Effects of Providing Known Associates on Intentional Vocabulary Learning: Comparing Synonyms, Co-hyponyms, and Lexical Collocations

A Dissertation

Submitted to the University of Tsukuba

In Partial Fulfillment of the Requirements for

the Degree of

Doctor of Philosophy in Linguistics

Go TADA

2019

Abstract

When learning English vocabulary, it is important to learn lexical items in a communicative context. However, paired-associate learning, such as utilizing word lists, has been recently re-evaluated because research has shown that it can efficiently develop learners' knowledge of words, which is applicable even for real-life communication (Elgort, 2011). Intentional vocabulary learning is considered a necessary component of vocabulary learning, along with incidental vocabulary learning (Nation, 2013). Therefore, it is important to seek ways to further improve the efficiency of intentional vocabulary acquisition and alleviate learners' burden of learning a large number of expressions.

For broadening English vocabulary, learners should notice what they cannot say (Swain, 1995), motivate themselves to learn unknown expressions (Laufer & Hulstijn, 2001), and process the new information so that they can integrate it into their knowledge and facilitate its retrieval (Stahl & Nagy, 2006). Of the many methods of intentional vocabulary learning, realizing the above processing in identifying and learning unknown associates of new words is an effective way to enhance the learning effect.

However, it should be noted that learning unknown related words at the same time often inhibits memorization and retrieval (e.g., Erten & Tekin, 2008; Hoshino, 2010; Tinkham, 1997). Therefore, it is recommended to utilize "known + unknown" combinations when finding and learning new words through association, in order to derive the maximum benefit from this method (Nation, 2000).

According to some previous studies on this topic (e.g., Kasahara, 2010; Webb, 2007), learning with "known + unknown" combinations generally facilitates learners' memorization or retrieval of target words. However, as there are few related studies, some points for maximizing the effect of this method are left unclear. Thus, as shown more specifically below,

the purpose of the study is to make information about the effects of known related words more concrete and generalizable.

First, it is unclear whether the facilitative effects reported in previous studies can be generalized to other types of related words. Second, although there are repeated experiments on the effect of known related words on the establishment of form-meaning link of target words, the effects on other aspects of vocabulary knowledge, such as collocational knowledge of target words, are not clear in some cases. Third, it is also unknown which type of known related words exerts greater effects on learners' memory. Fourth, more concrete information is needed to further enhance the effects of known related words (e.g., strength of relationship, number of related words presented, or instruction about the use of related words while learning). To clarify these questions, four experiments were conducted, as described below.

In Experimental Study 1, known categorical associates were given when learners were learning and retrieving the meaning of target words. Participants learned 30 target words through paired-associate learning using word lists. Three conditions were set out: (a) presenting 10 words with closely related words, (b) presenting 10 words with slightly related words, and (c) presenting 10 words alone. Thereafter, posttests were conducted 10 minutes after the learning phase and one week after the first session. Both posttests consisted of two parts. The first part was a plain recall test, which required participants to recall Japanese translations of target words without any hint. The second part was a cued-recall test, which provided the known related word of each target word as clues, which facilitated the retrieval of target words. As for the effects of semantic relatedness, the evidence was indecisive, but it was indirectly implied that higher relatedness led to a stronger effect of facilitating retrieval.

Experimental Study 2 focused on synonyms and clarified whether the information of synonyms can lend greater precision to learners' understanding of the meaning of target words. As in the previous experiment, participants learned 60 target words in paired-associate

learning. There were five conditions in this experiment. In the four experimental conditions, semantic similarity and the number of presented known synonyms were controlled. In addition, one control condition was prepared, in which unrelated words were presented. Posttests of the same format were conducted five minutes after the learning phase and one week later. The same tests as in Experimental Study 1 were used again to measure the knowledge acquired by participants. In addition, a multiple-choice format was prepared to assess learners' precise knowledge of meaning (Joe, 1998). The results showed that one semantically close synonym is enough for maximizing the learning effect. Additionally, regardless of the condition, performance on the multiple-choice test was high even after a one-week delay, implying that learners' attention to meaning or their processing of comparison contributed to precise understanding.

Experimental Study 3 focused on the comparison of categorical associates and collocates. In the experiment, participants engaged in paired-associate learning of 20 target words. In each condition, 10 words were given with known categorical associates, and 10 words were provided with known collocates. The same measurements as in Experimental Study 1 were used and a multiple-choice format was used to measure learners' collocational knowledge (Sonbul & Schmitt, 2013). Posttests were conducted twice, once immediately after the learning phase and again one week later. The results indicated that categorical associates were more effective on facilitating the retrieval of target words, while collocates were more effective on developing the collocational knowledge.

Experimental Study 4 focused on the comparison of instructions in learning target words with known collocates. In the experiment, participants engaged in paired-associate learning of 20 target words. In one condition, participants learned 10 words while seeing the known collocates as mere references. In the other condition, participants learned 10 words and tried to remember the given collocation at the same time. The measurements in Experimental

Study 3 were used again. Posttests were conducted twice, once immediately after the learning phase and again one week later. The results indicated that the effects on form-meaning link of target words were stronger when participants tried to remember the collocation. However, the effects on collocational knowledge did not differ between the two conditions.

On the basis of these findings, it could be concluded that known related words generally facilitated learners' retrieval when they were presented as cues. Especially, known categorical associates had a strong effect on learners' knowledge of form-meaning link of target words. Known collocates were more effective on developing collocational knowledge, but the effects on form-meaning link were weaker. Although more verification is needed, relatedness between target and known related words should be higher. Regarding the method to utilize known collocates, the collocation consisting of target words and known collocates should be remembered at the same time.

Acknowledgements

This dissertation is a product of my five years of my research life at the Graduate School of Humanities and Social Science, University of Tsukuba. To complete this dissertation, great amount of work was needed for establishing research questions, refining experimental designs, doing experiments, and analyzing or discussing the data. Thanks to the favorable environment at the University of Tsukuba, I could get enormous support from my advisers and colleagues throughout the long process of research.

First, I would like to express my deep gratitude to Professor Hirosada IWASAKI, my chief supervisor, for encouraging and advising me throughout the process of completing the present study. He also substantially improved my skill of academic writing and presentation in English from the level of novice.

Second, I could learn a lot from my vice supervisors. I am thankful to Professor Akira KUBOTA for cultivating my careful, critical, and objective attitude in discussion. He also gave me some hints for the direction of my study. In addition, Assistant Professor Yuko HIJIKATA gave me a great deal of fine-grained advice for improving this work. My deep gratitude goes to Kazumi AIZAWA at Tokyo Denki University, an external supervisor. His careful review and discussion from another standpoint enlightened me further on dealing with the research on vocabulary learning.

Furthermore, Professor Akiyo HIRAI and Associate Professor Rie KOIZUMI helped me to understand statistics that are quite useful in the field of English pedagogy. I am also grateful for their caring responses to my questions. Additionally, Assistant Professor Yuichi ONO’s intriguing topics and discussions in classes cultivated my understanding of CALL. It is thanks to the information about this kind of new and hot topics that I could always be interested on English pedagogy. I also learned a lot about reading and vocabulary from the

lectures of Professor Yuji USHIRO.

I also thank all of my colleagues, especially Dr. Yukino KIMURA, Dr. Akira HAMADA, Dr. Natsumi TANAKA, Dr. Masaya HOSODA, Dr. Yoshinobu MORI, Ms. Eleanor DOWSE, Ms. Yuka TOKUDA, Mr. Hiroaki NAKAGAWA, and Mr. Kozo KAMIMURA, and Mr. Takumi OJIMA for their various kinds of help including recruit of participants, ratings, and suggestions for improving my research.

Additionally, this project was supported by various people. The present study was supported by the Japan Society for the Promotion of Science from 2017 to 2019 (No. 17J00906). I am also thankful to the students at the University of Tsukuba, Chuo University, Toho University who participated in this research. And staffs at the University of Tsukuba have given me caring support throughout the clerical process.

Last, I want to thank my family for always encouraging and supporting me. It is thanks to them that I could have come this far.

Go TADA

Tsukuba, Japan

December 2019

Table of Contents

List of Tables

List of Figures

List of Appendices

Chapter 1
Introduction

1.1 Background of the Present Study

When people learn English or any other foreign language, they must learn a wide range of vocabulary in the target language. Vocabulary knowledge is the foundation for language learning skills and supports all four kinds of communicative skills (i.e., listening, speaking, reading, and writing). It is one of the most difficult but the most basic and important part of second language acquisition. Therefore, several methods have been developed by teachers and several types of research have been conducted to increase the efficiency of learning foreign language vocabulary.

Such studies of second language vocabulary acquisition can be divided into two main strands, namely, intentional and incidental learning (Nation, 2013; p. 348). *Intentional learning* refers to the method where learners consciously engage in memorizing target words. Learning with word lists is a typical example of this strand. *Incidental learning* refers to the method where learners learn words by chance, as the focus of their tasks or activities is not on learning vocabulary. Extensive reading is a typical example of this strand. Since both of these two approaches have their advantages and disadvantages, intentional learning and incidental learning should be utilized in a balanced way (Nation, 2013). This study focused on the method of enhancing the effects of intentional vocabulary learning, which can develop the most basic aspect of vocabulary knowledge—the link between L2 form and its meaning—with higher efficiency than incidental learning.

One of the possible methods to increase the efficiency of intentional vocabulary acquisition is to find and learn related words together with target words, such as synonyms (words that have similar meanings to each other; e.g., *thunder* and *lightning*) and collocations

(words that are used in combination; e.g., *side* and *effect*). This way of learning may sound practical and effective. Learners can increase the chance to notice through association what they cannot say and should learn all by themselves (noticing the gap; Swain, 1995). Also, many word lists provide learners with information of related words or example sentences including collocations.

However, it should be noted that many studies have shown that learning related unknown words together hinder memorization and retrieval of target words. For example, Hoshino (2010) demonstrated using four types of semantically related words for the study, and these words were less efficiently learned than the same number of unrelated target words. The cause of this inhibitory effect is the *interference of memory*. When a newly created memory that is similar or strongly related, this phenomenon happens. The related memories are more easily confused and forgotten than unrelated memories. Therefore, it is suggested that when teachers or learners use related words for vocabulary acquisition, they should try to teach or learn a set of words in the form of "known + unknown" word combination instead of "unknown + unknown", which are subject to the inhibitory interference of memory (Nation, 2000).

This approach may sound inefficient because a learner will learn only one word whereas using the "unknown + unknown" method may help learners learn more words. However, previous studies have highlighted several possible benefits of utilizing associates of known words, which cannot easily be obtained by the simple one-by-one intentional vocabulary learning using word lists.

First, as already mentioned, this method increases the possibility for learners to identify what they cannot express. This identification happens when learners come up with associates of known words on their own, as well as when they encounter similar or opposite expressions in word lists or textbooks.

Second, using known related words can lead to the facilitation of retrieval of target words (Kasahara, 2010, 2011, 2015; Webb, 2007). Considering that repeated recollection is needed before a lexical memory is firmly consolidated, helping the retrieval of target words on their own, which is known to be better for strengthening memory than simply receive the right answer (retrieval practice effect; Baddeley, 1997), will increase the efficiency of learning in the long run. Teachers can help learners achieve this through several types of tasks, such as review questions and vocabulary tests, and letting them identify important target words over and over again.

Finally, the known related words can help learners acquire some other aspects of vocabulary knowledge apart from the form-meaning link. As highlighted in previous studies (Channell, 1981; Nation, 2000), when learning words with known synonyms, the more precise meaning of words can be identified by comparing the words and understanding the similarity and difference between them. Also, in the case of learning words with known collocates, the link between target words and collocates is learned. This link in itself works as collocational knowledge and contribute to natural and fluent communication (Pawley & Syder, 1983). Since these deeper aspects of vocabulary knowledge are, generally speaking, thought to be developed mainly by incidental contextualized learning, it can be considered as important merit if learning words in "known + unknown" combination can help learners develop form-meaning link and gain other aspects of vocabulary knowledge at the same time.

For the reasons mentioned above, learning vocabulary in the form of "known + unknown" lexical combinations have some possible positive effects on learners' development of vocabulary knowledge.

However, since the research is limited to the "known + unknown" method of learning words, the whole picture of the method is still vague in many ways. In other words, it is not clear exactly what and how learners should utilize when using the "known + unknown"

combinations. Webb (2007) has reported that presenting known synonyms of target words facilitated learners' memory of some aspects of target words. Kasahara (2010, 2011, 2015) has also reported that target words presented with known collocates were recollected better in posttests. Learning unknown words with known related words was found helpful in the facilitation of memory in those studies, but the following several ideas are still unclear.

First, although the facilitative effects of known related words were found in the cases of synonyms and collocations, whether this effect can be generalized to other types of relationships, such as categorical associates, is still unknown. Second, how closely the target words and known words are related should be taken into consideration (Ishii, 2015). Closely related words may help more in integrating new words into the existing mental lexicon or retrieving target words. Third, it is still unclear which type of relationship is the most facilitative for learning using the "known + unknown" word sets. A semantic relationship and collocational relationship may have different effects on learners' memory when learning the target words. Fourth, other possible effects than just helping in the recollection of form-meaning link are indicated by previous research but very little verification of this point has been done. For example, a more precise understanding of meaning could be obtained by comparing with known synonyms (Channell, 1981; Nation, 2000), and collocational knowledge could be obtained if known collocates are presented together, as discussed above. Lastly, more concrete information is needed for enhancing the effects of the "known + unknown" combination. For instance, there is the question of whether more than one known related words enhance the learners' memory still more, or whether learning of collocations (i.e., target words and their known collocates) should be incidental or intentional.

Thus, elucidating these unknown factors help with the concrete understanding of the method, and this will provide learners or teachers with information about how "known + unknown" word pairs should be utilized.

1.2 Organization of This Dissertation

This dissertation is comprised of nine chapters: Introduction (Chapter 1), Literature Review (Chapter 2), The Present Study (Chapter 3), Experimental Study 1 (Chapter 4), Experimental Study 2 (Chapter 5), Experimental Study 3 (Chapter 6), Experimental Study 4 (Chapter 7), General Discussion (Chapter 8), and Conclusion (Chapter 9).

After the introduction part (Chapter 1), a review of previous research related to the current study follows (Chapter 2). In the review, the author first identifies why and how intentional methods of vocabulary learning are necessary for English as a Foreign Language (EFL) learning. Next, knowledge about word association (e.g., semantic relationship and collocations) is introduced. Then, some theories on vocabulary learning are reviewed. They can be referred to for evaluating whether effective mental processing is achieved or not in a certain leaning method. After that, the author discusses the pros and cons of utilizing related words in intentional vocabulary learning. Focusing on the cases of "known + unknown" word set, the discussion moves on to the question of how the effectiveness of the learning could be maximized.

In the next part introducing the present study (Chapter 3), the author summarizes what is already known or unknown in previous research. This chapter then clarifies what has been newly investigated in the present study. As explained above, the overall purpose of the present study is to expand on the results of previous research, to compare the effectiveness of learning methods between types of related words, and to find more concrete methods to enhance learning. The focus of the following experiments reflects these points, which previous research has left unclear up to now.

Chapter 4 explains Experimental Study 1, which focuses on the case of using known categorical associates (i.e., those words which belong to the same category, such as *dog* and *cat,* which fall in the category of animals). It additionally investigated whether *semantic*

relatedness (i.e., how close or easy to associate their meanings are) of target words and their known categorical associates affects efficiency in facilitating the memorization and retrieval of target words. The results showed that the same facilitative effect found in the case of known collocates (Kasahara, 2010, 2011, 2015) also applies to the case of known categorical associates, which has not been examined. However, since the question of whether higher semantic relatedness leads to greater facilitation was not decisively answered, this issue was carried over to the next experiment.

Next, Chapter 5 describes Experimental Study 2, which focuses on the case of synonyms. This experiment also studies whether the existence of known synonyms helps learners to make their understanding of target words' meaning more precise. In addition, to obtain more concrete information to make the most of this method, the experiment took the number of related words presented together with target words into consideration. This experiment afforded us detailed information on the preparation of known synonyms. It was clearly shown that preparing one higher-related synonym is the most efficient method. On the other hand, the precise understanding of meaning was not affected by the number or relatedness of synonyms presented.

Chapter 6 is Experimental Study 3, which focuses on the comparison of the effects on learning between known categorical associates and known synonyms. It also studies the effect on collocational knowledge by the two types of known related words. The results of this comparison suggest that known categorical associates are more efficient in facilitating the cued-recall of target words. However, in the case of collocational knowledge, known collocates could better facilitate recall.

Chapter 7 explains the last experiment, Experimental Study 4, which focuses on the comparison of the effects between learning collocations (i.e., the combination of target words and their known collocates) intentionally or incidentally. Using the same design as

Experimental Study 3, it also investigates the effect on collocational knowledge by the two types of known related words. The results of the delayed posttest showed that the target words were more effectively learned intentionally using “known + unknown” collocations.

In Chapter 8, a general discussion of the four Experimental Studies is attempted to draw answers to the research questions presented in Chapter 3 and to get a better understanding of this method. Some discussion is also made based on a comparison of the results of the Experimental Studies in order to determine which types of known related words and which manners of use lead to the most effective learning. Finally, Chapter 9 concludes the dissertation with some pedagogical implications, limitations of the current study, and suggestions of future research.

Chapter 2

Literature Review

2.1 Importance of Vocabulary for EFL Learners

Vocabulary is the foundation for developing proficiency in a foreign language. In all kinds of communications (i.e., listening, speaking, reading, and writing), almost nothing can be conveyed without vocabulary knowledge. As shown below, even the basic skills for daily communication requires learners a large vocabulary. Thus, the development of vocabulary knowledge is important in the process of language learning, and it widens the possibility of real-life language use and more choice in doing tasks or activities to acquire higher proficiency.

How many words and phrases should an English as a Foreign Language (EFL) learner learn? It depends on the objective of a learner, or what the learner wants to do using English. An educated, adult, native speaker of English can use nearly 20,000 word families (Goulden, Nation, & Read, 1990; Zechmeister, Chronis, Cull, D'Anna, & Healy, 1995), but the amount of vocabulary that an EFL learner needs is much smaller. In Nation (2013), it is suggested that 3,000–4,000 word families are at least needed for using English without much difficulty, and 9,000 word families (i.e., mid-frequency words) should be considered as the eventual goal of vocabulary learning. In addition to the eventual goal, several stages of development and corresponding ability of language use are summarized in Nation (2013, p. 39). For example, basic speaking skills require the knowledge of 1,200 word families, and daily conversations require about 6,000–7,000 word families.

There is a reason for this figure of 9,000 word families for the eventual goal. The most difficult of the four strands of communicational skills (i.e., listening, speaking, reading, and writing) in terms of vocabulary is reading. According to Nation (2013), reading authentic

texts without using a dictionary requires the largest size of vocabulary. To realize it by guessing the meaning of the unknown words, 98% of words in running texts must be known (Hu & Nation, 2000). Analyzing the large-scale corpus BNC (British National Corpus), Nation (2013) highlighted that this 98% of coverage can be realized with the knowledge of the most frequently used 9,000 words.

In addition to the large size of vocabulary, some of the formulaic sequences, such as the expressions that are consisted of more than two words, are also important for successful communication. Formulaic sequences are not necessarily needed for learners especially in the environment of English as a Lingua Franca (Kecskes, 2018); however, it is also considered that learners will have difficulty in the input (i.e., reading or listening) without the knowledge of the most important or known expressions (Grant, 2005). In the previous studies, approximately 100–300 expressions have been extracted as the most important formulaic sequences (Grant, 2005; Shin & Nation, 2008). Additionally, learning more than this minimum requirement will contribute to fluency in communication by forming more chunks in learners' mental lexicon that can be retrieved faster (Pawley & Syder, 1983). To sum up, more than 9,000 expressions should be eventually learned if formulaic sequences are also taken into consideration.

Moreover, just linking meanings and forms of vocabulary items (i.e., *breadth* of vocabulary knowledge) is not enough. For improving efficiency and accuracy of communication, having considerable knowledge about each word (i.e., *depth* of vocabulary knowledge) is also important. Although all aspects of the depth of vocabulary are difficult to cover, Nation (2013, p. 49) summarizes this aspect of vocabulary knowledge well. For example, collocational knowledge is one of the aspects of depth of vocabulary. It has almost the same meaning as the knowledge of formulaic sequences, and the question of why and how this knowledge is important is briefly explained above. Additionally, a detailed understanding

of meaning is important to appropriately differentiate the words that are in use. When learners' knowledge of lexical items is only partial or vague and they do not have enough control over connotation or nuance of expressions, their communication gets less natural or fluent. For avoiding such inconveniences, precise meaning of expressions should be understood by learners. This aspect of lexical knowledge is also considered to contribute to the fluency of reading (Perfetti, 2007). Therefore, when choosing or developing methods for vocabulary learning, depth of vocabulary knowledge should also be considered.

Thus, the development of vocabulary is quite a long and inevitable process of second language acquisition, and it is necessary to develop the methods to increase the efficiency of vocabulary learning and ease the learning burden of EFL learners.

2.2 Intentional Learning and Incidental Learning of L2 Vocabulary

As mentioned above, in vocabulary acquisition, learning methods can be divided into two major strands; intentional and incidental learning (Nation, 2013; p. 348).

Intentional learning comprises those methods by which learners learn vocabulary as the primary purpose of the task or activity. Paired-associate learning of lexical form and meaning by using a word list personally or by using a flashcard in a class is a typical example of intentional vocabulary learning. Of course, learning using some kind of context is also included in intentional learning if learners consciously regard vocabulary acquisition as the main goal.

On the other hand, incidental learning refers to those methods of study in which learners expand their vocabulary by chance while performing certain tasks that do not focus on vocabulary acquisition as the main objective. Incidental learning of vocabulary can happen during reading, with or without the help of a dictionary, or in the process of listening or speaking.

In the past, incidental learning of vocabulary drew more attention from researchers. It is mainly because implicit knowledge of vocabulary, which was attained by implicit learning, was regarded as better in the paradigm of the communicative approach. Compared to the knowledge obtained by unnatural decontextualized ways of learning, incidentally and implicitly learned knowledge was considered to be more effective, and these kinds of knowledge were considered to be more suitable for communicative use (Krashen, 1985). Incidental or implicit learning is similar to the situation when children learn their first language, which is natural and a better way of learning.

Since then, the strong points of incidental learning were widely investigated and recognized. When it comes to the depth of vocabulary knowledge, contextual use of vocabulary is regarded as inevitable even now. In fact, learners' intuition of collocation or the context in which each L2 expression can be used cannot be easily developed only by using word cards. Moreover, tackling tasks or activities motivates learners to learn vocabulary as it makes them realize the need for more vocabulary knowledge for practical communications (Nation, 2013).

However, the importance of intentional vocabulary learning was re-evaluated in recent studies, and now, it is also considered as inevitable in developing vocabulary. One of the biggest benefits of intentional learning is the high efficiency in broadening the knowledge of form-meaning link of lexical items. In other words, incidental learning alone cannot provide learners with a sufficient range of vocabulary. For example, much research has been conducted on extensive reading, but according to some reviews, it has been assumed that about 10 encounters are required to achieve consolidation of a word meaning (Grabe, 2009; Schmitt, 2008). Moreover, the correct meaning of unknown words is not always successfully guessed. The reviews also highlighted that only 5–15% of unknown words are learned through extensive reading. At this pace, at least 420 books are required for learning 2,000

word families. Therefore, it is highly disadvantageous to rely only on incidental vocabulary learning for effectively expand the breadth of vocabulary (Elgort, 2011; Mondria, 2003).

Furthermore, it has recently been proved that knowledge acquired through decontextualized intentional learning is not impractical in communicative use. In an experiment of a lexical decision task with priming, Elgort (2011) showed that implicit knowledge, which can be quickly accessed in learners' mental lexicon, was successfully developed by repeated paired-associate learning.

Therefore, since both intentional and incidental vocabulary learning have different advantages and disadvantages, it is now considered that incidental and intentional vocabulary learning should be used complementarily (Nation, 2013, p. 348). Although the best possible balance of intentional and incidental vocabulary learning is yet to be achieved, this study focused on intentional learning as the necessary first step to learn vocabulary for broadening the ability to do many kinds of activities, which in turn serve as opportunities for incidental vocabulary learning. It can be considered as a good approach to vocabulary learning to acquire a form-meaning link first before trying to develop the depth of vocabulary knowledge. This is because it is impossible to learn all aspects of word knowledge at once due to the limited attentional resources (Barcroft, 2002); other aspects of vocabulary knowledge than the form-meaning link can be better learned when the form-meaning link of the word is already well established in learners' mental lexicon (Bogaards, 2001). Therefore, the present research attempted to find a method to increase the efficiency of intentional learning of form-meaning link, while considering effects of enhancing other aspects of knowledge.

2.3 Learning Related Words Together

When learning unknown words intentionally, one of the possible strategies is to remember the target words with their related words, such as synonyms, antonyms, or categorically related words (hereafter, *categorical associates*). This method seems intuitively effective and could, in reality, be supported by some theoretical discussions.

First, learning related words together can be considered as effective in establishing the form-meaning link because it increases opportunities to identify expressions that are still unknown to learners (Swain, 1995). The process of finding and selecting words worth remembering is also beneficial because it encourages learners to feel intrinsically motivated to learn the words, which they select themselves (strong need; Laufer & Hulstijn, 2001).

Second, even after establishing the form-meaning link, the memory learned together would be benefited by the link among one another. As highlighted by Wharton and Race (1999), if a learner knows related words, encountering one member of a related word set could activate memories of other members. The facilitation of retrieval will result in the reinforcement of recollected memories.

Third, other aspects of knowledge could also be effectively learned. In the case of learning synonyms together, a more precise understanding of word meaning can be obtained by comparing the synonyms and understand the similar and different features of them (Channell, 1981). Additionally, knowledge of usage and collocation of related lexical items would be transferred (Webb, 2007). It is because the usage and collocation of synonyms are often similar to each other.

Thus, in theory, learning related word sets improves the efficiency of intentional vocabulary learning, both in terms of establishing the form-meaning link and other aspects of vocabulary knowledge.

However, the first point of discussion, which is the possible facilitative effect on establishing the form-meaning link, was disproved by the results of empirical studies. Contrary to this intuition, many pieces of studies have reported the inhibitory effects of learning related “unknown + unknown” word combinations together in many situations. In the studies of Higa (1963) and Tinkham (1997), the number of attempts required to learn target words increased when the target words had a semantic relationship, indicating an adverse effect on memorizing words.

After that, as summarized in Table 2.1, many similar results were replicated with different kinds of participants, related words, and learning methods. For example, Erten and Tekin (2008) experimented with 55 children, who were native Turkish speakers. In this experiment, the memorization of categorical associates and unrelated words were compared. A total of 80 target words were prepared and 20 of them were learned per trial. A group of related words comprised of 20 words. After the learning session with word lists and classroom instruction, an immediate posttest was conducted. After a week, a delayed posttest was also conducted. The test required the participants to choose a picture equivalent to the meaning of each target word. As a result, inhibitory effects were found in most of the tests, indicating an adverse effect on retention and retrieval of words.

Additionally, Papathanasiou (2009) further verified the effect in the case of categorical associates, as well as in the cases of synonyms and antonyms. 31 children and 32 adults, who were Greek native speakers, participated in the study. The study compared the learning of the related words with that of unrelated words. 120 target words were prepared and 10 of them were learned per trial. 10 words made a group of related words. After all the learning sessions were completed, in which the participants made a word list and did a confirmation task, posttests were conducted. Therefore, the first posttest was done several days to three weeks after the learning session. The second test was two weeks after the first test. The test required

the L1 translation of target words. The result indicated inhibitory effects of related words on adults but not on children.

A similar result is also confirmed in the case of Japanese EFL learners. Hoshino (2010), in the same way as Papathanasiou (2009), compared antonyms, categorical associates, and synonyms with unrelated words. 119 Japanese university students participated in the study. Target words were 100 in total, and 20 of them were learned at once. Two words made a group of related words. Three or four days after the learning in the form of homework, which can be done freely, a posttest, requiring L1 translation, was conducted. Except for the case of categorical associates, the inhibitory effect of related words was found.

Bolger and Zapata (2011) conducted a similar experiment on 66 native speakers of English. All of 32 target words were learned at once intentionally with context, and the cases of categorical associates and unrelated words were compared. A set of related words consisted of eight words. From day one to the sixth day, after the learning session, word recognition speed was measured as the posttest. The result again indicated the inhibitory effect for categorical associates.

Furthermore, this inhibitory effect is not limited to cases of semantic relationships. Laufer (1989) highlighted that orthographically similar words are more likely to confuse learners and make it difficult for them to learn those words (e.g., *comprehensive* and *comprehensible*).

The interference theory is often used as a mechanism to explain this phenomenon (Ishii, 2015). When learners attempt to learn similar or strongly associated memory that has not yet been established, the memories are subject to an interference effect. This is considered to be caused by a lack of clues for selectively activating each memory (Anderson & Reder, 1979; Baddeley, 1997).

Table 2.1

Conditions and Results of Major Experiments of Learning Related Words

	Erten and Tekin (2008)	Papathanasiou (2009)	Hoshino (2010)	Bolger and Zapata (2011)
Type of Related Words Examined	CAs	Antonyms, CAs, and synonyms	Antonyms, CAs, and synonyms	CAs
Age of Participants	Children	Adults and Children	Adults	Adults
Number of Target Words per Trial	20	10	20	32
Task for Learning	Word list and class	Making a word list	Free (homework)	Intentional learning in context
Format of Posttest	Choice of Pictures	L1 translation	L1 translation	Recognition speed
Result	Inhibition	Inhibition[a]	Inhibition[b]	Inhibition

Note. CA and IM stand for categorical associate and immediate posttest, respectively.

[a]Children were not inhibited. [b]CAs were not inhibited.

Since pieces of memory can be connected to many things, unnecessary memories conflicting with an important memory are inhibited when activated by a certain stimulus. This system facilitates the forgetting of unnecessary memories for obtaining an efficient structure of memory, which is known as retrieval-induced forgetting (Anderson, 2003). That is why related words tend to be confusing and are forgotten faster even if learners are not conscious of it.

This may also explain the tendency of weaker inhibitory effect by smaller sets of related words. For instance, Hoshino (2010) provided participants with pairs of related words and obtained facilitative effects for the case of categorical associates. In contrast, Bolger and Zapata (2011) clustered eight words, Erten and Tekin (2008) clustered 20 words, and Papathanasiou (2009) clustered 10 words. They all resulted in inhibitory effects as mentioned above.

To sum up, it is suggested by the interference theory that learners should avoid learning unknown related words at once. If this interference can be avoided, benefits such as facilitated retrieval of the form-meaning link (Wharton & Race, 1999) and subtle discrimination among synonyms (Channell, 1981) can be obtained, which will certainly improve the efficiency in intentional vocabulary learning.

To realize this, Nation (2000) suggested the method of learning words in “known + unknown” word sets because if related memories are already well consolidated, newly acquired memories are far less subject to the inhibitory interference. As discussed later in detail, this method can easily be integrated into classroom instructions and assignments (Aizawa & Mochizuki, 2010). Benefits of the knowledge of form-meaning link and other aspects have been reported in some of the previous studies in the case of “known + unknown” sets of synonyms (Webb, 2007) and collocations (Kasahara, 2010). Before delving into the detailed discussion about the most relevant studies on “known + unknown” word sets,

theories about word association in the mental lexicon and vocabulary learning will be organized in the following section.

2.4 Knowledge of Word Association

2.4.1 Paradigmatic and syntagmatic association

When considering the memorization of vocabulary, it should be noted that the memory of each word is not independently retained in a human's brain. There are links among the memory of words that affect one another.

As mentioned before, there are several kinds of relationships between words that affect one another via links in a learner's mental lexicon. These relationships can first be divided into two large groups: *paradigmatic associations* and *syntagmatic associations*. Paradigmatic association is a semantic relationship. If some words are so related that they can replace one another without making the sentence nonsensical, then these words are paradigmatically associated. Categorical associates (e.g., *cat* and *dog*), synonyms (e.g., *thunder* and *lightning*), and antonyms (e.g., *light* and *heavy*) exhibit paradigmatic relationships. Paradigmatic associates are generally the same part of speech. On the other hand, a syntagmatic association is almost equivalent to the relationship of collocations, although the meaning of *collocation* depends on its definition. If some words often co-occur in sentences (e.g., *cat* and *purr*, *heavy* and *rain*), then these words are syntagmatically associated.

The linking of paradigmatic and syntagmatic associations in learners' mental lexicons is regarded as one component of their depth of vocabulary knowledge (Nation, 2013; Webb, 2007). In fact, the ability to identify which words are associated with one another is used to indicate an examinee's proficiency. Native speakers' tendency of association shifts from *clang association* (i.e., association of phonologically similar words) to paradigmatic association as they get older. Hence, the ability to paradigmatically associate words is regarded as indicative

of a nativelike depth of vocabulary knowledge. Syntagmatic association is also indicative of proficiency, but is regarded as less proficient than paradigmatic association (see Wolter, 2001, for review).

2.4.2 Paradigmatic association

2.4.2.1 Definition of terms

As mentioned above, paradigmatic associates have some subcategories, such as synonyms and categorical associates. Basically, they are semantically related. For example, *antonyms* are most often defined literally as words that have opposite meanings (Hoshino, 2010).

In the current study, categorical associates and synonyms must be clearly defined, because they were used as experimental materials. First, while categorical associates are not uniformly defined, generally speaking they are considered words that belong to the same category, as in Bolger and Zapata (2011) or Hoshino (2010). The current study basically follows this definition. The term *co-hyponym* means almost the same thing, but emphasizes the hierarchical network of meanings in the definition. For example, *cat*, *dog*, and *mouse* belong to the same category of *animal*. In this case, *animal* is a *hypernym* of *cat*, *dog*, and *mouse*, and *cat*, *dog*, and *mouse* are *hyponyms* of *animal*. It should be noted that the words *cat*, *dog*, and *mouse* are on the same level of the semantic hierarchy. In the cases like this, *cat*, *dog*, and *mouse* are termed co-hyponyms. In the current study, not all categorical associates follow this definition, but most of the categorical associates are co-hyponyms.

Second, the meaning of the term *synonym* should be clarified. Synonyms can be defined as words that have the same meaning or a similar meaning. Little research has distinguished this subtle difference when defining synonyms, presumably because it is most often unnecessary. However, in the current study, the term *synonym* refers to words that have

similar but slightly different meanings, because one of the focuses of the current study is an understanding of the subtle differences of meanings between target words and their synonyms.

2.4.2.2 Significance of paradigmatic association

From the above discussion, it can be understood that proficiency and word association knowledge are surely correlated. However, correlation does not guarantee the relationship of cause and effect. In other words, there is still the question of how the ability to form paradigmatic and syntagmatic associations contributes to the skills of English communication.

