate images. We collected these images by engaging learners from different cultural backgrounds who were learning a foreign language.

As the first step of this experiment, we prepared a list of 83 frequently used abstract nouns of English. We used the same set of nouns (Table 3-10) that were previously used in the preparation of our AIVAS-ABST8300 image set.

Then we downloaded 30 top-ranked images by Google image search engine on 2017-10-23 for each of the listed nouns. We used Fatkun batch download image (Chrome, n.d.), a Google Chrome extension to download those images. For this experiment, we used the 30 top-ranked Google image search images for all the nouns except for the noun ‘position’, which yielded a large number of sexually explicit images. For the noun ‘position’, we selected image numbers 1, 2, 3, 5, 10, 11, 14, 15, 16, 18, 22, 23, 25, 27, 29, 30, 31, 32, 37, 38, 39, 40, 41, 47, 48, 49, 50, 51, 55, and 56.

We used both paper-and-pencil and on-line survey for this experiment. We displayed all the 30 images for an abstract noun and asked the participants to select those images that are considered appropriate to represent that particular noun in regard to foreign vocabulary learning. Participants who used on-line survey were allowed to enlarge an image and have a closer look at it. Figure 4-3 displays the user interface of the survey system developed for data collection.

Twenty-four participants from nine nationalities took part in this experiment. All the participants were university students and were actively learning foreign languages. Table 4-7 shows the number of participants and their nationalities.

**Table 4-7** Participant Details

<b>Nationality (Number of Participants)</b>
Japan(13), South Korea(3), Malaysia(2), Iran(1), Thailand(1), Vietnam(1), Poland(1), Taiwan(1), and Mongolia(1)

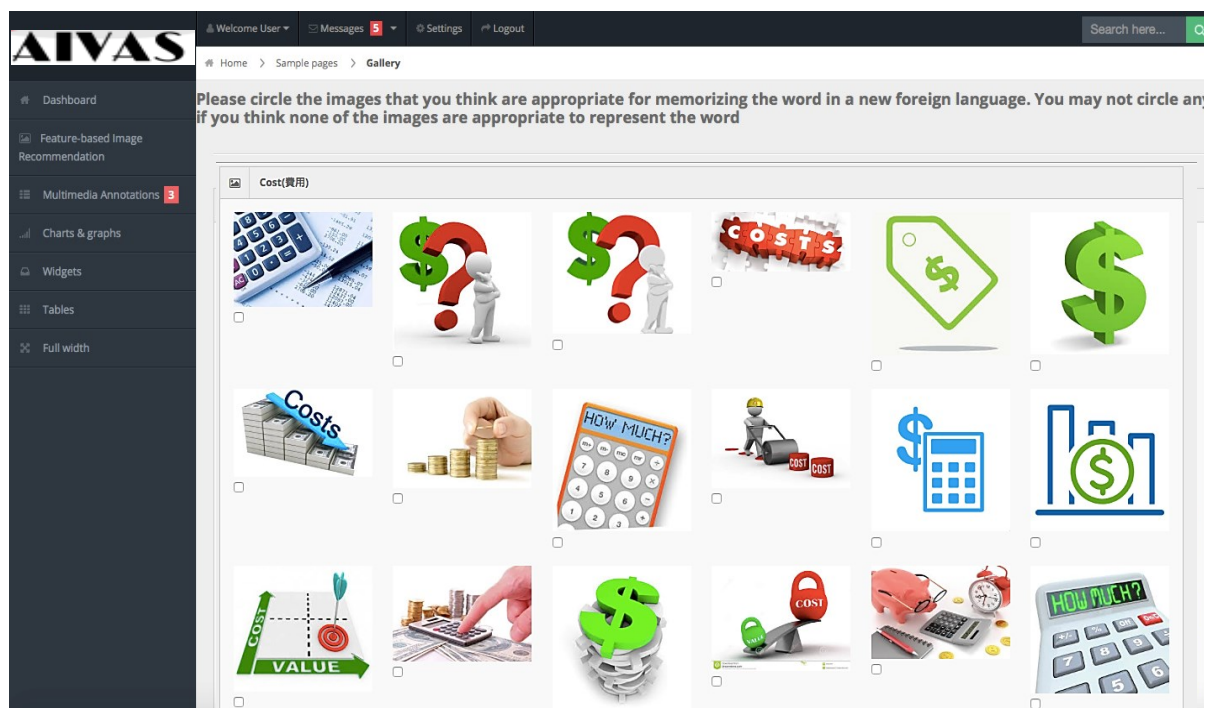


Figure 4-3 Survey Questionnaire

The purpose of the experiment was explained above. We described it to the participants both orally and through written instruction.

The written instructions were,

- *Please circle the images that you think are appropriate for memorizing the word in a new foreign language.*
- *You may choose to not circle any image if you think none of the images are appropriate to represent the word.*

There was no time limit placed on the participants, so they could spend as long as needed to submit the feedback.

In the analysis of the result, we have determined to consider a maximum of 10 images for representing an abstract noun. We decided to select the 10 images chosen by the highest number of participants. So we expected a total of 830 appropriate images. However, we were able to accumulate a total of 795 images only, because for some abstract nouns, fewer than 10 images were chosen by the participants. This clearly indicates that, Google recommended images are not always suitable to represent an abstract noun. There were a noticeable number of irrelevant images observed in the Google image search output.

This survey yielded a collection of 795 appropriate images for representing 83 frequently used English abstract nouns. Hence, AIVAS-ABST-LS795 image set is prepared with 795 appropriate images that are chosen by the learners of foreign languages.

### **AIVAS-ABST8300 Image Set**

AIVAS-ABST8300 Image set is a collection of 8300 still images representing 83 frequently used abstract nouns. One hundred images for representing each of the frequently used abstract nouns were accumulated to prepare this image set, which is designed for our future research activities. Therefore, we only report on the creation of this image set here.

To begin with, a list of frequently used English abstract nouns was prepared. To identify frequently used abstract nouns, we used the study of Paivio et al. (Paivio A. Y., 1968) as the reference. The list of the frequently used abstract nouns is displayed in Appendix B.

Next, preliminary image sets were prepared by downloading images from Google image search engine. Fatkun Batch Download Image (Chrome, n.d.), a Google Chrome extension was used to download images. A total of 19759 images were downloaded in the preliminary image sets for representing those 83 frequently used English abstract nouns. Those images were accumulated based on each word's relevancies as shown in Google image search outputs. Several relevancies were taken into consideration in the accumulation of images.

Once the preliminary image set for each word was prepared, further processing was performed to prepare the final image set. Irrelevant or inappropriate images were discarded from the preliminary image set by personal hand operations. Images to represent each abstract word were strictly limited to 100. That is, regardless of the number of the images in the preliminary image set, most appropriate 100 images were chosen. Two major preprocessing steps were followed in determining the final image set for each word: duplicate copy check, and the RGB values check. The RGB values check means whether or not RGB values from each image can be extracted. Finally, we recognized those 100 images as sample appropriate images for representing the word. Accumulated images are roughly around 350\*275. Full-color images, gray-scaled images, clip-art images, line-drawing images, images with text etc. were taken into consideration.

### 4.2.3 Determination of Feature Extraction Methods

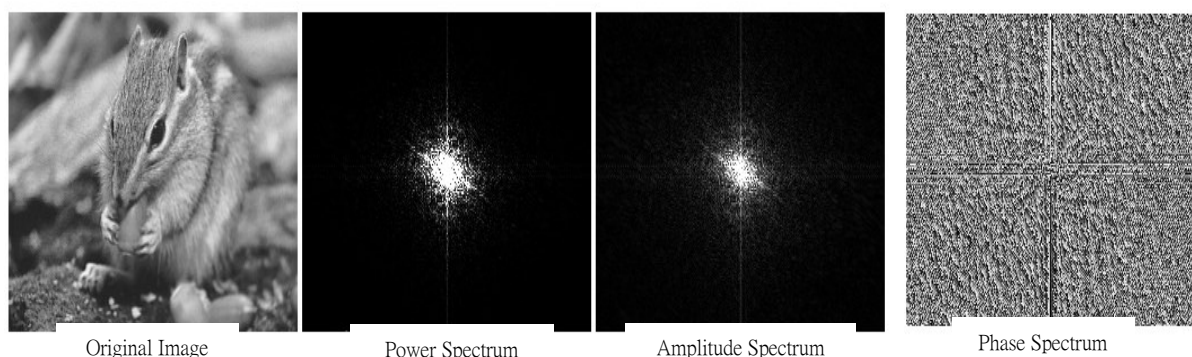
Image feature extraction is known to be a type of dimensionality reduction that systematically represents the interesting parts of an image as a compact feature vector (MATLAB, n.d.). Determination of an adequate feature extraction method was essential for the implementation of the AIVAS-IRA algorithms. Each feature extraction method has its own merits and limitations. We have adopted FFT-based feature extraction in power spectrum and unsupervised learning feature extraction using pretrained AlexNet neural network methods for our tasks. This section describes the roles of these two feature extraction methods.

#### Feature Extraction using FFT Algorithm in Power Spectrum

The Fourier transformation is known to be a powerful signal analysis tool that is applicable to a wide variety of fields. The domains of Fourier transformation include digital filtering, spectral analysis, acoustics, medical imaging, applied mechanics, modal analysis, seismography, numerical analysis, instrumentation, and communications (Fahy, 1993). In image and signal processing, FFT (Fast Fourier Transformation) is often used for deriving features in the frequency domain. Frequency domain-based features in power spectrum, amplitude spectrum, and phase spectrum of a signal is able to derive many features such as,

- Rudimentary statistical analysis of the spectrum gives some timbral properties such as spectral centroid, brightness, flatness, etc. (Lartillot, 2007)
- An approximation of roughness or sensory dissonance can be determined by appending the beats caused by each pair of energy peaks in the spectrum (Lartillot, 2007) (Terhardt, 1974)

Figure 4-4 shows a natural digital photo, its power, amplitude, and phase spectra. These example images were taken from the lecture notes on image processing offered by the Department of Computer Science, University of Auckland, New Zealand ([www.cs.auckland.ac.nz](http://www.cs.auckland.ac.nz), n.d.). Several algorithms in the field of image processing and signal processing have been implemented using FFT-based methods (Kuglin, 1975) (Horner, 1984) (Alliney, 1986) (Apicella, 1988).



**Figure 4-4** Frequency Domain-based Features ([www.cs.auckland.ac.nz](http://www.cs.auckland.ac.nz), n.d.)

The computation of FFT features in power spectra (Welch, 1967) (Fahy, 1993) (Maryland, n.d.) is a

powerful technique for measuring and analyzing the frequency content of stationary or transient signals. The FFT numerical algorithm is a fast and efficient method for computing the Fourier transform (G., 2014). The FFT features in the power spectrum are a plot of the power and variance of the time series as a function of frequency (Bendat, 2011). We adopted this method in the implementation of the AIVAS-IRA algorithm. The main advantage of using the Fast Fourier Transform algorithm in the spatial frequency domain power spectrum is a reduction in the number of computation required and a reduction of required core storage (Welch, 1967). Few other key advantages of using FFT in the power spectra are,

- Fewer computation are required than other conventional methods (Welch, 1967).
- The transformation of the sequences are shorter than the whole record. Therefore, this method is more advantageous when the computations are performed on a computer with a limited processing capability and minimal core storage (Welch, 1967).
- It has a direct impact on the time dimension, which is necessary for testing and measuring the level of nonstationarity (Welch, 1967).
- It shows excellent robustness against random noise (Reddy, 1996).
- The advantage of reducing the noise effect outweighs the disadvantage of reduced frequency resolution (G., 2014).

Another reason we adopted FFT feature extraction in the power spectrum is that we wanted to recommend images that will meet our proposed definition of an appropriate image, which is *‘an image having the actual object(s) located in the middle ground, with the highest proportion of the actual object(s) in the image frame.’* We found FFT-based feature extraction in the power spectrum most adequate for our tasks compared with other algorithms such as correlation method (Barnea, 1972), where image pixel values are directly used, low-level feature (e.g. edges or corners) extraction using feature-based methods (Brown, 1992), and high-level features (parts of objects or relations between features) using graph theoretic method (Brown, 1992). Furthermore, a study conducted by (Bao, 2004) reported excellent classification accuracy by employing a mixed set of time-domain and frequency-domain features (Preece, 2009). After carefully considering all these factors, we have decided to adopt the FFT-based feature extraction in the power spectrum for our tasks.

### **Unsupervised Learning Feature Extraction using Pretrained AlexNet Neural Network**

We have implemented Alexnet (Krizhevsky, 2012), a pre-trained neural network to perform unsupervised learning feature extraction from the sample of 795 appropriate images stored in the AIVAS-ABST-LS795 image set. The Alexnet model is trained on a subset of ImageNet (ImageNet, n.d.) database. This deep neural network was first introduced in the ImageNet Large-Scale Visual Recognition Challenge (ILSVRC), where it won the first prize (Russakovsky, 2015). The architecture of this model is capable of learning rich and distinct features from images with wide variations. We have used this model as an unsupervised learning feature extractor because of the wide variations of images in the AIVAS-ABST-LS795 image set.

Convolutional Neural Networks, known as CNNs or ConvNets, are biologically-inspired variants of multilayer perceptron in the artificial neural network (Learning, 2017). Generally speaking, typical

CNN architectures are alike to traditional neural networks. Traditional neural networks are constructed of neurons having learnable weights and biases. However, the main difference between deep CNNs compared and ordinary neural networks is that the deep CNN architecture is intelligent enough to predict the assumptions of an input image, which lets us encode and/or visualize certain properties into the architecture (Karpath, 2017). These then let the forward function become more proficient to implement the architecture, and reduce considerably the number of parameters in the architecture. A typical deep convolutional neural network is trained using large collections of diverse images. Therefore, deep learning utilizes the convolutional neural networks to extract (learn) rich features directly from images. These feature representations often outperform hand-crafted features such as FFT (fast Fourier Transformation (Welch, 1967)), HOG (Histograms of Oriented Gradients (Dalal, 2005)), LBP (Local Binary Patterns (Ojala, 2002)), MSER (Maximally Stable Extremal Regions (Matas, 2004) (Obdržálek, 2009)) or SURF(Speeded Up Robust Features (Bay, 2008)). One of the cutting-edge approaches is to leverage the power of deep convolutional neural networks. By leveraging the power of deep CNNs, we can use an existing pre-trained CNN as a feature extractor (MathWorks, n.d.) without spending much time and effort into training a new neural network. The key difference in this approach is that instead of using image features such as handcrafted FFT or HOG or SURF, features are extracted using a convolutional neural network. The superiority of CNN features over handcrafted features has been demonstrated in several studies. Moreover, a classifier trained with CNN features provides close to 100% accuracy, which is higher than the accuracy achieved using the bag of features and SURF (MathWorks, n.d.). The example images (taken from MathWorks) show some scenarios of CNN-based feature extraction and classification tasks.

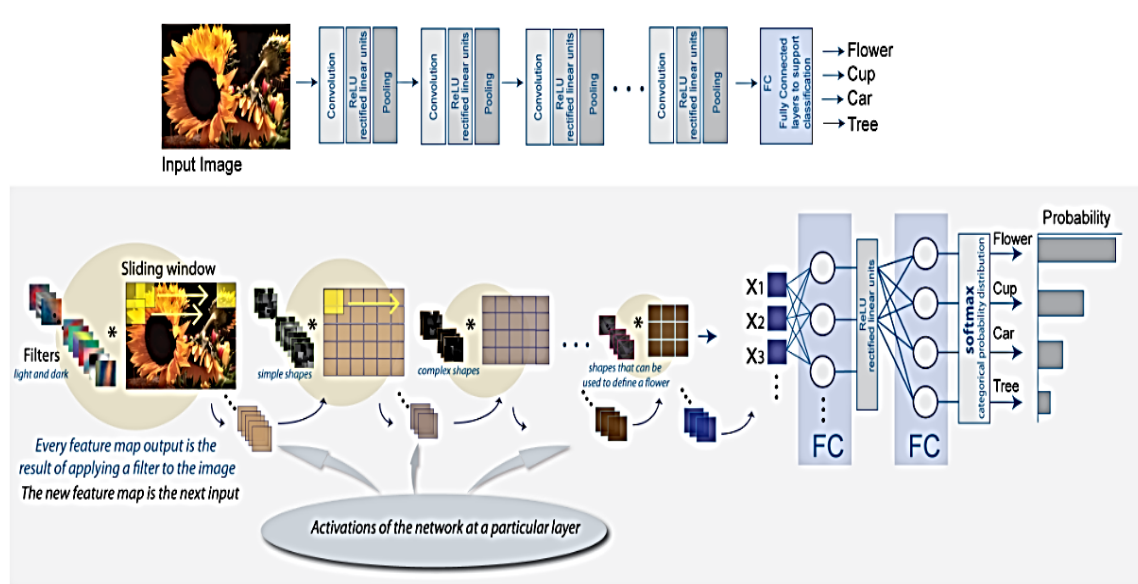


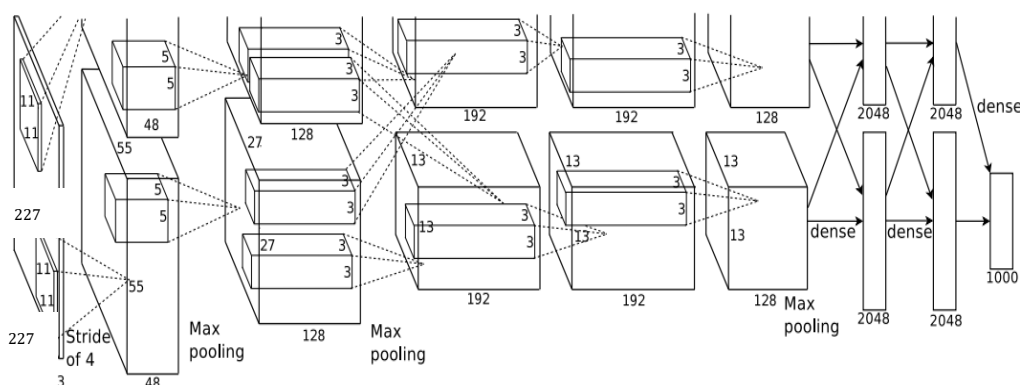
Figure 4-5 CNN-based Feature Extraction and Classification (MathWorks, n.d.)

With the revolution of deep learning, several convolutional neural network architectures have been proposed. Several case studies have also been reported by adopting these CNN architectures. In computer vision, commonly adopted CNNs are LeNet (LeCun, 1998), AlexNet (Krizhevsky, 2012), ZF Net (Zeiler, 2014), GoogLeNet (Szegedy, 2015), VGGNet (Simonyan, 2014) and ResNet (He,

2016) etc. For our unsupervised learning feature extraction tasks, we have used the pre-trained Alexnet architecture. We have considered the fact that our AIVAS-ABST8300 image set contains a wide variety of images. For an image set with diverse images, the key advantages of pre-trained Alexnet as feature extractor over other CNNs are:

- The Alexnet architecture is trained on the network of ImageNet database, which contains over 15 million annotated images from 22,000 categories.
- This pre-trained model has already learned distinct feature representations from a wide range of images, and it can classify images into 1,000 object categories.
- The architecture used ReLU for nonlinear functions. Moreover, ReLUs are several times faster in processing than the conventional hyperbolic tangent function. Therefore, this architecture, compared with the other architectures, decreases the feature extraction and training time.
- The architecture uses data augmentation techniques containing patch extractions, horizontal reflections, and image translations (Deshpande, 2016).
- In order to combat the problem of overfitting the training data, the model implemented dropout layers (Deshpande, 2016).
- The model is trained using batch stochastic gradient descent with specific values for momentum and weight decay (Deshpande, 2016).
- This model has been cited over 14,000 times, and is widely regarded as one of the most influential deep architectures for image feature extraction and recognition tasks. Moreover, this architecture works well for an image set with widely varying images.

The pretrained AlexNet architecture incorporates a total of 25 layers, with eight layers (5 convolutional layers and 3 fully connected layers) having learnable weights. The final layer does the classification tasks, classifying the input images into 1,000 object categories. Other layers in the architecture perform the feature extraction tasks. We have implemented this architecture to perform feature extraction from our image set. We have implemented the architecture without the last layer. Figure 4-6 shows the architecture of the pre-trained AlexNet neural network.



**Figure 4-6** The Architecture of the Pre-trained AlexNet Neural Network (Krizhevsky, 2012)

#### 4.2.4 Algorithms Implementation & Output Demonstration

This section describes the implementation of AIVAS-IRA algorithms: first by employing an FFT-based feature extraction method; and next by employing a CNN-based feature extraction method. FFT-based feature extraction method is employed for extracting an appropriate image representing a concrete noun. The CNN-based feature extraction method is applied for feature-based image category recommendation for representing an abstract noun.