Paradigmatic association is considered to have some possibility of contribution to fluency in communication. In theory, if a learner can associate a certain term with other relevant words, then vocabulary related to a certain topic can be activated through the associative link in his or her mental lexicon. This activation of useful words in a certain topic can contribute to the fluent processing of language, especially in productive processing (Meara, 1990). This ability has also been proven to be advantageous for reading comprehension and the maintenance of the global coherence of texts (Oakhill, Cain, & McCarthy, 2015).

However, it should be noted that paradigmatic association is not necessarily an essential component of vocabulary knowledge. Some suspect that it is not necessary for EFL learners to consciously try to develop the skill. The correlation between the ability to form paradigmatic associations and communicative skills such as reading might be a spurious correlation to some extent. In fact, Wolter (2001, 2006) pointed out that production of paradigmatic associates is limited by learners' breadth of vocabulary knowledge, referring to the fact that the paradigmatic associates produced by native speakers were often low-frequency words. In addition, although the experiments of Oakhill et al. (2015) are carefully controlled to reduce the possibility of spurious correlation, there is still the question of

whether the effect of paradigmatic association is so significant that learners should learn it intentionally. Wolter (2001), showed that advanced-level nonnative learners' association of firmly consolidated words were often reliant on syntagmatic association and argued that the learners' mental lexicon is not functionally inferior even if syntagmatic associates are more often produced in word association tests.

Therefore, as discussed in the next section in detail, acquiring the knowledge of syntagmatic association might be more important and should be prioritized. Collocational knowledge can presumably contribute to natural and fluent communication in a more direct way.

2.4.3 Syntagmatic association

2.4.3.1 Definition of terms

Before proceeding to the detailed information of syntagmatic associates, the definitions and taxonomy of terms should be clarified. In the past research, complex terms were used and the discussion on definition was notoriously complicated (Nation, 2013; Schmitt, 2010).

Many researchers have created different terms for multiword expressions, conforming to their focus of research. For example, the term *idiom* was used for expressions that have metaphorical meanings in Howarth (1996). Biber, Johansson, Leech, Conrad, and Finegan (1999) referred to a *lexical bundle* as conventionalized chain of words with a certain communicative purpose. In addition, according to Schmitt (2010), *collocation* tends to just refer to word pairs, while the definition of *prefabricated expressions* or *chunks* tends to emphasize that the expressions are stored holistically in learners' mental lexicon. As a comprehensive term, Nation (2013) uses *multiword unit*, while Schmitt (2010) introduces *formulaic sequence* from Wray (2002). As a result, according to Wray (2002), over 50 terms are current in research on multiword expressions. However, in the studies of the past several

years, two approaches have become widely accepted: *frequency-based approach* and *phraseological approach*.

The frequency-based approach is a method of defining what is a formulaic sequence and what is not. Some studies rely on the judgment of native speakers or collocation dictionaries when defining formulaic sequences. However, this method of definition still suffers from arbitrariness or subjectivity (Revier, 2009). In contrast, the frequency-based approach is the "least subjective and not dependent on native-speaker judgements, intuitions, or inter-rater agreement." (Siyanova-Chanturia, 2015) This is because it relies only on corpus and statistical analyses.

The standard of definition is whether the expressions really appear in the corpus with greater than random probability predicted by the individual frequencies of their component words. For evaluating this, t-scores or MI (mutual information) are used. To be more specific, Schmitt (2010) recommends the more concrete standard of border. First, the value of MI should be 3 or more. Second, the frequency of appearance in a corpus should be at least 5 or more in order to avoid the biases of MI values that might appear too high for formulaic sequences that are actually low-frequency. Both of these two conditions should be met. Although not all studies met the whole standard explained above, there are many studies adopting this approach for defining formulaic sequences (e.g., Siyanova-Chanturia, 2015; Webb et al., 2013; Wolter & Gyllstad, 2013).

The phraseological approach includes a taxonomy of formulaic sequences. In many cases, this approach is based on Howarth's Continuum Model (Howarth, 1996), and it is quite useful for classifying the types of formulaic sequences or discriminating semantically transparent and opaque expressions. Studies that take semantic transparency into consideration as a factor often use this approach in defining formulaic sequences (Gyllstad & Wolter, 2016; Revier, 2009).

Howarth (1996) classifies formulaic sequences into four categories. The first category is free combinations, whose meaning can be interpreted literally by their component words (e.g., *pay a bill*). The second category is restricted collocations, in which one of the component words is used in a figurative or technical meaning (e.g., *pay a visit*). The third category is figurative idioms, which have (as a whole) a literal meaning and a metaphoric meaning (e.g., *pay the price*). The fourth category is pure idioms, which only have a metaphoric meaning (e.g., *pay the piper*).

The author uses the term formulaic sequence to refer to all kinds of multiword expressions, including collocations and idioms. Hereafter, the author will call the former two categories *collocations* and the latter two categories *idioms*. Semantically transparent expressions are those whose meaning as a whole can be interpreted by combining the literal meanings of component words. Thus, free combinations are transparent collocations, and restricted collocations are opaque collocations.

2.4.3.2 Significance of syntagmatic association

As mentioned above, knowledge of syntagmatic association is almost equal to the knowledge of formulaic sequences. Many pieces of research showing the effectiveness of this knowledge have already been accumulated.

The importance of knowing formulaic sequences is well described by Pawley and Syder (1983). If a learner has little knowledge of formulaic sequences, the learner has to translate the words one by one. It inhibits the fluency of language use by increasing the search in mental lexicon, and it makes the produced language less natural because of the one-by-one literal translation without knowledge of collocation. Native speakers and advanced learners can reduce the burden of search by processing the expressions as chunks of collocation. In

reality, this effect of formulaic sequence on L2 speaking has been confirmed empirically in several studies (Boers, Eyckmans, Kappel, Stengers, & Demecheleer, 2006).

Then, some information about the learning of collocations should be briefly summarized. One of the major difficulties in learning collocations is that collocations are largely understood at the time of reading or listening, and that lack of collocation knowledge is not noticed by learners until they try to write or speak in a second language (Eyckmans, 2009). Therefore, the methods that draw a learner's attention to collocations are effective. For example, Sonbul and Schmitt (2013) highlighted that input enhancement, such as underlining, facilitates the learning of collocations. Of course, intentional learning was also effective (Boers, Demecheleer, & Eyckmans, 2004; Webb & Kagimoto, 2009). On the other hand, it should also be noted that the incidental learning of collocations was effective enough in cases where the component words were already known (Durrant & Schmitt, 2010; Sonbul & Schmitt, 2013).

Additionally, the factors that can affect the acquisition of collocational knowledge should be kept in mind. First, the frequency of a collocation has been found to affect its difficulty (Eyckmans, 2009; Wolter & Gyllstad, 2013). As is the case with single words, it is more difficult to learn low-frequency expressions that are rarely encountered. However, it should be noted that while the effect of frequency on the knowledge of words is predominant (Milton, 2009), correlation between frequency of formulaic sequences and frequency in the corpus is relatively low (Garnier & Schmitt, 2016). It is partly because the frequency of the component words (Revier, 2009; Wolter & Gyllstad, 2013) and MI (Ellis, Simpson-Vlach, & Maynard, 2008) also matter. A learner's attention tends to be drawn to interpreting or learning difficult low-frequency words (Siyanova-Chanturia, Conklin, & van Heuven, 2011).

Second, semantic factors also affect the difficulty of learning. It has been proven that semantically transparent expressions are easier to learn than opaque ones (Revier, 2009;

Gyllstad & Wolter, 2016). In addition, congruency affects the learning (Yamashita & Jiang, 2010; Wolter & Gyllstad, 2013). Congruency is the correspondence of expressions between L1 and L2. If the literal translation of the L2 component words conveys the same meaning as the original expression in L1, the expression is congruent. For example, *pay attention* is a congruent collocation, because the same meaning can be expressed in Japanese with the literal translation of the component words (*pay* [*払う*] + *attention* [*注意*] = *pay attention* [*注意を払う*]). On the other hand, kill time is incongruent (*kill* [*殺す*] + *time* [*時間*] ≠ *kill time* [*時間を潰す*]). In many cases, incongruent collocation is also semantically opaque, although there are such exceptions as *gain weight*, an incongruent and transparent expression, and *royal water*, a congruent and opaque expression, whose meaning is not literally the king's water but a certain mixture of strong acids.

Therefore, some effective approaches to learning collocations might be drawing learners' attention to collocations and focusing on important and easy (i.e., congruent and semantically transparent) expressions.

2.5 Theories of Vocabulary Learning

In this section, several theories of vocabulary learning are briefly introduced. Although the current study focuses on the learning in "known + unknown" combinations, and effects on both of form-meaning link and other aspects of vocabulary knowledge are concerned, these theories can be referred to for help when evaluating the extent and efficacy of the processing done by each method.

As discussed later, the current study's method requires learners to learn target words and to create links between target words and related words that they already know, which in turn helps them to retrieve the target words. Therefore, while attention and cognitive processing should be encouraged to the maximum in the process of learning, the appropriate

balance of attentional resource spent on target words and the links between "known + unknown" pairs should be taken into consideration. In this regard, the generative model (Slamecka & Graf, 1978) and the involvement load hypothesis (Laufer & Hulstijn, 2001) are informative. As these theories were developed from the *depth of processing* hypothesis of Craik and Lockhart (1972), they emphasize the importance of higher cognitive load, and detailed information to effectively realize it is well organized. As for the latter perspective, the type of processing-resource allocation (TOPRA) model (Barcroft, 2002) is especially important.

First, the generative model is a theory that regards the process of interpreting and integrating new information using prior knowledge as important for learning. To this end, this theory emphasizes the importance of generating learners' original output with target items utilizing related existing knowledge, which is called *generative use*. Although the methods in the current study did not require the participants to generate something on their own, use of known related words involves the integrative processing of known and unknown words. In this sense, this theory is partly helpful in evaluating the effects of known related words or instructions on learning.

Next, the involvement load hypothesis also regards it as important to make full use of a learners' cognitive resources and motivation. In this theory, to raise the *involvement load*, the learning burden effective for better retaining target items, three components of cognitive processing should be considered. First, *need* is the extent to which learners think the information of unknown words is needed for completing the task. Need is moderate when the information of unknown words is required by teachers or tasks, while it is strong when the student feels an intrinsic need to search for the information. Second, *search* is the process of looking for information on unknown words. The strength of the search does not change regardless of the source of information (e.g., teachers or dictionaries). Third, *evaluation* is the

process of confirming the correctness of the information found. If this process includes only the comparison of words given as choices, evaluation is moderate. If learners also evaluate meaning and usage of words so that the words fit into the context, evaluation is strong. This theory was originally put forward for evaluating the effectiveness of tasks for incidental vocabulary learning. However, Nation and Webb (2011) highlight that this theory can be applied to the evaluation of intentional vocabulary learning to some extent. Therefore, this theory is also helpful in discussing the effects of known related words or instructions on learning.

However, for acquiring the form-meaning link and other aspects of knowledge at the same time, the processing should not be interpreted straightforwardly. In this case, it is not that the higher the cognitive load is, the better the memory is retained, because instruction requiring too high a cognitive load has some risk of drawing learners' attention away from the target words. This view is exactly what the TOPRA model holds. In this model, a learner's attentional resource is assumed to be limited, and the ratio of attentional resource allotted for learning each aspect of vocabulary knowledge is directly reflected in the strength of that learning. In addition to this model, it should be noted that the main target of initial learning of unknown words should be the form-meaning link of the target words. As shown in Bogaards (2001), depth of vocabulary knowledge can more easily be acquired when the form-meaning link of the words is well established. For evaluating the effects of a certain learning method, this theory is also important.

2.6 Use of Known Related Words in Intentional Vocabulary Learning

2.6.1 Possibility of known related words

As explained above, when many pieces of related unknown information are learned, the memories are more easily forgotten than usual because of memory interference (Anderson &

Reder, 1979; Baddeley, 1997). However, if newly acquired knowledge is related to already consolidated memories, the new memory is far less subject to the inhibitory effect. In the same way, newly learned lexical memories do not suffer if they are associates of known words (Nation, 2000).

As for the establishment of memory, integration of the lexical memory, which is first stored in the form of unstable episodic memory, into learners' mental lexicon is the key. According to Lindsay and Gaskell (2010), spaced repetition and sleeping facilitates this integration, which could lead to a reduction of the inhibitory memory interference. Therefore, as also suggested by Nation (2000), related words should be introduced after at least several days after the first target words are introduced.

As long as known related words are selected, related words can be effectively integrated into methods for introducing or reviewing new vocabulary knowledge. For example, known related words, especially synonyms, can be simply used as hints in vocabulary quizzes or similar assignments to review vocabulary knowledge. Glosses in reading can utilize synonyms or antonyms. These glosses can also encourage students to think more than when they are just told L1 translations.

In addition, Aizawa and Mochizuki (2010) suggest a useful form of question for reviewing vocabulary knowledge as shown in Figure 2.1. Furthermore, semantic mapping can give students more opportunities to intuit unknown vocabulary by word association network (Figure 2.2). This method is typically used as a pre-reading activity to introduce and activate words in students' memory that may appear in texts (Ushiro, 2009). It can also be used at the stage of review after reading through texts and consolidating the newly introduced vocabulary; since several days after the first introduction of new words, it is expected to be consolidated and works as bases of association. One notable strong point of this activity is that it can involve students by having them find frequent or useful expressions through

association on their own. As mentioned above, this stimulation of intrinsic need for learning words can result in a fairly good effect on retaining memory (strong need; Laufer & Hulstijn, 2001). This can also be interpreted as a form of noticing the gap (Swain, 1995), because it is nothing but the process of finding what they cannot say in their L2. Thus, teachers and learners can make use of known related words for vocabulary learning in many ways.

Then, how effective is the method of learning in the form of "known + unknown" word sets? Compared to simple one-by-one vocabulary learning with word lists, this method has some potential effects on learners' lexical memory.

Select synonyms for word 1–7 from the box below and write it down in the parentheses.

1. multiple ( )		2. gradually ( )	
3. start ( )		4. actual ( )	
5. later ( )		6. glad ( )	
7. large ( )			

ability	after	always	begin	big	funny
happy	many	real	slowly		

Figure 2.1. Example of matching task utilizing known synonyms (adapted from Aizawa & Mochizuki, 2010, p. 115).

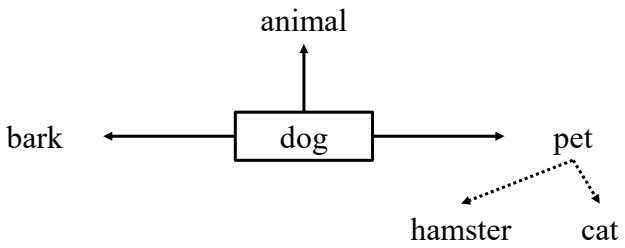

Figure 2.2. Example of semantic mapping (adapted from Tada, 2017, p. 15).

First, there are theoretical and empirical supports for the positive effects of establishing the form-meaning link of target words. In theory, it is assumed that associating target words with known related words benefits vocabulary knowledge by increasing the clues for retrieval in a learner's memory (Stahl & Nagy, 2006). As mentioned earlier, this process of associating new words with known related words is nothing but the integration of new knowledge utilizing prior knowledge, which is the basis of the generative model of learning (Slamecka & Graf, 1978). Some previous studies have reported the benefits on form-meaning link of "known + unknown" sets of synonyms (Webb, 2007) and collocations (e.g., Kasahara, 2010). In these studies, the memories of target words were retrieved better, especially when the known related words were presented to participants as clues. Kasahara suggests that this is because of links between target words and known related words. Similar to the explanation of Stahl and Nagy (2006), known related words are considered to facilitate the retrieval via the links established in learning in the form of "known + unknown" word sets.

Moreover, when the "known + unknown" pairs are prepared using collocations, as it was done in Kasahara's experiments (2010, 2011, 2015), there is a possibility of acquiring collocational knowledge along with the form-meaning link of target words. It is because the links between target words and collocates, which can become good clues for retrieving target words, are also considered to be an initial stage of collocational knowledge itself (Sonbul & Schmitt, 2013). When learning collocations, learners' attention should effectively be drawn to collocations in such ways as input enhancement (Sonbul & Schmitt, 2013). Since this method deliberately draw learners' attention to collocations, the effects on collocational knowledge can be expected.

Also, there is an additional possible benefit in the case of synonyms. As mentioned earlier, distinction among synonyms can be refined by comparing target words and synonyms and grasping the similarities and differences between them (Channell, 1981; Nation, 2000).

Without this deeper understanding, learners' expressions become vague or wrongly connotated and their communication becomes less natural or fluent.

Basically, other aspects of vocabulary knowledge than form-meaning link, such as collocational knowledge, are considered to be incrementally acquired through incidental learning in communicative contexts (e.g., Schmitt, 2008). Therefore, it would be useful if collocational knowledge and more precise meaning could be learned at the same time as the target words themselves.

2.6.2 Empirical studies of "known + unknown" word sets

Although the possibility of "known + unknown" word sets is highlighted in many pieces of previous studies, several points about this method are unknown both in terms of whole picture and detailed concrete information. To clarify what is known and what still requires further research, this section gives detailed information on the most relevant empirical studies.

First, Webb (2007) attempted to clarify the effects of known synonyms. In the experiment, 84 Japanese EFL learners memorized 20 words (10 words alone and the other 10 words with known synonyms) in a counterbalanced within-participants design. The participants intentionally learned target words in word lists or glossed sentences in eight minutes. The conditions of word lists and glossed sentences were designed as a between-participants factor.

Immediately after the learning phase, posttests were conducted to measure five aspects of lexical knowledge: (a) form, (b) form-meaning link, (c) grammatical function, (d) syntagmatic association, and (e) paradigmatic association. Each of these tests had two formats for measuring the receptive and productive levels of knowledge. Therefore, 10 formats of tests were carefully arranged so that former tests do not affect latter ones.

The productive version of the tests on form required the participants to write the spelling of target words by hearing the words. The receptive version asked the participants to choose the correct spelling of the target words from resembling distractors in multiple-choice format without any other hint.

For testing form-meaning link, the productive version required the participants to recall the L2 form by seeing the L1 translation. The receptive version was the opposite; it required the participants to recall the L1 translation by seeing the L2 form.

To test grammatical function, the productive version required the participants to write a sentence using target words. The forms of target words were provided and the answer was scored as correct if the part of speech of the target words was correctly reflected in the sentence. The receptive version required the participants to choose the sentence in which the part of speech of the target words was correct in multiple-choice format.

Assessing syntagmatic association, the productive version required the participants to write as many words as possible, which are commonly encountered in context with target words. The receptive version required the participants to choose the words that meet the same condition in multiple-choice format.

The productive version of the test for paradigmatic association required the participants to write as many words as possible that are paradigmatic associates, namely synonyms, antonyms, hypernyms, hyponyms, and co-hyponyms. The receptive version required the participants to choose the words that meet the same condition in multiple-choice format.

The result of the analysis indicated that the overall scores were significantly higher when the participants learned words with their known synonyms, while the context (i.e., the glossed sentences) did not significantly affect the results. In the 10 types of tests, productive knowledge of paradigmatic association, productive knowledge of syntagmatic association, and receptive knowledge of form were significantly facilitated by presenting known

synonyms. Although this study did not conduct a delayed posttest, it was implied that memorization of words is facilitated by the presence of known synonyms to some extent. In addition, knowledge of word association was presumably transferred from known synonyms.

Second, Kasahara (2010) demonstrated the effects of known collocates. In this study, 39 EFL students from a Japanese high school memorized 20 words (10 words alone and the other 10 in a form of "known + unknown" two-word collocation) in a within-participants design. The participants intentionally learned target words in word lists in five minutes.

After the learning phase, immediate posttests were conducted to measure their memorization of target words. This can be considered as the test of receptive knowledge, because it requires the participants to write L1 translations of given L2 forms. The posttest had two formats: a plain recall test of the target words alone and a cued-recall with known collocates given as clues. The target words learned alone did not have clues in the cued-recall format. In this research, delayed posttests of the same format were also conducted one week after the immediate posttest.

Since the results of the tests are a little complicated, they are summarized in Figure 2.3 and Figure 2.4. These results show that words learned alone were better retrieved in the tests of plain recall. This tendency was the same both in the immediate posttest and the delayed posttest. However, the difference between the scores of immediate and delayed posttests were bigger in the words learned alone, as indicated by the significant interaction. On the other hand, on the cued-recall test, target words learned with known collocates were better retrieved. This tendency was the same both in the immediate posttest and the delayed posttest. Again, the interaction was significant, indicating that the words learned alone were more prone to be forgotten.

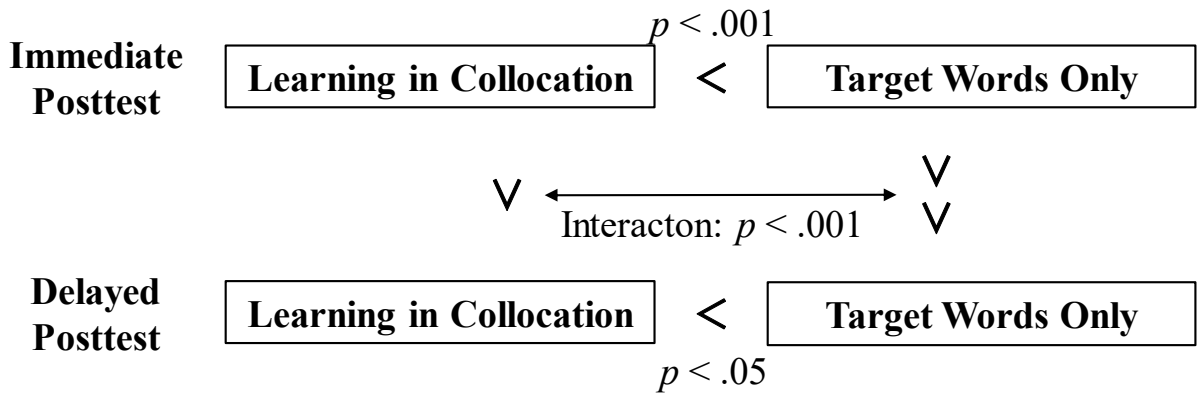

Figure 2.3. Results of plain recall test in Kasahara (2010).

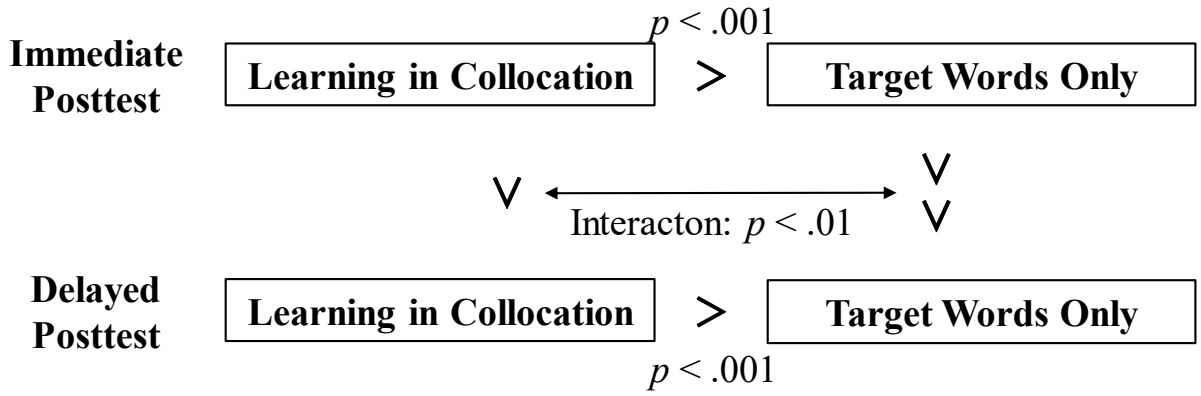

Figure 2.4. Results of cued-recall test in Kasahara (2010).

In a similar way, Kasahara (2011) conducted an experiment in which the condition of learning (i.e., target words alone or "known + unknown" two-word collocations) was the between-participants factor. In each of the conditions, the participants learned 20 of the target words. This experiment also had an immediate posttest and a delayed posttest conducted one week after the learning phase. The format of the tests corresponded to the conditions of learning: for the participants who learned the target words alone, plain recall tests were conducted, while the participants in the other condition took the cued-recall test with known collocates.

Results simply show that the target words learned with known collocates were better retrieved. This tendency was the same both in the immediate posttest and the delayed posttest. Again, the interaction was significant, indicating that the words learned alone were more prone to be forgotten.

Kasahara (2015) conducted yet another experiment to compare the effects of different types of collocations. For the target nouns, known verbs and adjectives that can collocate were prepared, and the impacts of these two types of known collocates were compared in a between-participants design. In each of the conditions, the participants learned 20 target words. This experiment also had an immediate posttest and a delayed posttest conducted one week after the learning phase. Both posttests had two formats: a plain recall test of target words alone and a cued-recall with known collocates given as clues.

The results basically supported the former studies (Kasahara, 2010, 2011), although in this study there was no condition where participants learned target words without collocates. The scores of cued-recall were better than those of plain recall, for both the immediate and delayed posttests. The type of clues did not make any significant difference in the scores.

Thus, the positive effect of presenting known collocates prior to cued-recall tests can be seen as reliable, because several studies have shown similar results.

This effect exerted only on cued-recall might sound very limited, but it can be considered fruitful for the following reasons. First, giving the cue and facilitating retrieval is important considering that retrieval benefits learners' lexical memories (Barcroft, 2007). Retrieval in itself strengthens memory of the target words more than just seeing or being taught the answers (*retrieval practice effect*; Baddeley, 1997). In addition, as indicated by the contrasting results of acquisition of implicit knowledge, learners must retrieve their new lexical memory several times before the memory is firmly established as implicit knowledge. While the repeated intentional learning of words led to implicit knowledge (Elgort, 2011),

infrequent opportunities for learning showed less impact on the implicit knowledge of collocations (Sonbul & Schmitt, 2013) Thus, giving clues repeatedly could help learners.

2.6.3 Limitations of the previous research

The previous research has showed the effect of presenting known related words on learning. It has the potential of improving both the acquisition of form-meaning link and other aspects of knowledge. However, since this line of research is limited up to now, many points about this method are yet unclear.

First, it is unknown whether the effects of known synonyms can be generalized to other types of paradigmatic associates. Teachers and learners could benefit from understanding whether known categorical associates are also effective, or which type of paradigmatic associates have the greatest impact. As the types of related words affected the inhibitory memory interference (Hoshino, 2010), it is possible that the types also affect the learning of "known + unknown" word sets. In addition, the degree of similarity among target items might affect results and should also be taken into consideration (Ishii, 2015).

Second, it should also be clarified which of paradigmatic and syntagmatic associates are better for facilitating the retrieval of target words.

Third, although these "known + unknown" word sets are considered to be effective on other aspects of vocabulary knowledge than the form-meaning link, little has been verified in past research. Although Webb (2007) included many types of tests to assess each aspect of vocabulary knowledge acquired, the study did not include the precise distinction between synonyms highlighted in Channell (1981). In addition, although Kasahara (2010, 2011, 2015) made use of the links between target words and their known associates, the link itself, collocational knowledge, was not measured. The study required the participants to recall the meaning of collocation as a whole, but the participants only needed to answer the meaning of

words already known from the beginning. Therefore, the cued-recall test in the experiment (Kasahara, 2010, 2011, 2015) did not measure the learners' knowledge of which words can collocate with the target words.

Fourth, more concrete ways of further enhancing the positive effects of known related words should also be investigated. Various methods of achieving this, according to the types of related words, include instruction to control learners' attentional resource, or the number of known related words presented for increasing the integrative processing (Slamecka & Graf, 1978) and clues for retrieval (Stahl & Nagy, 2006).

To investigate these questions, four Experimental Studies were prepared following the design of the previous study (Kasahara, 2010). Therefore, although it may be natural and even effective for learners to use example sentences when intentionally learning words, example sentences were not included in the materials. First, giving a context or example sentences has already been found in this line of research to be not so effective in intentional vocabulary learning (Webb, 2007). Second, there is a possibility that the characteristics of example sentences may confound the results if example sentences are included in the experiment. Therefore, the current study focused on the effects of known related words and the method of their use. The organization of the experiments in the current study will be clarified in the next chapter.

Chapter 3
The Present Study

As discussed above, presenting known related words in intentional vocabulary learning potentially leads to better memory of breadth and depth of vocabulary knowledge. However, since previous studies on learning "known + unknown" word sets are limited in number, the whole picture of this method, as well as concrete ways to further enhance the effect, are still unknown.

Therefore, the overall purpose of the present study is to get a clear picture of how known related words should be utilized for making the most of their positive effects. To be more specific, the purpose is to clarify, compare, and maximize the effects of known related words on form-meaning link and other aspects of target word knowledge between the cases of paradigmatic associates and syntagmatic associates. To this end, Figure 3.1 summarizes what each Experimental Study of the present study intends to clarify.

Experimental Studies 1 and 2 focus on the cases in which the known related words are paradigmatic associates. As reflected in Areas 1 and 2 in Figure 3.1, Experimental Study 1 was an attempt to confirm that the positive effects found in the case of known synonyms (Webb, 2007) are generalizable to the case of known categorical associates. As Ishii (2015) posits, the extent of how closely associates are related may affect the link in a learner's memory. Therefore, relatedness of target words and their known categorical associates was taken into consideration. Three conditions were made to compare test results: learning with high-relatedness categorical associates, learning with low-relatedness categorical associates, and learning target words alone.

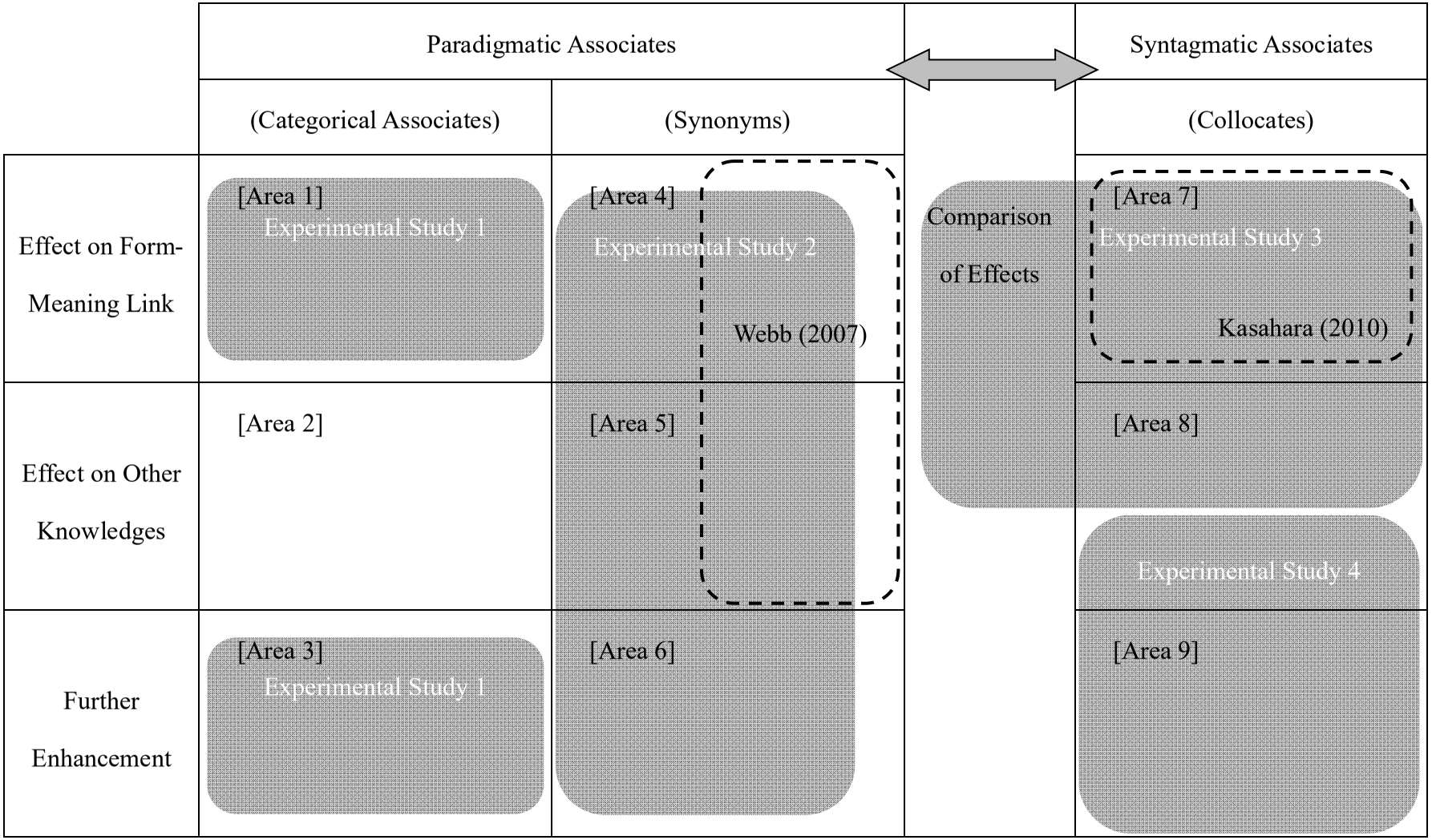

Figure 3.1. The role of each experimental study.

Experimental Study 2 delved into the effects of known synonyms on form-meaning link and other aspects of vocabulary knowledge. As depicted in Areas 4 and 5 in Figure 3.1, the points not clarified by Webb (2007) were investigated. More specifically, the effects on delayed posttests (i.e., retention and retrieval), cued-recall, and the precise understanding of meaning were not verified in the previous research. Experimental Study 2 also investigates a more concrete option to further enhance the facilitative effects of known synonyms by controlling the number of known synonyms given together (Area 6). To deny the possibility that the facilitative effect was caused by the similarity of situations between the learning phase and the posttest (*encoding specificity principle*; Thomson & Tulving, 1970), unrelated words were given to the participants in the control condition instead of target words alone. Whether relatedness affects the memory of target words is also focused on in Experimental Study 2 to further clarify the implications of Experimental Study 1. Including a number of presented related words, experimental conditions were divided into 2 (one or two synonyms) × 2 (semantic closeness of synonyms).

Experimental Study 2 uses synonyms instead of categorical associates, whose effects on other aspects of vocabulary knowledge than form-meaning link (Area 2) are still unknown. The reason for this is that categorical associates are presumed to only develop the ability of paradigmatic association, which may be correlated with communicative language fluency. However, as discussed above, it is not certain that paradigmatic association really contributes to fluency to the extent that learners should deliberately try to develop the knowledge of paradigmatic association. Therefore, this point (Area 2) was not examined in the current study.