##### FFT-based Method in Power Spectrum

In the initial phase, Fast Fourier Transform (FFT) in the power spectrum is applied to the set of appropriate images stored in the AIVAS-CNCRT59 image set to determine their center. First, FFT-based power spectrum features in the spatial frequency domain is obtained from each image by implementing the FFT algorithm. All the images in the AIVAS-CNCRT59 image set were gray-scaled in 256-gradation and resized to 32×32 dimension. The output of the power spectrum is regarded as a 1,024-element vector, and the center-point of all the vectors derived for all the sample appropriate images is determined. Then the derived vectors are converted into their unit vectors. Vector normalization method is applied to determine the unit vectors. Next, we performed cluster analysis with the derived vectors using the R clustering environment. The result of the cluster analysis suggested that all 59 images do not vary widely in their properties, and were assigned to one cluster. Next, we calculated the average of the generated vectors. Then the derived average vector was converted to a unit vector by vector normalization method, and recognized it as the centroid of sample appropriate images. We have considered the centroid of the appropriate image set as a measure of the appropriateness of images in the re-ranking phase. In other words, by using this centroid, we judge the appropriateness of the still images for all concrete nouns. For the sake of clarification, neither the system nor a learner requires preparing any new sample image when a new word is input.

In the intermediate phase, when an input (a concrete noun) query is received, a set of corresponding images for the word is downloaded. The algorithm supports downloading images from third-party API services (i.e. Flickr image search API, Yahoo image search API etc.) as well as from a browser extension (such as Fatkun batch downloader for Google chrome). Then FFT-based power spectrum features in the spatial frequency domain of all the downloaded images are extracted by applying FFT algorithm. Here, all images in the corresponding image set are gray-scaled (256-gradations) and resized to 32 × 32 dimension. The output is derived in the form of 1,024-element vectors. Afterwards, unit vectors of the derived vectors are determined. Finally, the Euclidean distance of each vector from the centroid (calculated in the initial phase) is computed.

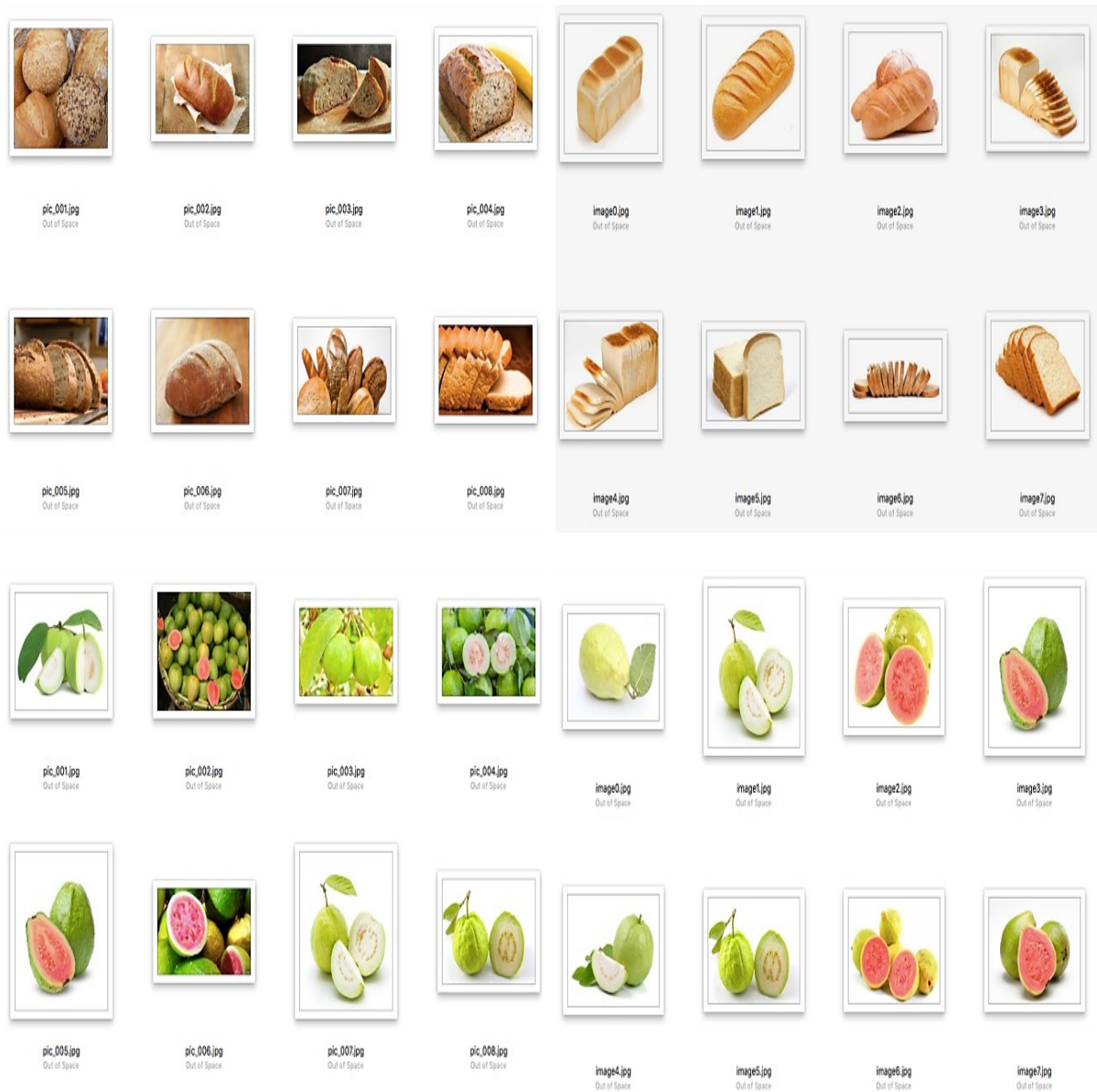
In the re-ranking phase, first the Euclidean distances found in the intermediate phase are compared. An image closest to its centroid is regarded as the most appropriate image, and is extracted.

The re-determine centroid(s) phase is performed manually when a significant number of appropriate images are observed in the database. Currently, no mechanism has been employed to automatically recalculate the centroid of appropriate images.



### Output Demonstration

In Figure 4-7, we demonstrate output of the algorithm for a concrete noun. To prepare this demonstration, we have downloaded a set of 35 images that are top-ranked by Google image search engine for each the noun ‘bread’ and ‘guava’. Figure 4-7(a) shows the 8 top-ranked images by Google image search engine for representing these two nouns. On the other hand, Figure 4-7(b) shows the 8 top-ranked images ranked by our algorithm. The image titled as image0.jpg is the most appropriate image recommended by our algorithm for learning the word.



(a) (b)  
**Figure 4-7** Output Demonstration for Concrete Nouns

## Pretrained AlexNet as the Feature Extractor

The steps involved in the implementation of the initial phase are as follows,

Step 1: All of the 795 appropriate images stored in the AIVAS-ABST-LS795 image set were resized to 227\*227 RGB images, because the pre-trained AlexNet can only process RGB images in this size.

Step 2: Apply pre-trained AlexNet as the feature extractor to these 795 images recommended by learners of foreign languages.

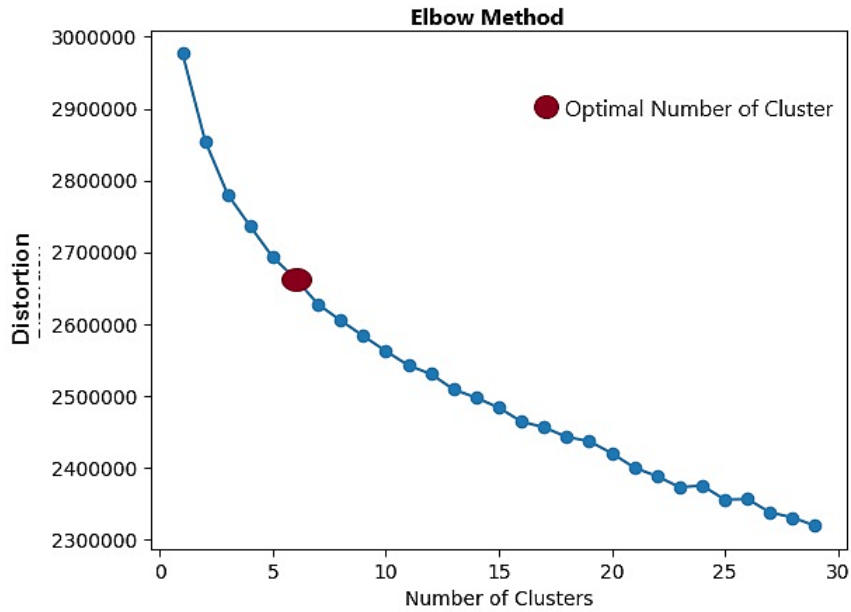
Layer activations as unsupervised learning features were used here. The Pre-trained AlexNet has 8 layers with learnable weights. Five of these are convolutional layers and three are fully connected layers. We have only used 7 layers (five convolutional layers and two fully connected layers) for our feature extraction tasks excluding the final fully connected layer (the FC8 layer), which performs the classification tasks. Each layer produces a response or activation to an input image.

Convolutional Layer 1 consists of three hidden layers: convolutional1, learn1, and maximum pooling1. The hidden convolutional1 filters a 227\*227 RGB image with 96 kernels of size 11\*11\*3 with a stride of 4 pixels and 0 padding. Hidden learn1 performs cross-channel normalization with 5 channels per element. Hidden maximum pooling1 performs 3\*3 max pooling with a stride of 2 and a padding of 0. Convolutional Layer 2, which takes as input the output of Convolutional Layer 1, also consists of three hidden layers, which are convolutional2, learn2, and maximum pooling2. Hidden convolutional2 filters the input with 256 kernels of size 5\*5\*48 with 1 convolutional stride and padding size of 2. Hidden learn2 performs cross-channel normalization with 5 channels per element. Hidden maximum pooling 2 performs 3\*3 max pooling with stride of 2 and padding 0. Convolutional Layer 3 is connected with Convolutional Layer 4 without any intervening maximum pooling or learning layers. These layers filter the input with 384 kernels of size 3x3x256 convolutions with stride 1 and padding 1. Convolutional Layer 5 has two hidden layers: convolutional5 and maximum pooling5. The hidden convolutional5 layer filters the input with 256 kernels of size 3x3x192 convolutions with stride 1 and padding 1. Hidden Maximum pooling5 performs 3\*3 max pooling with a stride of 2 and padding 0. Two fully connected layers in AlexNet architecture possess 4096 neurons each. The last fully connected layer outputs the feature vector of 4096 in length. The ReLU (Rectified Linear Units) non-linearity was used in the output of every convolutional and fully-connected layer as the layer activation function.

Step 3: Unsupervised learning features were obtained from each image as a feature vector after applying pre-trained AlexNet as a feature extractor. A total of 795 feature vectors were obtained, each having a length of 4096. We have named these high-dimensional deep CNN feature vectors as the feature map.

Step 4: Perform cluster analysis on the feature map to determine the true optimal number of clusters, using Elbow's method. We chose Elbow method because it is one of the oldest and efficient method to determine the true optimal number of clusters for K-means clustering (Kodinariya, 2013)

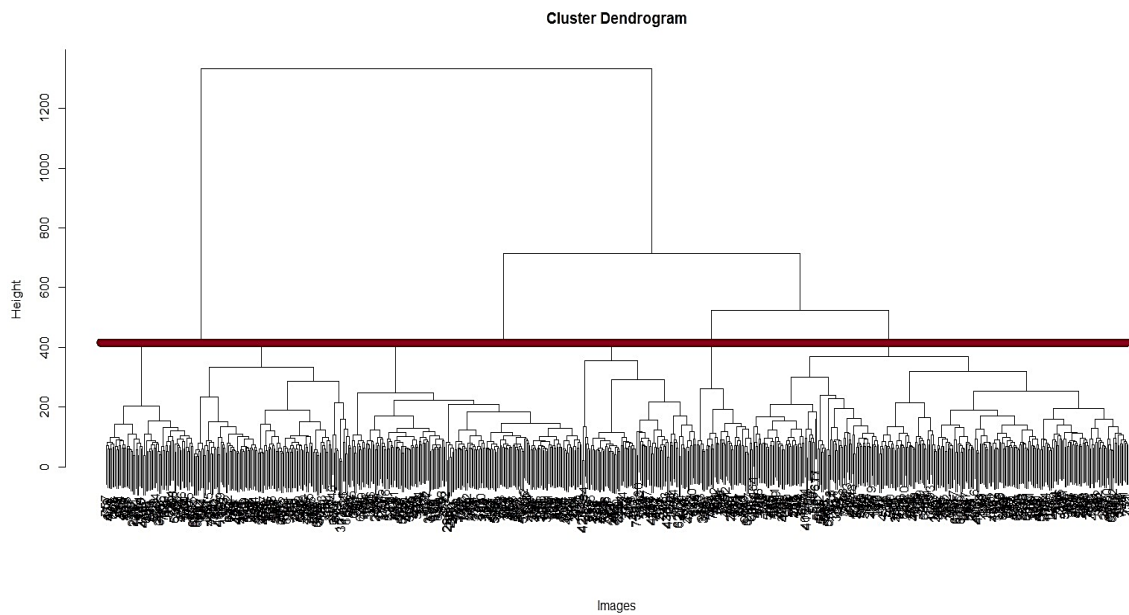
(Madhulatha, 2012) (Tibshirani, 2001). Based on result given by the Elbow clustering algorithm for the span for  $k = (1 \text{ to } 30)$ , we have determined that the optimal number of clusters is 6. Figure 4-8 displays the output graph of the Elbow method on our feature map derived in the previous step. In Figure 4-8, the distortion increases slowly onwards, and a rapid distortion is not observed until  $k=6$ . Therefore, we determined the optimal number of clusters to be 6.



**Figure 4-8** Identification of the Elbow Point

Step 5: we performed K-means clustering to assign an image to its appropriate cluster. K-means clustering algorithm (Hartigan, 1979) is a widely used technique that is simple and fast. Therefore, we used this technique to determine the appropriate number of images in each cluster. We report that Cluster 1, Cluster 2, Cluster 3, Cluster 4, Cluster 5, and Cluster 6 contain 93, 45, 260, 74, 195, and 128 images, respectively. We performed this task with the R software.

We also observed the similarities and dissimilarities in the extracted learning features using Ward’s hierarchical clustering method (Murtagh, 2011), which is considered the most appropriate for quantitative variables. It is used to determine similarities, dissimilarities, and distances among clusters. Figure 4-9 shows the dendrogram view of the images in 6 clusters.



**Figure 4-9** Distribution of the Images in 6 Clusters

Step 6: we determined the centroids of the six clusters: the average of the generated feature map in each cluster was calculated to determine the centroid of that cluster. We consider the distance from the centroid of a cluster to be the measure of the appropriateness of any image belonging to that cluster.

In the intermediate phase, when an abstract noun is input, firstly, a set of corresponding images for that noun is downloaded into a folder from an image search engine or through a browser extension. Secondly, the feature vectors of all the downloaded images are obtained by using the pre-trained AlexNet as a feature extractor on the transformed 227\*227 RGB images. Thirdly, the Euclidean distances of each feature vector from the centroids of the 6 clusters (determined in the initial phase) are calculated. Fourthly, the algorithm compares these 6 Euclidean distances for each image and based on the calculated distances, that image is assigned to its nearest cluster. In this way, the algorithm assigns all the images that were downloaded as the set of corresponding images into their appropriate clusters.

In the re-ranking phase, previously calculated Euclidean distances of the images belonging to the same cluster are compared. An image closest to its centroid is considered as the most appropriate image, and it is extracted and recommended. In this way, six most appropriate images are extracted from six clusters. These six images are the recommended appropriate images for an abstract noun. The ranking phase is also able to extract other appropriate images (the second most appropriate image, the third most appropriate image, and so on) belonging to the same cluster.

### Output Demonstration

Figure 4-10 displays a demonstration of the category-based recommendation of appropriate images for an abstract noun. To prepare this demonstration, we downloaded a set of 65 top-ranked images by the Google image search engine for the abstract noun *happy*. However, in Figure 4-10 we display only 36 of these images that are top-ranked by our algorithm: top 6 images representing each category are displayed here. Note that the first image in each category is the most appropriate image from that particular category.

- Dashboard
- Feature-based Image Recommendation
- Multimedia Annotations 3
- Charts & graphs
- Widgets
- Tables
- Full width

## Appropriate Image Recommendation System (AIRS)

(Categorical Recommendation)

Select an Appropriate Image For The Word '**HAPPY**' and Click on Create Button to Generate a Learning Material

**Category 1**

**Category 2**

**Category 3**

**Category 4**

**Category 5**

**Category 6**

**Create**

Figure 4-10 Categorical Recommendation for an Abstract Noun

### 4.3 Learning Material Creator

The acronym AIVAS-LMC spells out Appropriate Image-based Vocabulary Learning System-Learning Material Creator. It facilitates learners in creating their own vocabulary learning material by inputting the words that they intend to learn or memorize.

One basic question is: what constitutes learning material? The term *learning material* (sometimes called learning content) has been widely adopted by learning analytics, educational technologists, linguists, teachers, and researchers. Often the term learning material is confused with teaching material. Even though teaching material is also designed for learning, but the terminology *learning material* has been specifically used by many researchers (Tomlinson, 2008) (Mishan, 2005). Simply speaking, learning material is an organized body of educational resources prepared carefully for helping a learner to acquire knowledge. In the old days, vocabulary learning materials were prepared around text-books and dictionaries. As a result, reading was the main mode of acquiring vocabulary. Ever since the technology boom (often known as the dot-com boom) in the mid-1990s, several multimodal representations of learning material have been proposed and tested. Annotated (with text, sound, image, video, animation etc.) educational resources have been widely used in the implementation of multimedia-enriched learning systems. Image, text, translation, and sound of a word are the four major components of the learning material generated by the AIVAS-LMC.

AIVAS-LMC automatically creates sets of learning material, each of which consists of the spelling, the meaning, the pronunciation data together, and an appropriate image. A learner simply needs to input a word into the system that he/she wishes to learn. The learner may or may not be familiar with the word that he/she inputted into the system. Once the system receives a text-based query, four steps are involved in the creation of a set of learning material in the AIVAS-LMC system as explained below:

First, it looks for the meaning of the word from an external translation engine in the web.

Second, it sends the query to an external image search engine (such as Google image engine, Yahoo image search engine etc.) for extracting images corresponding to that word. Then, it downloads a set of images corresponding to that input word to a folder in the local drive. With the help of AIVAS-IRA algorithms, the set of images are ranked and the most appropriate image to represent the input word is selected. In the current version, image search on the web is performed based on the English translation of any non-English word. That is, all non-English input queries are translated into English downloading the set of corresponding images. A polysemic word is searched with its default translation set by any translation engine on the web.

Third, the system looks for the pronunciation data on the web and extracts it from a text-to-speech engine (such as, VoiceRSS or Google TTS services etc.).

Fourth, the spelling and the meaning are embedded as subtitles and the learning material is generated by superimposing the subtitles into the appropriate image while playing the pronunciation. After this,



the AIVAS-LMC displays the created learning material into the computer screen of the learner so that he/she can memorize the vocabulary. With the help of this AIVAS-LMC-created learning material, a learner is able to learn foreign vocabulary in an informal educational setting.

Figure 4-11 displays the learning material creation operation in the AIVAS-LMC.