On the other hand, Experimental Studies 3 and 4 examine cases in which the known related words are syntagmatic associates. Experimental Study 3 compares the effects of known paradigmatic and syntagmatic associates. In addition, as Kasahara (2010, 2011, 2015) did not focus on collocational knowledge of target words, the effect of known collocates on

collocational knowledge was also examined. These points are depicted in Areas 7 and 8 in Figure 3.1. Two conditions were made to compare test results: learning with categorical associates and learning with known collocates. Since it is already known from the previous research (e.g., Kasahara, 2010) and the preceding Experimental Studies that both known categorical associates and known collocates are effective for retrieving target words, a control condition was not made.

Finally, Experimental Study 4 delves into a more concrete possibility for further enhancing the facilitative effects of known collocates by comparing instruction on how to use them. This experiment is mainly focused on Area 9 in Figure 3.1. In this experiment, it was examined whether the "known + unknown" collocations should be intentionally learned as whole collocations or not. Two conditions were made to compare test results: learning collocations at the same time and learning target words while referencing the collocations as help. A control condition was not made due to the same reason as in Experimental Study 3.

The reason why Experimental Study 4 compares different instructions instead of the number of known related words given is that giving two related words was not found to improve the effect on learning in Experimental Study 2. In the following four chapters, each of the Experimental Studies are explained in detail.

Chapter 4

Experimental Study 1: Effects of Known Categorical Associates and Their Semantic Relatedness

4.1 Purpose and Research Questions

As highlighted in the previous sections, the inhibitory effect of "unknown + unknown" paired word learning has been verified in various types of relationships, including antonyms, categorical associates, synonyms, and thematic associates (Hoshino, 2010; Papathanasiou, 2009; Tinkham, 1997). On the other hand, the knowledge about the effects of "known + unknown" word pairs are not currently as generalizable. Specifically, the positive effects of known related words have been verified in cases of collocates and synonyms (Kasahara, 2010, 2011, 2015; Webb, 2007), but not in cases of other types of relationships (e.g., categorical associates). Therefore, one purpose of the current experiment is to clarify whether the memorization and retrieval of target words are enhanced by presenting known categorical associates.

In addition, it is still unknown whether and how relatedness (i.e., closeness of semantic relationship) of target and known related words affects learning. Therefore, to clarify the effect of relatedness, the current experiment also took this into consideration.

To approach these two questions, words were presented in pairs of target words and known related words in the current experiment. This was because "known + unknown" word pairs are the most effective form of studying. If more than two target words are associated with one known related word, the target words will probably end up bringing about memory interference because of their semantic relationship. Thus, the research questions (RQs) of this experiment are as follows:

RQ1: How are intentional learnings of target words affected when presented with known categorical associates?

RQ1-1: Are memorization and retrieval of target words facilitated if learners are presented with categorical associates?

RQ1-2: How are memorization and retrieval of target words affected by the closeness of the semantic relationship between target and related words?

Following Kasahara (2010), the current experiment was designed to measure and compare both the memorization and retrieval of target words in three conditions. In one condition, 10 target words were presented with closely related words. In another condition, 10 other target words were paired with weakly related words. The final condition was the control condition, in which 10 target words were given alone.

Posttests were conducted twice in order to get a better view of participants' memories. Immediate posttests were conducted soon after the learning phase to measure the memorization of target words. Additionally, delayed posttests were performed one week after the first session to gauge the retrieval of target words. The two posttests had the same format and were divided into two parts. The first format required participants to recall L1 translations of target words, whereas the second part required the same process while presenting known related words learned together as cues. Details of the experimental procedures and results are described in the following sections.

4.2 Method

4.2.1 Participants

In Experimental Study 1, a total of 22 Japanese EFL learners participated. All of them were graduate and undergraduate university students whose majors were diverse, including

English, psychology, pre-med, science, and engineering. Seventeen participants were male, and five were female. The participants had learned English for more than six years as a component of formal education in Japan. English proficiency of the participants was estimated by the self-reports of their qualifications, such as grades from the EIKEN Test in Practical English Proficiency (Grade 3 to Pre-1: Grade 3, *n* = 3; Grade Pre-2, *n* = 2; Grade 2, *n* = 5; Grade Pre-1, *n* = 2) or a TOEIC® listening and reading test score (450 to 790; average = 636.25). From this data, it can be assumed that the participants' English proficiencies ranged from elementary- to intermediate-level.

4.2.2 Materials

4.2.2.1 Target words and known categorical associates

As the target words for the current experiment, 30 low-frequency nouns were prepared so that none of the target words were known by participants prior. Each of the target words had one high-frequency related noun with higher relatedness and another with lower relatedness. The target words were chosen from level 8 or higher in the *JACET List of 8000 Basic Words* (The Japan Association of College English Teachers [JACET], 2003) except *flesh* (level 3) and *equilibrium* (level 5). The word *flesh* was given an untypical meaning (i.e., *肌色*), and *equilibrium* was included because it was considered unknown to the participants. Most of the words were, in reality, confirmed unknown to the participants in the pretest. As for the known categorical associates, 94.2% of them belonged to basic or derived forms of loanwords or level 4 or lower in *JACET List of 8000 Basic Words* (JACET, 2003). Either of the two related words was presented together with a target word according to experimental condition.

It might lead to relatively low ecological validity to learn low-frequency words, especially in the case of elementary-level participants, because low-frequency words are not

likely to be learned in the process of gradual development of lower-proficiency learners. Therefore, it might seem more valid to prepare pseudowords which have the meanings of high-frequency words as the target words. However, from the viewpoint of research ethics, participants should have merits of contributing to research (Schmitt, 2010), and deception should be avoided whenever possible. If participants can learn real words in the experiment, the acquired knowledge will be a merit to them, and presenting pseudowords as if they were real words can be considered a form of deception. As such, the author chose to use real words for the materials.

In the current experiment, all of the target words were nouns. This is because focusing on nouns is practical for the experimenter as well as learners and teachers. Nouns tend to have more categorical associates than words of other parts of speech because many nouns are included in hierarchical networks of meaning and tend to have hypernyms or hyponyms (Miller & Fellbaum, 1991). For example, *elephant*, *giraffe*, or *monkey* are included in the *animal* category. In this case, *animal* is a *hypernym* of those words, *monkey* is a hyponym of *animal*, and *elephant* is a co-hyponym of *monkey*. Taking this feature of nouns into consideration, it is sufficiently practical or ecologically valid to use nouns as target and related words in order to investigate whether learning vocabulary with known categorical associates facilitates learners' memories.

Additionally, for the purpose of making lexical features of materials less diverse, most of the categorical associates were co-hyponyms, which are, as explained above, on the same level in the hierarchical semantic networks.

4.2.2.2 Conditions and control

To investigate the effect of known categorical associates and their relatedness to target words, three conditions were prepared. In one experimental condition, categorical associates

with higher relatedness were presented when participants learned target words. In the second experimental condition, categorical associates with higher relatedness were presented in the same way. In the control condition, target words were provided alone. The semantic relatedness of target words and known categorical associates, high or low, was determined through a pilot test (see Section 4.2.2.3).

Of the 30 target words, 10 were allotted to each of the three conditions. As described later, Experimental Study 1 and all of the following Experimental Studies adopted within-participants' experimental designs. Therefore, each participant learned 30 target words in total through the three conditions described above.

Some lexical properties of target words were controlled among the three conditions in order to prevent those traits from affecting the results and make the experimental comparison more accurate. To be more specific, familiarity, frequency, and number of letters were controlled. Detailed information about the control of each property is explained in detail below.

First, regarding familiarity and frequency, the value of target words was controlled by referring to the *NTT database series* (Amano & Kondo, 1999, 2000), so that the average value of familiarity and frequency was not significantly different among the three conditions. Following the methods from previous research (Dunning, 1993; Leech, Rayson, & Wilson, 2001), the author used logarithms for the value when analyzing and averaging frequencies. This transformation was conducted because the distribution of the word frequency value follows Poisson distribution and deviates from a normal distribution, compromising the accuracy of statistical comparison (Wolter & Gyllstad, 2013).

In controlling familiarity and frequency, the author used the database of Japanese (participants' first language). This method was adopted because the study is more valid if the nature of L2 learners' mental lexicon is taken into account. According to Jiang (2000, 2002),

even in the case of advanced learners, memories of L2 word forms tend to be linked to the meanings of their L1 equivalents in their mental lexicons. Additionally, when participants interpret the meanings of the target words—which are supposed to be seen for the first time—they inevitably rely on the meanings of their L1 translations. Therefore, in the participants' processing of target words, the frequency and familiarity of L1 translations are assumed to be reflected in this study.

Secondly, word length can also affect the learnability of target words, as reported in such studies as Ellis and Beaton (1993). In fact, control of word length is also conducted in this line of research (Kasahara, 2010, 2011, 2015; Webb, 2007). Therefore, in the current experiment, the total length of ten L2 target words in each condition was controlled so that there was no significant difference of total word length among the three conditions.

In addition to the above control, the order of learning conditions was counterbalanced utilizing the Latin square design (Barcroft, 2002). As shown in Table 4.1, this counterbalancing also shuffled the combination of target words and learning conditions. To realize this, three different booklets were prepared for Experimental Study 1, and each of the three types of booklets was allotted to one third of the 22 participants. This process is important for reducing other possible confounding effects, such as a serial position effect. Serial position effect is the effect of the presentation order of materials. Some information presented at the beginning and some presented in the end tend to be memorized better than the rest, which are called the *primacy effect* and *recency effect*, respectively (Ushiro, 2009). In other words, without this counterbalancing, it is possible that the first condition or the last condition would yield unreasonably high score because of these effects. Thus, the possibility of a serial position effect was prevented by this counterbalancing.

However, because of this counterbalancing, it became necessary to control another trait of words among the conditions, namely the relatedness between target words and their related

words. This was because the comparison between conditions would have become unfair if the average relatedness of one condition had been higher or lower than that of other conditions.

Table 4.1

Counterbalancing of Conditions and Order

	Participant Group 1 (n = 7)	Participant Group 2 (n = 7)	Participant Group 3 (n = 8)
words 1–10	Presented with HRW	Presented with LRW	Presented Alone
words 11–20	Presented with LRW	Presented Alone	Presented with HRW
words 21–30	Presented Alone	Presented with HRW	Presented with LRW

Note. HRW and LRW stand for related words of high relatedness and related words of low relatedness, respectively. This table was adapted from Tada (2017, p. 26).

Table 4.2

Control of Word Features Among Three Groups of Target Words

	Words 1–10		Words 11–20		Words 21–30	
	M	*SD*	*M*	*SD*	*M*	*SD*
Relatedness to HRW	4.59	0.71	4.45	0.76	4.63	0.53
Relatedness to LRW	2.65	0.49	2.56	0.69	2.52	0.48
Number of letters	7.70	2.87	7.80	1.87	7.90	3.14
Frequency[a]	2.35	0.84	1.94	0.75	1.79	0.82
Familiarity	4.96	1.59	4.87	1.12	5.53	0.51

Note. [a]Logarithms for each frequency value were averaged. This table was adapted from Tada (2017, p. 27).

In summary, five aspects of word characteristics (i.e., log frequency, familiarity, word length, relatedness of lower-relatedness categorical associates, and relatedness of higher-relatedness categorical associates) were controlled for the preparation of materials for Experiment Study 1. The descriptive statistics of the values for lexical properties are illustrated in Table 4.2. The author confirmed that the average of these values was not significantly different among the three conditions using one-way analysis of variance (ANOVA): relatedness to closely related words, $F(2, 27) = 0.21$, $p = .818$, $\eta_p^2 = .015$; relatedness to slightly related words, $F(2, 27) = 0.15$, $p = .866$, $\eta_p^2 = .011$; word lengths, $F(2, 27) = 0.01$, $p = .986$, $\eta_p^2 = .001$; logarithm of frequency, $F(2, 27) = 1.32$, $p = .284$, $\eta_p^2 = .089$; and familiarity, $F(2, 27) = 0.95$, $p = .284$, $\eta_p^2 = .066$.

4.2.2.3 Pilot study: Evaluation of semantic relatedness

As mentioned above, the relatedness of categorical associates was evaluated in a pilot test. The method for evaluating semantic relatedness followed that of Jiang (2002), in which five native speakers rated the semantic relatedness of word pairs in a 5-point Likert scale. In the current study, a questionnaire listing 172 pairs of candidate word pairs was given to participants who did not participate in the main experiment. Then, the participants were asked to rate how close or easy it was to associate the paired words. The participants of this pilot test were Japanese EFL learners. Therefore, although the data of semantic relatedness yielded by this method are not perfectly objective, it is advantageous that the perceptions of L2 learners' mental lexicons could be reflected.

In the pilot test, a 6-point Likert scale was used for rating instead of Jiang's (2002) 5-point scale. This change was intended to prevent participants from rating at the middle of the scale when they were uncertain about their judgment. The word pairs were provided with their L1 translations in order to prevent participants from marking relatedness without knowing the

meanings of words. Specifically, low-frequency target words were likely to be unknown to participants. This presentation of L1 translations in the questionnaire will presumably not contaminate the result significantly. This is because, as described in Jiang (2000), semantic knowledge of L2 learners is dependent on their L1 translations in their mental lexicons, even in the cases of known words. The format of the questionnaire is shown in Appendix A-1. The pilot test lasted less than 60 minutes on average.

The participants of the pilot study were 13 graduate and undergraduate students who majored in English pedagogy. Similarly to the participants of the main experiment, they had learned English for at least six years and had not lived or studied in English-speaking countries for a long time. Therefore, the participants of the pilot study and the main experiment were recruited from almost the same population.

One thing that should be noted is that their major is related to English, which means that the participants of the pilot study were considered more proficient than the participants of the main experiment. However, this difference presumably did not make much difference when processing the meanings of material words. This is because the frequency of target words is so low that almost none of the participants had previous knowledge of them, and the frequency of categorical associates were high enough to already be known by most of the participants. Processing of low-frequency words is not significantly different regardless of learners' proficiencies, because the processing is highly dependent on L1 translations, even in the cases of proficient learners (Jiang, 2000, 2002). Processing of high-frequency categorical associates is also considered to almost be the same because most of the words were supposed to be known even by less proficient learners.

The ratings of the participants were averaged and adopted as values of relatedness in the main experiment. As a result of the pilot test, categorical associates with higher relatedness than the average (3.47) were considered higher-relatedness words, and those with

lower relatedness than this average were considered lower-relatedness words. In the current experiment, each target word must have one higher-relatedness word and one lower-relatedness word. Therefore, word pairs for the main experiment were selected from candidate pairs so that target words would have higher- or lower-relatedness categorical associates with significantly different relatedness values. As a result, the average values of higher and lower relatedness were 4.56 (3.63–5.92, SD = 0.65) and 2.58 (1.23–3.23, SD = 0.53), respectively, and the difference was significant: $t(58) = 12.75$, $p < .001$, Cohen's $d = 3.29$. Appendix E shows the complete list of target and related words.

4.2.3 Procedure

Experimental Study 1 was carried out from August to September 2015. The participants underwent the experiment individually for approximately 90 minutes between two sessions. The first session mainly required the participants to memorize the target words and take immediate posttests. The second session was conducted one week after the first, and participants took delayed posttests. Before starting the experiment, the purpose and procedure of the experiment and the security of personal information were explained to the participants, and informed consent was obtained. After finishing all processes of the experiment, the participants were paid for their participation.

Participants underwent all the experimental procedures following the instructions on booklets. For ease of understanding, an outline of the experimental procedure is provided in Figure 4.1. As previously explained, three versions of booklets were prepared to counterbalance the conditions and word presentation order (Table 4.1). One of those booklets was presented to each participant. Detailed information about each step was described in the following.

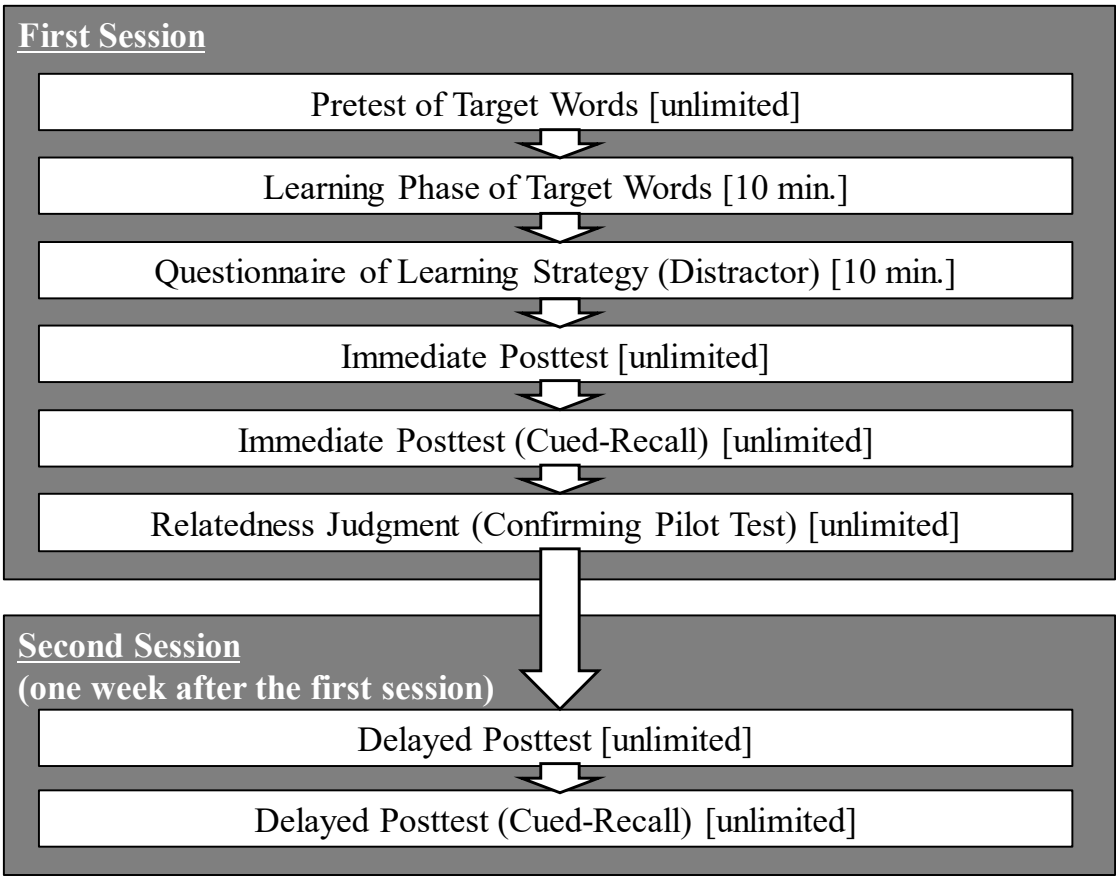

Figure 4.1. Procedure of Experimental Study 1. (Adapted from Tada, 2017, p. 32)

The first task for the participants was a pretest of target words and categorical associates. The test was a recall test of word meanings, which confirmed whether the target words were truly unknown and the categorical associates were known for the participants.

The second part was the learning phase, in which the participants were asked to memorize meanings (L1 translations) of target words within 10 minutes. When learning target words, the participants could use the learning strategy they preferred (e.g., self-testing or repeatedly writing down the target words). The only instruction given to participants was to associate the target words with their related words when provided. This instruction was intended to prevent the participants from ignoring the presented categorical associates and failing to pay attention to the relationships of the word pairs. This attention is important

because developing connections between known and unknown information makes memory more durable (Slamecka & Graf, 1978) and helps retrieval by increasing cues (Kasahara, 2010; Stahl & Nagy, 2006). Also, in the real situations where learners were taught or learned associates of known words, they were supposed to pay attention to the relationships. L1 translations of related words were also provided because it was possible that their meanings were unclear due to polysemy of words or vague knowledge of participants. The format of the word list for the learning phase is depicted in Appendix B-1.

After the learning phase, 10 minutes were taken for a questionnaire and an interview where the participants were asked about the learning strategies they use in their daily learning and how related words were utilized in the learning phase. This part was actually intended to serve as a distractor to prevent participants' short-term memories from affecting the performance of posttests. In other words, there is a risk that the short-term memories of the last few target words participants saw unfairly raise thescore of the condition that includes those words—if immediate posttests were conducted rightafter learning without this process. Therefore, this process enabled the posttests to more precisely gauge participants' knowledge acquired in their long-term memories.

Next, participants took the immediate posttests of target words. In both the immediate and delayed posttests, two measurements were taken following Kasahara (2010, 2011). The first test was a plain recall test for measuring memorization of receptive knowledge. In this part, participants wrote down L1 translations of target L2 words without any cue. The second test was a cued-recall test for measuring retrieval of receptive knowledge. This test added the known related words that were presented together in the learning phase as cues for retrieving the memories of target words. The cued-recall test was conducted after the plain recall test for the purpose of preventing the earlier test from affecting the result of the later one. The target words of the control condition, which were presented alone, were excluded from the cued-

recall test because there were no cues that could be used. The order of target words was randomized, and the pretest word list for the learning phase, immediate posttest, and delayed posttest presented words in different orders. This randomization was intended to prevent the effect of encoding specificity (Thomson & Tulving, 1970). Encoding specificity principle highlights that any situational or contextual information that accompanies an acquired memory can affect memory if it is present at the time of retrieval. Therefore, this randomization made the situation different than the same order of presentation. The results of these immediate posttests suggest how well the participants could memorize the target words.

The final task in the first session was the relatedness judgment, which had the same format as the pilot test. As explained above, the method of pilot test still has the drawback of subjectivity, and the perceptions of the participants of the main experiment might differ from the results of the pilot study. By conducting this task, the validity of the division between closely and slightly related words was confirmed. All of the 60 word pairs (i.e., 30 target words × 2 categorical associates) adopted for the main experiment were tested in this part. In the same way as the pilot test, average scores on the 6-point Likert scale were compared between higher- and lower-relatedness words.

The second session was scheduled one week after the date of the first session. In the second session, participants underwent delayed posttests of the same format as the immediate posttest. The results of the delayed posttests indicate how well the participants could retrieve the target words.

Throughout the experimental process, all oral instructions, answers to participants' questions, and explanations in booklets were presented in Japanese, which was the native language of the author and all participants. Because of this, the explanation to the participants was precisely conveyed.

4.2.4 Scoring and data analysis

The answers to the posttests were scored by awarding one point for each correct answer. If the answers were wrong, no points were given. Although it was seldom, some answers were clearly a result of mistakes of Chinese letters. If the answers were correct without the mistakes of Chinese letters, a point was given to the answers because they are considered not the result of forgetting the meanings of target words, just the wrong processing of Japanese.

Before averaging or analyzing the scores, some cases were discarded in situations where they might be problematic if they were included in the analysis. First, target words that were already known at the point of pretest were omitted (2.73%: 18 out of 660 cases). Second, since it had not been confirmed before the main experiment that the known related words were really known to the population of participants, this was also checked in the pretest. As a result, the cases where known related words that the participants did not know before the test were also discarded (23.41%: 103 out of 440 cases). After excluding those data, the rate of correct answers calculated from the remaining data was used for averaging and analyzing the scores of each condition.

The average scores of the test were analyzed with a two-way repeated measures ANOVA. The analysis was conducted twice because the plain recall test had three conditions whereas the cued-recall test had only two conditions. The plain recall test was analyzed with 3 (conditions) × 2 (immediate or delayed posttest) ANOVA, and the cued-recall test was also analyzed with 3 × 2 ANOVA, which included the average score of the control condition from the plain recall test. This inclusion of the control condition of the plain recall test followed the method of Kasahara (2010), which could clarify how effective giving cues is for retrieval.

4.3 Results

4.3.1 Relatedness judgment

The result of the relatedness judgment task for the participants of the main experiment is shown in Table 4.3. Overall, relatedness was rated similarly to the pilot test, and it can be concluded that the division of higher- and lower-relatedness categorical associates was valid.

Table 4.3

Descriptive Statistics of Relatedness Judgment for the Participants of the Main Experiment

	M	95% CI	*SD*
High relatedness	4.48	[4.22, 4.75]	0.71
Low relatedness	2.74	[2.45, 3.04]	0.79

The averages of higher- and lower- relatedness values were 4.48 and 2.74, respectively. Although this task was dependent on individuals' subjective perceptions, variation of rating, which was reflected on SDs, was not large. As a result of a two-sided *t*-test, the difference of relatedness value of higher- and lower-relatedness categorical associates was significant: $t(58) = 8.95$, $p < .001$, Cohen's $d = 2.31$.

4.3.2 Recall test without cues

The descriptive statistics of the plain recall test are summarized in Table 4.4. In the immediate posttest, the average score was around 70% in all conditions, whereas it was about 50% in the delayed posttest. The value of SDs was around 25, which is relatively large, presumably depending on each participant's ability. At the phase of immediate posttest, the score of control condition was about 1.2 times higher than the condition of higher-relatedness categorical associates. On the other hand, there were overlaps of 95% CIs between immediate

and delayed posttests in both experimental conditions. Therefore, giving known categorical associates may be disadvantageous for plain recall tests at first, but it may help retain and retrieve the memories of target words.

Table 4.4

Descriptive Statistics of Posttests Without Cues

	Immediate posttest			Delayed posttest		
	M	95% CI	*SD*	*M*	95% CI	*SD*
High relatedness	68.6	[55.3, 82.0]	30.1	50.0	[38.4, 61.6]	26.2
Low relatedness	72.3	[60.9, 83.6]	25.6	54.1	[43.3, 64.8]	24.2
Target only	75.5	[63.5, 87.4]	26.9	51.4	[39.9, 62.8]	25.9

Note. These values indicate percentages of correct answers in each condition. This table was adapted from Tada (2017, p. 34).

However, the result of ANOVA showed that the difference of descriptive statistics explained above was only a trivial one. The only significant difference was found between the overall scores of immediate and delayed posttests. As a matter of course, scores for the delayed posttests were lower than those of the immediate posttests: $F(1, 21) = 53.25$, $p < .001$, $\eta_p^2 = .72$. On the other hand, the three conditions in learning made no significant differences: $F(2, 42) = 0.85$, $p = .434$, $\eta_p^2 = .04$. The interaction was also found to be insignificant: $F(2, 42) = 0.98$, $p = .384$, $\eta_p^2 = .05$.

4.3.3 Cued-recall test

The overall results of the cued-recall test are summarized in Table 4.5. In the immediate posttest, the average score was 86.8 and 83.6 in the condition of higher and lower relatedness,

respectively. In the delayed posttest, the score was 82.3 and 71.8 in the condition of higher and lower relatedness, respectively, and the difference increased clearly. The SDs were mostly under 20.0 and looked smaller than in the plain recall test. However, this may be partly because of the ceiling effect. In the same way as the plain recall test, overlaps of 95% CIs were found between immediate and delayed posttests in the two experimental conditions.

Table 4.5

Descriptive Statistics of Posttests with Cues

	Immediate posttest			Delayed posttest		
	M	95% CI	*SD*	*M*	95% CI	*SD*
High relatedness	86.8	[79.7, 94.0]	16.2	82.3	[73.8, 90.7]	19.0
Low relatedness	83.6	[75.7, 91.6]	17.9	71.8	[61.6, 82.0]	23.0
Target only	75.5	[63.5, 87.4]	26.9	51.4	[39.9, 62.8]	25.9

Note. These values indicate percentages of correct answers in each condition. This table was adapted from Tada (2017, p. 35).

As for the significance test, the interaction between conditions and time was found to be significant this time: $F(2, 42) = 9.10$, $p = .001$, $\eta_p^2 = .30$. Combined with the descriptive statistics, this interaction indicates that target words learned with known categorical associates were better retrieved at the time of the delayed posttest, while this difference did not appear at the time of the immediate posttest. To take a closer look at the difference between conditions, the author proceeded to the subordinate analysis. Figure 4.2 and 4.3 are provided in the next section for a clearer view of the complex results explained below.

As a result of the analysis of simple main effects, the scores of the three conditions were found to be significantly different in the immediate posttest: $F(1.53, 32.22) = 5.00$, p

= .019, η_p^2 = .19, and in the delayed posttest, $F(2, 42) = 19.93$, $p < .001$, η_p^2 = .49. Simple main effects of time were also significant (*p*s < .01), except in the case of learning with higher relatedness: $F(1, 21) = 3.48$, $p = .076$, η_p^2 = .14.

To get more information on the difference among conditions, a Bonferroni post-hoc test was conducted for multiple comparisons. The results of immediate posttest yielded no significant difference (all *p*s > .016). In the delayed posttest, on the other hand, the facilitative effect of known categorical associates was clear both in the cases of higher relatedness ($p < .001$) and lower relatedness ($p = .003$) compared to the control condition. There was no significant difference between higher- and lower-relatedness categorical associates in facilitating retrieval of target words ($p = .109$).

4.4 Discussion

4.4.1 Recall test without cues

The analysis of the plain recall test showed that there was no significant difference among conditions, regardless of whether the test was conducted immediately after or one week after the learning phase. This result exhibits that association with known related words cannot enhance the memorization and retrieval of target words without any cue. This is not unnatural because previous research (Kasahara, 2010, 2011, 2015) also reported that known collocates did not affect the performance of plain recall test.

However, it is still premature to conclude that presenting known categorical associates has no effect on intentional vocabulary learning. In fact, the research on the effects of known synonyms demonstrated facilitative effects on some aspects of vocabulary knowledge in the immediate posttest without any cue (Webb, 2007). For some reasons explained below, the result of Experimental Study 1 alone cannot assure the effect of known related words.

First, as can be read by Table 4.4, it should be noted that the result of the immediate posttest in the control condition showed a ceiling effect (average + SD > full mark). The situation was similar in other conditions, although the statistics did not reach the standard of ceiling effect. This might imply that the immediate posttest was possibly a little too easy for the participants. In that case, scores of all conditions approached the full score, and the difference among the conditions might have become unfairly small.

Moreover, although the participants were instructed to pay attention to the relationships between target words and their categorical associates presented together, the experiment could not perfectly reproduce the real situation of intentional vocabulary learning. When learners learn words that are associates of known words, they notice the words themselves that are unknown and should be learned. This step is important for learners because this processing is equivalent to noticing what the learners cannot express in L2 (noticing the gap; Swain, 1995) and cultivating intrinsic motivation to learn them (strong need; Laufer & Hulstijn, 2001; Nation & Webb, 2011). Thus, the treatment of the present study, which just listed known related words with target words, might be a little less effective in stimulating learners' motivations and attention on known associates. This might have made the effect of known categorical associates smaller than in real pedagogical situations.

For the reasons described above, this experiment alone cannot perfectly deny the possibility that known related words have some positive effects on learners' memories, even if they are not presented at the time of retrieval. Specifically, as shown in Webb (2007), there are potentially some effects on other aspects of knowledge than form-meaning linkage of target words.

4.4.2 Cued-recall test

Unlike the plain recall test, the analyses of the cued-recall test showed a clear rise of the test score compared with the case where target words were presented alone. More specifically, the multiple comparisons among the three conditions in delayed cued-recall posttests yielded a distinct difference between the control condition and the two experimental conditions (see white boxes in Figure 4.2). This means that providing known categorical associates learned together can facilitate retrieval of target words, even if there are only slight relationships between target and related words. It can also be interpreted that finding unknown words by association or showing known categorical associates while introducing new vocabulary would make an effective hint for learners to recollect their memories.

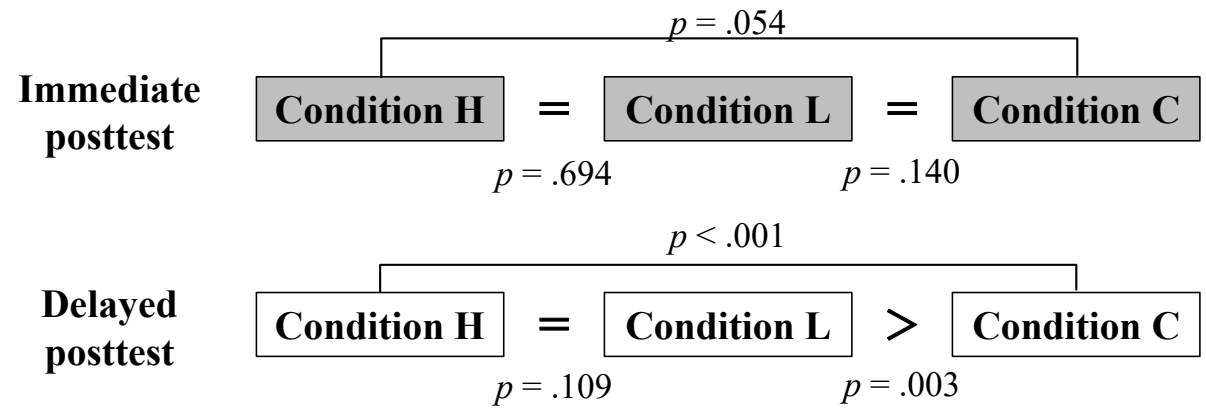

Figure 4.2. Results of multiple comparison of cued-recall among conditions. White boxes indicate the comparisons focused on in discussion. H and L represent high and low relatedness, respectively, while C represents control condition. (Adapted from Tada, 2017, p. 38)

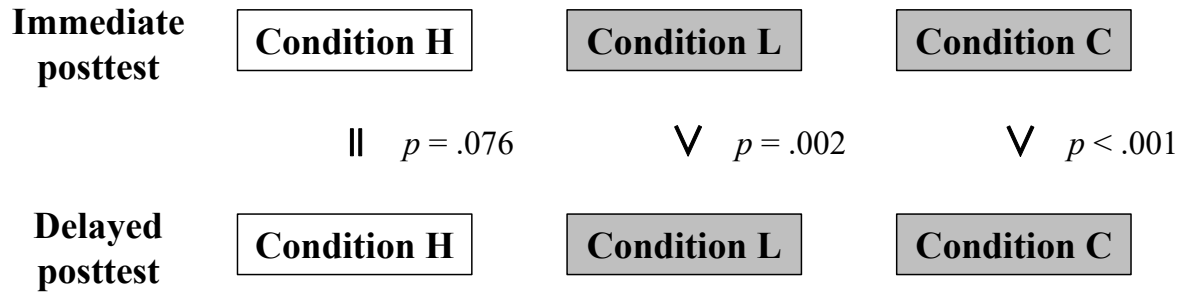

Figure 4.3. Results of multiple comparison of cued-recall between immediate and delayed posttests. White boxes indicate the comparisons focused on in discussion. H and L represent high and low relatedness, respectively, while C represents control condition. (Adapted from Tada, 2017, p. 38)

As explained in Chapter 2, retrieving memory by oneself more effectively strengthens the memory of target words than just being given the answer (Baddeley, 1997). Also, this result partly supports the possibility of positive effect on memory by activation (Wharton & Race, 1999). This maintains that words learned in semantic sets have a greater chance of activating the memories of other members in the sets when encountered in reading, writing, and any other tasks. Now that it has been demonstrated that known categorical associates can activate the memory of related target words, it is possible that the same thing happens in many situations other than just in the vocabulary tests. Thus, it was suggested that learning vocabulary with known categorical associates benefit learners' memories, especially when the known words serve as cues for retrieving target words.

Additionally, the possibility was suggested that the effectiveness of known categorical associates depends on their relatedness. It is true that the effect of higher- and lower-relatedness categorical associates does not seem different from the simple main effect of conditions (see white boxes in Figure 4.2). However, the simple main effect of time suggested that the retrieval of target words was not significantly different between immediate and

delayed posttest, only in the case of higher relatedness (see white boxes in Figure 4.3). Additionally, the result of delayed posttest again exhibited a ceiling effect in the case of higher-relatedness words. It presumably made the difference between the conditions of higher- and lower-relatedness categorical associates smaller than it should have been. From these discussions, it is possible to consider that there is still some possibility that closely related words have more of an effect on retrieval. Therefore, it is worthwhile to continue the investigation into the effects of relatedness in the following research.

As discussed above, known related words are considered to facilitate the retrieval of target words through the semantic link in learners' mental lexicons (Kasahara, 2010). However, it should be noted that this effect can be the result of encoding specificity. According to Thomson and Tulving (1970), recollection of memory could be facilitated if the situation of retrieval is similar to that of learning phase. In other words, anything in the learners' environment at the time of learning, such as the location of the experiment or the format of booklets, could be linked to their memory and stimulate retrieval to some degree, regardless of semantic relatedness. Therefore, due to its experimental design, the results of the current experiment alone cannot perfectly deny the possibility that the observed effect of known categorical associates was not a result of associative links or their strength in the participants' mental lexicons.

However, as explained in Chapter 4.2.3, the order of the target words was randomized in each step of measurement in order to prevent the effects of encoding specificity. This randomization prevented the facilitation of retrieval caused by an unexpected link in memories (e.g., the same word adjacent to a target word as in the learning section). Although the environment that can stimulate participants' retrieval is not limited to the order of target words, most of other situational factors were the same between the three conditions. Therefore, the situation of the posttests was supposed to be sufficiently different from that of the learning

phase. If the stronger effect of closer semantic relatedness is confirmed in the following study, it will also support the stronger effects of word association linkage over the mere effects of encoding specificity.

4.5 Summary of Experimental Study 1

In Experimental Study 1, the effects of presenting known categorical associates were examined in intentional vocabulary learning with a word list. The focus of the experiment was on the facilitative effects on retrieving target words (RQ1-1) and on the effect of different semantic relatedness (RQ1-2). The result of the experiment clarified some answers for these questions, while some parts of the discussion could not decisively be concluded.

Concerning RQ1-1, different results were yielded in the plain recall test and cued-recall test. The plain recall tests showed no significant difference of score among the three conditions, which indicates that there was almost no effect on the memorization or retrieval of the form-meaning link of the target words. The results of previous studies (Kasahara, 2010, 2011, 2015) was the same, where no effect was exerted by providing known collocates only in the learning phase. On the other hand, the result of cued-recall test showed the positive effect of giving known related words as cues for retrieval, both in the immediate posttest and delayed posttest conducted one week after learning.

Regarding RQ1-2, stronger effect of higher-relatedness categorical associates was implied from the cued-recall test. The score of immediate and delayed posttests had no significant difference only in the condition of higher-relatedness words. It can be interpreted as an indirect evidence that retention of the link between target words and known associates or the facilitation of retrieval is stronger when target words and their related words are semantically closer or easy to associate.

However, the current experiment left some issues unclear because of its limitations. Although detailed information about limitations are described in Section 9.3, some points are investigated again in Experimental Study 2. Therefore, limitations that are related to the next experiment are focused on here.

First, this experimental design could not clearly distinguish the effects of semantic link of word pairs and that of encoding specificity, the effect on memory caused merely by similar environment of learning and retrieval. Second, it was not decisively showed that whether known related words should be closely related or not. Third, the result of Experimental Study 1 is still not generalizable. Especially, it is not predictable whether the effect of semantic relatedness is different in the case of other types of semantic relationship (e.g., synonymy). Also, because the target words were limited to nouns in Experimental Study 1, it should be clarified whether similar results can be obtained when the target words are verbs or adjectives.

Now that it was proved that utilizing related words when expanding vocabulary results in positive effects on memory, it is worthwhile to investigate into the above controversial issues and get more concrete information on how to use known related words. Thus, in order to elucidate these issues further, the next experiment was designed.

Chapter 5

Experimental Study 2: Effects of Known Synonyms, Their Number, and Semantic Relatedness

5.1 Purpose and Research Questions

As highlighted in the previous studies, using known synonyms (Webb, 2007) or known collocates (Kasahara, 2010, 2011, 2015) has positive effects on the learning of target words. In addition, it was demonstrated in Experimental Study 1 that the facilitative effects of known related words on cued-recall could be generalized to the case of categorical associates. Therefore, the effect of known related words on retrieving form-meaning links of target words can now be considered to be widely applicable.

On the other hand, the effects of known related words on other aspects of vocabulary knowledge (e.g., collocational knowledge) have not been thoroughly investigated. In this regard, the effects of categorical associates are estimated to be mainly on the ability of paradigmatic association. However, as explained in Section 2.4.2, this knowledge does not necessarily contribute to fluency to the extent that it should be deliberately learned. In contrast, synonyms have some potential of enhancing learners' knowledge. Although the effects of known synonyms on depth of vocabulary knowledge has been widely examined in Webb's (2007) research, it is still unclear whether the presence of known synonyms makes the understanding of target word meaning more precise. Therefore, Experimental Study 2 focused on known synonyms and sought a more concrete method for enhancing their facilitative effects. More specifically, the following three questions are the focus of Experimental Study 2.

First, the effect of relatedness is still unknown in the case of synonyms. In Experimental Study 1, there was a high possibility that higher-relatedness categorical associates have a more facilitative effect on the retrieval of target words. In the same way,

semantically closer synonyms may have greater effects on the learning of target words. At the same time, however, there is also some possibility that the relatedness of target and related words has little to do with the facilitative effect on retrieval. According to Thomson and Tulving's (1970) encoding specificity principle, retrieval of memory can also be facilitated by a mere similarity of situation, such as the location of the experiment or the format of booklets. The problem is that previous research and Experimental Study 1 could not strongly refute the possibility that encoding specificity (Thomson & Tulving, 1970) confounded the result of cued-recall. Therefore, Experimental Study 2 aimed to prove that facilitation of retrieval is caused by the semantic link between target and related words, denying the influence of encoding specificity. In the current experiment, *similarity* of meaning was taken into consideration as the factor of semantic relatedness, because the related words of this experiment were synonyms.

Second, it has not been clarified how many related words should be presented with each target word. As explained in the literature review (Section 6.2.3), it is theoretically possible that two or more related words are more effective than one in facilitating retrieval of target words. This is because more related words may increase cognitive factors that are effective for learners' memory, such as integrative processing (Slamecka & Graf, 1978) and clues for retrieval (Stahl & Nagy, 2006). However, very few studies have experimented with the effects of numbers of known related words presented together.

Third, along with the development of form-meaning links, a precise understanding of word meanings can possibly be facilitated by contrasting target words and synonyms (Channell, 1981; Nation, 2000). It is also possible that the development of this profound knowledge is enhanced by the number or semantic similarity of known synonyms. If this effect on precise semantic knowledge is confirmed, presenting known synonyms will be demonstrated as an effective method for intentional vocabulary learning. This is because this

kind of profound knowledge is, in general, learned incrementally through incidental learning. However, this aspect of knowledge has not been focused on in previous research on this topic.

Thus, while focusing not only on form-meaning links but also on the precise understanding of meaning, Experimental Study 2 aimed to elucidate concrete information to enhance the facilitative effects of known synonyms. Clarifying an effective method of how to prepare related words is meaningful, especially for teachers, because teachers can make their tasks, review questions, or vocabulary tests more effective by applying such information. To this end, the research questions (RQs) of Experimental Study 2 were established as follows.

RQ2: How are intentional learning of form-meaning links and the precise semantic knowledge of target words affected when learners are presented with known synonyms?

RQ2-1: Does the semantic similarity of known synonyms presented together facilitate memorization and retrieval of target word meanings?

RQ2-2: Does the number of known synonyms presented together affect memorization and retrieval of target word meanings?

RQ2-3 How is a deeper understanding of target word meanings affected by the semantic similarity and the number of known synonyms presented together?

To approach these questions, similarly to Experimental Study 1, the author designed the present experiment to measure memorization, retrieval, and precise semantic knowledge of target words. Participants learned target words intentionally with a word-list under five conditions. Of these conditions, four were experimental, in which semantic similarity (high or low) and the number of presented synonyms per word (one or two) were controlled for. One was a control condition, in which unrelated words were presented alongside target words.

Since the current experiment also used a within-participants design, participants learned 60 target words in total (12 words × 5 conditions).

In the same way as Experimental Study 1, posttests were conducted twice, immediately after and one week after the learning phase. The format of the immediate and delayed posttests was the same, and immediate and delayed posttests were intended to measure memorization and retrieval of target words, respectively. Both posttests had three parts that were carefully arranged so that the former ones would not affect the latter ones. The first part was a plain recall test, which required participants to recall L1 translations of target words without any cue. The second part was a cued-recall test, which was the same format as the recall test but with known synonyms provided as cues. The third part was a multiple-choice test that measured precise understanding of meaning, following the format of Joe (1998). Detailed information about these procedures is provided below.

5.2 Method

5.2.1 Participants

Using the same population as in Experimental Study 1, 23 Japanese graduate and undergraduate students were recruited from various departments (e.g., comparative culture, humanities, industrial engineering, sociology, or urban environment). They had learned English for at least six years in an EFL environment. Thirteen participants were male and 10 were female. Many of the participants self-reported their proficiency of English in the form of a grade from the EIKEN Test in Practical English Proficiency (Grade 3 to Pre-1: Grade 3, n = 3; Grade Pre-2, n = 3; Grade 2, n = 6; Grade Pre-1, n = 1), a TOEIC® listening and reading test score (500 to 790; average = 642.50), or TOEFL ITP® test score (453 to 550; average = 494.33). From these reported qualifications, the participants' proficiency of English was estimated to range from elementary to intermediate.

5.2.2 Materials

5.2.2.1 Target words and known synonyms

Sixty low-frequency words were prepared as the target words, so that participants had no prior knowledge of them at the time of learning. The target words comprised 36 nouns and 24 verbs. The 6:4 ratio was adopted following previous research (Webb, 2007), because nouns and verbs appear in context in this proportional frequency (Kucera & Francis, 1967). In principle, the target words were selected from level 8 or higher in the *JACET List of 8000 Basic Words* (JACET, 2003). It was confirmed in the pretest that most of the target words were quite unknown to the participants.

Each of the target words had four high-frequency synonyms of the same part of speech. Two of them were higher-relatedness words that were semantically closer to the target words. The other two were lower-relatedness words, with relatively less similar meanings. As a standard of distinguishing synonyms and other words, synonyms for the target words were selected by referring to *The New Oxford Thesaurus of English* (Patrick, 2000). In order to select synonyms that were well-known to the participants, most of them (86%) were selected from words with sufficiently high frequency or familiarity. To be more specific, target words were level 4 or lower in the *JACET List of 8000 Basic Words*, loanwords often used in Japanese, or their derived forms.

The current experiment adopted a within-participants design. Therefore, the participants learned 60 target words at once. This design was conscientiously adopted, but there were admittedly some possible disadvantages. For example, there is a lower ecological validity, because in actual English education, learners are unlikely to be required to remember 60 words at once. As discussed in detail later (Section 9.3), there were also some risks of contaminating the data because of participants' fatigue or interference of memory. However, from the viewpoint of improving the statistical power of the comparison, this experimental

design was quite reasonable. First, a within-participants design is better for preventing individual differences of participants from affecting the comparison. Moreover, the difficulty of the posttests needed to be increased to prevent the ceiling effect observed in Experimental Study 1. Although it may seem to be a memory capacity overload for participants to learn 60 words at once, this is unlikely to happen considering that the memorization of 30 words was too easy in Experimental Study 1. Therefore, the increase of target words was expected to contribute to a higher difficulty than was found in Experimental Study 1. Thus, to elucidate concrete information by judiciously comparing many conditions, and to raise the difficulty of the tests and reduce the ceiling effect, the author decided to use 60 words for the experiment.

5.2.2.2 Conditions and control

As described above, the purpose of the study is to investigate the effect of semantic similarity and the number of presented known synonyms. To this end, four experimental conditions (number of synonyms [1, 2] × semantic similarity [high, low]) and one control condition were prepared. An example of presented words is shown in Figure 5.1. Each of the five conditions had 12 target words. The experiment was designed as a within-participants experimental design, and each participant underwent five conditions. Therefore, participants learned 60 words in total.

Semantic similarity between known synonyms and target words was determined through a pilot test (see Section 5.2.2.3). For each target word, four synonyms with different similarity were selected. In the condition of high-similarity synonym × 1, synonyms with the highest similarity value were presented with target words. Similarly, in the condition of low-similarity synonym × 1, synonyms with the lowest similarity value were presented. In the condition of high-similarity synonym × 2, synonyms with the highest and the second highest similarity value were presented with each target word. In the condition of low-similarity

synonym × 2, synonyms with the lowest and the second lowest similarity value accompanied target words. In the control condition, synonyms of each target word were shuffled, and unrelated words were presented alongside each target word.

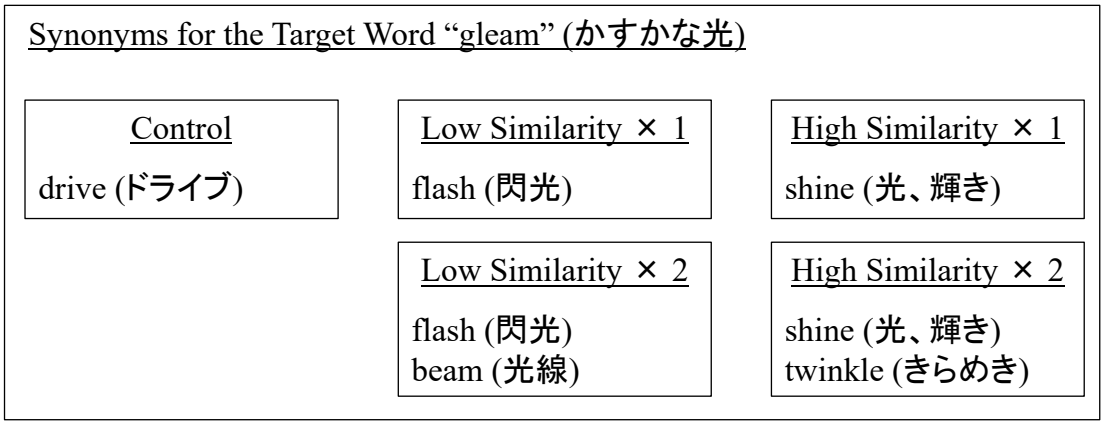

Figure 5.1. Examples of target words and known synonyms presented in each condition (with the translation equivalents in parentheses). (Adapted from Tada, 2017, p. 46)

The control condition in Experimental Study 2 is different from that of Experimental Study 1, where target words were presented alone. The purpose of this condition is to refute the possibility that the facilitation of retrieval by the existence of known related words is a mere effect of encoding specificity. In other words, the control condition was prepared in order to prove that the positive effects of known related words are not the result of the similarity of the environment between learning and testing. To this end, the only difference between the control condition and the experimental conditions was the existence of semantic similarity between target words and synonyms. Therefore, this control condition made it possible to elucidate whether the effect of known related words truly resulted from the semantic relatedness.

To control for difficulty among the experimental comparison, the characteristics of target words were controlled among the conditions, following Experimental Study 1. More

specifically, familiarity, frequency, number of letters, and part of speech were controlled, because they are expected to affect learning difficulty. The method of control is explained in detail below.

First, referring to the *NTT database series* (Amano & Kondo, 1999, 2000), the values of familiarity and frequency were averaged and confirmed not to be significantly different among conditions. Regarding frequency, logarithms for each frequency value were computed, following previous studies (Dunning, 1993; Leech et al., 2001). As explained in Section 4.2.2.2, this transformation was needed to make the distribution of the word frequency closer to normal distribution and improve the accuracy of the statistical comparison.

For the values of familiarity and frequency, data from Japanese (participants' first language) were used, because, according to Jiang (2000, 2002) L2 word forms are apt to be linked with the semantic information of their L1 equivalents even in the case of advanced learners. In the current experimental situation in particular, participants inevitably relied on the meaning of L1 translations of target words, because the target words are basically unknown to them. Therefore, the familiarity and frequency of L1 equivalents are reflected in processing.

Second, word length was controlled for so that the total length of the 12 L2 target words in each condition did not differ substantially. This is because word length also affects learnability of words (Ellis & Beaton, 1993).

Third, the parts of speech of the target words were controlled for; it has been reported in previous research that nouns and verbs have different learnability (Kawamura, 2004; Morgan & Bonham, 1944). For this purpose, target words were distributed to each condition so that the number of nouns and verbs did not differ significantly among the conditions. As a result, seven nouns and five verbs or eight nouns and four verbs were allotted to each condition.

In addition to controlling for the lexical properties, the order of learning conditions was counterbalanced with the Latin square design (Barcroft, 2002). As depicted in Table 5.1, the combination of target words and learning conditions was also counterbalanced using this procedure. As a result, five formats of word lists and posttests were made that were reflected in five different booklets. Each of the five booklets was given to one fifth of the 23 participants. As discussed in the chapter of Experimental Study 1 (Section 4.2.2.2), this process was important to reduce such effects as the serial position effect.

However, because of this counterbalancing, further control of similarity between target words and their related words was needed, because the impartiality of the comparison between the conditions would have been contaminated if the average relatedness of one condition had significantly differed from that of the other conditions.

In summary, seven aspects of word characteristics were controlled for in Experimental Study 2: (a) log frequency, (b) familiarity, (c) word length, (d) similarity of highest-similarity synonyms, € similarity of second-highest-similarity synonyms, (f) similarity of lowest-similarity synonyms, and (g) similarity of second-lowest-similarity synonyms. Table 5.2 shows the descriptive statistics of the values. As a result of a one-way analysis of variance (ANOVA), it was confirmed that these word features did not significantly differ among the five conditions: familiarity, $F(4, 55) = 0.13$, $p = .970$, $\eta_p^2 = .010$; log frequency, $F(4, 55) = 0.06$, $p = .992$, $\eta_p^2 = .005$; word length, $F(4, 55) = 0.33$, $p = .859$, $\eta_p^2 = .023$; similarity of highest-similarity synonyms, $F(4, 55) = 1.09$, $p = .373$, $\eta_p^2 = .073$; similarity of second-highest-similarity synonyms, $F(4, 55) = 0.37$, $p = .830$, $\eta_p^2 = .026$; similarity of lowest-similarity synonyms, $F(4, 55) = 0.88$, $p = .480$, $\eta_p^2 = .060$; and similarity of second-lowest-similarity synonyms, $F(4, 55) = 0.64$, $p = .636$, $\eta_p^2 = .045$.

Table 5.1

Counterbalancing of Conditions and Order

	Participant Group 1 (n = 5)	Participant Group 2 (n = 5)	Participant Group 3 (n = 5)	Participant Group 4 (n = 4)	Participant Group 5 (n = 4)
Words 1–12	Control	High Similarity × 1	High Similarity × 2	Low Similarity × 1	Low Similarity × 2
Words 13–24	High Similarity × 1	High Similarity × 2	Low Similarity × 1	Low Similarity × 2	Control
Words 25–36	High Similarity × 2	Low Similarity × 1	Low Similarity × 2	Control	High Similarity × 1
Words 37–48	Low Similarity × 1	Low Similarity × 2	Control	High Similarity × 1	High Similarity × 2
Words 49–60	Low Similarity × 2	Control	High Similarity × 1	High Similarity × 2	Low Similarity × 1

Note. This table was adapted from Tada (2017, p. 50).

Table 5.2

Control of Lexical Properties Among Five Groups of Target Words

	Words 1–12		Words 13–24		Words 25–36		Words 37–48		Words 49–60	
	M	SD	M	SD	M	SD	M	SD	M	SD
Familiarity	5.36	0.84	5.50	0.78	5.39	0.87	5.48	0.57	5.31	0.64
Frequency[a]	2.90	0.78	2.87	1.26	2.99	0.84	2.92	1.43	3.06	0.85
Number of letters	7.25	3.04	6.50	1.68	6.75	1.71	7.33	2.57	7.25	1.86
Similarity (highest)	4.29	0.44	4.01	0.45	4.34	0.53	4.28	0.37	4.13	0.47
Similarity (second highest)	3.73	0.43	3.66	0.40	3.88	0.50	3.74	0.52	3.73	0.47
Similarity (second lowest)	3.27	0.41	2.97	0.44	3.25	0.52	3.07	0.46	3.08	0.54
Similarity (lowest)	2.72	0.47	2.53	0.46	2.76	0.51	2.67	0.46	2.48	0.67

Note. [a] Logarithms for each frequency value were averaged. This table was adapted from Tada (2017, p. 51).

5.2.2.3 Pilot study: Evaluation of semantic relatedness

To evaluate the similarity of known synonyms, a pilot test was conducted. The method of the pilot study was basically the same as in Experimental Study 1. Using the same method as Jiang (2002), the relatedness of word pairs was rated by native speakers' intuition with a 5-point Likert scale. Utilizing a questionnaire listing candidate word pairs for the main experiment, the author asked participants to rate the closeness of meaning with their intuition. The Likert scale was changed to a 6-point scale to prevent participants from marking at the middle of the scale too often. Although this method cannot provide perfectly objective data of semantic similarity, there is merit in its reflection of perceptions of L2 learners' mental lexicon.

The format and scoring of the questionnaire followed that of Experimental Study 1. The averaged values of the rating were used as a similarity of each pair in the main experiment. To make participants rate similarity based on the correct knowledge of meaning, L1 equivalents of target words and related words were written in the format of questionnaire. It can be considered that this provision of L1 translations would not distort the rating, because, according to Jiang (2000), L2 forms are apt to be linked to the meaning of L1 equivalents in learners' mental lexicons. Around 60 minutes were needed to complete this pilot test.

Unlike the pilot study of Experimental Study 1, participants were also asked to mark the items if they did not know the meaning of synonyms. In this process, it was confirmed that the synonyms that were prepared as supposedly known words were actually known by the participant population. Therefore, this pilot test had three purposes: (a) examining semantic similarity of known synonyms, reflecting the perception of L2 learners' mental lexicon, (b) selecting known synonyms from the candidate pairs so that the semantic similarity value of high- and low-similarity synonyms became sufficiently separate, and (c) excluding synonyms that were often unknown to participants.

Ten graduate and undergraduate students participated in the pilot test. Unlike Experimental Study 1, the participants were recruited from almost the same population as the main experiment for the pilot test of Experimental Study 2. This was intended to increase the validity of judging similarity and checking whether the synonyms were known. Therefore, the participants of the pilot study also had various majors, such as English education or Japanese culture and language. In the same way as the main experiment, participants were graduate or undergraduate students who had learned English for more than six years in the EFL environment of Japan. However, the participants of the main experiment were not involved in the pilot test.

The averaged similarity values were used as the similarity of the main experiment. From the candidate word pairs, the author selected 240 pairs consisting of 60 target words that have two higher- and two lower-similarity synonyms each. To make sure that higher- and lower-similarity synonyms actually have different similarity values, the average values of similarity were statistically compared using a *t*-test; the average value of higher-similarity synonyms was 3.98 (2.6–5.1, $SD = 0.51$), and the value of lower-similarity synonyms was 2.88 (1.7–4.1, $SD = 0.55$). These were shown to be significantly different, $t(238) = 16.08$, $p < 0.01$, Cohen's $d = 2.07$.

There were only a few cases where the synonyms were unknown to the participants. All cases of checked synonyms added up to 2,400 cases (240 synonyms × 10 participants), and the synonyms were unknown in only 48 cases; the percentage of problematic cases was only 2.00%. Therefore, it can be assumed that most of the participants in the main experiment had sufficient prior knowledge of synonyms.

5.2.3 Procedure

An experiment was carried out on participants in August in 2016. Basically, the participants, individually, underwent the experimental treatment, which comprised of two sessions that took around 120 minutes. Before beginning the experiment, the author explained the purpose and procedure of the experiment as well as the security of personal information and took an informed consent from the participants. After finishing the second session, the participants were paid a fee.

The experimental procedure was paper-based and the participants followed the instructions given in the booklets. As mentioned above, five different versions of booklets were prepared to reflect the counterbalancing (Table 5.1). Participants were provided with one version of those booklets.

All the steps of the experimental procedure were depicted in Figure 5.2 and each step of the procedure explained in detail. The basic design of the current experiment followed that of Experimental Study 1 and the content and order of treatment was similar to the former experiment. However, Experimental Study 2 had three different points. First, the learning phase took a longer time and contained more words than Experimental Study 1. Second, multiple-choice test of precise meaning of target words was added to the immediate and delayed posttests. Third, relatedness judgement was placed after the delayed posttests in order to prevent the effect of this task on the memory of the participants, while Experimental Study 1 had this process after the first session.

As the first task of the experiment, a pretest of target words was conducted. Since it had already been confirmed in the pilot test that the synonyms prepared for the main experiment were appropriate, the main purpose of the pretest was to confirm that the participants had no prior knowledge of the target words.

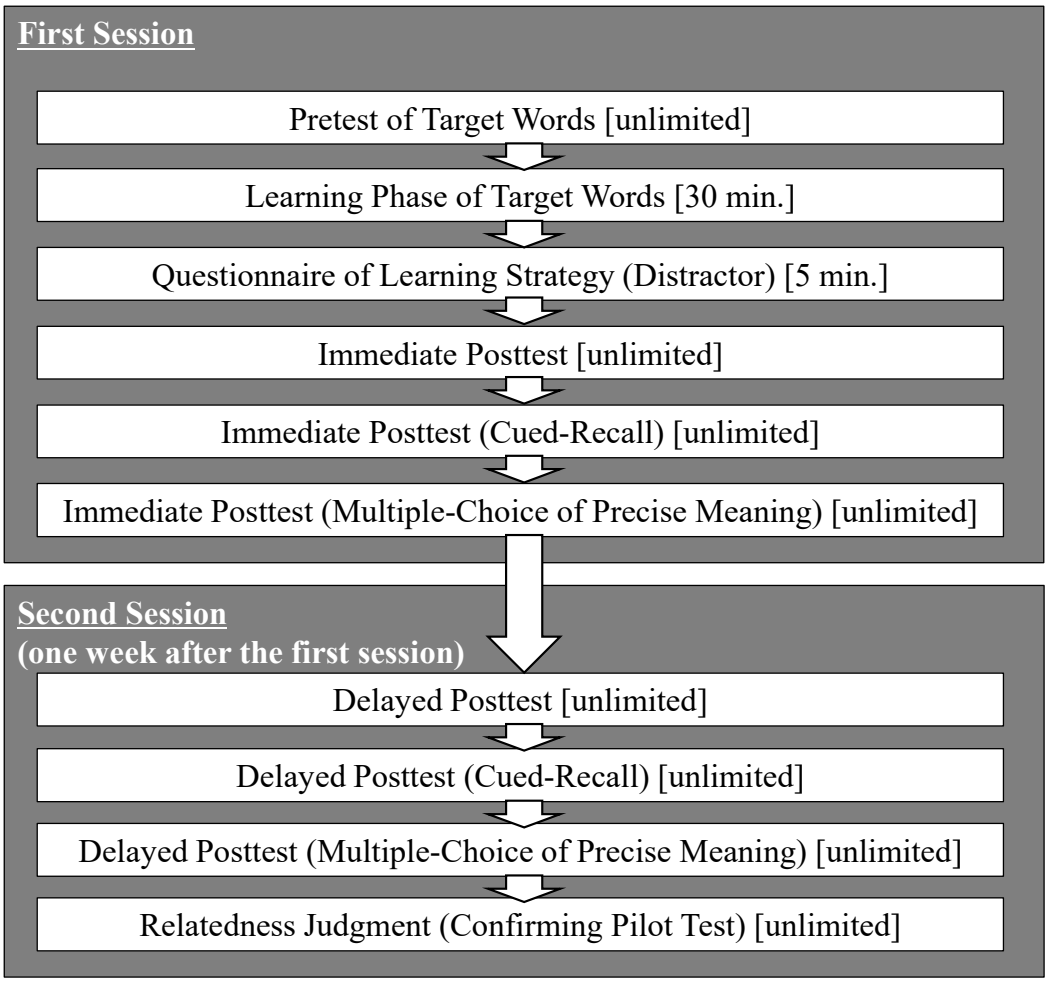

Figure 5.2. Procedure of Experimental Study 2. (Adapted from Tada, 2017, p. 55)

Second, the participants were required to learn the target words. In the learning phase of Experimental Study 2, the participants learned meanings (L1 translations) of 60 target words in 30 minutes. The participants were instructed to take care of the semantic difference between the target words and the presented synonyms. This instruction was intended so that participants make full use of the synonyms to distinguish the subtle difference in words and make their knowledge of meaning more precise. The attention to the known synonyms was also important for encouraging the cognitive processing that is effective on facilitating learning. It encourages integration of known and unknown information (Slamecka & Graf, 1978) and development of link between target words and known synonyms that helps retrieval

(Kasahara, 2010; Stahl & Nagy, 2006). As long as they followed this instruction, the participants were allowed to use any learning strategy they preferred. In the same way as Experimental Study 1, L1 translation were presented with synonyms in case their meanings were somehow unclear to the participants.

Third, the questionnaires were given to the participants. The contents of the questionnaires were the same as that of Experimental Study 1 and it asked the participants of the learning strategies they used when they learnt English vocabulary. This part took five minutes and was intended to eliminate target words from the participants' short-term memory. This process was needed because it could affect the result of posttests, the purpose of which was to examine what was acquired in the participants' long-term memory. However, the interview administered in Experimental Study 1 was not done in the current experiment. It was intended to alleviate the participants' burden, because Experimental Study 2 took much more time and had much more target words to be learnt than in the Experimental Study 1.

Fourth, immediate posttests of target words were conducted. Both immediate and delayed posttests included three measurements to answer the research questions. The first test was a plain recall test, with the format being the same as that of Experimental Study 1. It required the participants to write down the meanings of target words, in order to measure memorization of receptive knowledge. In this format, accurate answers were required to show that the participants could distinguish the meaning of target words from their synonyms, which had a similar, but slightly different meaning. The second test was a cued-recall test, which also had the same format as that of Experimental Study 1. This format presented all the known synonyms that were learned together as cues for retrieval. The third format was a multiple-choice test which measured the precise understanding of meaning. Following the format of Joe (1998), the meanings of distractors were made out to be similar to the correct choice. To realize this, L1 translations of the four synonyms were presented as distractors for

each target word. In the same way as Experimental Study 1, the order of test items was shuffled in each format of the posttests, in order to exclude encoding specificity (Thomson & Tulving, 1970).

Lastly, delayed posttests were implemented a week after the date of the first session. The three types of measurements had the same format as immediate posttests, but the order of the items was randomized. It is assumed that retrieval of target words is reflected on the results of the delayed posttests.

5.2.4 Scoring and data analysis

The method of scoring was basically the same as that of Experimental Study 1. If an answer was correct, one point was added, while no points were given for wrong answers. As mentioned above, the answers were carefully scored to ensure that points were awarded only when the participants recollected the meaning on their own. In other words, if the participants only wrote down the meaning of known synonyms, no points were awarded.

To make this scoring method less subjective, the author asked a rater to score 30% of the answers of cued-recall tests. The rater was a graduate student who majored in English pedagogy. After that, inter-rater reliability was computed, which was found to be 98.25%, and the author and the rater discussed to resolve all the discrepancies of the rating. Then, following the standard set through the discussion, the author rated the rest of the answers.

Some problematic cases were excluded from the analysis. Firstly, if the target words were already known to the participants before the experiment, these were excluded from the data (0.65%: nine out of 1,380 cases). Secondly, the material of the experiment erroneously included the cases where the same known synonyms were presented for two different target words. These cases were also discarded (4.06%: 56 out of 1,380 cases) because the effect on retrieval could be adverse if one cue was linked to two target words. From the rest of the data,

the rate of correct answers was calculated and analyzed with 5 (conditions) × 2 (immediate or delayed posttest) repeated measures ANOVA to check if the survey or the experiment results were significant.

5.3 Results

5.3.1 Recall test without cues

Table 5.3 shows the descriptive statistics of the plain recall test. The average score of immediate and delayed posttests were about 50% and 35%, respectively. As the difficulty of the test was raised, the scores were lower than those of Experimental Study 1 and there were no ceiling effects observed. The value of SDs was as small as around 5−6, indicating that most of the participants had difficulty in a similar way and that very few participants could achieve an overwhelmingly high score.

Table 5.3

Descriptive Statistics of Posttests Without Cues

	Immediate posttest			Delayed posttest		
	M	95% CI	*SD*	*M*	95% CI	*SD*
High similarity × 1	48.3	[37.4, 59.1]	5.23	36.2	[23.3, 49.2]	6.25
High similarity × 2	42.4	[30.9, 53.9]	5.54	37.6	[21.8, 53.3]	7.59
Low similarity × 1	50.5	[40.0, 61.0]	5.07	34.7	[21.0, 48.4]	6.59
Low similarity × 2	52.0	[39.2, 64.7]	6.13	33.9	[21.3, 46.5]	6.05
Control	45.7	[34.8, 56.7]	5.28	34.6	[21.4, 47.8]	6.35

Note. These values indicate percentages of correct answers in each condition. This table was adapted from Tada (2017, p. 58)

The results of the significance test were similar to that of Experimental Study 1. The average score of immediate posttests was significantly higher than immediate posttest, F (1, 22) = 18.96, $p < .001$, $\eta_p^2 = .46$. Other significant main effect or interaction was not found. The effect of five conditions of the learning phase was not exerted on the plain recall test, F (4, 88) = 0.14, $p = .968$, $\eta_p^2 = .006$. The interaction was also insignificant, F (4, 88) = 0.52, $p = .719$, $\eta_p^2 = .023$.

5.3.2 Cued-recall test

Table 5.4 shows the descriptive statistics of the cued-recall test. In contrast to the plain recall test, the results of the significance test were complicated as described below, especially in the delayed posttest. Therefore, to make it easy to understand, Figure 5.3 and Figure 5.4 were provided to organize the results of the posttests.