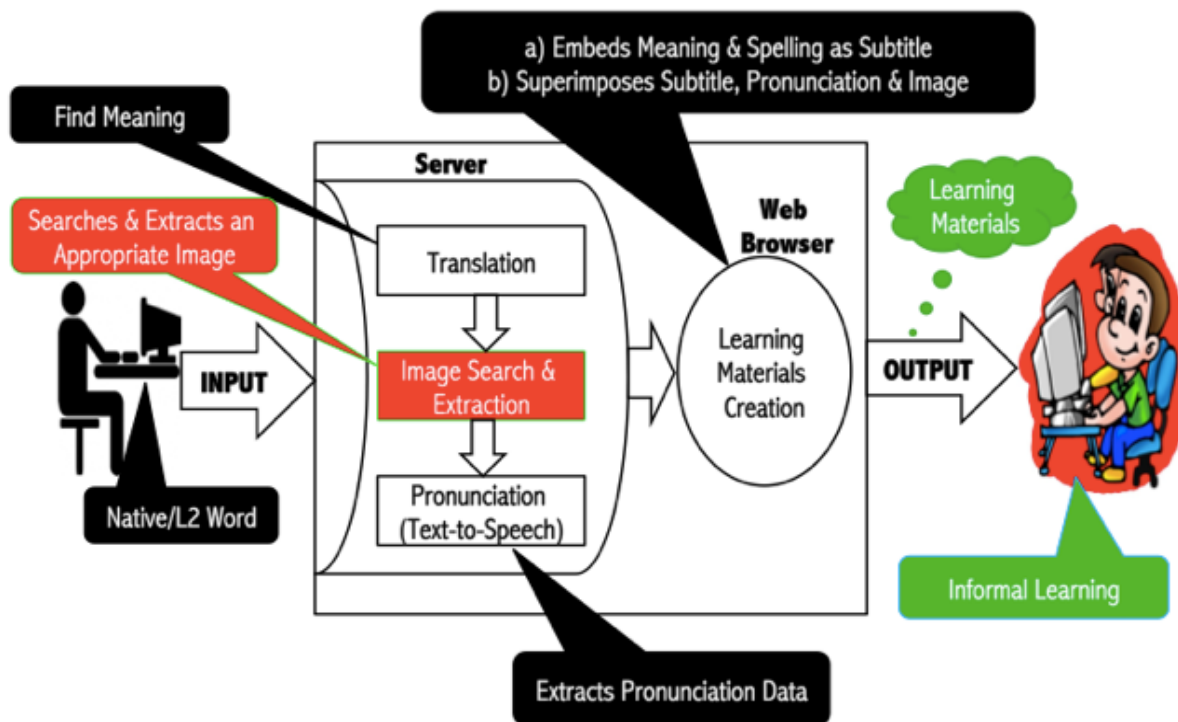


Figure 4-11 The Operation of the AIVAS-LMC

The interface of the AIVAS system has been designed for learners with inadequate IT skills. Figure 4-12 shows the user interface of the AIVAS-LMC.

The AIVAS-LMC interface consists of four simple fields:

- Field 1) Input a text-based query in this field (insert the word the learner wants to acquire),
- Field 2) Specify whether or not the entered word is in the learner’s spoken language
- Field 3) Specify the spoken language of the learner, and
- Field 4) Specify the target (foreign/spoken) language that the learner aims to acquire.

In order to create a set of learning material, all four fields need to be specified by the learner. Field 1 lets the learner input a word into the text box by typing or through copy-paste operation. The input must be a single word or a compound word with or without a hyphen. In Field 2, the learner needs to specify if the input word is in his/her spoken language or not just by clicking the appropriate radio button. AIVAS-LMC is programmed so that can create learning material for a word regardless of whether it is in the learner’s familiar or unfamiliar language. Based on the information from Field 2, the system determines the type of learning material to be created. That is, 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). After this, the learner needs to click on the ‘Create’ button, at which point the learning material will be created to acquire the target language.

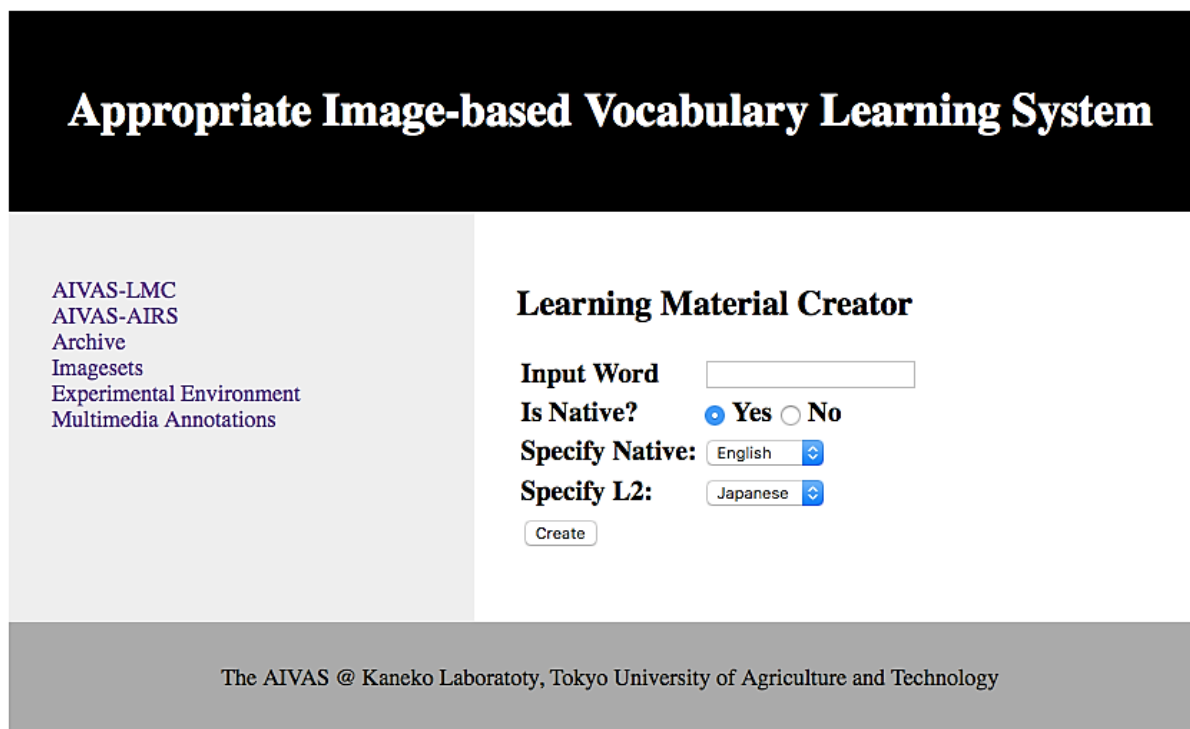


Figure 4-12 AIVAS-LMC Interface

However, if the input word is not from a language spoken by the learner, then the system automatically considers it as a word of the target language that the learner wants to acquire. Thereafter, upon specifying the spoken language (in Field 3) and the target language (in Field 4), the learning material will be generated to acquire that unknown word. Currently AIVAS-LMC is not programmed for automatic detection of languages. Hence, the learner is expected to select the spoken and the target languages in both Fields 3 and Field 4. At this moment, AIVAS-LMC supports eleven widely spoken languages: Chinese, English, French, German, Italian, Japanese, Korean, Polish, Russian, Spanish, and Swedish. A non-native learner of any of these languages is expected to be familiar with at least one of these languages to learn the vocabulary using AIVAS because the system currently does not support other languages.

As stated earlier, the learning material created by AIVAS-LMC is of 5-second duration. Throughout this 5-second interval, the appropriate image and the spelling of the word are displayed. The pronunciation of the word is repeated twice: on the first and the third second, respectively. The meaning is displayed on the screen two seconds after the beginning. The delay is designed to give the learner a brief time to look at the spelling of the word to determine whether they have already memorized the word or not. The learning material format is based on an earlier study by (Hasegawa K., 2007) (Kaneko, 2007).

Figure 4-13 displays the learning material format created by AIVAS-LMC. Figure 4-14 shows some samples of the learning material created by AIVAS-LMC. The system displays the created learning material (as shown in Figure 4-14) in the output window where the learner acquires foreign vocabulary.

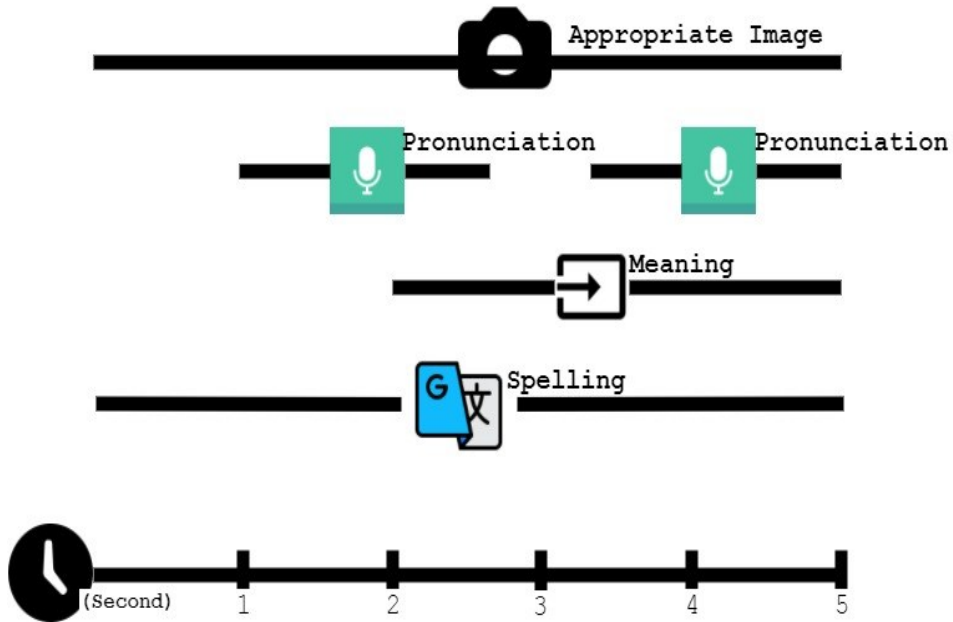


Figure 4-13 The Format of a Learning Material in the AIVAS System



Figure 4-14 Learning Material Samples

## 4.4 Experimental Environment

The subsystem AIVAS-Experimental Environment (AIVAS-EE) is built to support our experiments. In the current version, only the authors are allowed to use this subsystem. So far, this subsystem has been used to conduct the surveys (1) to prepare the AIVAS-ABST-LS795 image set (discussed in 4.2.2), (2) for the Learning Effect Investigation I (to be discussed in 5.2), (3) for the Image Evaluation Experiment II (to be discussed in 5.3), and (4) Learning Effect Investigation II (to be discussed in 5.4). Functionalities and the interfaces have been redesigned to handle each of those experiments. Figure 4-15 shows the interface of the AIVAS-EE used in Learning Effect Investigation I.

### Appropriate Image-based Vocabulary Learning System


[AIVAS-LMC](#)  
[AIVAS-AIRS](#)  
[Archive](#)  
[Imagesets](#)  
[Experimental Environment](#)  
[Multimedia Annotations](#)

#### Group 1

**Study Session: *For Non-native Japanese Speakers***

Please click a word to generate learning materials:

- [dog](#)
- [cat](#)
- [tiger](#)
- [pigeon](#)
- [mango](#)
- [kiwi](#)
- [melon](#)
- [grape](#)
- [cucumber](#)
- [potato](#)
- [carrot](#)
- [cabbage](#)
- [aircraft](#)
- [bedroom](#)
- [bookshelf](#)
- [earphone](#)
- [chair](#)
- [clock](#)
- [tree](#)
- [camera](#)



**капуста**  
**cabbage**

The AIVAS @ Kaneko Laboratory, Tokyo University of Agriculture and Technology

Figure 4-15 An Interface of the Experimental Environment

## 4.5 Technical Specifications

AIVAS system is currently an experimental system. Not all the functions can be used from the AIVAS interfaces. Basic understandings of the computer technology, particularly command line execution, is essential to take full advantage of this system. In this section, we briefly describe the technical details about AIVAS implementation.

AIVAS is a localhost-server application that can be connected locally. The client side can be accessed through a PC web browser. AIVAS source codes are written in PHP, HTML, Java, and Python languages. The system was coded using CSS and JavaScript libraries. To run the source code, one needs to ensure a computer with certain libraries installed. Furthermore, phpmyadmin and pgadmin administrators are required as the system used both MySQL and PostgreSQL databases. Table 4-8 provides the details of the development tools.

**Table 4-8** Development Tools

OS	macOS, Windows 10 Professional
Coding Toolkit	Frontend: HTML5, CSS, JavaScript, CreateJS, jQuery, Bootstrap Backend: PHP, Java, Python
Web Server	Apache
Database Server	MySQL
Cloud Services (External APIs)	Google Image Search API (currently depreciated), VoiceRSS API Microsoft Translation API(currently under Microsoft Cognitive Service API) Google Places API Flickr API
Clustering	R, EZR
Machine Intelligence Toolkits	Tensorflow MATLAB
Major Open Source Libraries	Libsvm, jTransforms, Numpy, PIL, caffe_classes, Python collections, Scikit-learn, CreateJS etc.

- **AIVAS-Login:** AIVAS-login system has been implemented with session written in PHP. AIVAS-login system works properly upon locating the source codes into htdocs, and then just changing the database settings, which refer to matching the database names. It can be done by editing programs. Currently, AIVAS login system is connected with the aivas\_login database. aivas\_login which is a MySQL database contains login table with seven fields, namely Id, first\_name, last\_name, email, password, hash, and active.
- **AIVAS-LMC:** As stated in 4.3, this subsystem handles the learning material creation task. It is implemented mainly in HTML5 (frontend) and PHP (backend). To perform a learning material creation task, a user is required to only enter a word in the input field. Therefore, the learning

material creation task is trouble-free. This subsystem is connected with the aivas\_lm database, which records the learning material created and stored by a learner. The aivas\_lm database contains information on nine fields, namely learner id, source\_language, destination\_language, query, english\_translation, destination\_translation, image\_url, date\_created, and rating score.

- **AIVAS-IRA:** AIVAS-IRA algorithms are implemented with PHP, Java and Python programming languages. Additionally, it requires Tensorflow to perform the image feature extraction tasks. AIVAS-IRA algorithms need to be run through command line instruction.
- **AIVAS-AIRS:** This subsystem is implemented with HTML5 and CSS in front-end, and PHP in the backend. The AIVAS-AIRS recommends appropriate images for a noun and lets a learner choose an appropriate image himself/herself to create the learning material. Figure 4-16 displays the current interface of AIVAS-AIRS. At the time of submitting this thesis, no database is implemented to distinguish between nouns. However, implementation of two name databases aivas\_concrete and aivas\_abstract are under consideration, which will contain the list of concrete and abstract nouns in the English language, respectively.

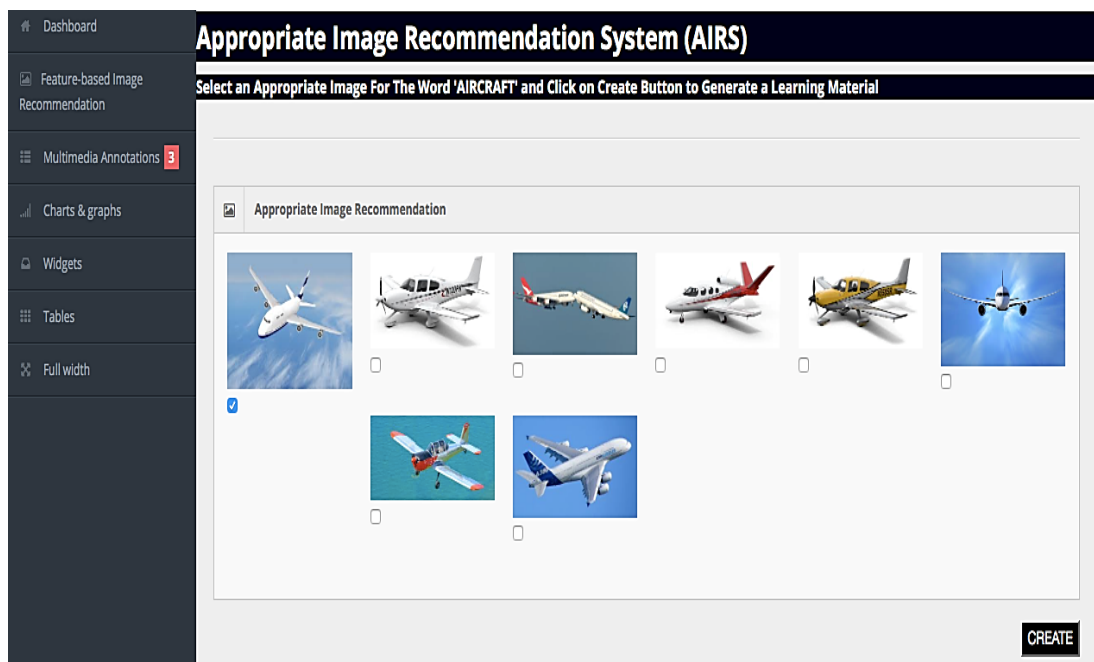


Figure 4-16 AIVAS-AIRS Interface

- **Archive:** This subsystem connects between learners, and allows them to have a look in their own logs. It also allows a registered learner to browse through the learning material created by other learners.
- **Image sets:** This subsystem lets a learner access the AIVAS image sets. Currently, it can be accessed through local connection. At the time of submitting this thesis, the image set is not open publicly.

## 4.6 Summary

In this chapter, we reported the key system developments. At the beginning of the chapter, we articulated the system architecture. Then we described our recommendation system (AIRS) for generating appropriate images. In Section 4.2, we described the design of the AIVAS-IRA algorithms and their implementation. We also provided details of the image sets that we prepared to test our algorithms. We explained the reasons for employing FFT-based feature extraction in power spectrum and CNN-based learning feature extraction methods. We also discussed the approach taken to determine the optimal number of clusters in our dataset, and demonstrated the output of our algorithms.

Next, in Section 4.3, we described the learning material creator (LMC) subsystem. We discussed how the learning material is created by AIVAS-LMC by providing some examples. We also explained the format of learning material.

Next, in Section 4.4, we briefly discussed the experimental environment (AIVAS-EE) that supports handling our experiments. Finally, we provided the technical details of AIVAS in Section 4.5.

## 5. Experiments

This chapter describes the evaluation experiments we carried out to assess the efficacy of appropriate images recommended by our system. We conducted four experiments: two of which are image evaluation experiments and two are learning effect investigations.

This chapter is organized as follows. Section 5.1 describes the Image Evaluation Experiment I that was carried out to assess the performance of our algorithm for concrete nouns. Section 5.2 describes the Learning Effect Investigation I that shows the immediate-after, mid-term and long-term learning effect of images. In Section 5.3, we describe the Image Evaluation Experiment II that assesses the performance of our algorithm for abstract nouns. Then we discuss the Learning Effect Investigation II for assessing the learning efficacy of images for abstract nouns in Section 5.4. Finally, we summarize the chapter in Section 5.5.

### 5.1 Image Evaluation Experiment I

The goal of this experiment was to assess the performance of AIVAS-IRA algorithm with regard to the extraction of an appropriate image for representing a concrete noun. We evaluated the performance based on a questionnaire survey where feedback from 30 participants was collected and analyzed. Experimental procedures, result analysis, and findings are described in this section.

#### 5.1.1 Approach

For evaluating the performance of AIVAS-IRA for suggesting an appropriate image to represent a concrete noun, first of all, we fixed a subset of concrete nouns. Although concrete nouns, in general, can be represented with a concrete object but generalizing thousands of concrete nouns is difficult. Therefore, we tested our algorithm with a subset of concrete nouns, which was limited to concrete nouns representing animals, fruits, vegetables, compound nouns, and objects. The reason is that novice (K1) EFL (English as Foreign Language) learners in many Asian countries often begin memorizing vocabulary with an English wordbook known as ABC book. An image of an ABC book is shown in Figure 5-1. Those ABC books often contain a picture gallery of words frequently and regularly used in one's native language, following the format of <alphabet, word-image> pair. Words listed in an ABC book often represent animals, fruits, vegetables, compound nouns, and objects. Hence, we decided to limit our subset to nouns that represent animals, fruits, vegetables, compound nouns, and objects.