Table 5.4

Descriptive Statistics of Posttests With Cues

	Immediate posttest			Delayed posttest		
	M	95% CI	*SD*	*M*	95% CI	*SD*
High similarity × 1	64.6	[54.4, 74.8]	4.92	50.7	[41.0, 60.5]	4.71
High similarity × 2	65.0	[53.2, 76.8]	5.70	51.4	[39.9, 62.9]	5.54
Low similarity × 1	57.4	[46.7, 68.2]	5.18	38.2	[29.0, 47.5]	4.46
Low similarity × 2	62.1	[49.8, 74.4]	5.93	43.2	[32.6, 53.8]	5.10
Control	46.3	[34.9, 57.8]	5.51	19.5	[12.1, 26.9]	3.57

Note. These values indicate percentages of correct answers in each condition. This table was adapted from Tada (2017, p. 59)

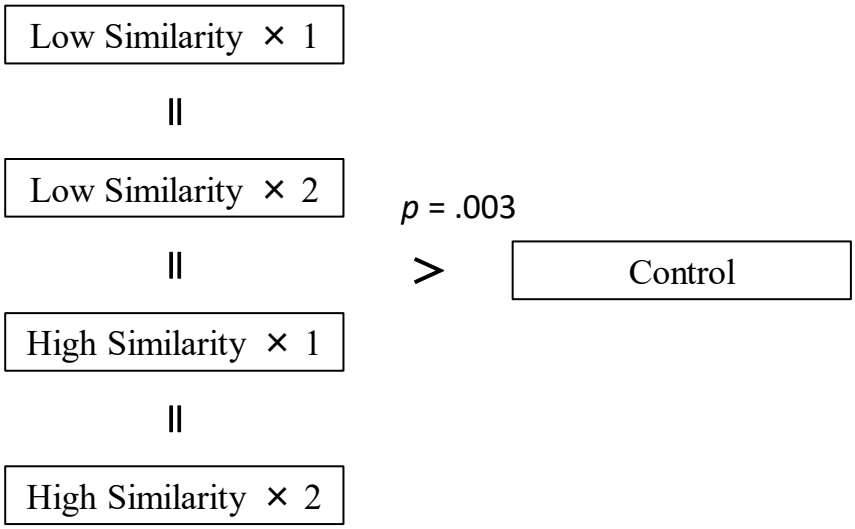

Figure 5.3. Results of multiple comparison among conditions in immediate cued-recall test.

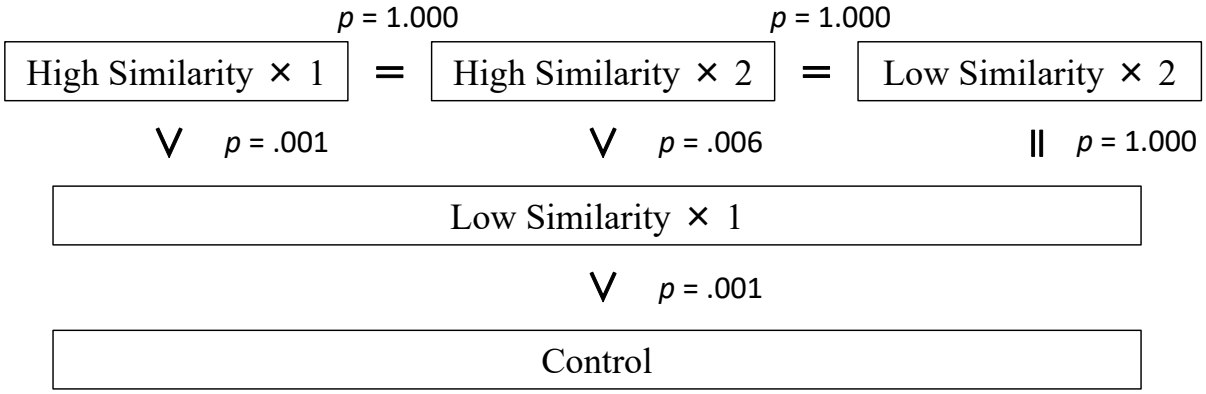

Figure 5.4. Results of multiple comparison among conditions in delayed cued-recall test.

(Adapted from Tada, 2017, p. 60)

The average score of the immediate posttest was around 50−60%. In the delayed posttest, the score was around 40−50%, except the 19.5% of the control condition. The SDs were small in cued-recall test as well, presumably reflecting the difficulty or lower average score of the test. From these descriptive statistics, it can be understood that the control condition was obviously disadvantageous and that lower-similarity synonyms resulted in slightly lower scores.

As a result of the ANOVA, the analysis yielded a significant interaction. It showed that the results of the analysis concerning the five conditions were different between immediate and delayed posttests. As for the analysis of simple main effects, the scores of immediate and delayed posttests were significantly different on all five conditions (all *p*s < .001). Additionally, the simple main effect of conditions was significant both in the cases of immediate and delayed posttests (all *p*s < .001).

Next, the difference of scores among the five conditions was further clarified through multiple comparisons using Bonferroni post-hoc test. In the immediate posttest, a significant difference was yielded between the condition of higher-similarity synonym × 1 and control condition ($p = .003$). However, 95% CIs of the two conditions showed an overlap and did not suggest this difference. The difference among other conditions were not present in this analysis.

In the delayed posttest, on the contrary, many significant differences were found among the five conditions and the complicated results were summarized in Figure 5.3. First, the score of the condition of lower-similarity synonym × 1 was significantly lower than both of the higher-similarity conditions (all *p*s < .01). However, giving one lower-similarity synonym was not disadvantageous compared with two lower-similarity synonyms ($p = 1.000$). Second, the result of control condition was even lower than that of the condition of lower-similarity synonym × 1 ($p < .001$). To sum up, the control condition and the condition of lower-similarity synonym × 1 were found to be more disadvantageous than others.

5.3.3 Multiple-choice test

Table 5.5 shows the overall results of the test of precise understanding of the meanings. The average scores were around 85% in the immediate posttest, whereas it was approximately 75% in the delayed posttest. Since the value of the SDs was small, no ceiling effects were

observed. From the descriptive statistics, there were obviously no big differences among the five conditions. However, considering that this test format required precise understanding and memory of target words, the scores were high even after the period of one week. The discussion is made from this viewpoint.

Table 5.5

Descriptive Statistics of Multiple-Choice Posttests

	Immediate posttest			Delayed posttest		
	M	95% CI	*SD*	*M*	95% CI	*SD*
High similarity × 1	87.9	[82.4, 93.4]	2.67	77.6	[69.6, 85.6]	3.87
High similarity × 2	84.0	[74.5, 93.4]	4.55	78.3	[68.5, 88.1]	4.73
Low similarity × 1	87.5	[79.6, 95.4]	3.81	82.4	[74.3, 90.5]	3.88
Low similarity × 2	82.5	[73.4, 91.6]	4.41	75.7	[65.8, 85.6]	4.78
Control	84.4	[76.6, 92.2]	3.75	78.5	[69.7, 87.4]	4.28

Note. These values indicate percentages of correct answers in each condition. This table was adapted from Tada (2017, p. 60)

As for the statistical analysis, the results were found to be simple. In the same way as the plain recall test, only the difference between immediate and delayed posttests was significant, F (1, 22) = 19.07, p < 001, η_p^2 = .46. The main effect of five conditions was insignificant, F (3.03, 66.62) = 1.13, p = .35, η_p^2 = .049. The interaction was also proved to be insignificant, F (4, 88) = 1.07, p = .38, η_p^2 = .046.

5.4 Discussion

5.4.1 Recall test without cues

The results of the plain recall test showed no significant effects except the difference between immediate and delayed posttests. As for the effect of known, related words on learning, the facilitation was not observed from the result. The result of the current experiment was relatively reliable because it did not show a ceiling effect or floor effect. Therefore, it was suggested that providing known synonyms only in the learning phase does not facilitate the performance in plain recall.

This result may be interpreted as a contradiction with the result of Webb (2007), which reported some facilitative effect on plain recall. It cannot be said that it is totally impossible, because there were some limitations in the present study, such as small sample size. However, this possibility is relatively low considering that the facilitative effect found in Webb (2007) other aspects of vocabulary knowledge rather than the retrieval of meaning.

Additionally, this result was consistent with the results of Experimental Study 1, which examined the effect of known categorical associates. This was also consistent with Kasahara (2010, 2011, 2015), which examined the case of giving known collocates. Thus, it is also highly possible that, regardless of the type of relationship, known and related words presented only when learning did not facilitate plain recall of target words.

The results did not show any effect of semantic similarity either. Since this result was consistent with that of Experimental Study 1, it is fairly possible that the relatedness of known and related words had little to do with the performance of plain recall test. However, it might be premature to conclude that relatedness has no effect regardless of the type of relationship. It is because very little research has confirmed this issue apart from the current experiment, especially in cases of types other than categorical associates and synonyms.

Concerning the number of presented known synonyms, it was theoretically possible that presenting two synonyms increased the integrative processing of new and known information, which was effective for enhancing memory of new knowledge (Slamecka & Graf, 1978). However, contrary to this expectation, presenting two synonyms did not improve the performance of plain recall. It can be anticipated that the number of known related words will not affect learning in the cases of other types of related words, because the effect on the cognitive processing is expected to be almost the same. However, in the same way as the above issues, this point needs to be verified in an experiment in the future.

5.4.2 Cued-recall test

Compared with the control condition design, the results of the experimental conditions indicated that known synonyms facilitated retrieval of target words in cued-recall. This was also consistent with Experimental Study 1 and previous studies (Kasahara, 2010, 2011, 2015). This result also partly supported the previous research that the words learned together in semantic sets can activate one another in the sets when encountered again (Wharton & Race, 1999).

The fact that the control condition yielded lower scores was also important for disproving the confounding effect of encoding specificity. Although it is theoretically possible that the cued-recall is facilitated by the similarity of environment between learning and testing (i.e., encoding specificity) rather than the semantic link between words, this possibility has not been directly refuted in previous research. In the control condition of Experimental Study 2, unrelated words were presented, in order to make the existence of semantic link the only one difference between the control and experimental conditions. Therefore, the lower score of the control condition denied the possibility of encoding specificity. It indicated that the

semantic link between target words and synonyms played an important role in improving the retrieval of target words.

From the comparison among the experimental conditions, more concrete information on how teachers should prepare known synonyms could be drawn. First, it was suggested that just one synonym was enough for each target word, because, regardless of the similarity of synonyms, the conditions of two synonyms did not result in significantly higher scores than the condition of one synonym. Second, synonyms with higher similarity should be prepared. It is because the score was lower than other experimental conditions when the participants were given only one lower-similarity synonym per one target word. To realize the same effect as higher-similarity synonym × 1, two lower-similarity synonyms were needed, which is not efficient. To sum up, from the viewpoint of efficiency, preparing one higher-similarity synonym was found to be the best way of facilitating cued-recall of target words.

When interpreting the effect of semantic similarity, it should be noted that the current experiment was designed to prevent some possible unfairness caused by different semantic similarity. For example, if the participants could have seen which synonyms were prepared as higher-similarity synonyms, they could have narrowed down the meanings to recollect more easily. Also, the participants might just have written down the meaning of the higher-similarity synonyms and those answers could have been marked as correct. However, in the process of Experimental Study 2, participants had no chance of knowing which synonyms were of higher- or lower-similarity. Moreover, the known synonyms were carefully selected so that their meanings were slightly different from those of the target words. The scoring was also carefully done to give no points for answers that were not distinct from synonyms. Therefore, the possibility of unfair factors between the different similarities was excluded in this experimental design and only the difference of semantic link in the participants' mental lexicon was considered to be reflected.

5.4.3 Multiple-choice test

Basically, the multiple-choice test required the same knowledge as the plain recall test, although the format was a recognition test. Therefore, the reason for the results can similarly be explained as the case of plain recall test.

However, as mentioned in the section of results, it is worth noting that the scores of the posttest were high, considering that precise knowledge of 60 words was required. In fact, compared with a study which measured recognition of target words after a period of one week (45.50% in the case of unrelated sets; Erten & Tekin, 2008), the score of the delayed posttest in Experimental Study 2 (78.51%) was much higher. This comparison alone cannot conclude so many things, because the two studies have many differences including the population of participants. However, it is possible to infer from this result that it had a positive effect on the memory of target words to direct learners' attention to subtle meaning of target words through comparison with known synonyms.

In the multiple-choice test, the score was generally rather high, though the difference in conditions was not observed. This could also be explained from the perspective of attentional resource. As mentioned in Section 2.5, deeper processing leads to more efficiency of learning (Craik & Lockhart, 1972). However, as suggested by Barcroft (2002), attentional resource is limited and should be properly allocated to important things to be remembered. From the results of Experimental Study 2, it can be inferred that the participants paid much more attention to comparison and precise understanding of meaning, presumably because the instructions led them to do so. Accordingly, the attentional resource spent on synonyms themselves may have been reduced, resulting in the little difference among the conditions. However, the effect on cued-recall was different, presumably because the semantic closeness served as an intrinsic difference among synonyms in the efficiency of forming links with target words.

5.5 Summary of Experimental Study 2

Focusing on the method of learning words in “known + unknown” pairs, Experimental Study 2 examined the effects of semantic similarity and the number of given synonyms on plain recall, cued-recall and precise semantic knowledge of target words. The focus of RQ2-1 was on the effect of semantic similarity; Q2-2 concerned the effect of the number of given synonyms; whereas RQ2-3 focused on the development of precise semantic knowledge. The three types of measurements clarified the answers for the questions and concrete information about how synonyms should be prepared for activities and assignments was clarified.

As for RQ2-1, the advantage of higher semantic similarity was suggested especially from the result of delayed cued-recall. It is also important that the effect of facilitating cued-recall was shown to be not the result of encoding specificity but the result of semantic link in the mental lexicon of the learners.

Concerning RQ2-2, giving one synonym per word was enough. From the summary of the results it can be suggested to teachers that preparing one higher-similarity synonym per word is the most effective and efficient way.

Regarding RQ2-3, precise understanding of lexical meanings can be facilitated by presenting known synonyms. However, it is possible that the effect is the result of the instruction to encourage learners to elaboratively compare the subtle meaning of words. Semantic closeness and the number of presented synonyms did not make any difference in this measurement.

From the results of Experimental Studies 1 and 2, it can be considered as highly possible that known paradigmatic associates can generally facilitate cued-recall of target words. The two experiments in the current study raised some tentative issues. However, before delving into detailed verification concerning the effects of paradigmatic associates, clarifying broader questions is important to get useful information as to the effects of “known

+ unknown" word pairs. Therefore, Experimental Study 3 focused on the comparison of the effects between known paradigmatic associates and known syntagmatic associates.

Chapter 6

Experimental Study 3: Comparing the Effects of Known Categorical Associates and Known Collocates

6.1 Purpose and Research Questions

While Experimental Studies 1 and 2 focused on cases in which the relationship between target words and known related words was paradigmatic, Experimental Studies 3 and 4 focused on cases of syntagmatic associates. As shown below in detail, Experimental Study 3's purpose was, in short, to compare the effect of known paradigmatic and syntagmatic associates. The current study as a whole aims to clarify what kind of known related words can be helpful in vocabulary learning, how they should be used, and in what way they can be helpful. Therefore, Experimental Study 3 plays quite an important role.

To be more specific, Experimental Study 3 focused on the points regarding the effects of known related words left unclear in previous studies, especially those conducted by Kasahara (2010, 2011, 2015). First, although Webb (2007) and Kasahara (2010, 2011, 2015) highlighted the effects of learning unknown words with known synonyms and known collocates, respectively, it is still nebulous which kind of known related words are better for learning. Second, in Kasahara's series of study, all of the benefits learners can gain from known collocates are not necessarily clarified. Kasahara (2010, 2011, 2015) focused on the effect of the cued-recall of target words, and the effect on other aspects of vocabulary acquisition were left unclear. Although Kasahara's series of study included the test form that required participants to recall the meaning of whole collocations, including target words, participants only had to write down the meaning of the known word in addition to the meanings of the target words. Therefore, this test did not necessarily clarify whether or not collocational knowledge was acquired at the same time by learning with known collocates.

To answer these questions, Experimental Study 3 focused on the comparison between the effects of known categorical associates and known collocates. In addition, to elucidate the effects of known collocates on learners' collocational knowledge, the author added a multiple-choice test that required participants to choose the one word that collocated with each target word. The reason that known categorical associates (words such as those used in Experimental Study 1) were used as paradigmatic associates is that the scoring was straightforward. More specifically, it was made easy so that the same standard of judging right or wrong answers could be adopted in both cases of categorical associates and collocates. If known synonyms had been adopted, participants would have had to accurately recall the slight difference between the known synonyms to get the right answer. This might have caused unfairness between conditions because target words with known synonyms do not require such accuracy for the right answer in a natural learning situation.

Thus, the main purposes of the experiment explained above converge in the following two research questions (RQs):

RQ3: Of known categorical associates and known collocates, which are more effective in facilitating the learning effects of target words?

RQ3-1: Are the memorization and retrieval of target words more facilitated when presented with known categorical associates or with known collocates?

RQ3-2: Is the acquisition of collocational knowledge of target words more facilitated when presented with known categorical associates or with known collocates?

To approach these questions, an experiment was designed to measure and compare both memorization and retrieval of word-list learning in two conditions. In each condition, 10 words were given with known categorical associates, and 10 words were provided with

known collocates. Posttests were conducted twice, once immediately after the learning phase and again one week later in order to get a better view of how the participants' memory of the target words changed under the possible effect of related words. The results of the immediate and delayed posttests indicated how successfully the participants memorized and retrieved the target words, respectively. The posttests were divided into three parts. The first part required participants to recall L1 translations of target words. In the second part, participants were given a multiple-choice test that asked them to choose the collocates of the target words. The third part again asked participants to write down L1 translations of the target words with known related words provided as clues. Posttest scores were compared using a 2 (conditions) × 2 (immediate or delayed posttest) ANOVA. Details of the experimental procedures are described below.

6.2 Method

6.2.1 Participants

A total of 35 Japanese undergraduate students participated in Experimental Study 3. Participants were recruited in the same way as in Experimental Studies 1 and 2, and care was taken so that none of the participants in Experimental Studies 1 and 2 were included in this experimental study. Their majors were diverse, including art and design, biological sciences, engineering sciences, engineering systems, geoscience, international studies, mathematics, medical science, physical education, health and sport sciences, physics, and policy and planning sciences. Of the participants, 16 were male, and 19 were female. They all had learned English for at least six years as a component of formal education in Japan. Many of the participants self-reported their English proficiency in the form of a grade from the EIKEN Test in Practical English Proficiency or a TOEFL ITP® score. Additionally, the Mochizuki Vocabulary Size Test was conducted in the experiment to gauge the participants' lexical

proficiency. The results showed an average vocabulary size of 5,034 words (*Min* = 4,154, *Max* = 5,846). As such, it can be assumed that the participants' English proficiency ranged from a pre-intermediate to an upper-intermediate level.

6.2.2 Materials

6.2.2.1 Target words and known related words

Twenty low-frequency nouns were chosen as target words, each of which had one high-frequency co-hyponym and one high-frequency word that could be collocated. The target words and their co-hyponyms were mostly chosen from Experimental Study 1 materials. However, some of the target words were newly chosen because all candidate word pairs (i.e., target words and their co-hyponyms) had similar relatedness values as a result of the pilot study. These were not eligible for use in Experimental Study 1 because the target words needed to have both higher-related and lower-related co-hyponyms. Table 6.1 shows the target words that were newly added for Experimental Study 3, and Appendix G shows the complete list of target and related words.

As mentioned in the explanation of Experimental Study 1, the target words were chosen from level 8 or higher in the *JACET List of 8000 Basic Words* (The Japan Association of College English Teachers [JACET], 2003). The word *equilibrium* was the only exception. Nouns were selected as the target words due to practical reasons both in the experiment and in the learners' or teachers' practice because nouns tend to have more categorical associates than other parts of speech. According to Miller and Fellbaum (1991), many nouns have hypernyms or hyponyms and are involved in hierarchical networks of meaning (e.g., *hawk*, *peacock*, or *penguin* are hyponyms of *bird* while *color* is a hypernym of *red*, *blue*, and *green*). This feature of nouns suggests that the target and related words used in Experimental Studies 1 and 3 are

sufficiently practical to investigate the effectiveness of learning vocabulary with known categorical associates.

Table 6.1

Target Words Added for Experimental Study 3

Target Words	Known Categorical Associates	Relatedness Values
cicada	butterfly	3.69
mica	glass	3.85
persimmon	apple	4.38
petal	leaf	5.15
prosecutor	court	5.23

For the known co-hyponyms, high-frequency words were selected, most of which belonged to basic or derived forms of loanwords, or level 4 and lower in the *JACET List of 8000 Basic Words* (JACET, 2003), so that the known related words were actually known in most cases, even by relatively less-proficient learners. To simulate the natural or ecologically valid situation in which teachers or learners try to make the most of the learning method utilizing known co-hyponyms, word pairs with higher relatedness values were selected for Experimental Study 3. In addition, to make the words' concreteness less diverse, the known related words were, in principle, categorized as co-hyponyms that belonged to the same level in hierarchical semantic networks, just as *hawk*, *peacock*, and *penguin* belong to the category of *bird* as its hyponyms.

For the known collocates of the target words, adjectives that made congruent collocations with the target nouns were selected from the COCA (Corpus of Contemporary

American English), the large-scale corpus of American English. The standards and reasons for selecting these collocates are as follows.

First, adjective + noun combinations were used in this study because this combination is one of the most important types of collocation. While it is true that verb + noun combinations are considered to be important in L2 learning because they are frequent (Nesselhauf, 2005), they are also difficult for L2 learners (Biskup, 1992). A verb and its object are indeed the most important parts in sentences for conveying information (Altenburg, 1993); however, when it comes to the frequency of use by L2 learners, adjective + noun combinations are most frequent, followed by verb + noun combinations (Barfield, 2009). Furthermore, from the viewpoint of fair control between the experimental conditions, verb + noun combinations, which could have an article between the verb and noun, seemed unsuitable for Experimental Study 3. It is known that Japanese learners of English tend to have difficulty with articles, especially when writing in English, because the concept of articles is not included in the grammatical system of the Japanese language (Butler, 2002; Takahashi, 2000). Articles might not affect participants' memorization and retrieval of knowledge when they only learn and examine lexical meanings, but this study, namely Experimental Studies 3 and 4, required participants to demonstrate how well they picked up the knowledge of how the presented words collocated. As such, to prevent articles from affecting the difficulty of learning collocations, adjective + noun combinations seemed better for this study. Thus, the author opted to use adjective + noun combinations for Experimental Studies 3 and 4.

Second, known collocates of target words were carefully selected so that all word pairs were semantically transparent and congruent collocations. This was because such collocations are easy for L2 learners to acquire. Transparent collocations are those expressions in which the whole meaning is the same as the sum of the component words' meanings. Previous

research has proven that transparent collocations are easier than opaque collocations for L2 learners to learn (Revier, 2009) and process (Gyllstad & Wolter, 2016). Congruent collocations are those expressions that make sense and have the same meaning in the original language when their component words are literally translated into the other language. Previous research has proven that congruent collocations are easier than incongruent collocations for L2 learners to process (Wolter & Gyllstad, 2011, 2013; Yamashita & Jiang, 2010). Therefore, when learning collocations, it is more efficient to focus on semantically transparent and congruent collocations, although this is not to say that opaque or incongruent collocations should be ignored. Furthermore, as highlighted by Schmitt (2008), learners should first focus on establishing the form-meaning link when learning unknown words because the acquisition of other aspects of knowledge, including collocations, is facilitated if learners have already consolidated the form-meaning link of the words they are learning (Bogaards, 2001). This establishment of the form-meaning link will be prohibited if learners spend too much of their cognitive resources on remembering other aspects of vocabulary knowledge (e.g., collocations) rather than focusing on remembering the form or meaning of unknown words. This is because a learner's cognitive resources are limited, and the learning effect of a certain type of vocabulary knowledge depends on how much of those cognitive resources are spent on learning the corresponding part of vocabulary knowledge (general version of the TOPRA model; Barcroft, 2002). Therefore, to prioritize the establishment of form-meaning links of target words, the author chose semantically transparent and congruent collocations to alleviate the learning burden for participants.

Last but not least, the standard for choosing collocations from the corpus, or the definition of collocation in this study, must be clarified. As mentioned earlier, many studies focusing on learners' collocational knowledge adopt a frequency-based definition of collocation (e.g., Webb et al., 2013; Wolter & Gyllstad, 2013). In this approach, the boundary

of what is and what is not a collocation is determined by its statistical strength of co-occurrence based on corpus data. This approach is least subjective (Siyanova-Chanturia, 2015) and has the advantage that the data are available for nonnative teachers and learners (Webb et al., 2013). For these reasons, the frequency-based approach is widely accepted and useful, as far as frequent and practical collocations are concerned. However, statistics based on corpus data are not always reliable and practical in use. In experiments like this, low-frequency words are used as target words for participants to remember, and collocations including low-frequency target words have very few samples, even in large-scale corpora. As such, if a frequency-based approach is strictly adopted in this study, no materials eligible as a collocate for target words can be picked up from a corpus. Thus, as the second-best solution, sequences of words are considered as collocations when the existence of the expression is confirmed in a corpus in this kind of experiment (Kasahara, 2010, 2011, 2015). In other words, the term collocation is operationalized here as a two-word pair that co-occurs often, as evidenced in a large-scale corpus, which was the approach also adopted by Sonbul and Schmitt (2013). In Experimental Studies 3 and 4, known collocates of target words were gathered following this method, and one of the two related words was presented with a target word according to the experimental condition.

6.2.2.2 Conditions and control

To investigate which kind of known related words—categorical associates or collocations—facilitate learners' memory more, two conditions were made and compared. In one experimental condition, known categorical associates were presented along with target words while, in the other condition, known collocates were presented. Ten target words were allotted to each of these conditions. Therefore, in the procedure of Experimental Study 3, each

participant learned a total of 20 target words, as described later in detail (a within-participants experimental design).

For the fairness of the experimental comparison, the target words' lexical properties that could affect learning difficulty, namely, familiarity, frequency, and number of letters, were controlled among the two conditions. First, referring to the databases by Amano and Kondo (1999, 2000), target words for Experimental Study 3 were carefully controlled so that their familiarity and frequency values were, on average, not significantly different among the conditions. Similar to the process in Experimental Study 1, frequency was controlled by averaging and analyzing the logarithms for each frequency value, following methods from past studies (Dunning, 1993; Leech et al., 2001). This is because the distribution of the word frequency value follows Poisson distribution and deviates from a normal distribution, compromising the accuracy of statistical comparison without this transformation (Wolter & Gyllstad, 2013). As discussed in the explanation of Experimental Study 1, the familiarity and frequency of L1 (Japanese) translations were controlled because this method is more valid, considering that even in the mental lexicons of advanced learners, memories of L2 word forms are apt to be linked with the meanings of their L1 equivalents (Jiang, 2000, 2002).

Second, the total length of the 10 L2 target words in each condition was controlled because a word's learnability also depends on its length (Ellis & Beaton, 1993).

Third, control for features of collocations should also be considered. Many previous studies examining L2 learners' knowledge or processing of collocations have shown that collocations' item difficulty is affected by several factors:

- frequency of collocations (Eyckmans, 2009; Wolter & Gyllstad, 2013)
- frequency of component words (Revier, 2009; Wolter & Gyllstad, 2013)
- mutual information (MI) (Ellis et al., 2008)

- number of letters (Sonbul & Schmitt, 2013)
- congruency (Wolter & Gyllstad, 2011, 2013; Yamashita & Jiang, 2010)
- semantic transparency (Gyllstad & Wolter, 2016; Revier, 2009)

However, as is often the case with research on single-word learning, it is quite difficult to control for all of the lexical properties. In reality, previous research has controlled for only those above factors that were considered important in each study. For example, Gyllstad and Wolter (2016), who investigated the effect of semantic transparency on L2 learners' processing of collocations, controlled for number of letters, frequency of component words, and congruency. Kasahara (2010, 2011, 2015), on the other hand, focused on the learning of collocations and controlled for the number of letters in the target words and the part of speech. The main purpose of the current study is to compare the learning of the target words and collocations between the conditions (learning with known categorical associates or with known collocates). Therefore, the author prioritized control for the features of the target words and controlled only for congruency and semantic transparency of the known collocates.

In addition to the control for lexical properties, the combination of target words and conditions as well as the order of learning conditions were counterbalanced. This process was intended to reduce other possible confounding effects such as the serial position effect. In the experiment, four different booklets, which reflected the counterbalancing of two groups of target words × the order of two learning conditions, were made and distributed to participants. The order of words and conditions as well as the number of participants who went through the experiment with each booklet is shown in Table 6.2.

As in Experimental Study 1, relatedness values between the two groups of target words (i.e., items 1–10 and items 11–20) were also controlled for fairness of comparison between conditions.

Table 6.2

Counterbalancing of Conditions and Order

Participant Group 1 (n = 9)	Participant Group 2 (n = 9)	Participant Group 3 (n = 9)	Participant Group 4 (n = 8)
1.Words 1–10 with known CAs 2.Words 11–20 with known collocates	1.Words 1–10 with known collocates 2.Words 11–20 with known CAs	1.Words 11–20 with known CAs 2.Words 1–10 with known collocates	1.Words 11–20 with known collocates 2.Words 1–10 with known CAs

Note. CA stands for categorical associate.

Table 6.3

Control of Word Features Between Two Groups of Target Words

	Words 1–10		Words 11–20	
	M	*SD*	*M*	*SD*
Relatedness of known categorical associates	4.74	0.71	4.52	0.65
Number of letters (target words)	7.50	2.80	8.60	2.50
Number of letters (known collocates)	6.20	2.49	5.60	1.90
Number of letters (whole collocations)	13.70	4.64	14.20	3.68
Frequency[a]	2.38	0.70	2.08	0.80
Familiarity	4.98	1.42	4.89	0.97

Note. [a]Logarithms for each frequency value were averaged.

In summary, the author controlled for several aspects of word characteristics in preparing the target words for Experiment Study 3, as illustrated in Table 6.3. A two-sided *t*-test was used to confirm that these word features were not different between the two

conditions. The analysis indicated that none of the characteristics were significantly different between the groups, including relatedness to known categorical associates, $t(18) = 0.73$, $p = .473$, $d = 0.33$; length of target words, $t(18) = -0.92$, $p = .367$, $d = -0.41$; length of known collocates, $t(18) = 0.61$, $p = .552$, $d = 0.27$; length of collocations, $t(18) = -0.27$, $p = .793$, $d = -0.12$; logarithm of frequency, $t(18) = 0.90$, $p = .382$, $d = 0.40$; and familiarity, $t(18) = 0.17$, $p = .865$, $d = 0.08$.

6.2.2.3 Pilot study

To prepare appropriate known related words, candidate materials went through both a pilot test for selecting categorical associates and an inter-rater check for selecting congruent and semantically transparent collocations.

To select categorical associates with proper relatedness, a pilot test was conducted in preparation for Experimental Study 1. As mentioned above, the target words and categorical associates for Experimental Study 3 were basically the same as those of Experimental Study 1. Therefore, the detailed information about this pilot test is explained in Chapter 5. In short, following Jiang (2002), the author used a questionnaire that listed a large number of candidate materials and asked 13 participants to rate how easy the paired words were to associate. In the pilot test, a six-point Likert scale was used for rating (Appendix A-1). Their ratings were averaged and adopted as values of relatedness in the main experiment. When providing the word pairs, L1 translations of the target words and related words were provided to prevent participants from marking relatedness without knowing the meanings of the low-frequency target words. All participants in the pilot test were graduate and undergraduate students and were not involved in the main experiment.

To select congruent and semantically transparent collocations, the author (whose native language, as well as the participants', is Japanese) carefully selected the collocates of target

words from the corpus. However, as Revier (2009) pointed out, only one native speaker's intuition regarding the judgment of meanings has some room for arbitrariness. Therefore, to ensure validity, a check and discussion with other raters was employed, as in previous studies (e.g., Gyllstad & Wolter, 2016; Yamashita & Jiang, 2010). In the current study, an additional rater checked the congruency and semantic transparency of 37 candidate collocations in a randomized list. This rater, a native speaker of Japanese with an upper-middle English proficiency, was a graduate student specializing in English pedagogy. In this rating, the standard of congruency was that the literal translation of the English collocation into Japanese was possible by paraphrasing the translated word in Japanese within the range of the same meaning. The operationalization of semantic transparency followed Howarth's Continuum Model (1996), which is widely accepted in the research that considers the transparency of collocations (Gyllstad & Wolter, 2016; Revier, 2009). As such, the rater checked whether the constituent words were used in a non-prototypical or figurative meaning and whether the collocations as a whole had a literal meaning (i.e., the same as the sum of the constituent words' meanings). When the constituent words were used in prototypical meanings and collocations as a whole had a literal meaning, the collocations were rated as semantically transparent. The agreement rate on these items was 93.8%, and disagreements were found in two cases. These disagreements were resolved through discussion, and as a result, all 37 candidate collocations passed this selection process.

6.2.3 Procedure

Experimental Study 3 was conducted from May to August in 2019. In principle, the experimental treatment was performed individually and lasted for approximately 120 minutes over a total of two sessions (e.g., the first session was mainly for memorization and immediate posttests of target words, followed by the second session of delayed posttests

conducted after one week). The purposes, procedures, and security of personal information were first explained to the participants, and informed consent was obtained prior to the experiment. They were paid for their participation after finishing all procedures in the second session.

All tasks in the experiment were presented in four different versions of booklets, which reflected the counterbalancing of conditions and word presentation order. Each participant was given one booklet, and the number of participants was randomly allocated to the four versions.

First, the participants were asked to memorize the meanings (L1 translations) of target words within five minutes (learning phase). They were allowed to make use of any strategy to learn these words (e.g., self-testing or repeatedly writing down the target words), but they were instructed on how to make use of the known related words in each condition. When the participants learned with known categorical associates, they were instructed to learn the target words and associate them with the related words when provided in order to prevent the participants from ignoring the related words and failing to be aware of how the words were related. This awareness is important because they are supposed to have it when they are taught or learn associates of known words. When the participants learned with known collocates, on the other hand, they were instructed to use the collocation example as a reference in order to encourage them to learn the collocation incidentally. The related words were presented with their L1 translations in case the participants did not know the meanings or the meanings were unclear as a result of polysemy or vague knowledge. The word list format is shown in Appendix B-1 and Appendix B-2.

In the next step, the participants were asked about the learning strategies they usually use in a questionnaire, which was intended to serve as a distractor. This task lasted 10 minutes and was administered to prevent participants' short-term memory from facilitating their

performance on the immediate posttests, thereby allowing the examination of what they actually acquired in their long-term memory. If immediate posttests were conducted right after the learning phase without this distractor task, the last few target words the participants tried to learn might have remained in their short-term memory, and the condition in which the last few target words were provided would have resulted in an unfairly high mark.

After finishing the questionnaire, participants moved on to the immediate posttests on the target words. Two tests measuring the knowledge of the target words (Kasahara, 2010, 2011, 2015) and one additional test measuring collocational knowledge were prepared to answer the research questions: (a) a plain recall test measuring memorization of receptive knowledge, which required the participants to write L1 translations of target L2 words, (b) a multiple-choice test that required the participants to choose the words that were most likely to appear in context with the target words (Webb, 2007), and (c) a cued recall test measuring the retrieval of receptive knowledge in which known related words learned together in the learning phase were provided as cues. To prevent the earlier test from affecting the participants' performance on the later test, the three measurements were implemented in this order. In the multiple-choice test, the correct answers were the words prepared as known collocates for each of the target words. All distracter words were adjectives, the same part of speech as the correct answers. Participants were required to circle one word from the four choices in the format shown in Figure 6.1. This multiple-choice test was intended to check whether a link had been built between the two component words of a collocation as an initial stage of collocational learning (Sonbul & Schmitt, 2013). Additionally, the order of test items was randomized to prevent the effect of encoding specificity (Thomson & Tulving, 1970) caused by the same order of items as in the learning phase. Overall, the results of these immediate posttests indicated how well the participants had memorized the target words, both in terms of form-meaning links and collocational knowledge.

学習した単語	選択肢			
avalanche	tall	wild	huge	thin
sleet	blue	strong	hot	sudden
persimmon	clear	real	green	adult
mane	wild	high	cheap	clean
petal	healthy	cool	basic	rose

Figure 6.1. Collocational knowledge multiple-choice test format

Finally, one week after the learning phase and immediate posttests, a delayed posttest was conducted in the same format as the immediate posttest. The results of these tests indicated how well the participants could retrieve the target words, both in terms of form-meaning links and collocational knowledge.

After the delayed posttest, a Mochizuki Vocabulary Size Test was conducted to obtain data regarding the participants' lexical proficiency. This test was developed specifically to measure Japanese learners' knowledge of English vocabulary (Aizawa & Mochizuki, 2010). It was confirmed that none of the materials used for Experimental Study 3 were included in the vocabulary size test, so the experimental procedure would not affect the score of the lexical proficiency measurement.

Throughout the experimental process, all oral instructions, answers to participants' questions, and explanations written in the booklets for completing each experimental task were conveyed in Japanese (the participants' and the author's native language) in order to ensure that the participants understood them all.

For ease of understanding, an outline of the experimental procedure is provided in Figure 6.2, and detailed information about each step has been described in this order.

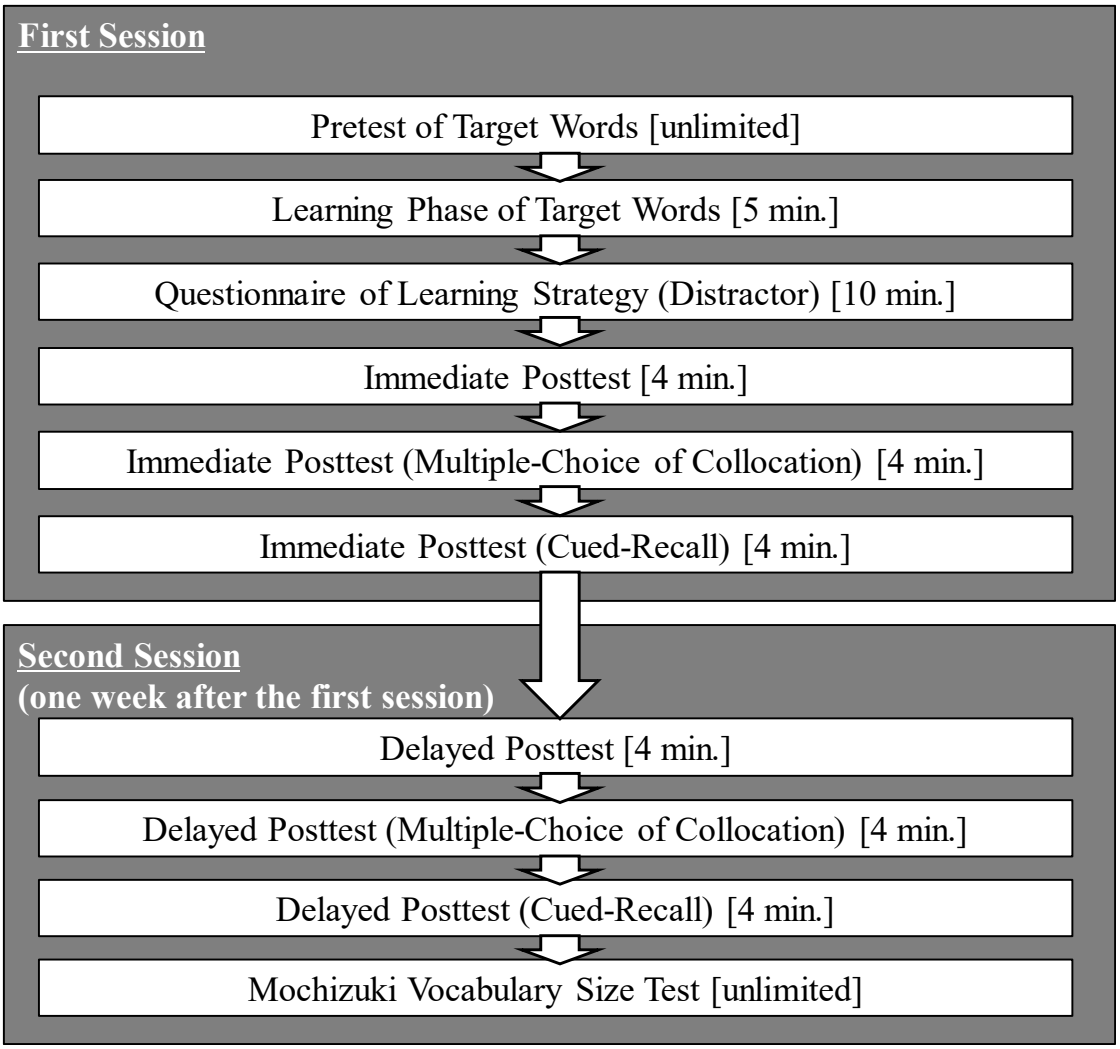

Figure 6.2. Procedure of Experimental study 3

6.2.4 Scoring and data analysis

To examine participants' memorization and retrieval by the scores on the immediate and delayed posttests, respectively, one point was awarded to each correct answer and zero points for incorrect ones. Although small in number, some answers were clearly a result of mistakes in the processing of Japanese, for Chinese letters in particular. These cases were marked as correct because they are not considered to have resulted from a failure to retrieve the semantic information of the L2 words.

As for the test score analysis, a two-way repeated measures ANOVA was conducted for each test—the plain recall of target words, multiple-choice collocational knowledge test, and cued-recall of target words—following Experimental Studies 1 and 2. All ANOVAs were 2 (known categorical associates or known collocates) × 2 (immediate or delayed posttest).

Before proceeding to the analyses, some problematic cases found in the scoring were omitted. First, target words that were already known at the point of pretest were discarded (0.43%: 3 out of 700 cases). Second, in Experimental Study 3, it had not been confirmed when the pilot test was conducted that the known related words were, in fact, known to the population of the participants. Therefore, the known related words that the participants did not know before the test were also checked by the pretest, and those cases were also discarded (22.42%: 157 out of 700 cases). After excluding those data, the rate of correct answers calculated from the remaining data was used in the following analyses.

6.3 Results

6.3.1 Recall test without cues

The overall results of the plain recall test are summarized in Table 6.4. As a matter of course, points for the delayed posttests were significantly lower than those of the immediate posttests, $F(1, 34) = 35.82$, $p < .001$, $\eta_p^2 = .51$. As in Experimental Studies 1 and 2, the kind of known related words learned together with target words made no significant differences in the test without cues, $F(1, 34) = 1.79$, $p = .190$, $\eta_p^2 = .05$, and there was no significant interaction, $F(1, 34) = 0.93$, $p = .343$, $\eta_p^2 = .03$.

Table 6.4

Descriptive Statistics of Posttests Without Cues

	Immediate posttest			Delayed posttest		
	M	95% CI	*SD*	*M*	95% CI	*SD*
Categorical	58.8	[48.5, 69.1]	29.9	39.3	[29.6, 48.9]	28.1
Collocation	51.2	[40.4, 62.1]	31.6	36.0	[26.0, 46.0]	29.2

Note. These values indicate percentages of correct answers in each condition. This table was adapted from Tada (2019).

6.3.2 Multiple-choice test

The overall results of the multiple-choice test on collocational knowledge of the target words are summarized in Table 6.5. The results simply show that the main effect of condition (learning with known categorical associates or known collocates) was significant, $F(1, 34) = 25.16$, $p < .001$, $\eta_p^2 = .43$. In contrast, the difference between the immediate and delayed posttests was insignificant, $F(1, 34) = 0.91$, $p = .35$, $\eta_p^2 = .03$, and the interaction was not significant, $F(1, 34) = 2.14$, $p = .15$, $\eta_p^2 = .06$.

Table 6.5

Descriptive Statistics of Multiple-Choice Posttests

	Immediate posttest			Delayed posttest		
	M	95% CI	*SD*	*M*	95% CI	*SD*
Categorical	45.5	[37.3, 53.6]	23.7	52.8	[44.4, 61.3]	24.6
Collocation	73.2	[62.4, 84.0]	31.4	72.5	[63.0, 82.0]	27.7

Note. These values indicate percentages of correct answers in each condition. This table was adapted from Tada (2019).

6.3.3 Cued recall test

The overall results of the test providing related words as cues are summarized in Table 6.6. The main effect of condition showed that learning with known categorical associates yielded significantly better results in the cued recall test, $F(1, 34) = 27.01$, $p < .001$, $\eta_p^2 = .44$. In addition, the difference between the immediate and delayed posttests was significant, $F(1, 34) = 18.82$, $p < .001$, $\eta_p^2 = .36$, and the interaction was not significant, $F(1, 34) = 0.79$, $p = .38$, $\eta_p^2 = .02$.

Table 6.6

Descriptive Statistics of Posttests With Cues

	Immediate posttest			Delayed posttest		
	M	95% CI	*SD*	*M*	95% CI	*SD*
Categorical	86.4	[78.2, 94.6]	23.9	81.2	[72.5, 89.8]	25.2
Collocation	65.4	[54.3, 76.4]	32.2	57.3	[46.6, 67.9]	30.9

Note. These values indicate percentages of correct answers in each condition. This table was adapted from Tada (2019).

6.4 Discussion

6.4.1 Recall test without cues

As suggested by the significance test, the measurements of the plain recall test providing no clues for the participants yielded no significant difference, regardless of whether the test was conducted just after the learning phase or one week later. As in Experimental Studies 1 and 2, this exhibits that the memorization and retrieval of target words were not enhanced by only associating them with known related words. This consequence was

consistent with this line of research, such as in the studies conducted by Kasahara (2010, 2011, 2015).

Although significance testing is not always reliable and informative because of the dependence of the *p* value on sample size (Plonsky, 2015), the results of this experiment can be considered as relatively reliable because the weak points in the design and results of Experimental Studies 1 and 2 were improved in Experimental Study 3. First, to ensure the ecological validity of the experimental procedure, the number of target words was reduced to 20, which can be considered as the number of words L2 learners would naturally learn at once. Second, 35 participants were involved in this study, and this number is generally large enough to stabilize the statistical analyses. Third, no ceiling effect was included in the result of the plain recall test.

Therefore, the results of Experimental Study 3 lend support to the view that utilizing related words only once at the stage of learning unknown words does not much facilitate the retrieval of the unknown words without clues. Although Webb (2007) reported the facilitative effects of known synonyms on some aspects of vocabulary knowledge in his immediate posttest without clues, the retrieval of meaning was not significantly facilitated. Some positive effects observed in Webb's (2007) study might well seem like a potential of known related words, especially when other types of known related words are used. However, the present result, reproducing Webb's (2007) and Kasahara's (2010, 2011, 2015) meaning retrieval tests, suggests a small effect of known related words on meaning retrieval without clues, irrespective of the relationship between the target words and related words.

6.4.2 Multiple-choice test

The data analysis of the multiple-choice test yielded a significant difference in scores between the two conditions. The condition of known collocates scored better. This is just a

natural result because the questions were all based on what the participants had learned in the condition of known collocates. They were just required to choose the collocates that had been presented with the target words. However, to answer the question of whether collocational knowledge can be enhanced by presenting known collocates of target words, this result can be positive evidence.

What should be noted here is the retention of the link between the target words and their known collocates. There was no significant difference between the scores on the immediate posttest and delayed posttest. As mentioned above, in the condition of known collocates, this link is the initial stage of collocation knowledge (Sonbul & Schmitt, 2013) and, at the same time, the link in the mental lexicon to facilitate the retrieval of target words in cued recall (Kasahara, 2010). If this knowledge is retained long even by incidental learning, it is worthwhile to utilize collocations when learning unknown words.

6.4.3 Cued-recall test

In the cued-recall test, on the other hand, the data analysis demonstrated that providing known categorical associates that had been presented together during the learning phase facilitated retrieval of learned lexical memory more than known collocates did. It was already known that both known collocates (Kasahara, 2010, 2011, 2015) and known categorical associates (Experimental Study 1), learned together with target words, have some positive effects when presented at the stage of retrieval. However, the difference in the extent to which the two types of known related words have an effect on retrieval was newly suggested by this result. In other words, the current result implies that finding and learning unknown words by association or using known categorical associates in introducing new vocabulary would provide a more effective hint for learners to remember the target words than would introducing new lexical items with collocations, as far as a form-meaning link is concerned.

To further interpret this result, information about the theories concerning the effects of cognitive processing on vocabulary learning should be organized. As mentioned in Chapter 2.4, theories have long been developed, suggesting that learners' cognitive load or depth when processing the target words should be higher. The generative model (Slamecka & Graf, 1978) suggests the importance of processing to integrate existing knowledge and newly learned knowledge. Similarly, involvement load hypothesis (Laufer & Hulstijn, 2001) and technical feature analysis (Nation & Webb, 2011) basically support the importance of how much and how well learners think about the words when learning, although they have more concrete and fine-grained scales that describe exactly what and how learners should think and that enable teachers to evaluate teaching methods or tasks involving newly introduced words.

However, in the current study, learners tried to establish more than one aspect of vocabulary knowledge, namely, form-meaning links and collocation. In this kind of learning method especially, the limitation and trade-off of attentional resources should be carefully taken into consideration (TOPRA model; Barcroft, 2002). When learning unknown words, learners should prioritize establishing the form-meaning link (Schmitt, 2008) because the acquisition of other aspects of knowledge, including collocations, is facilitated if learners have already consolidated the form-meaning link of the words to be learned (Bogaards, 2001). When an appropriate cognitive load is spent for related words, learners will benefit by enhancing their knowledge of word association, which also works as a link to facilitate retrieval when cues are available (Kasahara, 2010). However, if learners' attention is drawn to related words by more than an appropriate extent, the efficiency of establishing the form-meaning link will deteriorate.

Based on the above, the current result, especially the question of why known collocates are not as effective as known categorical associates, can be further investigated. As mentioned, the target words in both conditions were fairly given to the participants. Careful control and

counterbalance were established in order to prevent the characteristics of the words in a word group from affecting the result of the experiment. The only possible differences that could contribute to the result are the types of known related words and the accompanying instruction of how to use them. When the participants were presented with known categorical associates, they were instructed to learn the target words and associate them with their related words. On other hand, when the participants were presented with known collocates, they were instructed to use the collocation example as a reference. Therefore, the following discussion will focus on these two differences.

First, regarding the types of known related words, known collocates were not linked to the participants' existing knowledge or well-integrated into their mental lexicon, compared to the known categorical associates. In other words, the generative process was not so effective in the condition of known collocations. More specifically, Kasahara's (2010) discussion sounds the most relevant to the situation of cued recall. It can be considered that categorical associates could create a stronger link to the target words in a learner's mental lexicon. Otherwise, at the cued recall phase, the known categorical associates had a stronger effect on narrowing down the range of meaning the participants had to recall. This could be due to the relatedness or the ease of association, which had some effect on the cued recall of target words in Experimental Studies 1 and 2.

Second, the difference between the instruction cannot be ignored. While the condition of known collocates required the participants to simply use it as a reference, simulating the incidental learning of collocations, the condition of known categorical associates required the participants to associate the target words with the related words, which sounds rather like the intentional learning of paradigmatic word association. These instructions were needed to simulate a learning situation in which the effect can presumably be maximized. In the condition of known categorical associates, the instruction was intended to prevent the

participants from ignoring the related words and failing to be aware of how the words were related. In the condition of known collocates, the instruction to learn intentionally meant learning the target words and collocations together at the same time. This is possibly more of a burden for learners than associating target words with known categorical associates. If this instruction had been given, it might have resulted in a learning burden too heavy for the participants, and learners' attentional resources might have been drawn away from the target words. When this happens, according to the TOPRA model, target word learning is prohibited at the expense of making a strong link between the target words and known collocates.

However, if the instruction in one condition encourages the learners to intentionally establish the link between target words and known related words, it might be advantageous in cued recall. Additionally, target word learning would possibly be facilitated by encouraging lexical processing with a high cognitive load. Judging from the current result, there is a possibility that the condition of known collocates would not encourage participants to make full use of their cognitive resources and would result in disadvantageous learning. In other words, it might be more effective to encourage the learners in further processing to memorize target words and to make the link between target words and known collocates stronger when learning with collocations, such as learning collocations intentionally at the same time. In addition, as highlighted in previous studies (Chapter 2.3), learning collocations is far from a waste of time and energy. Collocation knowledge contributes to making communication more natural and fluent, especially when learners use lexical items in writing or speaking (Pawley & Syder, 1983; Wolter, 2006). Therefore, it would be worthwhile to further investigate this possibility in future research.

6.5 Summary of Experimental Study 3

In Experimental Study 3, the effects of providing learners with known categorical associates were compared with those of giving known collocates (RQ3-1). At the same time, the effects of presenting known collocates on learners' collocational knowledge were also examined (RQ3-2). With repeated-measure ANOVAs of the three tests, some answers for these questions were obtained.

Concerning RQ3-1, the difference in effects was mainly observed in the cued recall tests. The results of the plain recall tests showed almost no difference in effects between the two conditions. Taking the results of Experimental Studies 1 and 2 into consideration, almost no effect was found to be exerted on the memorization or retrieval of the form-meaning links of the target words by merely providing categorically related words at the learning stage. However, when it came to cued recall, known categorical associates clearly yielded a better result, both in the immediate and delayed posttests, suggesting that categorical associates make better cues to facilitate the learners' retrieval on their own.

As for RQ3-2, although not surprisingly, collocational knowledge (i.e., the link between target words and their collocates) was better acquired in the condition of known collocates. Additionally, it was suggested that the link between target words and known collocates, which enhances the cued recall of target words, was not easily attenuated, even after the gap of one week.

However, regarding the weaker effects of known collocates, some issues should be further investigated. The difference in effects could be explained by the intrinsic efficiency of collocates and categorical associates in facilitating retrieval of the target words. On the other hand, it can also be explained by the difference of how to use the known related words, which suggests the possibility of enhancing the effect of known collocates by increasing the cognitive load in learning.

Thus, to obtain more concrete pedagogical implications, further investigation of this issue was attempted in Experimental Study 4.

Chapter 7

Experimental Study 4: Comparing the Effects of Known Collocates When Learning Collocations Incidentally or Intentionally

7.1 Purpose and Research Questions

As discussed in the last chapter, some questions were left about the use of known collocates in learning target words. While the effects on cued-recall were found to be not as strong as that of known categorical associates, there is still a possibility that the effectiveness could be enhanced by increasing the learners' cognitive load. Collocation will not only be able to facilitate cued-recall when it is learned in the forms of known-unknown word pairs (Experimental Study 3; Kasahara, 2010, 2011, 2015), but it is also important in itself for natural and fluent communication (Pawley & Syder, 1983). Therefore, it is worthwhile to further investigate better methods which will make the most of known collocates in a learning vocabulary.

In Experimental Study 3, one important point was left unclear. Whereas the condition of known categorical associates required the participants to intentionally associate target words and their related words, the condition of known collocates only required them to incidentally learn the collocation. Therefore, the question is whether the effect on learning increases or not under the instruction to intentionally learn collocations at the same time.

From the knowledge gained from previous research, contradictory views can be extracted for anticipating the effect of intentionally learning collocations of target words. In reference to the above-mentioned study, more cognitive load while learning target words might facilitate the memory of target words. This is because if learners spend more of their attentional resources on known collocates, the integration of unknown and known information can be stimulated (generative process; Slamecka & Graf, 1978), and the link for retrieving

target words (Kasahara, 2010) can become stronger. On the other hand, there is a risk of drawing too much attentional resources away from target words. If this happens, according to the TOPRA model, the learning of target words would be prohibited.

This point seems to be important for both teachers and students. When students try to only learn vocabulary, their English textbooks or wordbooks contain the information regarding collocations—or at least sentence examples. Teachers can utilize the information of known collocates by encouraging their students to pay attention to the combinational usage of words or by repeatedly presenting it in their tasks, review questions, or vocabulary tests. In these learning situations, knowing exactly how the collocations should be processed or whether the collocations should be incidental or intentional makes a clear difference after the cumulative experience of learning.

Thus, given these facts, Experimental Study 4 focused on the following research questions (RQs) left by the previous experiment.

RQ4: In which case is the learning of target words more facilitated by known collocates when learning collocations intentionally or incidentally?

RQ4-1: In which case are the memorization and retrieval of target words more facilitated by known collocates when learning collocations intentionally or incidentally?

RQ4-2: In which case are the acquisition of collocational knowledge of target words more facilitated by known collocates, when learning collocations intentionally or incidentally?

To approach these questions in a similar way as Experimental Study 3, an experiment was designed for measuring memorization and retrieval as well as collocational knowledge of target words. In this experiment, two conditions were made to compare the effect of incidental

and intentional learning of collocation in the form of known + unknown word pairs. Ten target words were allotted to each condition along with their collocates. Posttests were conducted twice, immediately after and then one week after the learning phase. This was done in order to get a better view of how the participants' memory of target words changed under the possible effect of related words. Results of both the immediate and delayed posttests indicate how successful the memorization and retrieval of target words were, respectively. Parts of the posttests were all the same as Experimental Study 3. The first part was a plain recall test, which required the participants to recall L1 translations of target words. The second part was the multiple-choice test asking them to choose the collocates of the target words. The third part was a cued-recall test, again asking the participants to write down L1 translations of target words with known related words provided as clues. Scores of the posttests were compared using 2 (conditions) × 2 (immediate or delayed posttest) Analysis of variance (ANOVA). Details of the experimental procedures are described in the following sections.

7.2 Method

7.2.1 Participants

In Experimental Study 4, a total of 39 Japanese undergraduate students participated in the experiment. This experiment was conducted three times in conjunction with the students' weekly regular university course. The class consisted of 54 students, but the data from 39 students who actually correctly completed the experimental procedure were used for analysis. None of the participants participated in Experimental Studies 1, 2, or 3. Their majors were diverse, including biology, biomolecular science, chemistry, environmental science, information science, and physics. Of the participants, 30 were male and nine were female. They all had known English for at least six years as a result of their formal education in Japan. Many of the participants self-reported their English proficiency in the form of a grade from

the EIKEN Test in Practical English Proficiency or a TOEIC® listening and reading test score. Additionally, a Mochizuki Vocabulary Size Test was conducted in the experiment to gauge their lexical proficiency. As described later, a Mochizuki Vocabulary Size Test was carried out in the first session, and the other main experimental procedure was in the second and third session. Four participants who were absent from the first session were included in the data. As a result of the remaining 35 participants, the average vocabulary size was 4,899 (*Min* = 2,462, *Max* = 6,731). From these data, it can be assumed that their proficiency of English ranged from elementary to intermediate level.

7.2.2 Materials

7.2.2.1 Target words and known related words

The same words from Experimental Study 3 were used as the materials for the current experiment. More precisely, the target words and their known collocates in Experimental Study 3 were reused. Therefore, the target words for the participants to remember were 20 low-frequency nouns, each of which had one high-frequency word which can be collocated. At the same time, the target words were almost the same as that of Experimental Study 1, although five words were substituted by the candidate words which went through the pilot test of Experimental Study 1. This is described in Chapter 6.2.2.1. Appendix G where it shows the complete list of target words and their known collocates.

As stated in the above-mentioned explanation of Experimental Study 1, the target words were chosen from Level 8 or higher in the *JACET List of 8000 Basic Words* (The Japan Association of College English Teachers [JACET], 2003). The word *equilibrium* was the only exception. All of the selected target words were nouns because the part of speech of target words and collocations can affect the difficulty of learning and should be controlled. As discussed later in detail, the part of speech of collocations used in the current study is

adjective + noun. Also, there were practical reasons both in the experiment and in the learners' or the teachers' practice of using nouns as target words for Experimental Studies 1 and 3, in which known categorical associates of target words are included in the materials. If the teachers of learners try to learn unknown words utilizing known categorical associates, they can most often realize it when they learn nouns because they tend to have more categorical associates than other words (see Chapter 4.2.2.1 for details).

For the known collocates of target words, adjectives which make congruent collocations with the target nouns were selected from the large-scale Corpus of Contemporary American English (COCA). Adjective + noun combinations were used in the study because of their importance. This type of collocation is the most frequently used by L2 learners (Barfield, 2009). Also, this type of combination can easily be controlled in terms of the number of words in a collocation. There are few cases in which articles appear between an adjective and a noun. Therefore, when adjective + noun combinations are used as materials for the experiment, two-word collocations can easily be gathered without the risk of being difficult for the participants in regard to articles when they are learning collocations. It should also be noted that all collocations used in Experimental Studies 3 and 4 were semantically transparent and congruent. This is because such collocations are easy for L2 learners to acquire and do not utilize too much of their cognitive resources needed for remembering target words. The standard for choosing collocations from the corpus was whether the existence of the expression is confirmed in a corpus or not (Kasahara, 2010, 2011, 2015). Therefore, the term collocation is operationalized here as a two-word pair that often co-occurs as evidenced in a large-scale corpus, which is the approach also adopted in Sonbul and Schmitt (2013). In the Experimental Studies 3 and 4, known collocates of target words were gathered following this method. Detailed discussions concerning the selection of collocations are transcribed in Chapter 6.2.2.1.

7.2.2.2 Conditions and control

In the experiment, two conditions were made in order to investigate whether Japanese English-learners can effectively comprehend unknown words and their collocations with or without the intention to memorize the collocations. In one condition, participants were instructed to use the information from a collocation example as a reference in order to encourage them to learn the collocation incidentally, which is the same instruction for that condition of known collocations used in Experimental Study 3. In the other condition, participants were instructed to learn words and the given examples of collocations at the same time in order to encourage them to intentionally learn the collocation. Hereafter, the former and the latter conditions are referred to as *incidental condition* and *intentional condition*. Ten target words were allotted for each condition. Therefore, in the procedures of Experimental Study 4, each participant learned 20 target words in sum as described later in detail (within-participants experimental design).

Lexical properties of materials that can affect the result of learning were controlled between these two conditions. However, the materials for Experimental Study 4 are the same as Experimental Study 3; the method of controlling was also the same with Experimental Study 3. As such, this section provides only the summary of the information about control. For a detailed discussion, see Chapter 6.2.2.2.

For the fairness of the experimental comparison, the lexical properties of target words that could affect learning difficulty—familiarity, frequency, and number of letters—were controlled among the two conditions. First, referring to the databases by Amano and Kondo (1999, 2000), target words for Experimental Study 4 were carefully controlled so that the familiarity and log frequency (i.e., logarithms for each frequency value) of target words were, on average, not significantly different among the conditions. Transformation of frequency value into log frequency is a widely adopted method for making statistical calculations

accurate. It is because distribution of the word frequency value follows Poisson distribution and deviates from a normal distribution, compromising accuracy of statistical comparison without this transformation (Wolter & Gyllstad, 2013). As discussed in the explanation of Experimental Study 1, the reason why the familiarity and frequency of Japanese L1 translations were controlled was that this method is more valid when taking into account that even in mental lexicons of advanced learners, memories of L2 word forms are apt to be linked with the meanings of their L1 equivalents (Jiang, 2000, 2002). Additionally, the length of target words (Ellis & Beaton, 1993) and collocations (Sonbul & Schmitt, 2013) were considered to affect their learnability. As shown in Table 7.1, lengths of target words, their known collocates, and collocations as a whole are controlled in Experimental Study 4. As for the control of the features of collocations, congruency and semantic transparency of known collocates were also controlled. According to the previous studies, L2 learners' knowledge or processing of collocations are affected by several factors:

- frequency of collocations (Eyckmans, 2009; Wolter & Gyllstad, 2013)
- frequency of component words (Revier, 2009; Wolter & Gyllstad, 2013)
- mutual information (MI) (Ellis et al., 2008)
- number of letters (Sonbul & Schmitt, 2013)
- congruency (Yamashita & Jiang, 2010; Wolter & Gyllstad, 2011, 2013)
- semantic transparency (Revier, 2009; Gyllstad & Wolter, 2016)

However, it is quite difficult to control for all of these properties of expressions and much of the previous research has not controlled for all of the above factors (e.g., Gyllstad & Wolter, 2016; Kasahara, 2010, 2011, 2015). Therefore, following Kasahara (2010, 2011, 2015) and prioritizing the control for the features of target words, the author controlled only for

congruency and semantic transparency of known collocates. Following previous studies, in selecting congruent and semantically transparent collocations, checks and discussions with other raters were employed (e.g., Gyllstad & Wolter, 2016; Yamashita & Jiang, 2010). This method is employed in the present study because judgment of meanings with only one native speaker's intuition has some room for arbitrariness (Revier, 2009). In this study, the author and an additional rater checked the congruency and semantic transparency of 37 respondent collocations in a randomized list. The rater was a graduate student specializing in English pedagogy who was a native speaker of Japanese with an upper-middle proficiency in English. When literal translations of the English collocations into Japanese is possible by paraphrasing the translated words into Japanese within the range of the same meaning, the collocations were rated as congruent. When the constituent words are used in prototypical meanings and collocations as a whole, they have literal meaning and the collocations were rated as semantically transparent (Howarth, 1996). As a result, the agreement rate was 93.8% of the items with disagreements being found in two cases. These disagreements were resolved through discussion. The validity of the above control was statistically guaranteed a by two-sided *t*-test. The analysis indicated that none of the characteristics were significantly different between the groups, including length of target words, $t(18) = -0.92$, $p = .367$, $d = -0.41$; length of known collocates, $t(18) = 0.61$, $p = .552$, $d = 0.27$; length of collocations, $t(18) = -0.27$, $p = .793$, $d = -0.12$; logarithm of frequency, $t(18) = 0.90$, $p = .382$, $d = 0.40$; and familiarity, $t(18) = 0.17$, $p = .865$, $d = 0.08$.

For reducing other possible confounding effects, the combination of target words and conditions as well as the order of learning conditions were counterbalanced. In the experiment, four different kinds of booklets, which reflects the counterbalancing of two groups of target words × the order of two learning conditions, were made and distributed to the participants.

The order of words and conditions as well as the number of participants who went through the experiment with each booklet is shown in Table 7.2.

Table 7.1

Control of Word Features Between Two Groups of Target Words

	Words 1–10		Words 11–20	
	M	*SD*	*M*	*SD*
Number of letters (target words)	7.50	2.80	8.60	2.50
Number of letters (known collocates)	6.20	2.49	5.60	1.90
Number of letters (whole collocations)	13.70	4.64	14.20	3.68
Frequency[a]	2.38	0.70	2.08	0.80
Familiarity	4.98	1.42	4.89	0.97

Note. [a]Logarithms for each frequency value were averaged.

Table 7.2

Counterbalancing of Conditions and Order

Participant Group 1 (n = 10)	Participant Group 2 (n = 10)	Participant Group 3 (n = 9)	Participant Group 4 (n = 9)
1.Words 1–10 in Incidental Condition	1.Words 1–10 in Intentional Condition	1.Words 11–20 in Incidental Condition	1.Words 11–20 in Intentional Condition
2.Words 11–20 in Intentional Condition	2.Words 11–20 in Incidental Condition	2.Words 1–10 in Intentional Condition	2.Words 1–10 in Incidental Condition

7.2.3 Procedure

Experimental Study 4 was conducted in December of 2018. It was conducted three times as part of the students' weekly regular university course. The whole process took approximately 120 minutes within a total of three sessions. The first session, which was included in the first class of the second semester (i.e., September), was mainly for the Mochizuki Vocabulary Size Test. This test also functioned as a proficiency measurement and a goal-setting guide for the students in the course. The second session involved memorization and the immediate posttests of target words. The final session was scheduled just one week after the second session, and the participants tackled the delayed posttests. The purposes, procedures, and security of personal information were first explained to them and their informed consent was obtained before the experiment.

All tasks in the experiment were presented in four different versions of booklets, which reflected the counterbalancing of conditions and word presentation order. One of those booklets was given to each participant and the number of participants was randomly allocated to the four versions.

The first session of the above-mentioned experiment was mainly for the Mochizuki Vocabulary Size Test. The order of this lexical proficiency measurement in the experimental procedure was the one and only difference from the methods of Experimental Study 3. Although the version of the vocabulary size test used for Experimental Study 4 included one of the target words (*avalanche*) this measurement presumably did not affect the later processes of experiment. It is because there was a long period between the first and the second session. In fact, none of them correctly answered the Japanese meaning of *avalanche* at the pretest stage.

In the second session, the participants were first asked to memorize meanings (L1 translations) of target words within five minutes (learning phase). Basically, they were

allowed to make use of any strategy to learn these target words (e.g., self-testing or repeatedly writing them down), but they were instructed in how to make use of known collocates in each condition. In one condition, which was the same with Experimental Study 3, participants were instructed to use the information from a collocation example as a reference in order to encourage them to learn the collocation incidentally. In the other condition, participants were instructed to learn words and the given examples of collocations at the same time in order to encourage the participants to intentionally learn the collocation. The related words were presented with their L1 translations in case their meanings were unknown to the participants or unclear as a result of polysemy or the vague knowledge of participants. The format of the word list is shown in Appendix B-1 and Appendix B-2.