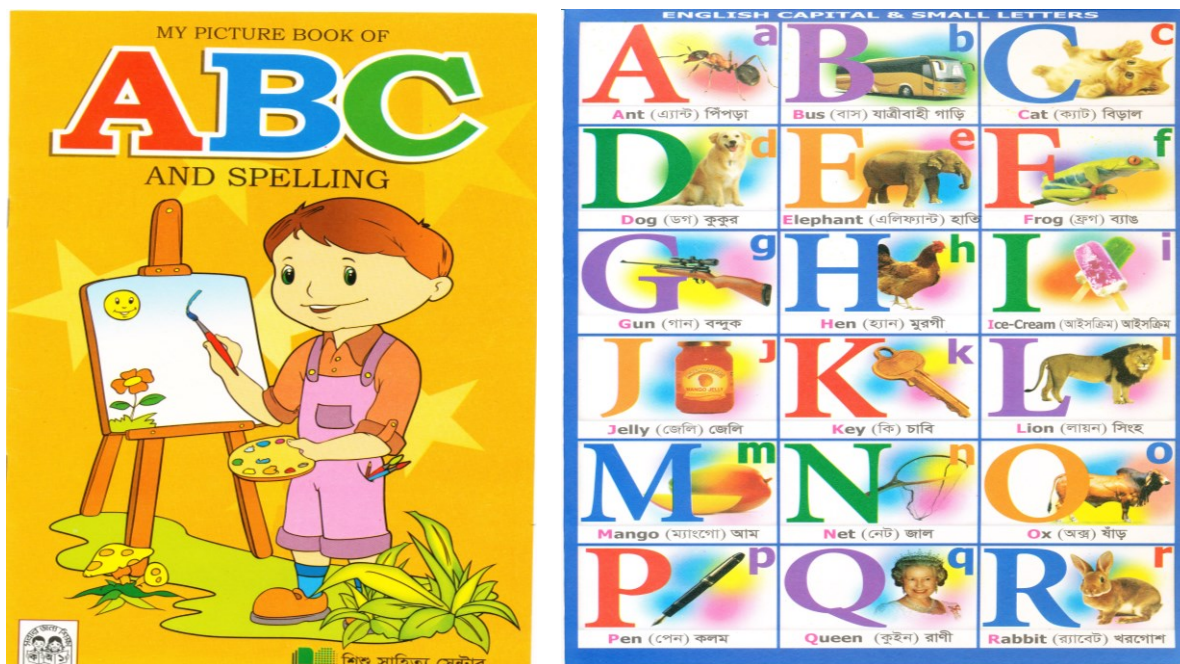


Figure 5-1 An ABC Book Gallery

Next, a list of ten English concrete nouns belonging to the chosen subset was prepared. As the subset consists of 5 different noun categories, we decided to select two nouns for each category. Table 5-1 shows the list of the words used for this experiment.

Table 5-1 Words and Their Corresponding Category

		Word List
Category 1 (Animals' Name)	Word 1	Penguin
	Word 2	Rabbit
Category 2 (Fruits' Name)	Word 3	Grape
	Word 4	Mango
Category 3 (Vegetables' Name)	Word 5	Tomato
	Word 6	Cabbage
Category 4 (Compound Nouns)	Word 7	Policeman
	Word 8	Water-bottle
Category 5 (Objects' Name)	Word 9	Stone
	Word 10	Computer

Then a survey questionnaire was prepared. To prepare the questionnaire, at first, 24 images for each English word were extracted from Google image search engine. The reason for choosing 24 images for every selected word was that we noticed that due to the limitations of the image search engine, irrelevant images might be included if a large number of images are downloaded.



Thirty students (both foreign and Japanese), enrolled in undergraduate and graduate programs, participated in the survey. The students were studying foreign languages. The purpose of the experiment was described above.

Participants were asked to state their preferences by ranking top 10 images out of 24 provided images for every word. The written instruction was,

‘Considering image-based vocabulary learning, state your image preference ranking’

The oral instructions were as follows:

‘if you are learning vocabulary in an image-based vocabulary learning system, which ten images would you like to use? Please rank the images you choose’

There was no fixed time limit for this task. As the participants came from different cultural backgrounds, they were requested to use their imagination in considering image-based vocabulary learning.

### 5.1.2 Result

As stated earlier, participants were requested to rank their top-ten image preferences for every single word. One was the most preferred image while ten was the least preferred one. We assigned eleven to images not chosen by the participants. In analyzing the results, we ranked the set of images using AIVAS-IRA algorithm. In ranking the corresponding images downloaded from Google image search engine, we used our AIVAS-CNCRT59 image set as the source of sample appropriate images. FFT-based features in the power spectrum were extracted. Then we assessed participant’s ratings of the AIVAS-IRA-extracted images, as the key focus of the experiment was to assess appropriateness of images.

Table 5-2 displays the average scores for AIVAS-IRA-suggested images compared with Google image search API-suggested top-ranked images. This comparison is based on calculating the average number of images included in the participants’s top-ten preferences.

The survey results show that our proposed algorithm met the user expectation almost equal or slightly superior to Google for Category 1, Category 2, and Category 3 words. On the other hand, our proposed algorithm was found less effective in fulfilling the user expectation for Category 4 and Category 5 words. Despite not being able to demonstrate superiority over Google images in Category 4 and 5, we decided to proceed with learning effect investigation using the appropriate images suggested by AIVAS-IRA.

**Table 5-2** Result of Image Evaluation Experiment I

Word Name	Average scores of AIVAS-IRA-extracted Appropriate Image	Average scores of Google- suggested Top-ranked Image
Penguin	3.2	9.63
Rabbit	10.1	9.3
Grape	9.27	7.07
Mango	8.01	9.07
Tomato	9.03	6.2
Cabbage	10.07	6.1
Stone	7.07	3.96
Computer	7.83	7.83
Water bottle	4.93	9.87
Policeman	4.83	8.1

### 5.1.3 Discussion

We considered reasons for failure of our proposed algorithm. We noticed that the proposed algorithm is unable to distinguish between a natural image and an illustrated image, and ranked many illustrated images very highly. But when the participants selected their preferences, they paid more attention to natural images instead of illustrated images. Therefore, we suggest that the performance of our algorithm can be increased if a mechanism is implemented to filter the illustrated images and give them less priority compared to natural images. We also noticed that the top-ranked images by Google image search engine for compound nouns and object-names are generally illustrated images. Hence, it remains a challenge to find natural images to represent those types of concrete nouns.

## 5.2 Learning Effect Investigation I

The goal of Learning Effect Investigation I was to evaluate whether AIVAS-IRA-extracted images for representing concrete nouns have any significant learning effect. Google image search engine-suggested top-ranked images were compared with AIVAS-IRA-extracted images for memorizing a set of concrete nouns. We measured learning progress by observing memory retention rates in immediate-after, mid-term and long-term delay conditions by a posttest 1, a delayed posttest, and an extended delayed posttest.

### 5.2.1 Approach

The Learning Effect Investigation I was carried out with the same set of concrete nouns as in the Image Evaluation Experiment I that was described above.

A native Russian speaker was asked to prepare 20 Russian-English word pairs (shown in Table 5-3) that would be appropriate for Russian learners at the introductory level.

**Table 5-3** List of Russian-English Word Pairs

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	книжная полка- bookshelf	наушники-earphone
тул-chair	будильник-clock	дерево-tree	камеры-camera

For each word, two sets of learning material were created: one with the AIVAS-IRA-extracted image, and the other with Google image search engine-suggested top-ranked image. Thus, 40 sets of learning material were created.

In selecting the AIVAS-IRA-extracted images, twenty-four images from the Google image search engine were downloaded and then passed onto AIVAS-IRA for selecting the most appropriate one. We used Google image search engine as the source of images because Google contains over 10-billion images which are much more compared to other image search engines. Moreover, Google has maximum accessibility, provides image details, ensures image protection policies, and so on. Additionally, Google's tweaks image search algorithm and SafeSearch option ensure showing less explicit contents (Fergus, 2005) (Hansell, 2007) (Tene, 2008).

Figure 5-2 shows two sets of learning material for the word 'cucumber'. It is important to remember that Google top-ranked images may change time-to-time.



**орыпец  
cucumber**



**орыпец  
cucumber**

**Figure 5-2** Comparison of Learning Materials

Fifty-two participants, from 21 nationalities and with ages between 17 and 29 (except one who was 39 years old) took part in this investigation. All participants were students enrolled in high-school, college, undergraduate, and graduate-level education.

Familiar representation of an object depends on the cultural background of the participants. Hence, establishing data uniformity is often challenging when collecting data from human participants. Taking this fact into account, the experiment was conducted in a language that is entirely new to the participants. To ensure the participants' lack of knowledge of the Russian language, a pretest questionnaire consisting of 20 preselected Russian words was given to the participants. The participants were asked to mark all the words with which they were familiar. The pretest result indicated that none of the participants were familiar with any word in the list. Therefore, we have set the pretest scores of all the participants to be zero.

Participants were randomly placed into Experimental Group (EG) and Control Group (CG), with each group containing 26 participants. We even tried to balance the distribution of the participants' nationalities across the two groups as shown in Table 5-4.

**Table 5-4** Distribution of Participant Nationalities

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

Participants were informed about the experimental task. EG participants studied the 20 Russian words with the learning material created with the AIVAS-IRA-generated images. On the other hand, CG participants studied the same words with the learning material created with the top-ranked images

from the Google image search.

After the pretest, a study session and three Post-tests were conducted as shown in Figure 5-3.

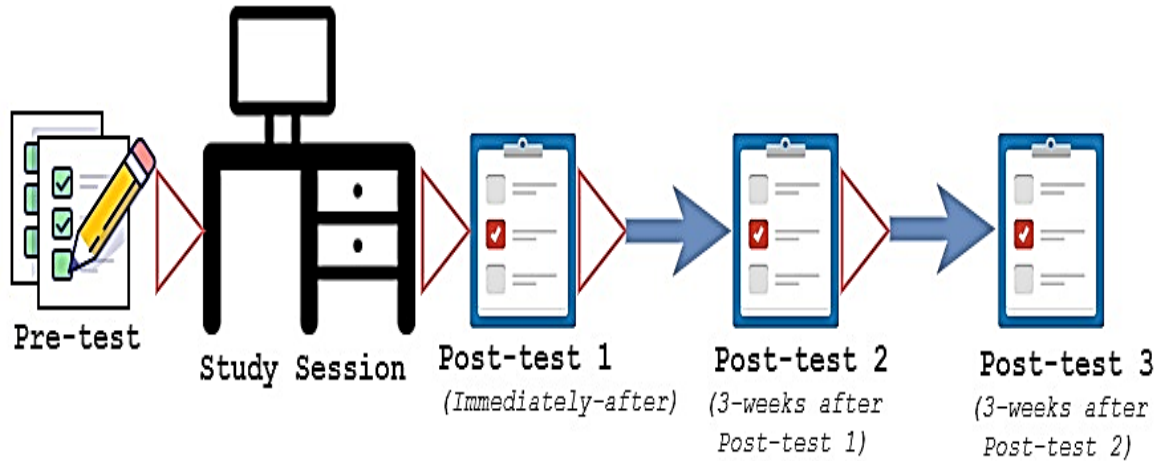


Figure 5-3 Flow of Learning Effect Investigation I

For the EG, AIVAS-EE was used to manage the learning material during the study session. The study session was set to 10 minutes. The study session was equipped with a laptop or a PC with a high-speed internet connection, a headphone, and a mouse. A pen and a sheet of white paper were also provided to support the study session. Participants were given the option to take notes during the 10-minutes study session. However, those notes were not allowed to be used during the posttests. We asked the participants to study the 20 Russian words using the created learning materials.

Immediately after the 10-minutes long study session, the Post-test 1 (PT1) was conducted via a printed questionnaire. The participants were asked to answer the questions within 10 minutes. In the PT1, PT2, and PT3 questionnaires, we listed all the 20 Russian words and asked the participants to write down their meanings in Japanese or English. Participants were asked not to do self-study during the intervals between the posttests.

After three weeks from the study session and PT1, all of the participants were given the second Post-test (PT2). The goal of PT2 was to measure the mid-term memory retention rates of the newly learned words. In PT2, we provided the participants the questionnaire that contained all 20 Russian words, and they were given 10 minutes to answer the questions. We asked the participants to not engage in self-study of those words, and conducted PT3 in a way similar to PT2 after another 3-week interval from PT2.

## 5.2.2 Result

Table 5-5 shows the results of the Learning Effect Investigation I.

**Table 5-5** Result of the Learning Effect Investigation I

	Average score of PT1 (S.D.)	Average score of PT2 (S.D.)	Average score 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)	0.12	0.07	0.02

The statistical analysis revealed no significant difference between the EG and CG in the immediate-after condition (t-test,  $p = 0.12$ ) and in the 1<sup>st</sup> delayed session (t-test,  $p = 0.07$ ). However, a significant difference in the average score by the participants in EG over the participants in CG in the long-term memory retention rates ( $p = 0.02$ ) was observed in the 2<sup>nd</sup> delayed session. Therefore, we conclude that our algorithm is able to extract and recommend images with a higher learning effect with respect to long-term memory retention.

We also report on the average distance from the center point of the 20 images top-ranked by AIVAS-IRA in comparison with the Google top-ranked images, and provide the comparative standard deviation of them. The statistical analysis of t-test revealed significant difference ( $p = 1.54 \times 10^{-7}$ ) between the average distances of the images from the cluster centroid. Table 5-6 shows the result of this analysis.

**Table 5-6** Result of the Analysis

	Distance from the Cluster Centroid(Avg.)	Standard Deviation
Appropriate Image	0.200	0.116
Google Top-ranked Images	0.478	0.148
	$p = 1.54 \times 10^{-7}$	

From this analysis, we conclude that our algorithm is able to extract and recommend better images than the Google image search engine for representing concrete nouns.

### 5.2.3 Discussion

We observed that the female participants in our experiment had a higher memory retention rate than the male participants for both the 1<sup>st</sup> and the 2<sup>nd</sup> delayed post-tests. However, the memory retention rates of both the male and female participants were almost equal in the immediate-after post-test session (PT1). In this experiment, among the EG participants, there were 18 male participants and 8 female participants. On the other hand, in CG, there were 21 male and 5 female participants. Table 5-7 shows the average scores and the standard deviation of the EG participants in their immediate-after, 1<sup>st</sup> and 2<sup>nd</sup> delayed post-tests. Table 5-8 displays the result of a similar analysis for the CG participants. Finally, Table 5-9 compares the post-test1, post-test 2 and post-test 3 results among the male and female participants.

**Table 5-7** Male -vs-Female Participants in Experimental Group

	Posttest 1 Average (S.D.)	Posttest 2 Average (S.D.)	Posttest 3 Average (S.D.)
Male (N=18)	12.44 (4.56)	4.33 (2.42)	3.55 (2.40)
Female (N=8)	14.87 (3.39)	7.5 (2.39)	7.5 (1.88)

**Table 5-8** Male -vs-Female Participants in Control Group

	Posttest 1 Average (S.D.)	Posttest 2 Average (S.D.)	Posttest 3 Average (S.D.)
Male (N=21)	11.80 (4.29)	3.95 (1.90)	3.19 (2.11)
Female (N=5)	9.4 (1.67)	4.4 (2.96)	2.2 (2.28)

**Table 5-9** Male -vs- Female Participants

	Posttest 1 Average	Posttest 2 Average	Posttest 3 Average
Male(N=39)	12.12	4.14	3.37
Female (N=13)	12.13	5.95	4.85

We also collected oral feedback from the participants, which was optional. Generally, we asked open-ended questions about their experience with our system and the experimental methods. As the feedback was optional, not all the participants provided the feedback. In Table 5-10, we provide a list of some significant comments provided by the participants. In addition to these comments, many participants commented that the system is very effective to learn Russian words. Also, many of them commented that the images recommended by our system give a clearer mental picture than the Google images.

Table 5-10 Feedbacks

Comment No	Description of the Comment	No of the Participants
1	Images used in the experimental group looks better than images in control groups (This comment received after posttest 3)	8
7	Three weeks' time interval is too long to recall learned words	5
2	Providing a list of English word in the posttests would have been good. Because they forgot some of the English words they have acquired.	2
3	Some sounds were not clear to hear	1
4	10 minutes study session was short in length	1
5	Prefer memorizing the words with his native translation (not English)	1



### 5.3 Image Evaluation Experiment II

The Image Evaluation Experiment II was conducted to assess the performance of the AIVAS-IRA algorithm for abstract nouns. The purpose of this experiment was to evaluate Approach 1 for representing abstract nouns (described in 3.3.2), namely “an image having physical or concrete existence positioned in the central position in the image frame” will be considered an appropriate image for representing an abstract noun. In this experiment, AIVAS-ABST-LS68 image set was used as sample appropriate images. FFT features in power spectrum were extracted from the images.

#### 5.3.1 Approach

Six English abstract nouns were selected for the purpose of this experiment, which were a subset of the targeted abstract nouns. To briefly recapitulate, our targeted abstract nouns were limited to three types: basic abstract nouns representing social contexts between humans; abstract nouns relating to feeling and/or emotion; and abstract nouns representing social and religious beliefs. Two English abstract nouns were chosen to represent each of those three categories, as shown in Table 5-11.

Table 5-11 List of the Words

Type	Word 1	Word 2
Basic abstract nouns representing social contexts between humans	Recognition	Idea
Abstract nouns relating to feeling and/or emotion	Sadness	Pain
Abstract nouns representing social and religious beliefs.	Hell	Angel

After this, a survey questionnaire was prepared as follows. First, 16 top-ranked images (.jpg formatted images only) for representing each noun were downloaded using Google Image Search APIs. Secondly, the downloaded images were re-ranked using AIVAS-IRA. Note that, when we run the program, we used AIVAS-ABST-LS68 image set as the source of sample appropriate images. We have extracted FFT features in power spectrum from the sample appropriate images.

Finally, a survey questionnaire was prepared for comparing AIVAS-IRA-recommended appropriate images with Google-recommended top-ranked images and with Yahoo UK-recommended top-ranked images for representing abstract nouns. The top-ranked images from both Google and Yahoo UK for the selected abstract nouns were gathered on the same day. The top-ranked images in both Google and Yahoo UK change regularly.

A total of 23 participants, with different cultural background, took part in this experiment. The participants have learned or were actively learning one or more second languages. Table 5-12 gives the details of the participants.

Table 5-12 Participant Details

Participant Id	Nationality/Native Language	Second Language	Affiliation
1	KSA/Arabic	English and Spanish	University of Brighton
2	Venezuela/ Spanish	English, Japanese, Portuguese, Italian	TUAT
3	Kazakhstan/Kazakh	Russian, English, Japanese	University of Waterloo
4	Bangladesh/Bengali	English	University of Brighton
5	Iraq/Arabic	English	University of Brighton
6	British/English	Italian	University of Brighton
7	KSA/Arabic	English	University of Brighton
8	Azerbaijan/Russian	Azerbaijanis, English, Japanese	---
9	Japan/Japanese	English	University of Brighton
10	Japan/Japanese	English	TUAT/University of Cock
11	Iran/Persian	English	University of Brighton
12	Germany/German	Spanish, English, Japanese	Sussex University
13	British/English	Japanese	--
14	South Korea/Korean	Japanese, English	TUAT
15	Japan/Japanese	English	TUAT
16	Vietnam/Vietnamese	English, Japanese	TUAT
17	Japan/Japanese	English	TUAT
18	Japan/Japanese	English	TUAT
19	Japan/Japanese	English	TUAT
20	Japan/Japanese	English	TUAT
21	Japan/Japanese	English	TUAT
22	Japan/Japanese	English	TUAT
23	Japan/Japanese	English	TUAT

During the experiment, both printed and online feedback forms were used. The participants were asked to rate each image on a 5-point scale based on their assessment of two factors as shown below.

First factor (Q1): Appropriateness of the image to the corresponding noun.

Second factor (Q2): Effectiveness of the image for memorizing the corresponding noun.

As each factor was scored on a 1-5 scale, a maximum of 10-points were assigned to each word.

No time restriction was set for answering the survey questionnaire. On average, the participants took about 5-7 minutes to answer the questions.

### 5.3.2 Result

The Kruskal-Wallis test was used for comparing the multiple measurements were taken during the image evaluation experiment. A statistical analysis revealed no significant differences between the AIVAS-IRA versus Google, and AIVAS-IRA versus Yahoo for Q1 ( $p = 0.95$  and  $p = 0.07$ , respectively). Additionally, no significant difference is shown between the AIVAS-IRA versus Google for Q2 ( $p = 0.52$ ). However, a significant difference between the AIVAS-IRA and Yahoo for Q2 ( $p = 0.02$ ; post-hoc analysis by the Steel-Dwass method) was observed. Table 5-13 shows the results of this experiment.

**Table 5-13** Result of the Image Evaluation Experiment II

	Q1			Q2		
	AIVAS-IRA	Google	Yahoo	AIVAS-IRA	Google	Yahoo
Avg	3.33	3.28	2.97	3.07	2.88	2.58
S.D.	1.42	1.50	1.51	1.48	1.54	1.60
Kruskal-Wallis test: $p = 0.06$ Multiple Comparison by the Steel-Dwass Method - AIVAS-IRA versus Google: $p = 0.95$ - AIVAS-IRA versus Yahoo: $p = 0.07$ - Google versus Yahoo: $p = 0.15$				Kruskal-Wallis test: $p = 0.02$ Multiple Comparison by the Steel-Dwass Method - AIVAS-IRA versus Google: $p = 0.52$ - AIVAS-IRA versus Yahoo: $p = 0.02$ - Google versus Yahoo: $p = 0.19$		

### 5.3.3 Discussion

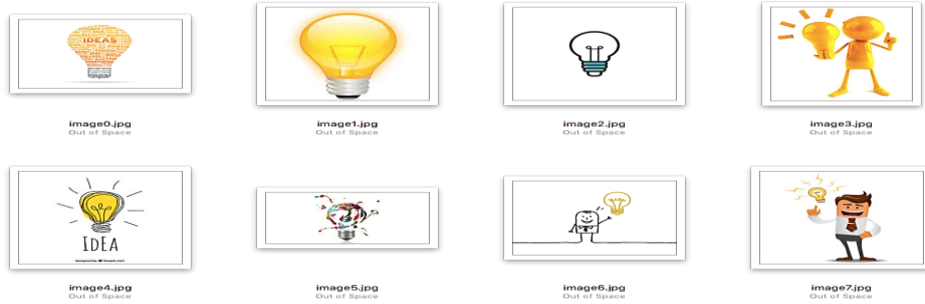
We considered the reasons for the failure of our algorithm to be superior over Google for extracting appropriate images for representing abstract nouns. We feel that the ranking output of the algorithm is not satisfactory for all of the nouns that were selected for the experimental purposes. More precisely, for some abstract nouns, the algorithm failed to meet our hypothesized standard for an appropriate image to picturize an abstract noun. Although for some nouns, the algorithm outputs a satisfactory image, by which we mean that an image that meets our hypothesized definition. Figure 5-4 shows ranked images for the abstract noun ‘idea’, which we consider satisfactory. On the other hand, Figure 5-5 shows ranked images for the abstract noun ‘hell’, which we consider as unsatisfactory. To evaluate these examples, we downloaded a set of 16 images for each of the abstract nouns that are top-ranked by the Google image search engine. Figure 5-4(a) and Figure 5-5(a) show top-8 images that are ranked by Google for the words ‘idea’ and ‘hell’. Figure 5-4(b) and Figure 5-5(b) show the 8 top-ranked images by AIVAS-IRA algorithm. The two images titled as image0.jpg are the appropriate images that are recommended by AIVAS-AIRS system to represent those two abstract nouns.

We also considered that the sample of appropriate images in AIVAS-ABST-LS68 may be a reason for the failure of the Image Evaluation Experiment II, because this sample contains learner-suggested images showing wide variations. Another possible reason for the failure may be the FFT-based feature extraction method in power spectrum. As the 68 sample-appropriate images vary widely, FFT-based feature extraction in power spectrum may not be the most suitable method to apply.

After further considerations of our results, we propose the following steps to improve the performance of our system with respect to abstract images. First, we propose to prepare a rather large image set for abstract nouns containing more sample appropriate images recommended by learners of foreign languages. With this goal in mind we prepared AIVAS-ABST-LS795 image sets. Second, we propose to employ deep CNN-based feature extraction method because the images for representing abstract nouns show a wide variation. Third, we propose to extract not just one appropriate image to represent an abstract noun, but instead follow the method of feature-based categorical recommendation of appropriate images. In this way, a learner will be able to decide the most appropriate image by himself/herself.

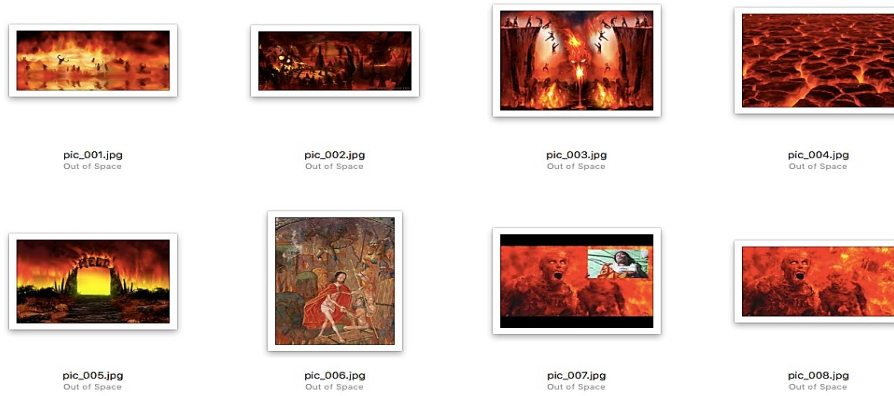


(a) Top-ranked Image in Google Image Search Engine



(b) Top-ranked Appropriate Images by AIVAS-IRA Algorithm

Figure 5-4 Example of a Satisfactory Output



(a) Top-ranked Image in Google Image Search Engine



(b) Top-ranked Appropriate Images by AIVAS-IRA Algorithm

Figure 5-5 Example of an Unsatisfactory Output

## 5.4 Learning Effect Investigation II

Learning effect investigation II was designed to assess the learning efficacy of the appropriate images for memorizing abstract nouns in a completely new language. This experiment was carried out with frequently used abstract nouns in English. In assessing the memory retention rates, top-ranked images suggested by the Google image search engine were compared with the AIVAS-IRA-extracted images.

### 5.4.1 Approach

This experiment was carried out with frequently used abstract nouns of English. The Polish language was chosen as the new language to be taught. To begin with, we asked a native Polish speaker to prepare a list of 20 English-Polish word pairs. We provided a list of 83 frequently used abstract nouns of English (listed in Appendix B) to the native Polish speaker. We asked her to choose a list of 20 words that can be taught to learners at an introductory level. The written instruction was,

*‘Please choose 20 words from the given list that you think a beginner of Polish language should memorize’*

We also asked her to crosscheck the pronunciations of those words. The written instruction was,

*‘After you choose those 20 words, please cross-check their pronunciation at <http://www.voicerss.org/api/demo.aspx>. Please insert the word in the box; then select the language to Polish, and listen to check if the pronunciation is accurate. If you find an inaccurate pronunciation, please replace the word with another one’*

Table 5-14 shows the list of the English-Polish word pairs that were chosen by the native Polish speaker, which were used in our experiment.

**Table 5-14** List of English-Polish Word Pairs

Appearance-Wyglad	Fun-Zabawa	Knowledge-Wiedza	Silence-Cisza
Crime-Przestepstwo	Happiness-Szczescie	Life-Zycie	Thought-Mysl
Death-Smierc	Health-Zdrowie	Love-Milosc	Time-Czas
Discovery-Odkrycie	Hour-Godzina	Month-Miesiac	Trouble-Klopot
Fact-Fakt	Interest-Zainteresowanie	Opinion-Opinia	Truth-Prawda

Next, for each abstract noun in Table 5-14, we determined an appropriate image to represent it as follows. We downloaded top-ranked 24 images from the Google image search engine corresponding to each abstract noun. The only exception was ‘truth’, for which we downloaded 48 images because the top 24 images contained the text ‘truth’ itself in the image. We assumed that an image containing visual text need to be eliminated from the selection of an appropriate image. This is because all the participants understand English, and so an image with visual text may bias the memory retention rate. Note that in preparing the set of corresponding images, we used the Google image search engine in English. Once the set of corresponding images were prepared, we ranked these images using the AIVAS-IRA algorithm, which was implemented for determining an appropriate image for

representing an abstract noun. The AIVAS-ABST-LS795 image set, which contains learner-suggested appropriate images, was used to determine the centroid of each cluster. Although the AIVAS-IRA algorithm implemented with the AIVAS-ABST-LS795 image set contains optimal clusters (discussed earlier in 4.2.4), we considered an image having the least distance from any of the six clusters to be the most appropriate image for representing the corresponding noun. This is because all the six clusters contain samples of appropriate images previously suggested by the learners; hence we assume that an image that is closest to any of the six clusters may count as the appropriate image to represent that particular noun. Figure 5-6 illustrates this approach with an example. Here, we have chosen pic\_003.jpg as the appropriate image to represent this particular noun because pic\_003.jpg has the least distance from any of the six cluster centroids (Cluster 3 in this occasion).

```
C:\Users\Nehal\Desktop\LS795 mED>py mED_final_mod.py
*****FINAL_OUTPUT*****
pic_003.jpg is the closest to Cluster 3 Centroid and its Euclidean distance is 42.174389682604435
pic_017.jpg is the closest to Cluster 5 Centroid and its Euclidean distance is 44.6703196203912
pic_021.jpg is the closest to Cluster 5 Centroid and its Euclidean distance is 45.90430764993964
pic_020.jpg is the closest to Cluster 3 Centroid and its Euclidean distance is 46.92123395506816
pic_002.jpg is the closest to Cluster 5 Centroid and its Euclidean distance is 47.797340370521255
pic_015.jpg is the closest to Cluster 3 Centroid and its Euclidean distance is 47.976555986475
pic_011.jpg is the closest to Cluster 5 Centroid and its Euclidean distance is 49.96442058204756
pic_016.jpg is the closest to Cluster 3 Centroid and its Euclidean distance is 50.772744301592006
pic_018.jpg is the closest to Cluster 5 Centroid and its Euclidean distance is 50.801220818442125
pic_023.jpg is the closest to Cluster 5 Centroid and its Euclidean distance is 51.26502934359308
pic_009.jpg is the closest to Cluster 5 Centroid and its Euclidean distance is 53.10983184999146
pic_022.jpg is the closest to Cluster 5 Centroid and its Euclidean distance is 53.65137801997753
pic_013.jpg is the closest to Cluster 5 Centroid and its Euclidean distance is 54.58302635948633
pic_019.jpg is the closest to Cluster 6 Centroid and its Euclidean distance is 54.83652339856727
pic_001.jpg is the closest to Cluster 3 Centroid and its Euclidean distance is 55.1703522257052
pic_014.jpg is the closest to Cluster 5 Centroid and its Euclidean distance is 55.584561082777064
pic_008.jpg is the closest to Cluster 5 Centroid and its Euclidean distance is 56.25847811742609
pic_006.jpg is the closest to Cluster 5 Centroid and its Euclidean distance is 59.617371765470516
pic_010.jpg is the closest to Cluster 4 Centroid and its Euclidean distance is 59.74363548237135
pic_024.jpg is the closest to Cluster 1 Centroid and its Euclidean distance is 62.649077930144955
pic_004.jpg is the closest to Cluster 5 Centroid and its Euclidean distance is 63.112358663758314
pic_012.jpg is the closest to Cluster 3 Centroid and its Euclidean distance is 63.48834264849886
pic_005.jpg is the closest to Cluster 2 Centroid and its Euclidean distance is 71.80240536646161
pic_007.jpg is the closest to Cluster 4 Centroid and its Euclidean distance is 72.4999596326758
```

Figure 5-6 Determine an Appropriate Image

After this, we used AIVAS-LMC to prepare two sets of learning material for each of the 20 abstract nouns, resulting in 40 sets of learning material. Among these, 20 sets were created with Google-suggested top-ranked images, and the rests with AIVAS-IRA-recommended appropriate images. In selecting a Google top-ranked image and an AIVAS-IRA-recommended image, we eliminated those images that contained visual texts stating the word itself in the image frame. Table 5-15 provides the details of the images used to create the learning material.

Table 5-15 Details on Google Top-ranked-vs-Appropriate Images

Words (Alphabetically)	Google Top-ranked Image			Appropriate Image		
	Ranking Order	Cluster No.	Distance from Centroid	Ranking Order	Cluster No.	Distance from Centroid
Appearance	1	6	55.416140	3	6	49.975620
Crime	3	1	60.388373	1	3	38.651120
Death	1	3	52.870525	1	3	40.725422
Discovery	1	3	54.630746	1	3	38.639247
Fact	16	1	49.074706	6	1	49.074706
Fun	1	2	55.282640	1	3	48.194325
Happiness	1	6	54.998695	1	3	39.671531
Health	2	6	63.264432	1	6	41.610776
Hour	2	2	61.196672	1	5	42.659013
Interest	2	6	52.952017	5	6	49.953851
Knowledge	1	6	61.359059	1	6	42.498769
Life	3	3	43.378054	1	3	43.378054
Love	1	4	68.271992	3	3	48.907341
Month	1	3	76.848540	1	5	49.372736
Opinion	1	4	49.785245	1	4	49.785245
Silence	1	3	54.219771	5	3	45.783707
Thought	5	2	71.802405	19	4	59.743635
Time	1	3	56.209057	1	2	45.106487
Trouble	9	1	72.117991	1	6	43.130687
Truth	25	3	47.507363	2	5	43.523167

Then we developed additional systems for AIVAS-EE to support both native Japanese speakers and non-native Japanese speakers. In particular, we built a system to assist native Japanese speakers in the acquisition of abstract nouns with the help of Japanese language (as the known language). At the same time, we developed another system to assist non-native speakers of the Japanese language to acquire these abstract nouns in English (as the known language). Since the 20 English nouns used in this experiment are frequently-used, we expected that non-native speakers of English would be familiar with these nouns. Figure 5-7(a) and Figure 5-7(b) show two examples of learning material used for native and non-native speakers of Japanese, respectively.



Please Click On A Word to Create A Learning Material

[Appearance](#)  
[Crime](#)  
[Death](#)  
[Discovery](#)  
[Fact](#)  
[Fun](#)  
[Happiness](#)  
[Health](#)  
[Hour](#)  
[Interest](#)  
[Knowledge](#)  
[Life](#)  
[Love](#)  
[Month](#)  
[Opinion](#)  
[Silence](#)  
[Thought](#)  
[Time](#)  
[Trouble](#)  
[Truth](#)



a) Learning Material for Non-native Japanese Speakers

Please Click On A Word to Create A Learning Material

[外観](#)  
[犯罪](#)  
[死](#)  
[発見](#)  
[事実](#)  
[楽しい](#)  
[幸福](#)  
[健康](#)  
[時間](#)  
[興味](#)  
[知識](#)  
[生活](#)  
[愛](#)  
[意見](#)  
[沈黙](#)  
[思想](#)  
[時間](#)  
[トラブル](#)  
[真実](#)



b) Learning Material for Native Japanese Speakers

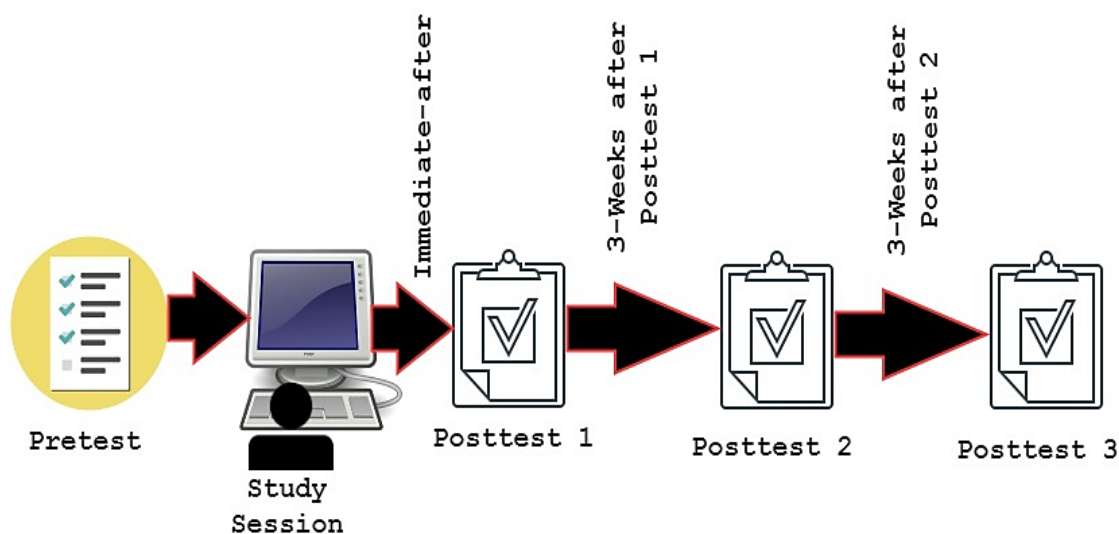
**Figure 5-7** Examples of Learning Material

Twenty-four participants from nine nationalities participated in the experiment. Twenty-two participants were between 21 to 24 years of age, and the remaining two were between 26 to 28 years of age. Participants were enrolled in undergraduate, graduate or exchange programs without having any prior knowledge of the Polish language. We divided them into two groups: an experimental and a control group. Table 5-16 shows the distribution of the participants based on their nationalities among experimental (EG) and control (CG) groups.

**Table 5-16** Distribution of Participant Nationalities

Group	Nationality (Number of Participants)
EG	Japan(6), Malaysia(2), Bangladesh(1), Iran(1), Vietnam(1), and Rwanda(1)
CG	Japan(5), Vietnam(2), Thailand(1), Pakistan(1), Bangladeshi(1), Malaysia(1), and Germany(1)

We designed the experiment with a pretest, a 10-minutes study session followed by the Post-test 1, the Post-test 2 and the Post-test 3. Figure 5-8 shows the flow of the Learning Effect Investigation II experiment.



**Figure 5-8** Flow of Learning Effect Investigation II

In the pretest, a questionnaire containing twenty preselected Polish words was provided to the participants. Participants were asked to mark the words that they know and/or familiar with. The pretest scores confirmed that none of the participants had any knowledge of the Polish language. As a result, the pretest scores of all the participants were set to zero. Based on this result, the participants were randomly assigned into EG and CG groups.

Then we asked the participants to take part in a 10-minutes study session. The participants in the experimental group were asked to study those words with the help of appropriate images. On the other hand, the control group participants were asked to study with the help of Google-suggested top-ranked images. The study session was equipped with a pen, a paper, and a headphone. Participants were allowed to use the pen and paper during the study session. However, they were not allowed to use them in Post-test 1, which was conducted immediately after the study session. Post-test 1 recorded the memory retention rates of the participants immediately after the study session.

The Post-test 2 and Post-test 3 were conducted after a 3-week interval from Post-test 1- and another 3-week interval from Post-test 2, respectively. Our goal in Post-test 2 and Post-test 3 was to record the mid-term and long-term memory retention rates of the participants, respectively.

#### 5.4.2 Result

Although twenty-four participants took part in the experiment, one participant from the Control Group could not take part in the Post-test 3 so his data was discarded from the final analysis. A two-way ANOVA statistical analysis revealed no significant differences ( $p \geq 0.05$ ) in the memory retention rates of the EG participants over CG participants. However, the average scores of the Post-test 2 and Post-test 3 were higher for AIVAS-IRA-suggested images compared to the Google-suggested top-ranked images. Table 5-17 shows the t-test results.

**Table 5-17** Result of the Learning Effect Investigation II

	Average score of PT1 (S.D.)	Average score of PT2 (S.D.)	Average score PT3 (S.D.)
EG (N = 12)	15.16 (3.71)	6.33 (3.60)	6.75 (3.46)
CG (N = 11)	15.90 (4.65)	6.09 (2.73)	6.45 (3.61)
p-value (t-test)	0.34	0.43	0.42

We compared the average distance from their centroids of the 20 top-ranked images suggested by AIVAS-IRA with the 20 Google top-ranked images. The t-test (shown in Table 5-18) revealed a significant difference ( $p \leq .00001$ ) between the average distances of the images from their appropriate cluster centroids.

**Table 5-18** Result of the Analysis

	Distance from the Cluster Centroid (Avg.)	Standard Deviation
Appropriate Image	45.51	5.15
Google Top-ranked Images	58.07	8.82
	$p \leq 1 \times 10^{-5}$	

### 5.4.3 Discussion

Further data analysis was carried out to observe whether male or female participants performed better. The analysis revealed that the male participants performed better in retaining newly learned abstract nouns in the immediate-after, mid-term delay and long-term delay conditions. A comparison of the Learning Effect Investigation II with the Learning Effect Investigation I is carried out with a small group of participants. In this experiment, there were 12 (11 male and 1 female) EG participants, and 11 CG participants (9 male and 2 female). Since the number of female participants was significantly lower than the male participants, we cannot draw any significant conclusion.

Along with the learning data analysis, we also recorded the participants' feedback, which was optional. Table 5-19 shows the comments that we have received in the form of feedbacks.

**Table 5-19** Feedbacks

<b>Feedback No.</b>	<b>Comments</b>
1	Because the orders of the words are together, rather than remembering the word, there may be a side that I remembered.
2	There were several words that I could remember because the order of the words remained same in the posttest.
3	It was really interesting.
4	Words are in alphabetical order, and same order in test too
5	Some of them look quite new. Many of them are familiar with their shape but I can't remember the meaning.
6	It was very exciting and It gave me insight of how I can swiftly memorize new vocabularies in just a short time

## 5.5 Summary

In this chapter, we reported the evaluation experiments that we conducted to assess the efficacy of appropriate images.