The remaining procedure is almost the same as the one in Experimental Study 3. After the learning section, the participants answered the questionnaire about learning strategy for 10 minutes, which served as a distraction. This task was intended to prevent their short-term memory from unfairly facilitating their performance in the immediate posttests. If the last few target words they tried to learn remained in their short-term memory, the condition in which the last few target words were provided would result in an unfairly high mark.

After finishing the questionnaire, the participants moved on to the immediate posttests of the target words. The format of the posttests was the same as the one in Experimental Study 3. The first type is a plain recall test measuring memorization of receptive knowledge, which required the participants to write L1 translations of target L2 words (Kasahara, 2010, 2011, 2015). The second type is a multiple-choice test which required them to choose the words that were most likely to appear in context with the target words (Webb, 2007). In this test, the correct answers were the words prepared as known collocates for each of the target words. All distractor words were adjectives, which were the same parts of speech as the correct answers. Participants were required to circle one word from the four choices in the format shown in

Figure 6.1 in the previous chapter. This test checks whether a link is built between the two component words of a collocation as initial stages of collocational learning (Sonbul & Schmitt, 2013). The third type is a cued-recall test measuring retrieval of receptive knowledge, in which known collocates learned together in the learning phase were provided as cues (Kasahara, 2010, 2011, 2015). To prevent the earlier test from affecting the participants' performance in the later test, the three kinds of measurements were implemented in this order. Additionally, the order of test items was randomized in each test to prevent the effect of encoding specificity (Thomson & Tulving, 1970) that could be caused by the same order of items as in the learning phase. Overall, the results of these immediate posttests indicate how well the participants could memorize the target words, both in terms of form-meaning link and collocational knowledge.

The third session was conducted just one week after the second section. On that day, a delayed posttest was conducted in the same format as the immediate posttest. The results of these tests indicated how well the participants could retrieve the target words, both in terms of form-meaning link and collocational knowledge.

Throughout the experimental process, all of the oral instructions, answers to participants' questions, and explanations written in booklets for completing each experimental task were conveyed in Japanese (the participants' and the author's native language) in order to ensure that the participants understood everything.

For ease of understanding, an outline of the experimental procedure is provided in Figure 7.1. Detailed information about each step was described in this order.

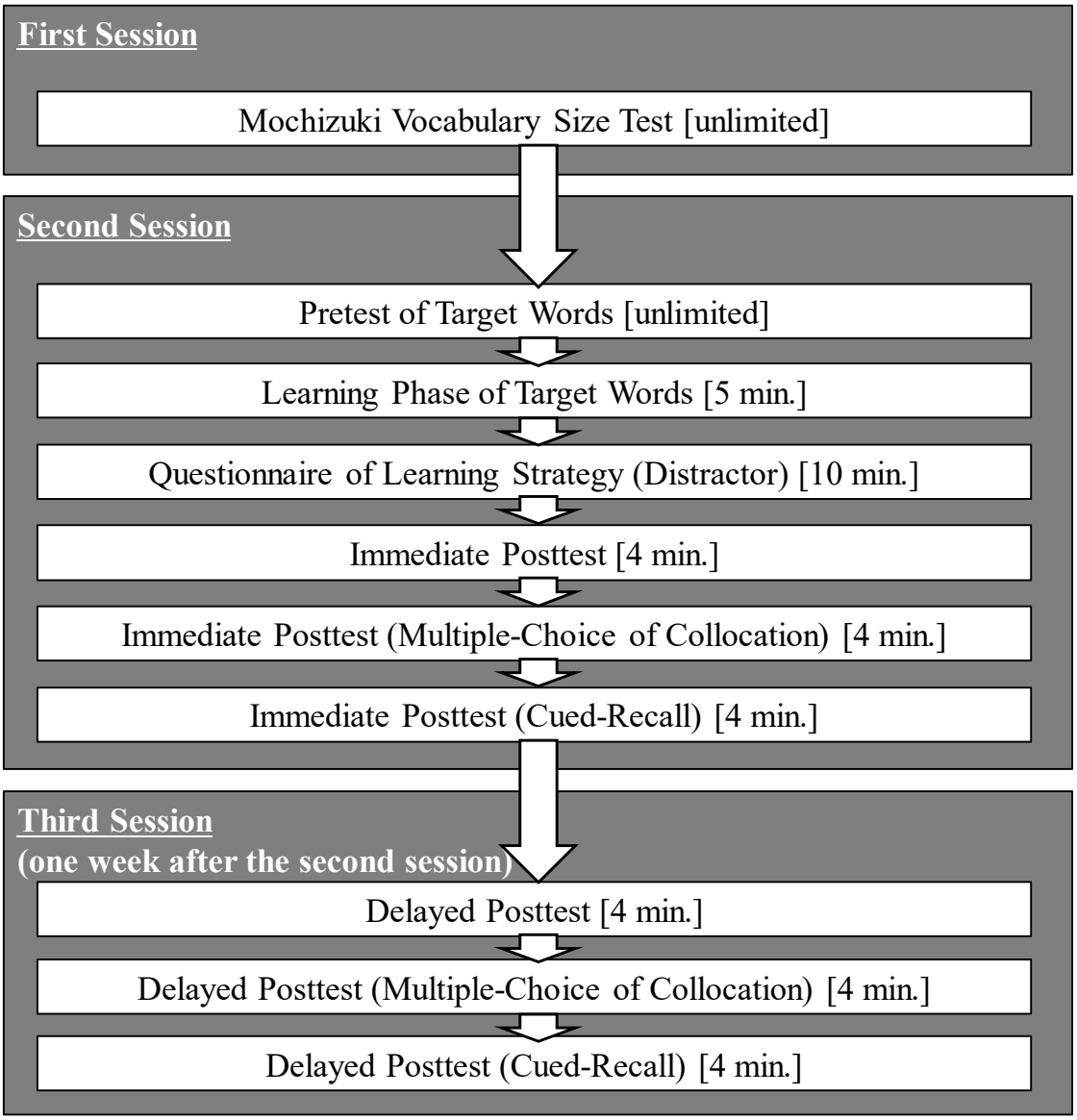

Figure 7.1. Procedure of Experimental Study 4.

7.2.4 Scoring and data analysis

Through the examination of the participants' memorization and retrieval by the scores of immediate and delayed posttests, respectively, one point was awarded to each correct answer and zero was added to an every incorrect one. As done in the previous experiments, mistaken Chinese letters were marked as correct, as long as the meanings in Japanese (L1) were accurately recalled.

The statistical method for analysis was also the same with Experimental Study 3. A two-way repeated measures ANOVA was conducted for each of the three types of tests. All ANOVAs were 2 (known categorical associates or known collocates) × 2 (immediate or delayed posttest).

Before proceeding to the analyses, some problematic cases found in the scoring were omitted just as in Experimental Study 3. First, target words already known at the point of pretest were discarded (1.15%: nine out of 780 cases). Second, known related words not recognized by the participants before the test were also checked via the pretest because they had not been confirmed at the stage of pilot test. These cases were also removed from the data (13.08%: 102 out of 780 cases). After excluding those data, the rate of correct answers calculated from the remaining data was used in the following analyses.

7.3 Results

7.3.1 Recall test without cues

The overall results of the plain recall test are summarized in Table 7.3. This time, the interaction between condition and time was found to be significant, $F(1, 37) = 5.11$, $p = .030$, $\eta_p^2 = .12$. Apparently, it can be interpreted from the descriptive statistics that the participants performed slightly better in incidental conditions at the time of the immediate posttest, whereas intentional conditions slightly outperformed at the stage of delayed posttest. To further clarify these “slight” effects, simple main effects were analyzed. To make it easy to understand, Figure 7.2 was also provided.

Table 7.3

Descriptive Statistics of Posttests Without Cues

	Immediate posttest			Delayed posttest		
	M	95% CI	*SD*	*M*	95% CI	*SD*
Incidental	51.3	[41.0, 61.6]	31.4	22.4	[17.1, 27.6]	16.1
Intentional	43.5	[34.7, 52.3]	26.7	24.6	[18.6, 30.7]	18.3

Note. These values indicate percentages of correct answers in each condition.

Naturally, simple main effects of time were significant on both of the conditions: incidental condition, $F(1, 37) = 43.13$, $p < .001$, $\eta_p^2 = .54$; and intentional condition, $F(1, 37) = 19.16$, $p < .001$, $\eta_p^2 = .34$. As for the simple main effect of condition, it was insignificant in the delayed posttest, $F(1, 37) = 0.62$, $p < .434$, $\eta_p^2 = .02$. However, the difference between the two conditions (i.e., the "slight" advantage of incidental condition) was marginally significant in the immediate posttest, $F(1, 37) = 3.06$, $p = .089$, $\eta_p^2 = .08$.

7.3.2 Multiple-choice test

The overall results of the multiple-choice test of collocational knowledge of target words are summarized in Table 7.4. The results indicated that the main effect of condition (incidental or intentional) was not significant, $F(1, 37) = 1.07$, $p = .309$, $\eta_p^2 = .03$. On the other hand, the difference between immediate and delayed posttest was significant, $F(1, 37) = 13.15$, $p < .001$, $\eta_p^2 = .26$, and the interaction was not significant, $F(1, 37) = 1.04$, $p = .314$, $\eta_p^2 = .03$.

Table 7.4

Descriptive Statistics of Multiple-Choice Posttests

	Immediate posttest			Delayed posttest		
	M	95% CI	*SD*	*M*	95% CI	*SD*
Incidental	82.3	[76.7, 87.9]	17.0	68.0	[58.8, 77.2]	27.9
Intentional	83.1	[74.4, 91.7]	26.3	73.0	[63.4, 82.7]	29.3

Note. These values indicate percentages of correct answers in each condition.

7.3.3 Cued-recall test

The overall results of the test providing related words as cues are summarized in Table 7.5. Again, the interaction between condition and time was significant, F (1, 37) = 6.96, p = .012, η_p^2 = .16. This means that the advantage of intentional condition became much bigger at the stage of delayed posttest. To further delve into the effect of the learning methods, simple main effects were analyzed. Figure 7.3 was also provided to organize the results.

Table 7.5

Descriptive Statistics of Posttests With Cues

	Immediate posttest			Delayed posttest		
	M	95% CI	*SD*	*M*	95% CI	*SD*
Incidental	57.2	[48.2, 66.1]	27.3	37.9	[30.3, 45.5]	23.2
Intentional	60.9	[52.0, 69.7]	26.9	49.4	[42.5, 56.4]	21.1

Note. These values indicate percentages of correct answers in each condition.

Naturally, the simple main effects of time were significant on both of the conditions: incidental condition, $F(1, 37) = 39.09$, $p < .001$, $\eta_p^2 = .51$; and intentional condition, $F(1, 37) = 24.86$, $p < .001$, $\eta_p^2 = .40$. As for the simple main effect of condition, it was insignificant in the immediate posttest, $F(1, 37) = 0.03$, $p = .379$, $\eta_p^2 = .02$, while the score of intentional condition was significantly better in the delayed posttest, $F(1, 37) = 12.28$, $p = .001$, $\eta_p^2 = .25$.

7.4 Discussion

7.4.1 Recall test without cues

For better comprehension, the results of the simple main effect analysis were illustrated in Figure 7.2. As suggested by the analysis, in the immediate posttest, it is possible to interpret that incidental condition was slightly better in learning target words than the other condition. Although the *p*-value is over the significance level, and the results of the analysis are rather reliable if the statistical power supported by the sample size of 39 people is taken into consideration, interpretation of the results depending only on significance testing is not always reliable and informative (Plonsky, 2015). In fact, from the descriptive statistics, it can be calculated that incidental condition yielded an almost 1.2 times more efficient result. This difference is, if any, exactly what could be anticipated with the TOPRA model (Barcroft, 2002), which assumes learners' attentional resources as limited and their learning effect of a certain method on each aspect of vocabulary knowledge as dependent on the distribution of the attentional resource in learning. Therefore, it should be kept in mind that intentional learning of collocations might have drawn the attention of learners away from target words and prohibited their memorization of them.

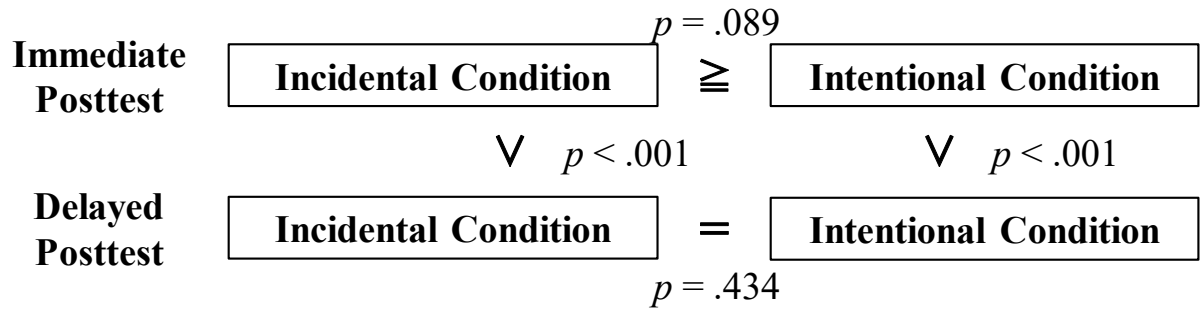

Figure 7.2. Results of simple main effect analysis of plain recall test between conditions.

On the other hand, it should also be noted that the scores of the two conditions were almost the same at the phase of delayed posttest. In the long run, intentional learning of collocation will result in an effect not inferior to that of incidental condition, even if there might be a slight disadvantage in attentional resource. The reason why this result appeared can be interpreted by taking the possible effects of cued-recall into consideration. Because of the stronger link between target words and their known collocates which developed from the extra emphasis on learning collocation, cued-recall was firmly supported so that its score was equivalent to that of incidental condition (Table 7.5, Figure 7.3). As highlighted in the section of previous studies, a learner's memory improves more when they retrieve it on their own than by simply being given the right answer (retrieval practice effect; Baddeley, 1997). Therefore, at the point of immediate posttest, this effect might have enhanced the memory of target words in intentional condition.

However, the overall results of the plain recall test again lend support to the view that utilizing the related words at the stage of learning unknown words only once will not facilitate the plain recall of the target words as much. This also occurs even if learners try to learn the collocation of target words at the same time. It is consistent with the previous Experimental Studies and previous studies in this line (e.g., Kasahara, 2010; Webb, 2007).

7.4.2 Multiple-choice test

The data analysis of the multiple-choice test showed no significant difference of score between the two conditions. Although the links between the target words and known collocates were presumably stronger as suggested by the better result of cued-recall in the delayed posttest (Table 7.5, Figure 7.3), it was not reflected in the results of this multiple-choice test. A possible explanation for this result is that the multiple-choice test adopted in the present study measures only the establishment of link between target words and collocates, which is the initial level of collocational knowledge (Sonbul & Schmitt, 2013). In the test set of Webb (2007), this test form is also considered to be an easy one because it is utilized to measure receptive knowledge. If a more difficult type of test, such as the recall or production of syntagmatic associates (test of productive knowledge of syntagmatic association; Webb, 2007) or lexical decision task with priming (test of implicit knowledge; Sonbul & Schmitt, 2013) had been used, the difference of score between conditions might have been reflected. However, from this result, it is most natural to consider that the difference of the condition is, if any, a slight one.

It should also be noted here that, contrary to Experimental Study 3, the retention of the link between target words and known collocates significantly attenuated at the phase of delayed posttest. It is unlikely that the difference of participants explains the discrepancy of the results between Experimental Studies 3 and 4, because there was little difference in proficiency of the participants between the studies (5,034 and 4,899 in Mochizuki Vocabulary Size Test on average, respectively). Then, what might explain this discrepancy is the number of expressions to be remembered. Experimental Study 4 required the participants to remember up to 40 expressions (20 target words and 20 collocations). It is possible that the increased learning burden resulted in lower performance and more confusion of word combinations, just as the overall results of Experimental Study 2 with 60 target words were lower than

Experimental Study 1. Regardless, the fact should be noted that around 70% of the knowledge was retained even after a week. It lends support to the knowledge obtained from Experimental Study 3 that even by incidental learning, the durable link between target words and their known collocates can be formed.

7.4.3 Cued-recall test

For better comprehension, the results of the simple main effect analysis were illustrated in Figure 7.3. Participants in the intentional condition performed better in the long run. As discussed earlier, it sounds the most natural to interpret this difference as the result of stronger link between target words and their known collocates created by more attention and cognitive load in intentional condition. This interpretation is consistent with the knowledge of previous research. In the TOPRA model, learning effect on each aspect of vocabulary knowledge is considered as dependent on the distribution of limited attentional resource in learning (Barcroft, 2002). Thus, intentional condition is presumably better for making stronger links between target words and their known collocates. As discussed in the previous section, this also explains why there was a tendency of lower performance in plain recall of target words in intentional condition.

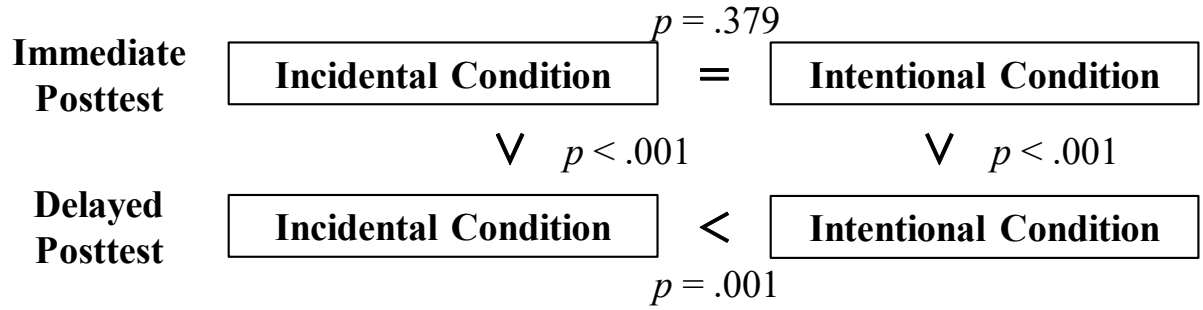

Figure 7.3. Results of simple main effect analysis of cued-recall test between conditions.

It is also important to interpret the result from the viewpoint of the total amount of cognitive load that contributed to the development of vocabulary knowledge. In other words, although it is important to strike the balance of cognitive load spent for learning each aspects of vocabulary knowledge, it is also vital to question whether a learner's attentional resource or cognitive resource is fully and effectively exerted in a certain method. Some theories of previous research commonly have this viewpoint. For instance, the generative model (Slamecka & Graf, 1978) considers that processing to integrate the existing and new knowledge is important. Similarly, both involvement load hypothesis (Laufer & Hulstijn, 2001) and technical feature analysis (Nation & Webb, 2011) basically support the importance of how much and how well learners process the words in learning.

From the current result, it is also possible to consider that the intentional condition more successfully exerted learners' cognitive resources in processing target words and their collocates. While the disadvantage in the memorization of target words were vague and incidental condition had nearly 1.2 times the efficiency at most, the advantage was clear (i.e., more than 1.3 times the efficiency, according to the descriptive statistics) in the retrieval of target words after a week. By clearly motivating participants to learn collocations (*need* in involvement load hypothesis; Laufer & Hulstijn, 2001), the integration of known and unknown knowledge that is considered as effective (Slamecka & Graf, 1978) was stimulated between unknown words and their known collocates, which also worked as the help for the retrieval of target words at the phase of cued-recall (Kasahara, 2010). In this way, the current result can be interpreted as the support for the view that when learning words with known collocates, it will be more efficient to memorize the target words in the form of collocations as a whole.

7.5 Summary of Experimental Study 4

In Experimental Study 4, the effects of providing learners with known collocates were compared between the case where the collocates were only given as references and the case where learners tried to learn collocation at the same time (RQ4-1). In addition, the effects on collocational knowledge were also compared in the same way (RQ4-2). As a result of the analysis, some answers for these questions were obtained.

First of all, as the answer for RQ4-1, information of known collocates should be utilized so that the target words are memorized with their collocation. In this method, the link between target words and known collocates are further strengthened and help the cued-recall more in the long run. Although it is basically better for retrieval to learn intentionally in the form of collocations, it should also be noted that this method requires high cognitive load for processing. If this method causes overload for learners, there is a risk of deteriorating the learning effect of target words, as implied, although not significantly, by the results of the immediate posttest regarding the plain recall. Therefore, when learners try to memorize new words in the form of known + unknown collocations, materials or conditions should carefully be controlled so as not to pose too much burden on them. For example, it might result in failure to learn too many words and collocations at the same time or to learn incongruent or semantically opaque collocations together with target words. Just in the same way as the previous Experimental Studies, plain recall of target words was not facilitated by merely providing known collocates only at the stage of learning, even if they were learned in the form of collocations.

As for RQ4-2, the initial level of collocational knowledge (i.e., the link between target words and their collocates) was fairly memorized and retained in both of the conditions. No matter whether the known collocates were used as references or intentionally remembered

together with target words, the link between target words and known collocates, which facilitates the cued-recall of target words, were found to be durable.

In the intentional condition of Experimental Study 4, the association between target words and known collocates were intentionally learned in a similar way to the known categorical associates in Experimental Study 3. However, the difference of effect seems not to be totally compensated with the instruction or cognitive load of processing. The difference of effects could presumably be explained by the intrinsic efficiency of collocates and categorical associates in facilitating retrieval of target words. A more detailed discussion concerning the four Experimental Studies was attempted in the Chapter of General Discussion.

Chapter 8

General Discussion

8.1 Overview of Findings

8.1.1 Brief summary of overview

As described in Chapter 3, the purpose of the current study is to clarify, compare, and maximize the effects of known related words in the cases of paradigmatic and syntagmatic associates. Before comparing the effects observed in each experimental study, the effects of each classification of known related words and ways to further enhance them are organized herein.

The results of Experimental Study 1 demonstrated the effect of known categorical associates on the facilitation of retrieval during cued-recall. The effect of relatedness was statistically insignificant, but was still weakly or indirectly implied.

Similarly, Experimental Study 2 confirmed that known synonyms have an effect on facilitating retrieval during cued-recall. The effect of relatedness was more clearly suggested, and there should be only one known synonym when the relatedness of the words is high. The precise meaning of the target words was well acquired after one week, regardless of the number and relatedness of the known synonyms, suggesting the effect of instruction in facilitating a careful comparison of the meanings of related words.

The results of Experimental Study 3 demonstrated that the known categorical associates were more effective than the known collocates in terms of their effect on facilitating retrieval during cued-recall. Naturally, the effect on collocational knowledge was greater when known collocates were used.

The results of Experimental Study 4 demonstrated that learning collocation intentionally facilitated retrieval during cued-recall more effectively than when known collocates were

used for reference. Meanwhile, the two types of instruction had nearly the same effect on collocational knowledge.

8.1.2 Effects of known related words on immediate and delayed posttests

At first, it was unclear whether the effect of known collocates on cued-recall, especially retrieval in the delayed posttest, was generalizable to cases of known paradigmatic associates. In Webb (2007), which examined the effects of known synonyms, a delayed posttest was not included in the experimental design. The results were relatively stable among the several conditions for learning evaluated in the present study, regardless of whether a plain recall or cued-recall test was performed.

First, the results of the plain recall tests were basically consistent across all the experimental studies. Regardless of differences in the conditions implemented in the four experiments (e.g., types of related words, types of instructions, number of target words, or part of speech of target words), there were no significant differences among the conditions in either the immediate or delayed posttests.

The only exception to the statement above is the lower score observed in the intentional condition of Experimental Study 4. This marginally significant difference only appeared in the immediate posttest, which implies that the intentional learning of collocations might introduce the risk of overload for learners, which could cause the distraction of their attentional resource. However, the absence of any difference in the score of the delayed posttest implies that the effect of the stronger link between the target words and known collocates, which is developed in the intentional condition, can overcome a negative effect by facilitating the learners' own retrieval.

Thus, giving known related words only at the phase of learning seemingly does not affect the memorization and retrieval of the target word's meanings. Although Webb (2007)

reported the positive effects of known synonyms on some aspects of vocabulary knowledge (i.e., productive knowledge of paradigmatic association, productive knowledge of syntagmatic association, and receptive knowledge of form), the result is consistent with the current research, in that the knowledge pertaining to the form-meaning link was not significantly benefited.

Second, the results of the cued-recall measurements were also stable. In both Experimental Studies 1 and 2, the experimental conditions in which known related words were presented resulted in significantly better scores in the posttests for cued-recall. This tendency is also consistent with Kasahara (2010, 2011, 2015)'s result. One exception was the immediate posttest in Experimental Study 1, in which the difference between the condition of words with high-relatedness and the control condition was marginally significant. However, this result could be interpreted as partial evidence of the effect of known related words on the immediate posttest because the result indicated a ceiling effect, and also because the number of participants was small, both of which lead to lower statistical power. Considering the presence of many differences between the experiments, this result can be regarded as somewhat reliable in various conditions.

8.1.3 Effects of semantic relatedness

In terms of semantic relatedness, the results of Experimental Studies 1 and 2 were indicative of a more facilitative effect prompted by closely related words. While the difference suggested by Experimental Study 2 was clear, the same effect was only indirectly suggested by Experimental Study 1 because the scores of the immediate and delayed posttests of cued-recall were not significantly different for the condition of words with high-relatedness. If the ceiling effect and small sample size of the experiment are considered, it is still possible

that known words that are semantically close facilitate retrieval more effectively, even in the case of categorical associates.

Regarding the semantic relatedness or similarity, it should be noted that a perfect standard does not exist. Therefore, the pilot tests used in Experimental Studies 1 and 2 cannot provide perfectly objective data or a criterion for semantic relatedness. This point should, in part, be interpreted as a limitation because it is possible that the perception of semantic relatedness varies among different learners. However, it could be interpreted as an advantage that this method for determining relatedness is capable of reflecting a perception related to the mental lexicon of Japanese EFL learners. One benefit to the reflection of L2 learners' mental lexicon is that the standard of relatedness or similarity can also easily be used by learners or teachers. In fact, as confirmed in Experimental Study 1, the results of relatedness judgments in the pilot test and the main experiment were not different. Therefore, for example, preparing closely related known words in review questions relying on teachers' perception or encouraging students to find close associates of known words worth memorizing relying on students' perception is a valid approach, albeit not completely reliable.

It seems relatively reasonable to deny the possibility of encoding specificity (Thomson & Tulving, 1970). The facilitative effect of known related words was not due to mere similarities in the situation between the learning and test phases but is instead owing to the link between the target words and their known related words. The order of the target words was randomized in the posttests in all four of the experimental studies, and Experimental Study 2 included a control condition, in which participants were presented with unrelated words to remove the factor of encoding specificity. In fact, the scores of the posttests were different among the conditions in Experimental Studies 2 and 3, presumably due to differences in the semantic similarity of the known related words. In addition, especially in terms of the observed difference in Experimental Study 2, in which the only difference

between the experimental conditions and that of the control was whether the word sets had semantic similarity, the possibility of encoding specificity is low.

When interpreting the results of cued-recall in Experimental Study 2, the possible presence of a misleading effect in the control condition should be considered because the average score of the plain recall measurement (34.6%) was higher than that of the cued-recall measurement (19.5%) in the delayed posttests. This result indicates that the participants were misled by the unrelated words, and perceived them as synonyms. In fact, approximately 8.7% (24 cases out of 12 target words × 23 participants) of the answers were considered a result of this dynamic.

If the result of the plain recall measurement in the delayed posttest is compared with that of the other experimental conditions in the delayed cued-recall measurement, as in Experimental Study 1, the score of the control condition (34.6%) is slightly, albeit not significantly, lower than that of the condition of the low-similarity synonym × 1 (38.2%). Conversely, for the immediate posttests in Experimental Study 2, the score for plain recall was 45.7% while cued-recall resulted in 46.3%, and the "misled" answers comprised approximately 2.2% of the answers (six cases out of 12 target words × 23 participants).

Thus, although the possibility that semantic similarity enhanced cued-recall is rather high, further confirmation, especially in the case of the known categorical associates, is needed to reach a definitive conclusion in this regard.

8.1.4 Effects of the number of related words presented

Based on the result of the delayed posttests in Experimental Study 2, it was suggested that presenting two known synonyms exerts few effects on learners' memory of target words, regardless of the semantic similarity of the synonyms. In both the cases of the high-similarity

and the low-similarity synonyms, the differences in the test scores were all revealed as insignificant.

In theory, it is possible that increasing the number of known related words benefits learners' knowledge by increasing the clues that facilitate retrieval in learners' memory (Stahl & Nagy, 2006) as well as increasing the integration of new knowledge that utilizes prior knowledge (Slamecka & Graf, 1978). However, presumably owing to limitations in learners' attentional resource (Barcroft, 2002), their attention, processing, and the resulting links that were established did not double those in the conditions of the single known synonyms.

Since a small effect was demonstrated for the low-similarity synonyms, it is highly likely that providing more than one known related word of another kind also causes a small effect, although further verification is needed to generalize this insight.

8.2 Comparing the Methods and Types of Known Related Words

8.2.1 Comparison of categorical associates and synonyms

Experimental Studies 1 and 2 were conducted according to similar experimental designs, and the results also revealed some similarities. However, before comparing the results, several differences between Experimental Studies 1 and 2, as well as their limitations, are relevant to the discussion and should be highlighted.

First, as a matter of course, the factors of focus in the two experiments were different. Although the basis of the concept of relatedness is shared between the cases of categorical associates and synonyms, strictly speaking, the definitions of relatedness in Experimental Study 1 and similarity in Experimental Study 2 are different. Thus, it is incorrect to say that the closeness of the relationship between the target and related words in Experimental Studies 1 and 2 is directly comparable.

Second, it cannot be ignored that the two experiments used words that are different parts of speech. Experimental Study 1 involved nouns alone, while both nouns and verbs were used in Experimental Study 2. Since some studies, such as that of Ellis and Beaton (1993), state that verbs are more difficult to learn than nouns, this difference might have contributed slightly to the lower scores in Experimental Study 2.

Third, the procedures comprising Experimental Studies 1 and 2 were different. As seen in a comparison of Figures 4.1 and 5.1, the relatedness judgments used to confirm the pilot test were implemented in different stages. In Experimental Study 1, the judgement was situated just after the immediate posttest, whereas it was placed after the delayed posttest in Experimental Study 2. The author decided to change the order of this experimental procedure because this relatedness judgment requires participants to process the meaning of the target words, which might have resulted in the facilitation of memory and contributed to the ceiling effect observed in Experimental Study 1. Moreover, the timing of the interval was different between Experimental Studies 1 and 2. In Experimental Study 1, the interval lasted for 10 minutes after the learning phase, whereas it lasted for five minutes in Experimental Study 2. This change was made to avoid overburdening the participants in Experimental Study 2, because it included the largest number of target words, and the time required for the entire experimental procedure was accordingly long. These changes in the procedure might also have affected the scores.

Fourth, the instructions given to the participants were different. The participants in Experimental Study 1 were asked to associate the target and related words, while the participants in Experimental Study 2 were told to compare the similar meanings of the target words and the known synonyms.

Fifth, in a comparison of Experimental Studies 1 and 2, the most different point was the number of target words the participants were required to learn. While Experimental Study 1

prepared 30 target words, Experimental Study 2 included as many as 60 target words. It is quite natural to consider that the lower scores resulting in Experimental Study 2 can be attributed to this significant contrast rather than to a difference in the known related words (i.e., categorical associates vs. synonyms).

Considering these differences, it is unfortunately almost impossible to directly compare the effects of the types of relationship (i.e., categorical associates vs. synonyms), because many confounders could be involved. This is because the focus of this work is not the comparison of known related words within paradigmatic associates, in which the difference of effects is less expectable. Rather, the focus of Experimental Study 3 was a comparison of the effects of paradigmatic and syntagmatic associates; more implications may be derived from the results by drawing a comparison with Experimental Study 4. Therefore, a reanalysis and discussion pertaining to Experimental Studies 3 and 4 follows in the subsequent section.

8.2.2 Comparison of categorical associates and collocates

A comparison of the effects between the categorical associates and collocates was nearly completed in Experimental Study 3 through its thoroughly-controlled design. However, it is possible that the instruction used to facilitate learning with the known collocates was unable to prompt the full use of the learners' cognitive resource. In fact, the intentional condition in Experimental Study 4 yielded better results overall. Therefore, the result of the intentional condition should also be compared to the effect of the categorical associates.

This comparison is inevitably a between-participants design. If the participants of the two experimental studies had been too different in their level of proficiency, the comparison would have been unfair. However, according to the results of the Mochizuki vocabulary size test, their lexical proficiency, or the assumed development of their mental lexicon, reveals almost no difference. In Experimental Study 3, the average learner vocabulary size was 5034

(*Min* = 4154, *Max* = 5846), while the average learner vocabulary size was 4899 (*Min* = 2462, *Max* = 6731) in Experimental Study 4. Therefore, as in Experimental Studies 3 and 4, a 2 (known categorical associates or known collocates) × 2 (immediate or delayed posttest) repeated measures ANOVA was conducted.

First, the overall results of the plain recall test are summarized in Table 8.1. As a matter of course, the points for the delayed posttests were significantly lower than those of the immediate posttests, $F(1, 71) = 40.92$, $p < .001$, $\eta_p^2 = .36$. In this comparison, the type of known related words that were learned together with the target words resulted in significant differences in the test without cues, $F(1, 71) = 7.98$, $p = .006$, $\eta_p^2 = .10$, and there was no significant interaction, $F(1, 71) = 0.01$, $p = .905$, $\eta_p^2 = .00$.

Even without cues, the effects of the known categorical associates surpassed those of the known collocations. This result clearly supports the superiority of categorical associates in terms of the effects of the form-meaning link or target words on learner memory. Categorical associates can be more easily linked to the memory of the target words, and can therefore facilitate retrieval more effectively.

Table 8.1

Descriptive Statistics of Posttests Without Cues

	Immediate posttest			Delayed posttest		
	M	95% CI	*SD*	*M*	95% CI	*SD*
Categorical	58.8	[48.5, 69.1]	29.9	39.3	[29.6, 48.9]	28.1
Collocation	43.5	[34.7, 52.3]	26.7	24.6	[18.6, 30.7]	18.3

Note. These values indicate the percentages of the correct answers for each condition.

However, since this comparison yields less precise results than those within the well-controlled experimental studies, it does not guarantee that the categorical associates had a facilitative effect on plain recall. In fact, Experimental Study 1 suggested almost no effect, compared to the control condition. This should be considered as another indirect piece of evidence supporting the superiority of categorical associates in learning the form-meaning link for target words.