The first evaluation experiment, Image Evaluation Experiment I, was aimed at evaluating the appropriateness of images recommended by AIVAS-IRA algorithm for representing a concrete noun. This survey revealed that AIVAS-IRA algorithm proved least effective in recommending appropriate images for concrete nouns that are compound-noun and object-names. However, it performed almost on par or slightly better compared to Google for recommending appropriate images for concrete nouns representing animals, fruits, and vegetables.

The second experiment, Learning Effect Investigation I, was to assess memory retention over immediate-after, mid-term and long-term time intervals. This experiment revealed no significant difference in memory retention in immediate-after and mid-term in the acquisition of the concrete nouns in a new language. However, there was a significant difference noted in long-term memory retention rate. Hence, we conclude that appropriate images recommended by our system can be used for vocabulary learning.

The third experiment, Image Evaluation Experiment II, was designed to assess our hypothetical definition of an appropriate image proposed for a subset of abstract nouns. This experiment revealed that AIVAS-IRA algorithm is able to suggest appropriate images that are accepted as the learning resource to memorize new foreign words compared to Yahoo-suggested top-ranked images. But, it failed to prove superior to Google-suggested top-ranked images.

The fourth experiment, Learning Effect Investigation II, was aimed to assess the efficacy of appropriate images in learning abstract nouns. The experiment failed to show any significant learning superiority of appropriate images over the Google-suggested top-ranked images. The reason for this can be a small number of participants. However, the average mid-term and long-term memory retention rates were higher for AIVAS-IRA suggested appropriate images compared to the Google-suggested top-ranked images. We plan to repeat this experiment with more participants in the future.

## 6. Aspects of Image Appropriateness in Vocabulary Learning

This chapter discusses the appropriateness of still images for vocabulary learning for other parts of speech (such as verbs, adverbs, adjectives etc.).

Linguists have identified five basic aspects of language: Phonology, morphology, syntax, semantics, and pragmatics. These five components are found in all the major languages (Fromkin, 2013), and language learning involves all of them. Phonology refers to the study of speech structure inside a language which includes both the patterns of basic speech units and the accepted rules of pronunciation. In linguistics, morphology refers to the memorization of words, hence the knowledge of the morphology of our language development is critical to vocabulary development and reflects the smallest building blocks for comprehension (Aronoff, 2011). Syntax refers to how individual words and their most basic meaningful units are combined to create sentences, that is, how well a sentence is constructed. In communication, semantics refers to the ways in which a language conveys meaning. Pragmatics refers to the ways the members of the speech community achieve their goals using language. From the definitions of these components, it is clearly noticed that vocabulary acquisition falls under morphology. Research indicates that, effective vocabulary instruction must provide learners with multiple and varied encounters with words (Stahl, 1986). Because of that, the learners should memorize other parts of speech too besides noun. The English language has eight parts of speech: nouns, pronouns, verbs, adjectives, adverbs, conjunctions, prepositions, and interjections). Therefore, mastering the other seven parts of speech along with noun is equally important.

Moving to the main focus of this study, the recommendation of appropriate images for vocabulary learning, we only investigated concrete and abstract nouns. Although other parts of speech are equally important, we could only reveal approach for nouns. It is because we had to consider few challenging facts that are involved in the development of vocabulary skills.

First, the key elements involved in vocabulary acquisition.

Second, the type of images that can be addressed as appropriate to the other parts of speech such as verbs, pronouns, adverbs etc.

Third, the complexities involved in developing systems.

First, we provide some necessary information about our concerns on the key elements involved in gaining significant vocabulary skills. We also mention how difficult finding appropriate image(s) can be for vocabulary learning. First thing to consider is the multidimensionality of a word, which consists of qualitatively different types of understanding. For instance, a noun is generally defined as a person, a place, or a thing, but ideas are also nouns. This multidimensionality of a word poses a challenge to find an appropriate image to represent a word. Second is the homonymy, which refers to words that can have multiple meanings, even when spelled exactly the same (for instance, bear as a noun (the animal) and bear as a verb (to carry a load)). As a result, finding one appropriate image

to represent a homonymic word often becomes a matter of debate. Third is the interrelatedness. Word knowledge is dependent on understanding of other words. Generally, learners must learn that words are not isolated units of meaning. Learners benefit from linking new knowledge to prior knowledge. Therefore, a mastery of previous relationships among concepts facilitates learning new words. Due to this fact, recommending appropriate images is more complex for certain parts of speech such as verbs or adverbs. For example, memorizing verbs often refers to nouns and memorizing adverbs often refers to verbs. Besides, multidimensionality, homonymy, and interrelatedness of words, incrementality and heterogeneity are considered to be the key components of vocabulary acquisition. Given the complexity introduced by all these factors, applying our current approach to these other parts of speech has not made.

Second, linguistics research varies on the type of images that can be considered as appropriate. Consider nouns, for example. Research indicates that still images have a significant learning effect on memorizing nouns, as was discussed in earlier chapters. However, still images may not always work well for verb memorization. Word-learning mechanisms may be different for nouns and verbs (Storkel, 2003). Different types of visual aids have different effects on our cognitive processes. Research indicates that verbs are often memorized effectively with videos and/or moving objects. The most appropriate form of visual aid for a regular verb is considered to be a video clip or a moving object showing the action corresponding to the verb. Research shows that verbs are generally harder to learn than nouns, and that moving objects have more impact on human cognitive processes compared to still images (Gentner, 1982) (Tardif, 1996). Research also suggests that verbs typically represent ephemeral actions, whereas nouns tend to label concrete objects (e.g., car) (Gentner, 1982). Actions are more abstract and fleeting, and are often labeled before or after the action has taken place (Tomasello, 1992). Another factor is that verbs are more polysemous (tend to have multiple meanings) than nouns, which have more restricted meanings. Objects can exist independent of actions, while actions require either an agent or an object. As a result, children, on hearing action labels, are faced with the problem of determining whether the label maps to the object or to the action. Finally, verbs can encode several components of an action, including, but not limited to, path (or the trajectory of agent; e.g. come, approach, enter), manner (or the way in which an agent moves; e.g. walk, dance, swagger, sway, stroll), result (e.g. open, close), and instrument (e.g. hammer, shovel), thereby making the task of finding the referent harder (Pruden, 2004). Due to all these factors, moving objects or video clips are more effective learning aids for memorizing regular verbs over compared to still images. Besides regular verbs (such as walk, talk, sleep etc.), there are irregular verbs like awake, begin, blow etc. Most English learners often find difficulty in memorizing these irregular verbs and be able to conjugate them. Therefore, repetitive exercise using music or games is a commonly adopted approach for teaching irregular verbs in elementary school. Considering all these facts, we suppose our proposed approach for recommending images for representing nouns may not be suitable for recommending images for verbs.

Moving to pronoun memorization, pronouns are used to replace nouns. Pronouns can be tricky to learn quickly because every language has a slightly different way of grouping people. For instance, most European languages have separate words for Singular-You (you) and Plural-You (you all). In English, 'we' groups together four handy pronouns: 'you and me', 'someone else and me', 'several

other people and me’, and ‘you, me, and one or more other people’. Images are not often used for memorizing pronouns. However, many linguists suggest images for learning subject pronouns (I, you, he, she, it, we, and they), because for such pronouns (such as first person singular, first person plural, second person plural etc.) human brain can create images. For other types of pronouns (for example it, this, whose etc.), it is difficult to create images in the human brain. As a result, performing with music is a technique that many teachers use to teach children. Adjectives, adverbs, prepositions and conjugations are very difficult to express with images. For each part of speech, there are different linguistic perspectives that are helpful in learning them. But it has been observed that images are not the most commonly used technique to teach pronouns.

AIVAS system can be configured to recommend appropriate images (in animated or in graphical representation or slow sync flash of a still image) to represent a verb. However, we have to consider the limitation of technologies. We cannot change the fact that each technology has its own limitations especially when we consider converting a still image to a graphical interexchange format (GIF). GIF images are compressed using Lempel-Ziv-Welch (LZW) technique to reduce the file size without degrading the visual quality. Conceptually, a GIF file describes a fixed-sized graphical area (the "logical screen") populated with zero or more images. However, transforming a still image to a GIF image is problematic. Because, GIF is almost never used for true color images, so image decoding is a challenge. Additionally, assigning the right movement of the objects (in an image frame) is extremely challenging.

Considering the above-mentioned facts, we have limited the current study to nouns only. However, in future, we intend to investigate further on the related aspects of the recommendation of appropriate images for the other parts of speech.



## 7. Conclusion

This thesis addresses a key problem in the field of image-based vocabulary learning. We focused on the area of recommending appropriate images for vocabulary. We also contributed to creating self-created vocabulary learning materials, and systems that offer automatic accumulation of multimedia annotations (such as text, audio, video, sound etc.).

In order to solve these problems, we worked on the following research questions:

- 1) Can a definition be proposed to identify an appropriate image for visualizing a noun?
- 2) Can an algorithm be designed to evaluate still images and extract only one appropriate image for representing a concrete noun?
- 3) Can a system be build to extract the most appropriate image and recommend other appropriate images so that the learners can select their own appropriate image?
- 4) Can a system be build to categorize images based on the image features? Considering the diversity involved in images for representing abstract nouns, this might be helpful.
- 5) Can a technical framework be built to extract multimedia annotations automatically from cloud services?
- 6) Can a single system be built to support learning many languages?
- 7) Can a learner learn new words in his/her native language, as well as in an unknown language?
- 8) Collect learning data from global learners and analyze them.
- 9) Adopt adequate evaluation methods for our developments.

Our objective was to provide learners with an effective environment for learning foreign vocabulary with the help of appropriate images together with text, translations and voice data. For this, we built an image recommendation system to determine appropriate images for representing nouns. In this thesis, we first proposed the definition of an appropriate image for representing a concrete noun for assisting foreign language learners in memorizing new vocabulary. Secondly, we worked on an algorithm to extract an appropriate image for a concrete noun and proposed our AIVAS-IRA algorithm to evaluate still images and extract the most appropriate one. Thirdly, we worked on two approaches to extract appropriate images for representing abstract nouns. Here, we first proposed a definition and tested our AIVAS-IRA algorithm to evaluate it. Then, we followed an image feature-based categorization approach to recommend images for representing abstract nouns. These systems were implemented and evaluated. The image recommendation system not only decides the most appropriate image, but also let a learner choose his/her own preferred appropriate image. In implementing these systems, three image sets (AIVAS-CNCRT59, AIVAS-ABST-LS68, and AIVAS-ABST-LS795) were created. Another image set named AIVAS-ABST8300 is created for the future research.

We utilized the modern cloud services (the third-party API services) to build our AIVAS system. Our system creates learning material for the words that a learner wants to acquire. AIVAS-LMC-generated learning materials are 5-second long. AIVAS automatically generates five-second learning

material upon receiving a text-based query that contains the spelling, the meaning, the pronunciation data of the word along with an appropriate image. In order to create the learning material using AIVAS-LMC, a learner only needs to specify four simple fields (Field 1, where the word to be learned needs to be input; Field 2, where a learner needs to specify whether the input word is his/her spoken language or not; Field 3, where a learner needs to specify the spoken language manually; and Field 4, where to specify the target language), followed by clicking on the 'create' button. As a result, operating the system is easy for people with/without adequate IT skills.

This thesis is based on the assessment of 6 experiments (2 pedagogical investigations, 2 image evaluation surveys, and 2 learning effect investigations).

The Pedagogical Investigations I and II were preliminary investigations carried out to define an appropriate image for representing a concrete noun. The pedagogical investigation I assessed whether or not different types of images play an important role in recalling a learned vocabulary in a new language. Bengali was the new language that we experimented with 20 participants. We compared the role of an image that contains just one object on a white background with an image that contains multiple objects in the background/foreground. Mann-Whitney's U-test revealed no significant difference ( $U = 35$ ,  $p = 0.07$ ) in the memory retention rates of the learners. From this, we conclude that background objects in the image frame of still images do not play a significant role in memorizing foreign vocabulary. Then, we proceeded to pedagogical investigation II, which investigated the object representation in an image frame. Based on the feedback from 26 participants, we found that the best placement of the highlighted object is in the center of the image frame (ANOVA ( $F_{9,250} = 170.1$ ,  $p < 0.01$ ), and the post Steel-Dwass multiple comparison tests outperformed other 9 types of object representations used.

Image Evaluation Experiments I and II were conducted to evaluate the performance of AIVAS-IRA algorithms with regards to the extraction of an appropriate image for representing a concrete and an abstract noun, respectively. Image Evaluation Experiment I was carried out with a set of concrete nouns limited to 5 categories, namely animal (Category 1), fruit (Category 2), vegetable (Category 3), compound noun (Category 4), and object names (Category 5). Thirty participants took part in this experiment. Results show that our algorithms failed to perform for nouns in Category 4 and Category 5. However, they matched or exceeded Google search results for nouns in Category 1, Category 2, and Category 3. Image Evaluation Experiment II was conducted with 23 participants to evaluate the performance of AIVAS-IRA algorithm for choosing an appropriate image for representing an abstract nouns that represent 1) social contexts between humans, 2) feeling and/or emotion, and 3) our social and religious beliefs. The result revealed that AIVAS-IRA-extracted images did not perform well ( $p \geq 0.05$ ) compared to Google search top-ranked images. However, it significantly ( $p = 0.02$ ) outperformed Yahoo search top-ranked images. Nevertheless, our algorithm failed to demonstrate significant performance over Yahoo and Google in the appropriateness of representing abstract nouns. From these failures, we concluded that we need to follow an alternative approach to recommend images for abstract nouns. We tried another approach, where image feature-based recommendation method for 83 frequently used abstract nouns was followed.

Learning Effect Investigations (I & II) looked at the efficacy of our system-recommended images in memorizing new words. Learning Effect Investigation I assessed the learning efficacy of the appropriate images in acquisition of concrete nouns by measuring the memory retention of participants after a considerable time delay. Fifty-two participants from different cultural background participated in this experiment. With a posttest, a delayed posttest, and an extended delayed posttest, we measured memory retention rates of the learners. Between the delayed and the extended delayed posttests there was a 3-week interval. The results did not find any significant differences in the immediate-after and the mid-term memory retention rates of the learners. However, a significant difference in the memory retention rates of the learners was observed in the extended delayed posttest ( $p = 0.02$ ) conducted 6-weeks after the study session. In the Learning Effect Investigation II, we conducted a posttest, a delayed posttest and an extended delayed posttest to measure the memory retention rates for newly acquired abstract nouns in the immediate-after, mid-and long-term memory of the learners. Twenty-four participants initially took part in this experiment. However, one participant was absent during the posttest 3, so that data was discarded in the final analysis. Based on an analysis of the data from the remaining 23 participants, no significant difference was observed between mid-term ( $p \geq 0.05$ ) and long-term ( $p \geq 0.05$ ) memory retention.

Two surveys were conducted to prepare our AIVAS-ABST-LS68 and AIVAS-ABST-LS795 image sets. In the first survey, a total of 73 images representing 14 English abstract nouns were collected as a sample of appropriate images chosen by 6 participants. However, 5 out of 73 images were discarded due to format mismatch. In the second survey, there were 24 international participants. We collected a sample of 795 appropriate images for representing 83 frequently used English abstract nouns.

This study concludes that AIVAS-IRA algorithms are able to determine the most appropriate image for representing a concrete noun and recommend it to the learners. These images can be considered as appropriate learning resources to acquire concrete nouns of a new language. Moreover, the proposed system AIVAS has proved an effective way to learn foreign vocabulary. Although our intention was to support learners in informal learning, we believe our system can be used in a classroom environment by both learners and instructors. Furthermore, instructors, in the creation of vocabulary learning material, can use this system to determine an appropriate image to represent a noun. We believe this contribution is able to recommend educational images for both concrete and abstract nouns that will help foreign language learners in quick memorization and long-term memory retention.

## 8. Limitations and Future Directions

In this chapter, we point out some limitations of this study along with directions to overcome them in the future. Also, we discuss some new research directions that this study may lead in the future.

First, the system is unable to distinguish between concrete and abstract nouns. Hence, a manual operation through the command-line interface is necessary to determine the appropriate images for nouns. To overcome this, a database containing the noun categories will be added in the future.

Second, the system is blind to polysemic words. To deal with polysemic words, additional functions need to be implemented.

Third, in the implementation of AIVAS-IRA algorithm for concrete nouns, we only counted FFT features in the power spectrum of gray-scaled images. That is, we only refer to the power spectrum that counts the level of noise existing in each image pixel as the periodogram. To make the algorithm more proficient, we need to investigate other parameters such as object features and image color distributions. In future, we plan to adapt neural network-based solutions to make the algorithm perform better.

Fourth, when we implemented the algorithm for abstract nouns, we used pre-trained Alexnet architecture. Although Alexnet is a widely accepted neural network, it has several limitations. Hence, other neural networks will be applied in the future to compare the extraction of learning features and compare the performance of the algorithm. In addition, we analyzed the features derived from fully connected 7 (FC7) layer assuming that the features in the final layer will be more precise and accurate. However, it may not always be true for every single image. Visualization of the image features in other layers are important. Hence, in future, we plan to count features extracted from other layers and perform a comparative analysis. In this way, we will be able to determine the best features to use for our algorithm.

Fifth, this thesis does not report on the performance of our algorithm for AIVAS-ABST8300 image set. A comparative analysis of the algorithmic performance of AIVAS-ABST-LS795 and AIVAS-ABST8300 image sets is planned for future studies. We are also considering other deep neural networks such as R-CNN or Faster R-CNN to our image sets to determine which deep architecture suits most of our data.

Sixth, at present the algorithm is unable to detect instances of real-world objects (such as face, fruits, buildings, animals etc.) from an image frame. We believe the performance of AIVAS-IRA algorithms will be better if the algorithms can detect actual objects from the image-frame. By detecting the instances of real objects, the irrelevancies from the corresponding image set can be reduced. In future, we plan to implement methods such as Tensorflow object detection API, deep learning for object detection, and so on.

Finally, the current prototype supports only 11 languages. We plan to implement new functions so that the system lets learners learn over 140 languages. Moreover, three features will be implemented shortly: a geographical location database containing the location information of the learner; an independent database for multimedia annotations; and a neural network-based intelligent dashboard. We also plan to make the system context aware in the future.

## Appendices

### Appendix A: The List of Abstract Nouns

Alphabetic Order	Abstract Nouns
A	Abasement, Abdication, Abduction, Aberration, Ability, Adage, Advantage, Adversity, Advice, Affection, Afterlife, Agility, Agony, Agreement, Allegory, Amazement, Amount, Amour, Anger, Animosity, Anxiety, Appearance, Aptitude, Array, Atmosphere, Attitude, and Attribute
B	Banality, Belief, Bereavement, Betrayal, Blandness, Blasphemy, Blessing, Boredom, Bravery, and Brutality
C	Capacity, Causality, Centennial, Chance, Chaos, Charm, Cleanness, Clemency, Comedy, Comparison, Competence, Competition, Comradeship, Concept, Confidence, Conquest, Context, Contribution, Cooperation, Cost, Courtship, Creator, Crime, Crisis, Criterion, and Custom
D	Dalliance, Death, Debacle, Deceit, Deduction, Delirium, Democracy, Demon, Destruction, Determination, Development, Devil, Devotion, Diffusion, Direction, Discipline, Disclosure, Disconnection, Discovery, Discretion, Disparity, Disposition, Distinction, Distraction, Drama, Dream, and Duty
E	Eccentricity, Economy, Effort, Ego, Elaboration, Emancipation, Embezzlement, Emergency, Encore, Engagement, Enterprise, Episode, Equity, Essence, Event, Exactitude, Exclusion, Excuse, Exertion, Exhaustion, Explanation, Expression, and Extermination
F	Facility, Fact, Fallacy, Fantasy, Fate, Fault, Feudalism, Figment, Flexibility, Foible, Footwear, Forethought, Formation, Fortune, Franchise, Freedom, Fun, and Functionary
G	Gaiety, Gender, Genius, Ghost, Gist, Glory, Goddess, Gratitude, Gravity, Greed, and Grief
H	Hankering, Happiness, Hardship, Hatred, Health, Hearing, Heaven, Heredity, Heroism, Hierarchy, Hindrance, Hint, History, Homicide, Honor, Hope, Hospitality, Hour, Humor, and Hypothesis
I	Idea, Idiom, Ignorance, Illusion, Immunity, Impact, Impotency, Impropriety, Impulse, Inanity, Incident, Inclemency, Increment, Inducement, Inebriety, Ingratitude, Insolence, Intellect, Interest, Interim, Intimate, and Investigation
J	Irony, Jealousy, Jeopardy, Joke, Joviality, Joy, and Justice
K	Kindness and Knowledge
L	Law, Length, Life, Limelight, Loquacity, Love, and Loyalty
M	Madness, Magnitude, Majority, Malady, Malice, Management, Marriage, Mastery, Memory, Menace, Mercy, Method, Mind, Miracle, Mirage,

	Mischief, Misconception, Misery, Moment, Month, Mood, Moral, and Multiplication
<b>N</b>	Namesake, Necessity, Nonsense, and Northwest
<b>O</b>	Obedience, Obsession, Occasion, Onslaught, Opinion, Opportunity, Origin, Outcome, and Ownership
<b>P</b>	Pacifism, Pact, Panic, Passion, Pep, Perception, Perjury, Phantom, Pleasure, Pledge, Position, Poverty, Power, Prayer, Predicament, Present, Pressure, Prestige, Pride, Profession, Promotion, Prosperity, and Proxy
<b>Q</b>	Quality, Quantity, and Quest
<b>R</b>	Rating, Reaction, Recognition, Reflection, Reminder, Replacement, Research, and Ritual
<b>S</b>	Sadness, Safety, Satire, Savant, Science, Semester, Sensation, Sentiment, Series, Session, Shame, Shock, Silence, Simile, Situation, Sobriety, Soul, Spirit, Spree, Strength, Style, Subtraction, Supplication, and Suppression
<b>T</b>	Temerity, Tendency, Theory, Thought, Time, Tragedy, Tribute, Trouble, Truce, and Truth
<b>U</b>	Unbelievable, Unification, Unreality, and Upkeep
<b>V</b>	Vacuum, Vanity, Velocity, Victory, Vigilance, Vigor, Violation, Virtue, Vision, and Vocation
<b>W</b>	Warmth, Welfare, and Wistfulness

**Appendix B:** List of the 83 Frequently-used Abstract Nouns

Name	Frequency	Name	Frequency	Name	Frequency
Advice	A	Happiness	A	Passion	A
Advantage	A	Health	AA	Pleasure	AA
Amount	AA	Heaven	AA	Position	AA
Anger	A	History	AA	Power	AA
Appearance	A	Honor	AA	Prayer	A
Attitude	A	Hope	AA	Present	AA
Chance	AA	Hour	AA	Pride	A
Confidence	A	Idea	AA	Quality	A
Cost	AA	Interest	AA	Quantity	A
Crime	A	Joy	AA	Safety	A
Custom	AA	Justice	A	Science	A
Death	AA	Knowledge	AA	Series	A
Devil	A	Law	AA	Shame	A
Direction	AA	Length	AA	Shock	A
Discovery	A	Life	AA	Silence	AA
Dream	AA	Love	AA	Situation	A
Duty	AA	Majority	A	Soul	AA
Effort	AA	Marriage	A	Spirit	AA
Event	A	Memory	A	Strength	AA
Excuse	A	Method	AA	Style	A
Expression	A	Mind	AA	Theory	A
Fact	AA	Moment	AA	Thought	AA
Fate	A	Month	AA	Time	AA
Fault	A	Moral	A	Trouble	AA
Fortune	A	Necessity	A	Truth	AA
Freedom	A	Occasion	A	Victory	A
Fun	A	Opinion	AA	Virtue	A
Glory	A	Opportunity	A		



## Bibliography

- Agca, R. K., & Özdemir, S. (2013). Foreign language vocabulary learning with mobile technologies. *Procedia-Social and Behavioral Sciences*, Vol. 83, No. 4, pp.781-785.
- Al Seghayer, K. (2001). The effect of multimedia annotation modes on L2 vocabulary acquisition: A comparative study. *Language Learning & Technology*, Vol. 5, No. 1, pp.202-232.
- Allen, V. F. (1983). *Techniques in teaching vocabulary*. New York: Oxford University Press.
- Alliance, V. T. (2014). *Professional development for primary, secondary & university educators/administrators*. Retrieved from Professional Development for Primary, Secondary & University Educators/Administrators
- Alliney, S., & Morandi, C. (1986). Digital image registration using projections. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Volume PAMI-8, Issue 2, pp.222-233.
- Amemiya, S., H. Hasegawa, K., Kaneko, K., Miyakoda, H., & Tsukahara, W. (2007). Long-term memory of foreign-word learning by short movies for iPods. *ICALT*, pp. 561-563.
- Andrews, E. (1990). *Markedness theory*. Duke University Press.
- Anonathanasap, O. He., C., Takashima, K., Leelanupab, T., & Kitamura, Y. (2014). Mnemonic-based interactive interface for second-language vocabulary learning. *Proceedings of the Human Interface Society'14*.
- Apicella, A. N., Nagel, J. H., & Duara, R. (1988). Fast multimodality image matching. *In Engineering in Medicine and Biology Society, Proceedings of the Annual International Conference of the IEEE*, Vol.1092, pp.414-415.
- Aronoff, M., & Fudeman, K. (2011). What is morphology? (Vol. 8). John Wiley & Sons.
- Atkinson, R. C., & Raugh, M. R. (1975). An application of the mnemonic keyword method to the acquisition of a Russian vocabulary. *Journal of Experimental Psychology: Human learning and memory*, Vol.1, No.2, pp. 126-133.
- Aw, G. P., Wong, L. H., Zhang, X., Li, Y., & Quek, G. H. (2016). MyCLOUD: A seamless Chinese vocabulary-learning experience mediated by cloud and mobile technologies. *In Future Learning in Primary Schools*, pp.65-78.
- Baddeley, A. D. (1997). *Human memory: theory and practice*. Psychology Press.
- Bao, L., & Intille, S. (2004). Activity recognition from user-annotated acceleration data. *Pervasive Computing*, Second International Conference, PERVASIE 2004, pp.1-17.
- Barcroft, J. (2009). Strategies and performance in intentional L2 vocabulary learning. *Language Awareness*, Vol.18, No.1, pp.74-89.
- Barnea, D. I., & Silverman, H. F (1972). A class of algorithms for fast digital image registration. *IEEE Transactions on Computers*, Vol.100, No.2, pp.179-186.
- Bay, H. E., Ess, A., Tuytelaars, T., & Van Gool, L. (2008). Speeded-up robust features (SURF). *Computer Vision and Image Understanding*, Vol.110, No.3, pp.346-359.
- Bendat, J. S., & Piersol, A. G. (2011). *Random data: analysis and measurement procedures*. Vol. 729, John Wiley & Sons.
- Ben-Haim, N. B., Babenko, B., & Belongie, S. (2006). Improving web-based image search via content based clustering. *In Computer Vision and Pattern Recognition Workshop (IEEE)*, 6pages.

- Bereiter, C., & Scardamalia, M. (1989). Intentional learning as a goal of instruction. *Knowing, learning, and instruction: Essays in honor of Robert Glaser*, pp.361-392.
- Bird, H. F. (2001). Age of acquisition and imageability ratings for a large set of words, including verbs and function words. *Behavior Research Methods*, Vol.33, No.1, pp.73-79.
- Brown, L. G., Franklin, S., & Howard, D. (1992). A survey of image registration techniques. *ACM Computing Surveys (CSUR)*, 325-376.
- Burmark, L. (2002). *Visual literacy: Learn to see, see to learn*. Association for Supervision and Curriculum Development.
- Cabrera, J. S., Frutos, H. M., Stoica, A. G., Avouris, N., Dimitriadis, Y., Fiotakis, G., & Liveri, K. D. (2005). Mystery in the museum: collaborative learning activities using handheld devices. *In Proceedings of the 7th international conference on Human Computer Interaction with Mobile Devices & Services*, ACM, pp. 315-318.
- Carlson, S. (2002). Are personal digital assistants the next must-have tool? *Chronicle of Higher Education*, Vol.49, No.7, pp.A33-A33.
- Carpenter, S. K., & Olson, K. M. (2012). 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.
- Cavus, N., & Ibrahim, D. (2009). 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.
- Chan, T. W., Roschelle, J., Hsi, S., Kinshuk, Sharples, M., Brown, T., & Soloway, E. (2006). One-to-one technology-enhanced learning: An opportunity for global research collaboration. *Research and Practice in Technology Enhanced Learning*, Vol.1, No.1, pp.3-29.
- Chang, C. Y., Sheu, J. P., & Chan, T. W. (2003). Concept and design of ad hoc and mobile classrooms. *Journal of Computer Assisted Learning*, Vol.19, No.3, pp.336-346.
- Choo, L. B., Lin, D. T. A., & Pandian, A. (2012). Language learning approaches: A review of research on explicit and implicit learning in vocabulary acquisition. *Procedia-Social and Behavioral Sciences*, Vol.55, pp.852-860.
- Chrome, G. (n.d.). <https://chrome.google.com/webstore/detail/fatkun-batch-download-ima/nnjahlikiabnchcpehckdeckfgnohf?hl=en>. Retrieved from <https://chrome.google.com/webstore/detail/fatkun-batch-download-ima/nnjahlikiabnchcpehckdeckfgnohf?hl=en>
- Chun, D. M., & Plass, J. L. (1996). Effects of multimedia annotations on vocabulary acquisition. *The Modern Language Journal*, Vol.80, No.2, pp.183-198.
- Coady, J., & Huckin, T. (1997). *Second language vocabulary acquisition: A rationale for pedagogy*. Cambridge University Press.
- Colligan, L. P., Potts, H. W., Finn, C. T., & Sinkin, R. A. (2015). Cognitive workload changes for nurses transitioning from a legacy system with paper documentation to a commercial electronic health record. *International Journal of Medical Informatics*, Vol.84, No. 7, pp. 469-476.
- Dalal, N., & Triggs, B. (2005). Histograms of oriented gradients for human detection. *In Computer Vision and Pattern Recognition*, Vol.1, pp.886-893.
- Deshpande, A. (2016). *A beginner's guide to understanding convolutional neural networks*. Los Angeles: UCLA.

- Deshpande, A. (2016). *The 9 deep learning papers you need to know about*. Retrieved from <https://adeshpande3.github.io/>: <https://adeshpande3.github.io/adeshpande3.github.io/The-9-Deep-Learning-Papers-You-Need-To-Know-About.html>
- Dewar, G. (2014). *The cognitive benefits of play: Effects on the learning brain*. Retrieved from [parentingscience.com: http://www.parentingscience.com/benefits-of-play.html](http://www.parentingscience.com/benefits-of-play.html)
- El-Bishouty, M. M., Ogata, H., & Yano, Y. (2007). PERKAM: Personalized knowledge awareness map for computer supported ubiquitous learning. *Journal of Educational Technology & Society*, Vol.10, No.3, pp.122-134.
- Ellis, N. C. (1994). *Implicit and explicit learning of languages*, pp.79-114, philpapers.org.
- Fahy, K. (1993). *Fast fourier transforms and power spectra in LabVIEW®*. Citeseer.
- Fromkin, V., Rodman, R., & Hyams, N. (2013). *An introduction to language*. Cengage Learning.
- Fergus, R. F.-F. (2005). Learning object categories from Google's image search. Tenth IEEE International Conference on Computer Vision: IEEE. Vol.2, pp. 1816-1823.
- G., M. J. (2014). *Lab 9: FFT and power spectra*, Keck Science Department
- Gentner, D. (1982). *Why nouns are learned before verbs: Linguistic relativity versus natural partitioning*. Center for the Study of Reading Technical Report, No.257.
- Gorjian, B. (2012). Teaching vocabulary through web-based language learning (WBLL) approach. *Procedia Technology*, Vol.1, pp.334-339.
- Gutierrez, K. (2014). *Studies confirm the power of visuals in eLearning*. Retrieved from <http://info.shiftelearning.com/blog/bid/350326/studies-confirm-the-power-of-visuals-in-elearning>
- Hamamrad, A. (2016). Integrating CALL into language teaching: Implementing WBLL technique to teach English language to EFL learners in a secondary school in Kurdistan region. *Journal of Education and Practice*, Vol.7, No.1, pp.38-47.
- Hansell, S. (2007). *Google keeps tweaking its search engine*. New York Times.
- Hartigan, J. A. & Wong, M. A. (1979). Algorithm AS 136: A k-means clustering algorithm. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, Vol.28, No.1, pp.100-108.
- Hasegawa, K., Amemiya, S., Kaneko, K. I., Miyakoda, H., & Tsukahara, W. (2007). MultiPod: A multi-linguistic word learning system based on iPods. *In Proceedings of the Second International Conference on Task-Based Language Teaching*.
- Hasegawa, K., Amemiya, S., Ishikawa, M., Kaneko, K., Miyakoda, H., & Tsukahara, W. (2007). 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.
- Hashemi, Z., & Hadavi, M. (2015). Investigation of vocabulary learning strategies among EFL Iranian medical sciences students. *Procedia-Social and Behavioral Sciences*, Vol.192, pp.629-637.
- Hasnine, M. N., Hirai, Y., Ishikawa, M., Miyakoda, H., & Kaneko, K. (2014). A vocabulary learning system by on-demand creation of multi-linguistic materials based on appropriate images. *Proceedings of the 2014 International Conference on e-Commerce, e-Administration, e-Society, e-Education, and e-Technology*, pp. 343-356.
- Hasnine, M. N., Hirai, Y., Kaneko, K., Ishikawa, M., & Miyakoda, H. (2015). Learning effects investigation of an on-demand vocabulary learning materials creation system based on appropriate images. *Proceedings of the 2015 International Conference on 4th ICT-ISPC*.

- Hasnine, M. N., Hirai, Y., Ishikawa, M., Miyakoda, H., Kaneko, K. & Pemberton, L. (2016). An image recommender system that suggests appropriate images in creation of self-learning items for abstract nouns. *23rd International Conference on Education and E-Learning (ICEEL)*, Oxford, Vol.2, No.5, pp. 8-14.
- Hasnine, M. N., Ishikawa, M., Hirai, Y., Miyakoda, H., & Kaneko, K. (2017). An algorithm to evaluate appropriateness of still images for learning concrete nouns of a new foreign language. *IEICE Transactions on Information and Systems*, Vol.100, No.9, pp.2156-2164.
- Hatch, E. & Brown, C. (1995). *Vocabulary, semantics, and language education*. New York: Cambridge University Press.
- Hayati, A. J., Jalilifar, A., & Mashhadi, A. (2013). Using short message service (SMS) to teach English idioms to EFL students. *British Journal of Educational Technology*, Vol.44, No.1, pp.66-81.
- He, K. Z., X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. *In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp.770-778.
- Herron, C. A., Hanley, J. E., & Cole, S. P. (1995). A comparison study of two advance organizers for introducing beginning foreign language students to video. *The Modern Language Journal*, Vol.79, no.3, pp.387-395.
- Hirschel, R. & Fritz, E. (2013). Learning vocabulary: CALL program versus vocabulary notebook. *System*, Vol.41, No.3, pp.639-653.
- Holzinger, A. N., Nischelwitzer, A., & Meisenberger, M. (2005). Mobile phones as a challenge for m-learning: examples for mobile interactive learning objects (MILOs). *In Pervasive Computing and Communications Workshops, 2005. PerCom 2005 Workshops*, IEEE, pp. 307-311.
- Horner, J. L., & Gianino, P. D. (1984). Phase-only matched filtering. *Applied Optics*, Vol.23, No.6, pp.812-816.
- Huang, Y. M., Huang, Y. M., Huang, S. H., & Lin, Y. T. (2012). 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.
- Huckin, T., & Coady, J. (1999). Incidental vocabulary acquisition in a second language: A review. *Studies in Second Language Acquisition*, Vol.21, No.2, pp.181-19.
- Hudson, T. (1982). The effects of induced schemata on the short circuit in L2 reading: non-decoding factors in L2 reading performance. *Language Learning*, Vol.32, No.1, pp.1-33.
- Hwang, G. J., Chin-Chung, T., & Yang, S. J. (2008). Criteria, strategies and research issues of context-aware ubiquitous learning. *Journal of Educational Technology & Society*, Vol.11, No.2, pp.81-91.
- ImageNet. (n.d.). *ImageNet dataset*. Retrieved from <http://www.image-net.org>
- Ishikawa, M. K., Kaneko, K., Miyakoda, H., & Tsukahara, W. (2007). Design and implementation of a database system for foreign-word learning materials by iPods. *In Information Technology Interfaces*, IEEE, pp. 351-356.
- Jain, V., & Varma, M. (2011). Learning to re-rank: query-dependent image re-ranking using click data. *Proceedings of the 20th International Conference on World Wide Web*, ACM, pp. 277-286.

- Joseph, S. B., Binsted, K., & Suthers, D. (2005). PhotoStudy: Vocabulary learning and collaboration on fixed & mobile devices. *In Wireless and Mobile Technologies in Education*, IEEE, pp.1-5.
- Kalyuga, M., Mantai, L., & Marrone, M. (2013). Efficient vocabulary learning through online activities. *Procedia-Social and Behavioral Sciences*, Vol.83, pp.35-38.
- Kaneko, K., Hasegawa, H., Amemiya, S., Ishikawa, M., Miyakoda, H., & Tsukahara, W. (2007). Multi-linguistic learning materials for vocabulary acquirement based on universal images. *In Proc. Sixth International Internet Education Conference*.
- Karpathy, A. (2017). *CS2321n Convolutional neural network for visual recognition*. Retrieved from <http://cs231n.github.io/>: <http://cs231n.github.io/convolutional-networks/>
- Katz, Y. J. (2013). *SMS-based learning in tertiary education: Achievement and attitudinal outcomes*. International Association for Development of the Information Society, pp.118-125.
- Kellogg, G. S. & Howe, M. J.(1971). Using words and pictures in foreign language learning. *Alberta Journal of Educational Research*, Vol.17, No.2, pp.89-94.
- Khokhlova, N. (2014). Understanding of abstract nouns in linguistic disciplines. *Procedia-Social and Behavioral Sciences*, Vol.136, pp.8-11.
- Kintsch, W. (1972). Abstract nouns: Imagery versus lexical complexity. *Journal of Verbal Learning and Verbal Behavior*, Vol.11, No.1, pp.59-65.
- Kodinariya, T. M., & Makwana, P. R. (2013). Review on determining number of cluster in K-Means clustering. *International Journal of Advance Research in Computer Science and Management Studies*, Vol.1, No.6, pp.90-95.
- Kritikou, Y. P., Paradia, M., & Demestichas, P. (2014). Cognitive web-based vocabulary learning system: The results of a pilot test of learning Greek as a second or foreign language. *Procedia-Social and Behavioral Sciences*, Vol.141, pp.1339-1345.
- Krizhevsky, A. S., Sutskever, I., & Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. *In Advances in Neural Information Processing Systems*, pp.1097-1105.
- Kruse, K. (2005). *Introduction to instructional design and the ADDIE model*. docshare01.docshare.tips.
- Kuglin, C. D. (1975). The phase correlation image alignment method. *In Proc. International Conference on Cyber-netics Society*, pp. 163-165.
- Lam, W. Y., & Renandya, W. A. (Eds.) (2002). *Methodology in language teaching: An anthology of current practice*. Cambridge University.
- Lartillot, O., & Toiviainen, P. (2007). A Matlab toolbox for musical feature extraction from audio. *In International Conference on Digital Audio Effects*, pp. 237-244.
- Learning, D. (2017). *Convolutional neural networks*. Retrieved from <http://deeplearning.net/>: <http://deeplearning.net/tutorial/lenet.html>
- LeCun, Y., Bottou, L., Bengio, Y., & Haffner, P. (1998). Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, Vol.86, No.11, pp.2278-2324.
- Levenshtein, V. I. (1966). Binary codes capable of correcting deletions, insertions, and reversals. *In Soviet Physics Doklady*, Vol. 10, No. 8, pp.707-710.
- Lu, M. (2008). Effectiveness of vocabulary learning via mobile phone. *Journal of Computer Assisted Learning*, Vol.24, No.6, pp.515-525.

- Lind, M., Simonsen, H. G., Hansen, P., & Holm, E. (2012). *Name relatedness and imageability*. Cork: International Clinical Linguistics and Phonetics Association.
- Machine, W. (1997). Retrieved from [http://www.3m.com:80/meetingnetwork/files/meetingguide\\_pres.pdf](http://www.3m.com:80/meetingnetwork/files/meetingguide_pres.pdf)
- Madhulatha, T. S. (2012). An overview on clustering methods. . *arXiv preprint arXiv*;, Vol.1205 No.1117, pp.719-725.
- Malpani, A. R., Ravindran, B., & Murthy, H. (2011). Personalized intelligent tutoring system using reinforcement learning. *FLAIRS Conference*, pp.561-562.
- ManyThings.Org. (n.d.). Retrieved from <http://www.manythings.org/vocabulary/lists/a/words.php?f=3esl>
- Marton, W. (1977). Foreign vocabulary learning as problem no. 1 of language teaching at the advanced level. *Interlanguage Studies Bulletin*, pp.33-57.
- Maryland, U. o. (n.d.). *The FFT and power spectrum*. Retrieved from <http://www.ece.umd.edu/~tretter/commlab/c6713slides/>: <http://www.ece.umd.edu/~tretter/commlab/c6713slides/ch4.pdf>
- Mashhadi, F. & Jamalifar, G. (2015). Second language vocabulary learning through visual and textual representation. *Procedia-Social and Behavioral Sciences*, Vol.192, pp.298-307.
- Matas, J. C., Chum, O., Urban, M., & Pajdla, T. (2004). Robust wide-baseline stereo from maximally stable extremal regions. *Image and Vision Computing*, Vol.22, No.10, pp.761-767.
- MathWorks. (n.d.). *Image category classification using deep learning*. Retrieved from MathWorks: <https://www.mathworks.com/help/vision/examples/image-category-classification-using-deep-learning.html>
- MATLAB. (n.d.). *Feature extraction*. Retrieved from <https://www.mathworks.com/discovery/feature-extraction.html>
- McGriff, S. J. (2000). *Instructional system design (ISD): Using the ADDIE model*.
- Mishan, F. (2005). *Designing authenticity into language learning materials*. Intellect Books.
- Murtagh, F. &. (2011). Ward's hierarchical clustering method: clustering criterion and agglomerative algorithm. *arXiv preprint arXiv*, Vol.1111, No.6285.
- Nakata, T. (2008). English vocabulary learning with word lists, word cards and computers: Implications from cognitive psychology research for optimal spaced learning. *ReCALL*, Vol.20, No.1, pp. 3-20.
- Nation, I. S. (2001). Learning vocabulary in another language. *Ernst Klett Sprachen*.
- Nicholes, J. (2016). Measuring the impact of language-learning software on test performance of Chinese learners of English. *TESL-EJ*, Vol.20, No.2, pp.1-20.
- Obdržálek, D. B., Basovník, S., Mach, L., & Mikulík, A. (2009). Detecting scene elements using maximally stable color regions. *In International Conference on Research and Education in Robotics* Berlin, Heidelberg. Springer, pp. 107-115.
- Ogata, H. &. & Yano, Y. (2004). Context-aware support for computer-supported ubiquitous learning. *In Wireless and Mobile Technologies in Education*, pp.27-34.
- Ogata, H., Li, M., Hou, B., Uosaki, N., El-Bishouty, Moushir M., & Yano, Y. (2011). SCROLL: Supporting to share and reuse ubiquitous learning log in the context of language learning. *Research & Practice in Technology Enhanced Learning*, Vol. 6, No.2, pp.69-82 .

- Ojala, T. P., Pietikainen, M., & Maenpaa, T. (2002). Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol.24, No.7, pp.971-987.
- Omaggio, A. C. (1979). Pictures and second language comprehension: Do they help? *Foreign Language Annals*, Vol.10, no.2, pp.107-116.
- O'malley, J. M., & Chamot, A. U. (1990). *Learning strategies in second language acquisition*. Cambridge: Cambridge University Press.
- Paivio, A., & Csapo, K. (1973). Picture superiority in free recall: Imagery or dual coding? *Cognitive Psychology*, Vol.5, No.2, pp.176-206.
- Paivio, A. (1969). Mental imagery in associative learning and memory. *Psychological Review*, Vol.76, No.3, pp.241-263.
- Paivio, A., Rogers, T. B., & Smythe, P. C. (1968). Why are pictures easier to recall than words? *Psychonomic Science*, Vol.11, No.4, pp.137-138.
- Paivio, A., Yuille, J. C., & Madigan, S. A. (1968). Concreteness, imagery, and meaningfulness values for 925 nouns. *Journal of Experimental Psychology*, Vol. 76, No. (1, pt.2), pp.1-25.
- Preece, S. J., Goulermas, J. Y., Kenney, L. P., & Howard, D. (2009). A comparison of feature extraction methods for the classification of dynamic activities from accelerometer data. *IEEE Transactions on Biomedical Engineering*, Vol.56, No.3, pp.871-879.
- Pruden, S. M., Hirsh-Pasek, K., Maguire, M., & Meyer, M. (2004). Foundations of verb learning: Infants categorize path and manner in motion events. In Proceedings of the 28th Annual Boston University Conference on Language Development, pp. 461-472.
- R. Russell, a. M. (1918). *USA Patent No. Soundex, US Patent 1*.
- Reddy, B. S., & Chatterji, B. N. (1996). An FFT-based technique for translation, rotation, and scale-invariant image registration. *IEEE transactions on Image Processing*, Vol.5, No.8, pp.1266-1271.
- Rogers, Y., Price, S., Randell, C., Fraser, D. S., Weal, M., & Fitzpatrick, G. (2005). Ubi-learning integrates indoor and outdoor experiences. *Communications of the ACM*, Vol.48, No.1, pp.55-59.
- Russakovsky, O. D., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., & Berg, A. C. (2015). Imagenet large scale visual recognition challenge. *International Journal of Computer Vision*, Vol.115, No.3, pp.211-252.
- Sahrir, M., & Alias, N. A. (2012). A design and development approach to researching online Arabic vocabulary games learning in IIUM. *Procedia-Social and Behavioral Sciences*, Vol.67, pp.360-369.
- Sakamura, K., & Koshizuka, N. (2005). Ubiquitous computing technologies for ubiquitous learning. *In Wireless and Mobile Technologies in Education*, IEEE, pp. 11-20.
- Santos, M. E. C., Taketomi, T., Yamamoto, G., Rodrigo, M. M. T., Sandor, C., & Kato, H. (2016). Augmented reality as multimedia: the case for situated vocabulary learning. *Research and Practice in Technology Enhanced Learning*, Vol.11, No.1, pp.4-27.
- Schmitt, N. (1997). *Vocabulary learning strategies*. Vocabulary: Description, Acquisition and Pedagogy.
- Simonyan, K., & Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv*., Vol.1409, No.1556.

- Stahl, S. A., & Fairbanks, M. M. (1986). The effects of vocabulary instruction: A model-based meta-analysis. *Review of Educational Research*, Vol.56, No.1, pp.72-110.
- Storkel, H. L. (2003). Learning new words II: Phonotactic probability in verb learning. *Journal of Speech, Language, and Hearing Research*, Vol.46, No.6, pp.1312-1323.
- Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., & Rabinovich, A. (2015). Going deeper with convolutions. *In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, IEEE, pp. 1-9.
- Tardif, T. (1996). Nouns are not always learned before verbs: Evidence from Mandarin speakers' early vocabularies. *Developmental Psychology*, Vol.32, No.3, pp.492-504.
- Tene, O. (2008). *What google knows: Privacy and internet search engines*. Utah L. Rev, pp.1434-1492.
- Terhardt, E. (1974). On the perception of periodic sound fluctuations (roughness). *Acta Acustica united with Acustica*, Vol.30, No.4, pp.201-213.
- Tessmer, M. (1993). *Planning and conducting formative evaluations: Improving the quality of education and training*, Psychology Press.
- Thornbury, S. (2006). *How to teach vocabulary*. Pearson Education India.
- Thornton, P., & Houser, C. (2005). Using mobile phones in English education in Japan. *Journal of Computer Assisted Learning*, Vol.23, No.3, pp.217-228.
- Tibshirani, R., Walther, G., & Hastie, T. (2001). Estimating the number of clusters in a data set via the gap statistic. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, Vol.63, No.2, pp.411-423.
- Ting, R. Y. (2005). Mobile learning: Current trend and future challenges. *ICALT*, IEEE. pp. 603-607.
- Tomasello, M., & Kruger, A. C. (1992). Joint attention on actions: Acquiring verbs in ostensive and non-ostensive contexts. *Journal of Child Language*, Vol.19, No.2, pp.311-333
- Tomlinson, B., (2008). *English language learning materials: A critical review*. Bloomsbury Publishing.
- Tummala, H., & Jones, J. (2005). Developing spatially-aware content management systems for dynamic, location-specific information in mobile environments. *Proceedings of the 3rd ACM International Workshop on Wireless Mobile Applications and Services on WLAN Hotspots*, ACM, pp. 14-22.
- Uosaki, N., & Ogata, H. (2009). Supporting communicative English class using mobile devices. *Proceedings of mLearn*, pp. 94-102.
- Uosaki, N., Ogata, H., Sugimoto, T., Li, M., & Hou, B. (2012). Towards seamless vocabulary learning: How we can entwine in-class and outside-of-class learning. *International Journal of Mobile Learning and Organization*, Vol.6, No.2, pp.138-155.
- Uosaki, N., Ogata, H., Mouri, K., & Choyekh, M. (2017). Implementing sustainable mobile learning initiatives for ubiquitous learning log system called SCROLL. *In Mobile Learning in Higher Education in the Asia-Pacific Region*, Vol.40, pp.89-114.
- Vi, C. T., Takashima, K., Yokoyama, H., Liu, G., Itoh, Y., Subramanian, S., & Kitamura, Y. (2013). D-FLIP: Dynamic and flexible interactive photoshow. *In Advances in Computer Entertainment*, Springer, pp.415-427.



- Wang, Y. K. (2004). Context awareness and adaptation in mobile learning. *In Wireless and Mobile Technologies in Education*, IEEE, pp. 154-158.
- Ward, M., & Newlands, D. (1998). Use of the web in undergraduate teaching. *Computers & Education*, Vol.31, No.2, pp.171-184.
- Webber, N. E. (1978). Pictures and words as stimuli in learning foreign language responses. *The Journal of Psychology*, Vol.98, No.1, pp.57-63.
- Welch, P. (1967). The use of fast Fourier transform for the estimation of power spectra: a method based on time averaging over short, modified periodograms. *IEEE Transactions on Audio and Electro-acoustic*, Vol.15, No.2, pp.70-73.
- Wilkerson, M., Griswold, W. G., & Simon, B. (2005). Ubiquitous presenter: increasing student access and control in a digital lecturing environment. *ACM SIGCSE Bulletin*, Vol. 37, No. 1, pp.116-120.
- Wilkins, D. A. (1972). Linguistics in language teaching. *E. Arnold*.
- Wong, L. H., & Looi, C. K (2010). Vocabulary learning by mobile-assisted authentic content creation and social meaning-making: two case studies. *Journal of Computer Assisted Learning*, Vol.26, No.5, pp.421-433.
- Wright, A. (2005). *Pictures for language learning*. Cambridge University Press.
- Wu, Q. (2015). Designing a smartphone app to teach English (L2) vocabulary. *Computers & Education*, Vol.85, pp.170-179.
- www.cs.auckland.ac.nz. (n.d.). *Spatial frequency domain*. Retrieved from <https://www.cs.auckland.ac.nz/courses/compsci773s1c/lectures/ImageProcessing-html/topic1.htm>
- Yang, S. J. (2006). Context aware ubiquitous learning environments for peer-to-peer collaborative learning. *Educational Technology & Society*, Vol.9, No.1, pp.188-201.
- Yang, W., & Dai, W. (2011). Rote memorization of vocabulary and vocabulary development. *English Language Teaching*, Vol.4, No.4, pp.61-64.
- Yeh, Y., & Wang, C. W. (2003). Effects of multimedia vocabulary annotations and learning styles on vocabulary learning. *Calico Journal*, Vol.21, No.1, pp.131-144.
- Zeiler, M. D., & Fergus, R. (2014). Visualizing and understanding convolutional networks. *In European Conference on Computer Vision*, Cham: Springer, pp. 818-833.
- Zhang, G., Jin, Q., & Lin, M. (2005). A framework of social interaction support for ubiquitous learning. *In Advanced Information Networking and Applications*, IEEE, Vol.2, pp. 639-643.

## Related Publications

### Main Papers

Mohammad Nehal Hasnine, Masatoshi Ishikawa, Yuki Hirai, Haruko Miyakoda, and Keiichi Kaneko: *An Algorithm to Evaluate Appropriateness of Still Images for Learning Concrete Nouns of a New Foreign Language*, IEICE Transactions on Information and Systems, Volume E100-D, No.9, pp.2156-2164, September 2017.

Mohammad Nehal Hasnine, Yuki Hirai, Masatoshi Ishikawa, Haruko Miyakoda, Keiichi Kaneko, and Lyn Pemberton: *An Image Recommender System that Suggests Appropriate Images in Creation of Self-Learning Items for Abstract Nouns*, International Journal of Management and Applied Science, ISSN: 2394-7926, pp.38-44 Volume-2, Issue-5, May 2016.

### Refereed International Conference

Mohammad Nehal Hasnine, Yuki Hirai, Masatoshi Ishikawa, Haruko Miyakoda, Keiichi Kaneko, and Lyn Pemberton: *An Image Recommender System that Suggests Appropriate Images in Creation of Self-Learning Items for Abstract Nouns*, ISERD 23rd International Conference on Education and E-Learning (ICEEL), pp. 8-14, Oxford, United Kingdom (2016.03.15).

Mohammad Nehal Hasnine, Yuki Hirai, Masatoshi Ishikawa, Haruko Miyakoda, and Keiichi Kaneko: *Learning Effects Investigation of an On-demand Vocabulary Learning Materials Creation System based on Appropriate Images*, Proceedings of the 2015 International Conference on 4th ICT-ISPC, Tokyo, Japan (2015.05.23-05.24). First Prize Award (E-Applications, Poster and Demo Session)

Mohammad Nehal Hasnine, Yuki Hirai, Masatoshi Ishikawa, Haruko Miyakoda, and Keiichi Kaneko: *A Vocabulary Learning System by On-demand Creation of Multi-linguistic Materials based on Appropriate Images*, Proceedings of the 2014 International Conference on e-Commerce, e-Administration, e-Society, e-Education, and e-Technology (e-CASE & e-Tech 2014) - Fall Session, Tokyo, Japan, pp. 343-356 (2014.11.12-11.14). Distinguished Paper Award

## Acknowledgement

This Ph.D. thesis contains the experimental results conducted during my study period (April 2013 to March 2015 as a Master course student; and April 2015 to March 2018 as a doctoral course student) at the Tokyo University of Agriculture and Technology.

First of all, I would like to thank MEXT (Ministry of Education, Culture, Sports, Science and Technology) for awarding me its prestigious Monbukagakusho scholarship.

My deepest gratitude to Professor Keiichi Kaneko for introducing me with his old projects (PSI and SIGMA), suggesting me the right research topic, and his kind supervision over the last 5 years and 6 months. Also, I am thankful to him for introducing me with Associate Professor Masatoshi Ishikawa (of Tokyo Seitoku University). Without Associate Professor Masatoshi Ishikawa help and support, the outcome my Ph.D. thesis wouldn't be possible. I also want to thank Assistant Professor Yuki Hirai for his guidelines during his tenure at Tokyo University of Agriculture and Technology.

I would like to thank Dr. Lyn Pemberton (of University of Brighton, UK) for hosting me as a visiting scholar at her laboratory.

I am thankful to the International Center of Tokyo University of Agriculture and Technology for its continuous support on Japanese language and culture. It helped me a lot to understand classroom lectures, presentations in the lab, and interacting with other lab mates smoothly.

I would like to thank the reviewers of my thesis. Their comments and suggestions have helped me a lot to improve this manuscript.

I am thankful to all participants who helped me to evaluate my developments.

Finally, I am grateful to my family and friends for inspiring me to peruse this Ph.D. study program.

# Index

## A

Abstract noun, iii, iv, 17, 19, 47, 48, 49, 55, 56, 57, 58, 59, 60, 61, 62, 63, 67, 69, 71, 72, 73, 74, 75, 80, 84, 85, 93, 95, 105, 108, 110, 111, 112, 116, 117, 118, 121, 122, 123, 124, 132, 133

Adverb, 118, 119, 120

AIVAS, 19, 20, 63, 64, 65, 66, 67, 68, 69, 70, 71, 73, 74, 75, 76, 77, 79, 80, 82, 87, 88, 89, 90, 91, 92, 93, 94, 95, 97, 98, 99, 100, 101, 102, 105, 107, 108, 109, 110, 111, 112, 115, 117, 120, 121, 122, 123, 124

AIVAS image sets, 68, 69

AIVAS-ABST8300, 68, 69, 73, 75, 79, 121, 124

AIVAS-ABST-LS68, 68, 69, 71, 105, 108, 121, 123

AIVAS-ABST-LS795, 68, 69, 73, 74, 77, 82, 91, 108, 111, 121, 123, 124

AIVAS-AIRS, 20, 64, 66, 67, 93, 108

AIVAS-CNCRT59, 68, 69, 70, 80, 97, 121

AIVAS-EE, 20, 64, 91, 94, 101, 112

AIVAS-IRA, 19, 20, 67, 68, 76, 77, 80, 87, 93, 94, 95, 97, 98, 99, 100, 102, 105, 107, 108, 109, 110, 111, 115, 117, 121, 122, 123, 124

AIVAS-LMC, 20, 64, 87, 88, 89, 90, 92, 94, 111, 121

AlexNet, 68, 76, 77, 78, 79, 82, 84

Algorithms, 17, 80

API, 20, 37, 47, 71, 80, 92, 97, 105, 121, 124

Approach 1, 57, 105

Approach 2, 61

Approaches, iii, iv, 16, 18, 26, 30, 31, 37, 46, 48, 49, 55, 78, 121, 130

Appropriate image, iii, iv, 16, 17, 18, 19, 20, 47, 48, 49, 54, 55, 57, 58, 59, 60, 61, 62, 63, 64, 66, 67, 68, 69, 70, 72, 73, 74, 75, 77, 80, 81, 82, 84, 85, 87, 89, 93, 94, 95, 97, 105, 108, 110, 111, 114, 117, 118, 120, 121, 122, 123, 124, 131, 132

## B

Bengali, 50, 51, 106, 122

## C

CNN, iii, 68, 77, 78, 80, 82, 94, 108, 124

Cognitive, 12, 15, 16, 17, 20, 23, 24, 26, 27, 46, 119, 131, 134

Concrete noun, iii, iv, 15, 17, 19, 47, 48, 49, 50, 53, 54, 55, 62, 63, 67, 80, 81, 95, 96, 98, 99, 102, 117, 121, 122, 123, 124, 132

## E

Euclidean, 67, 80, 84

## F

Feature extraction, iii, 67, 68, 76, 77, 78, 79, 80, 82, 93, 94, 108, 133, 135

FFT, iii, 20, 68, 76, 77, 78, 80, 94, 97, 105, 108, 124, 134, 135

Foreign language, iv, 12, 13, 14, 16, 17, 18, 23, 27, 36, 46, 53, 62, 63, 69, 71, 73, 74, 82, 97, 108, 121, 123, 130, 132, 133, 137

Formal educational settings, 13

## G

GIF, 120

Google, 37, 50, 57, 58, 59, 71, 72, 73, 74, 75, 80, 81, 85, 87, 92, 96, 97, 98, 99, 101, 102, 103, 105, 107, 108, 109, 110, 111, 112, 114, 115, 117, 122, 131

## I

Image evaluation, 91, 95, 98, 99, 105, 107, 108, 117, 122

Immediate-after, 95, 99, 102, 103, 116, 117, 123

Informal learning, 14, 16, 19, 23, 39, 63, 123

Informal settings, 13

## L

Learning effect, 91, 95, 99, 101, 102, 110, 114, 115, 116, 117, 123, 138

Learning feature, 77

Learning material, iv, 16, 17, 19, 28, 29, 30, 31, 32, 33, 34, 36, 37, 39, 47, 50, 51, 63, 64, 87, 88, 89, 90, 92, 93, 94, 99, 100, 101, 111, 112, 121, 123, 131, 132, 133, 134, 136

Long-term, iii, 15, 24, 28, 30, 33, 47, 95, 99, 102, 115, 116, 117, 123

## M

Mid-term, 47, 95, 99, 101, 115, 116, 117, 123

Multilingual, 12

## N

Noun imageability, 16, 48  
Nouns, 14, 48, 55, 56, 57, 58, 61, 68, 81, 96, 126, 136, 138

## P

Pedagogical investigation, 50, 51, 53, 60, 122  
Polish, 89, 110, 113, 114  
Power spectrum, 76, 80  
Pronoun, 19, 44, 118, 119  
Proposed Definition, 54

## R

Russian, 36, 41, 71, 89, 99, 100, 101, 103, 106, 129

## T

Technology-enhanced, 14, 18, 23, 28, 130

## U

Unsupervised learning, 77

## V

Verb, iv, 14, 16, 18, 19, 31, 37, 43, 48, 55, 118, 119, 120,  
130, 131, 135, 136  
Vocabulary learning, iv, 12, 13, 44, 133, 136, 137