Second, the overall results of the multiple-choice test pertaining to the collocational knowledge of the target words are summarized in Table 8.2. This time, the interaction between condition and time was significant: $F(1, 71) = 8.42$, $p = .004$, $\eta_p^2 = .11$, which simply means that while the collocational knowledge pertaining to the condition of the known categorical associates only changed at random, the knowledge acquired pertaining to the condition of the known collocates was not fully retained. As suggested by descriptive statistics, the simple main effect of time was insignificant for the condition pertaining to categorical associates, $F(1, 34) = 2.98$, $p = .093$, $\eta_p^2 = .08$, and significant for the condition pertaining to collocates, $F(1, 37) = 5.70$, $p < .022$, $\eta_p^2 = .13$. The simple main effect of the condition was naturally significant in the immediate posttest $F(1, 71) = 40.93$, $p < .001$, $\eta_p^2 = .37$, and in the delayed posttest, $F(1, 71) = 10.13$, $p = .002$, $\eta_p^2 = .12$.

Table 8.2

Descriptive Statistics of Multiple-Choice Posttests

	Immediate posttest			Delayed posttest		
	M	95% CI	*SD*	*M*	95% CI	*SD*
Categorical	45.5	[37.3, 53.6]	23.7	52.8	[44.4, 61.3]	24.6
Collocation	83.1	[74.4, 91.7]	26.3	73.0	[63.4, 82.7]	29.3

Note. These values indicate the percentages of the correct answers for each condition.

Third, the overall results of the cued-recall test are summarized in Table 8.3. This time, the interaction between the condition and time was insignificant, $F(1, 71) = 3.54$, $p = .064$, $\eta_p^2 = .05$. The main effect of the condition revealed that learning with known categorical associates yielded significantly better results in the cued-recall test, $F(1, 71) = 27.47$, $p < .001$, $\eta_p^2 = .28$. In addition, the difference between the immediate and delayed posttest was significant, $F(1, 71) = 25.56$, $p < .001$, $\eta_p^2 = .26$. Even if this interaction is considered significant, the interpretation of Experimental Study 4 necessarily remains the same. Based on the descriptive statistics, the declination of the score was smaller for the condition pertaining to the known categorical associates.

Table 8.3

Descriptive Statistics of Posttests With Cues

	Immediate posttest			Delayed posttest		
	M	95% CI	*SD*	*M*	95% CI	*SD*
Categorical	86.4	[78.2, 94.6]	23.9	81.2	[72.5, 89.8]	25.2
Collocation	60.9	[52.0, 69.7]	26.9	49.4	[42.5, 56.4]	21.1

Note. These values indicate the percentages of the correct answers for each condition.

Thus, the overall results of this reanalysis again supported the superiority of known categorical associates over known collocates in facilitating learner memory of the form-meaning link of the target words. Known collocates were once again found to be more effective in facilitating collocational knowledge. The implication of this result is that the link between the target words and the known related words can be more easily established in the case of categorical associates, or categorical associates are superior in their ability to efficiently narrow the meaning that the learners must retrieve. Therefore, when the main

purpose of vocabulary learning is establishing a form-meaning link, the known categorical associates can be more optimally utilized. When learners also intend to develop their collocational knowledge, learning that uses “known + unknown” collocations will contribute to both the form-meaning link and collocational knowledge.

Chapter 9
Conclusion

9.1 Summary of the Present Study's Findings

The present study was undertaken to clarify, compare, and maximize the effects of known related words for paradigmatic and syntagmatic associates. More specifically, the following research questions were formulated for each of the four experimental studies.

Experimental Study 1

RQ1: How are intentional learnings of target words affected when presented with known categorical associates?

RQ1-1: Are memorization and retrieval of target words facilitated if learners are presented with categorical associates?

RQ1-2: How are memorization and retrieval of target words affected by the closeness of the semantic relationship between target and related words?

Experimental Study 2

RQ2: How are intentional learning of form-meaning links and the precise semantic knowledge of target words affected when learners are presented with known synonyms?

RQ2-1: Does the semantic similarity of known synonyms presented together facilitate memorization and retrieval of target word meanings?

RQ2-2: Does the number of known synonyms presented together affect memorization and retrieval of target word meanings?

RQ2-3 How is a deeper understanding of target word meanings affected by the semantic similarity and the number of known synonyms presented together?

Experimental Study 3

RQ3: Of known categorical associates and known collocates, which are more effective in facilitating the learning effects of target words?

RQ3-1: Are the memorization and retrieval of target words more facilitated when presented with known categorical associates or with known collocates?

RQ3-2: Is the acquisition of collocational knowledge of target words more facilitated when presented with known categorical associates or with known collocates?

Experimental Study 4

RQ4: In which case is the learning of target words more facilitated by known collocates when learning collocations intentionally or incidentally?

RQ4-1: In which case are the memorization and retrieval of target words more facilitated by known collocates when learning collocations intentionally or incidentally?

RQ4-2: In which case are the acquisition of collocational knowledge of target words more facilitated by known collocates, when learning collocations intentionally or incidentally?

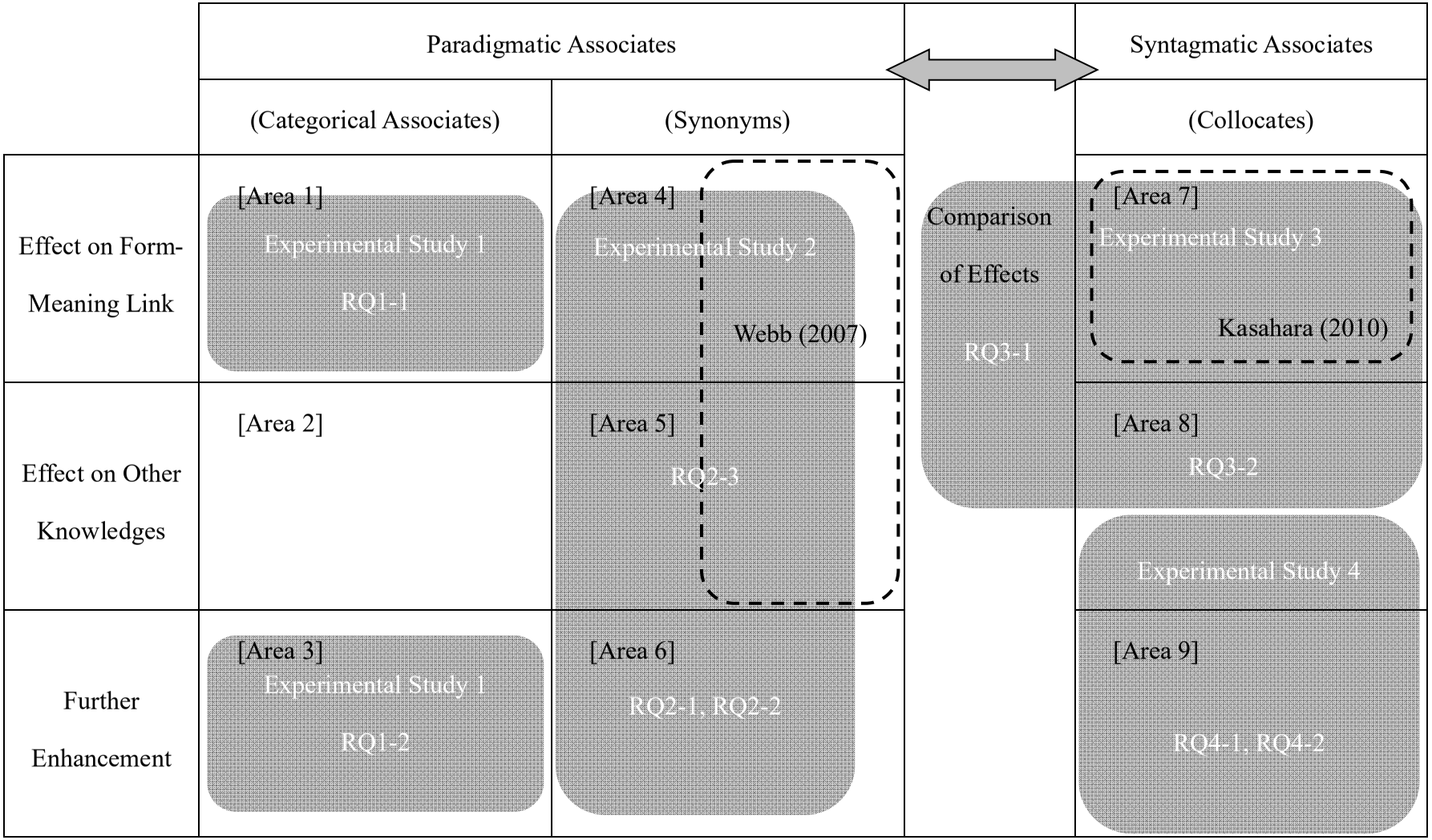

Figure 9.1. The role of each research question.

The role of each research question in the present study as a whole is depicted in Figure 9.1. The facilitative effects of the known categorical associates were verified in Experimental Study 1. For RQ1-1, the results indicated that the facilitative effect of presenting related words as clues for retrieval was generalized to the case of the categorical associates (Area 1). For RQ1-2, the possibility that the effect on the form-meaning link of the target words could be enhanced by greater relatedness was rather high, although the evidence was merely indirect (Area 2).

Experimental Study 2 followed a similar experimental procedure to determine the number of synonyms and their semantic similarity that should be presented with the target words. In addition to confirming the facilitative effect of the known synonyms on the form-meaning link (Area 4), the experiment yielded more detailed information about the effects themselves and ways to enhance them. For RQ2-1, semantic similarity played a significant role in strengthening the facilitative effect pertaining to remembering the form-meaning link of the target words. Regarding RQ2-2, conversely, the number of known synonyms made no significant difference in the effect. In summary, it was suggested that preparing just one semantically close synonym is the most efficient way to enhance cued-recall (Area 5). Concerning RQ2-3, learner performance on the multiple-choice tests was generally high, even after a one-week delay, but no dependence on the number or similarity of synonyms was revealed. It was implied that instructing learners to deliberately and carefully compare and understand the meanings of words may have a positive effect on the preciseness of learners' lexical memory (Area 6).

Experimental Study 3 focused on a comparison of the effects between the cases of the known categorical associates and the known collocates. For RQ3-1, it was clearly demonstrated that the categorical associates more effectively strengthened the form-meaning

link. Concerning RQ3-2, on the other hand, it was clearly shown that known collocates more effectively broadened learners' collocational knowledge (Area 8).

Experimental Study 4 was intended to maximize the effect of the known collocates. Regarding RQ4-1, intentional learning of the "known + unknown" collocations more effectively facilitated the form-meaning link of the target words in the long run. For RQ4-2, however, the conditions did not impact the effects on learners' collocational knowledge. Rather, it is instead prudent to intentionally learn the "known + unknown" collocations when utilizing the known collocates in vocabulary learning. However, care must be taken to avoid overburdening the learners by increasing the number of lexical items they must acquire (Area 9).

Taken together, these results suggest that the known related words work well as cues in learners' retrieval, and this effect is stable in various conditions. Considering that learners must repeatedly retrieve the target words before firmly consolidating them, the known related words, especially the paradigmatic associates, are useful as cues that learners may use to facilitate retrieval on their own, which leads to more optimal memory (Baddeley, 1997).

Regarding semantic relatedness or similarity, higher relatedness or similarity is recommended. In addition, it is almost certain that semantic relatedness, rather than encoding specificity, is the source of the positive effect. Unless learners aim to broaden their collocational knowledge, utilizing the known paradigmatic associates leads to more optimal facilitation of memory.

9.2 Pedagogical Implications

Based on the information revealed through the experimental studies, the current study provides some pedagogical implications. The suggestions are mainly applicable to circumstances in which L2 learners deliberately learn words found by association or when

teachers design class activities or review questions with a focus on vocabulary. For example, when semantic mapping is performed, which is a preparatory task in which a teacher and students activate known and unknown words important for a task that follows, they should make greater use of known expressions while decreasing the number of unknown words. Known related words can also be utilized in the form of notations in texts or be directly introduced by teachers when they explain new vocabulary. This can provide cues that facilitate learners' ability to retrieve lexical memory on their own, which can be utilized during vocabulary tests administered in class or in a task entailing the matching of newly learned words and known synonyms (Aizawa & Mochizuki, 2010). When preparing known related words for the activities explained above, preparing just one high-relatedness word is sufficient, as discussed above.

Known collocates should be introduced when knowing the collocation is especially important. If the frequency or MI of the collocation is especially high or the collocation of a target word is often mistaken, known collocates can benefit learners. It should be noted that when the collocations are incongruent or semantically opaque, this may result in an overload for learners. Therefore, in cases of congruent but easily mistaken expressions (e.g., *small room*, which is often mistakenly phrased as *narrow room*), this method could benefit learners' collocational knowledge.

Based on a comparison of the results from Experimental Studies 1 and 2, it should also be noted that learning too many words at once will not yield good results. The test scores of Experimental Study 2 were clearly lower than those of Experimental Study 1, and this was presumably due to the differing numbers of target words. Thus, it is not recommended that teachers or textbooks introduce too many words at once.

Further, drawing learners' attention to elaborative comparisons and their understanding of meanings might even enhance learners' lexical memory in terms of its preciseness. Thus,

for example, vocabulary tests that are administered in class and include a format that requires learners' precise understanding of meaning (as in the current research) could increase students' attention and processing in their efforts to learn the meanings of words.

Finally, information pertaining to a given strategy should be clarified for students. Students are often unaware that it is inefficient to learn too many words or unknown related words at once. Instead, noticing unknown expressions based on their association with known words is recommended. As students must spend large amounts of time learning incidental vocabulary, which typically uses word lists; thus, it is recommended that teachers advise their students accordingly and facilitate their independent study.

9.3 Limitations and Suggestions for Further Research

Although the present study revealed ample information regarding how teachers and learners should manage known related words to improve the efficiency of their learning, these experimental studies also have several limitations.

First, a small number of participants undermined the power of the statistical test. Although there were no between-participants factors in the experimental designs, the number of participants was relatively small in Experimental Studies 1 and 2. A total of 22 participants was included in Experimental Study 1 and 23 participated in Experimental Study 2. Therefore, some of the statistical comparisons might not have demonstrated the significant differences that should have been revealed.

Additionally, the material used in Experimental Studies 1 and 2 was potentially flawed. Experimental Study 1 included some lexical items with word parts that may have encouraged a recollection of the translations (e.g., *eye* in *eyelash*, *light* in *lighthouse*), although the experimental design carefully included counterbalancing to prevent such unexpected word characteristics from acting as confounders. Also, some of the target and related words in

Experimental Study 2 might have been insufficient in terms of their semantic difference. In other words, some of the words were clearly different and inappropriate for use in accessing more precise knowledge in a multiple-choice format. Thus, if a similar experiment is conducted, the validity of the experiment would be increased by using fewer target words that are appropriate for investigating deeper semantic knowledge.

Further, direct experimental comparison should be used to confirm whether the deeper understanding of the target word meanings was really facilitated by learner attention to differences between the target words and their synonyms.

The learning phase also has some room for improvement. In the learning phase, the learners could learn the target words freely, provided they followed the instructions. Therefore, the time spent on each condition by a given participant was not controlled. Judging from the same results obtained by Kasahara (2010) and Kasahara (2011), the latter of which avoided this lack of control by implementing a between-participants design, this factor does not significantly affect the result of the tests. However, as found in Kasahara (2011), confirmation of this fact is important.

Finally, although the experimental design is appropriate in terms of the relative comparison among conditions in an experiment, the large number of target words in Experimental Study 2 might have compromised the statistical tests. First, Experimental Study 2 was not ecologically valid. In other words, it was not a good reproduction of real-life vocabulary learning. Some participants' memory of the target words was more difficult to maintain than that of others, because similar or related combinations could have been included and caused a negative interference (Tinkham, 1997). This might have resulted in a larger *SD* for the test results and a lower statistical power. It is also possible that participants could not maintain their concentration throughout the long learning phase, thereby resulting in a stronger serial position effect. Therefore, these limitations could have affected the results,

despite the inclusion of counterbalancing in Experimental Study 2, which was intended to alleviate such an issue.

In summary, future research in this vein should overcome the limitations of sample size, types of materials, and ecological validity. In addition, other viewpoints might further reveal how teachers or learners should use known related words in intentional vocabulary learning.

First, although the effect of the known words on intentional vocabulary learning has been verified in some kinds of relationships, other possible relationships among words exist, such as that of antonyms or orthographic similarity. Research on these unknown types of word pairs would be informative, because it is also unclear whether other forms of relationships exert a more facilitative effect on the retrieval of the target word.

Furthermore, in an attempt to obtain more ecological validity, future relevant experiments should include various kinds of tasks that can be conducted as an assignment or classroom activity. Among the studies regarding memory of "unknown + unknown" word sets, Bolger and Zapata (2011) suggested that reading tasks, along with intentional vocabulary learning, weakened the inhibitory effects of unknown related words. In the same way, some tasks that include the development of vocabulary might stimulate learners' motivation or deeper processing, which leads to more optimal memory (Craik & Lockhart, 1972). Therefore, seeking the most appropriate activities using known related words will provide more useful information to maximize their effect.

References

Aizawa, K., & Mochizuki, M. (2010). *Eigo goi shidou no jissen aidia shu* [Practical handbook for English vocabulary teaching]. Tokyo, Japan: Taishukan Shoten.

Altenberg, B. (1993). Recurrent verb-complement constructions in the London-Lund Corpus. In J. Aarts, P. de Haan, & N. Oostdijk (Eds.), *English language corpora: Design, analysis and exploitation* (pp. 227–245). Amsterdam: Rodopi.

Amano, S., & Kondo, T. (1999). *NTT database series "Nihongo no goi tokusei" dai 1-kan tango shinmitsudo* [NTT database series "lexical properties of Japanese words" vol.1 word familiarity]. Tokyo, Japan: Sanseido.

Amano, S., & Kondo, T. (2000). *NTT database series "Nihongo no goi tokusei" dai 7-kan hindo* [NTT database series "lexical properties of Japanese words" vol.7 word frequency]. Tokyo, Japan: Sanseido.

Anderson, M. C. (2003). Rethinking interference theory: Executive control and the mechanisms of forgetting. *Journal of Memory and Language*, *49*, 415–445. doi:10.1016/j.jml.2003.08.006

Anderson, J. R., & Reder, L. M. (1979). An elaborative processing explanation of depth of processing. In L. S. Cermak & F. I. M. Craik (Eds.), *Levels of processing in human memory* (pp. 385–403). Hillsdale, NJ: Lawrence Erlbaum.

Baddeley, A. D. (1997). *Human memory* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum.

Barcroft, J. (2002). Semantic and structural elaboration in L2 lexical acquisition. *Language Learning*, *52*, 323–363. doi:10.1111/0023-8333.00186

Barcroft, J. (2007). Effects of opportunities for word retrieval during second language vocabulary learning. *Language Learning*, *57*, 35–56. doi:10.1111/j.1467-9922.2007.00398.x

Barfield, A. (2009). Exploring productive L2 collocation knowledge. In T. Fitzpatrick & A. Barfield (Eds.), *Lexical processing in language learners* (pp. 95–110). Bristol: Multilingual Matters.

Biber, D., Johansson, S., Leech, G., Conrad, S., and Finegan, E. (1999). *Longman Grammar of Spoken and Written English*. Harlow: Longman.

Biskup, D. (1992). L1 influence on learners' renderings of English collocations: A Polish/German empirical study. In P. Arnaud & H. B´ejoint (Eds.), *Vocabulary and applied linguistics* (pp. 85–93). Amsterdam: John Benjamins.

Boers, F., Demecheleer, M., and Eyckmans, J. (2004). Cross-cultural variation as a variable in comprehending and remembering figurative idioms. *European Journal of English Studies*, *8*, 375–388. doi:10.1080/1382557042000277449

Boers, F., Eyckmans, J., Kappel, J., Stengers, H., & Demecheleer, M. (2006). Formulaic sequences and Perceived oral proficiency: Putting a lexical approach to the test. *Language Teaching Research*, *10*, 245–261. doi:10.1191/1362168806lr195oa

Bogaards, P. (2001). Lexical units and the learning of foreign language vocabulary. *Studies in Second Language Acquisition*, *23*, 321–343. doi:10.1017/S0272263101003011

Bolger, P., & Zapata, G. (2011). Semantic categories and context in L2 vocabulary learning. *Language Learning*, *61*, 614–646. doi:10.1111/j.1467-9922.2010.00624.x

Butler, Y. G. (2002). Second language learners' theories on the use of English articles: An analysis of the metalinguistic knowledge used by Japanese students in acquiring the English article system. *Studies in Second Language Acquisition*, *24*, 451–480. doi:10.1017.S0272263102003042

Channell, J. (1981). Applying semantic theory to vocabulary teaching. *ELT Journal*, *35*, 115–122. doi:10.1093/elt/XXXV.2.115

Craik, F. I. M., & Lockhart, R. S. (1972). Levels of processing: A framework for memory

research. *Journal of Verbal Learning and Verbal Behavior*, *11*, 671–684. doi:10.1016/S0022-5371(72)80001-X

Dunning, T. (1993). Accurate methods for statistics of surprise and coincidence. *Computational Linguistics*, *19*, 61–74. Retrieved from http://www.mitpressjournals.org/loi/coli

Durrant, P., & Schmitt, N. (2010). Adult learners' retention of collocations from exposure. *Second Language Research*, *26*, 163–188. doi:10.1177/0267658309349431

Elgort, I. (2011). Deliberate learning and vocabulary acquisition in a second language. *Language Learning*, *61*, 367–413. doi:10.1111/j.1467-9922.2010.00613.x

Ellis, N., & Beaton, A. (1993). Psycholinguistic determinants of foreign language vocabulary learning. *Language Learning*, *43*, 559–617. doi:10.1111/j.1467-1770.1993.tb00627.x

Ellis, N. C., Simpson-Vlach, R., & Maynard, C. (2008). Formulaic language in native and second language speakers: Psycholinguistics, corpus linguistics, and TESOL. *TESOL Quarterly*, *42*, 375–396. doi:10.1002/j.1545-7249.2008.tb00137.x

Erten, I. H., & Tekin, M. (2008). Effects on vocabulary acquisition of presenting new words in semantic sets versus semantically unrelated sets. *System*, *36*, 407–422. doi:10.1016/j.system.2008.02.005

Eyckmans, J. (2009). Toward an Assessment of Learners' Receptive and Productive Syntagmatic Knowledge. In A. W. Barfield & H. Gyllstad (Eds.), *Researching collocations in another language: Multiple interpretations* (pp. 139–152). Palgrave Macmillan.

Goulden, R., Nation, P., & Read, J. (1990). How large can a receptive vocabulary be? Applied Linguistics, 11, 341–363. doi:10.1093/applin/11.4.341

Grabe, W. (2009). *Reading in a second language: Moving from theory to practice*. Cambridge University Press.

Grant, L. (2005). Frequency of 'core idioms' in the British National Corpus (BNC). *International Journal of Corpus Linguistics*, *10*, 429–451. doi:10.1075/ijcl.10.4.03gra

Gyllstad, H., & Wolter, B. (2016). Collocational processing in light of the phraseological continuum model: Does semantic transparency matter? *Language Learning*, *66*, 296–323. doi:10.1111/lang.12143

Higa, M. (1963). Interference effects of intralist word relationships in verbal learning. *Journal of Verbal Learning and Verbal Behavior*, *2*, 170–175. doi:10.1016/S0022-5371(63)80082-1

Hill, M., & Laufer, B. (2003). Type of task, time-on-task and electronic dictionaries in incidental vocabulary acquisition. *International Review of Applied Linguistics in Language Teaching*, *41*, 87–106. doi:10.1515/iral.2003.007

Hoshino, Y. (2010). The Categorical facilitation effects on L2 vocabulary learning in a classroom setting. *RELC Journal*, *41*, 301–312. doi:10.1177/0033688210380558

Howarth, P. (1996). *Phraseology in English academic writing*. Tübingen, Germany: Max Niemayer Verlag.

Hu, M., & Nation, I. S. P. (2000). Unknown vocabulary density and reading comprehension. *Reading in a Foreign Language*, *23*, 403–430.

Ishii, T. (2015). Semantic connection or visual connection: Investigating the true source of confusion. *Language Teaching Research*, *19*, 712–722. doi:10.1177/1362168814559799

Japan Association of College English Teachers [JACET]. (2003). *JACET list of 8000 basic words*. Tokyo, Japan: Author.

Jiang, N. (2000). Lexical representation and development in a second language. *Applied Linguistics*, *21*, 47–77. doi:10.1093/applin/21.1.47

Jiang, N. (2002). Form-meaning mapping in vocabulary acquisition in a second language. *Studies in Second Language Acquisition 24*, 617–637. doi:10.1017/S0272263102004047

Joe, A. (1998). What effects do text-based tasks promoting generation have on incidental vocabulary acquisition? *Applied Linguistics*, *19*, 357–377. doi:10.1093/applin/19.3.357

Kasahara, K. (2010). Are two words better than one for intentional vocabulary learning? *Annual Review of English Language Education in Japan*, *21*, 91–100.

Kasahara, K. (2011). The effect of known-and-unknown word combinations on intentional vocabulary learning. *System*, *39*, 491–499.

Kasahara, K. (2015). Which cues, adjectives or verbs, provide most assistance for remembering new nowns? *Annual Review of English Language Education in Japan*, *26*, 317–332.

Kawamura, K. (2004). Methodological options for enhancing vocabulary retention: In search of an effective combination of activities. *Annual Review of English Language Education in Japan*, *5*, 199–208.

Kecskes, I. (2018). Formulaic Language and its Place in Intercultural Pragmatics. In A. Siyanova-Chanturia & A. Pellicer-Sanchez (Eds.), *Understanding formulaic language: A second language acquisition perspective* (pp. 132–150). London, UK: Routledge.

Krashen, S. D. (1985). *The input hypothesis: Issues and implications*. Torrance, CA: Laredo Publishing Company.

Laufer, B. (1989). A factor of difficulty in vocabulary learning: Deceptive transparency. *AILA Review*, *6*, 10–20.

Laufer, B., & Hulstijn, J. (2001). Incidental vocabulary acquisition in a second language: The Construct of task-induced involvement. *Applied Linguistics*, *22*, 1–26. doi:10.1093/applin/22.1.1

Leech, G., Rayson, P., & Wilson, A. (2001). *Word frequencies in written and spoken English: Based on the British National Corpus*. Harlow, UK: Pearson Education Limited.

Lindsay, S., & Gaskell, M. G. (2010). A complementary systems account of word learning in

L1 and L2. *Language Learning*, *60*, 45–63. doi:10.1111/j.1467-9922.2010.00600.x

Meara, P. (1990). A note on passive vocabulary. *Second Language Research*, *6*, 150–154. doi:10.1177/026765839000600204

Miller, G. A., & Fellbaum, C. (1991). Semantic networks of English. *Cognition*, *41*, 197–229. doi:10.1016/0010-0277(91)90036-4

Milton, J. (2009). *Measuring second language vocabulary acquisition*. Bristol: Multilingual Matters.

Mondria, J. A. (2003). The effects of inferring, verifying and memorizing on the retention of L2 word meanings. *Studies in Second Language Acquisition*, *25*, 473–499. doi:10.1017.S0272263103000202

Morgan, C. L., & Bonham, D. N. (1944). Difficulty of vocabulary learning as affected by parts of speech. *Journal of Educational Psychology*, *35*, 369–377.

Nation, I. S. P., (2000). Learning vocabulary in lexical sets: Dangers and guidelines. *TESOL Journal*, *9*, 6–10. doi:10.1002/j.1949-3533.2000.tb00239.x

Nation, I. S. P., (2006). How large a vocabulary is needed for reading and listening? Canadian Modern Language Review, 63, 59–82. doi:10.3138/cmlr.63.1.59

Nation, I. S. P., (2013). *Learning vocabulary in another language* (2nd ed.). Cambridge University Press.

Nation, I. S. P., & Webb, S. (2011). *Researching and analyzing vocabulary*. Boston, MA: Heinle.

Nesselhauf, N. (2005). *Collocations in a learner corpus*. Amsterdam: Benjamins.

Oakhill, J., Cain, K., & McCarthy, D. (2015). Inference processing in children: The contribution of depth and breadth of vocabulary knowledge. In E. J. O'Brien, A. E., Cook, & R. F. Lorch (Eds.), *Inferences during reading* (pp. 140–159). Cambridge University Press.

Papathanasiou, E. (2009). An investigation of two ways of presenting vocabulary. *ELT Journal*, *63*, 313–322. doi:10.1093/elt/ccp014

Patrick, H. (Ed.). (2000). *The New Oxford Thesaurus of English*. Oxford University Press.

Pawley, A., & Syder, F. H. (1983). Two puzzles for linguistic theory: Nativelike selection and nativelike fluency. In Richards, J. C. & Schmidt, R. W. (Eds.), *Language and communication* (pp. 191–225). London, England: Routledge.

Perfetti, C. (2007). Reading ability: Lexical quality to comprehension. *Scientific Studies of Reading*, *11*, 357–383. doi:10.1080/10888430701530730

Plonsky, L. (2015). Statistical power, *p* values, descriptive statistics, and effect sizes: A "Back-to-Basics" approach to advancing quantitative methods in L2 research. In Plonsky, L. (Eds.). *Advancing quantitative methods in second language research*. New York: Routledge.

Revier, R. L. (2009). Evaluating a new test of whole English collocations. In A. W. Barfield & H. Gyllstad (Eds.), *Researching collocations in another language: Multiple interpretations* (pp. 125–138). Palgrave Macmillan.

Schmitt, N. (2008). Review article: Instructed second language vocabulary learning. *Language Teaching Research*, *12*, 329–363. doi:10.1177/1362168808089921

Schmitt, N. (2010). *Researching vocabulary: A vocabulary research manual*. Palgrave Macmillan: London, UK.

Shin, D., & Nation, I. S. P. (2007). Beyond single words: The most frequent collocations in spoken English. *ELT Journal*, *62*, 339–348. doi:10.1093/elt/ccm091

Siyanova-Chanturia, A. (2015). Collocation in beginner learner writing: A longitudinal study. *System*, *53*, 148–160. doi:10.1016/j.system.2015.07.003

Siyanova-Chanturia, A., Conklin, K., & van Heuven, W. J. B. (2011). Seeing a phrase "time and again" matters: The role of phrasal frequency in the processing of multiword

sequences. *Journal of Experimental Psychology*, *37*, 776–784. doi:10.1037/a0022531

Slamecka, N. J., & Graf, P. (1978). The generation effect: Delineation of a phenomenon. *Journal of Experimental Psychology: Human Learning and Memory*, *4*, 592–604. doi:10.1037/0278-7393.4.6.592

Sonbul, S., & Schmitt, N. (2013). Explicit and implicit lexical knowledge: Acquisition of collocations under different input conditions. *Language Learning*, *63*, 121–159. doi:10.1111/j.1467-9922.2012.00730.x

Stahl, S., & Nagy, W. (2006). *Teaching word meanings*. Mahwah, NJ: Lawrence Erlbaum Associates.

Swain, M. (1995). Three functions of output in second language learning. In G. Cook & B. Seidlhoffer (Eds.), *Principle and practice in applied linguistics: Studies in honor of H. G. Widdowson* (pp. 97–114). Oxford University Press.

Tada, G. (2017). Intentional vocabulary learning with known associates: Focusing on Number and relatedness (Unpublished master's thesis). University of Tsukuba, Tsukuba, Japan.

Tada, G. (2019). Effects of known related words on vocabulary learning: Comparing categorical associates and collocates. Manuscript submitted for publication.

Takahashi, T. (2000). The effect of count/uncount distinction on the choice of the indefinite English article measured by sentence verification tasks. *Annual Review of English Language Education in Japan*, *11*, 1–10.

Tinkham, T. (1997). The effects of semantic and thematic clustering on the learning of second language vocabulary. *Second Language Research*, *13*, 138–163. doi:10.1191/026765897672376469

Thomson, D. M., & Tulving, E. (1970). Associative encoding and retrieval: Weak and strong cues. *Journal of Experimental Psychology*, *86*, 255–262. doi:10.1037/h0029997

Ushiro, Y. (Ed.). (2009). *Eigo ridingu no kagaku: "Yometa tsumori" no nazo wo toku*

[Science of English reading: Solving the puzzle of “though I thought I could have read”]. Tokyo, Japan: Kenkyusha.

Webb, S. (2007). The effects of synonymy on second-language vocabulary learning. *Reading in a Foreign Language*, *19*, 120–136. Retrieved from http://nflrc.hawaii.edu/rfl/

Webb, S., & Kagimoto, E. (2009). The effects of vocabulary learning on collocation and meaning. *TESOL Quarterly*, *43*, 55–77. doi:10.1002/j.1545-7249.2009.tb00227.x

Webb, S., Newton, J., & Chang, A. (2013). Incidental learning of collocation. *Language Learning*, *63*, 91–120. doi:10.1111/j.1467-9922.2012.00729.x

Wharton, S., & Race, P. (1999). *500 tips for TESOL teachers*. London, England: Kogan Page.

Wilkins, D. A. (1972). *Linguistics in Language Teaching*. London: Arnold.

Wray, A. (2002). Formulaic sequences in second language teaching: Principle and practice. *Applied Linguistics*, *21*, 463–489.

Wolter, B. (2001). Comparing the L1 and L2 mental lexicon: A depth of individual word knowledge model. *Studies in Second Language Acquisition*, *23*, 41–69. doi:10.1017/S0272263101001024

Wolter, B. (2006). Lexical network structures and L2 vocabulary acquisition: The role of L1 lexical/conceptual knowledge. *Applied Linguistics*, *27*, 741–747. doi:10.1093/applin/aml036

Wolter, B., & Gyllstad, H. (2011). Collocational links in the L2 mental lexicon and the influence of L1 intralexical knowledge. *Applied Linguistics*, *32*, 430–449. doi:10.1093/applin/amr011

Wolter, B., & Gyllstad, H. (2013). Frequency of Input and L2 Collocational Processing: A Comparison of Congruent and Incongruent Collocations. *Studies in Second Language Acquisition*, *35*, 451–482. doi:10.1017/S0272263113000107

Yamashita, J., & Jiang, N. (2010). L1 influence on the acquisition of L2 collocations:

Japanese ESL users and EFL learners acquiring English collocations. *TESOL Quarterly*, *44*, 647–668. doi:10.2307/27896758

Zechmeister, E. B., Chronis, A. M., Cull, W. L., D'Anna, C. A., & Healy, N. A. (1995). Growth of a functionally important lexicon. *Journal of Reading Behavior*, *27*, 201–212. doi:10.1080/10862969509547878

Appendices

Appendix A-1: Relatedness Judgment in Experimental Study 1

語彙学習実験 予備調査

以下の「単語A」と「単語B」について、関連度の高低を「1～6で」評価して下さい。
2語が**意味的・イメージ的に近いなど、連想しやすければ高関連度（大きい数字）**です。

1	2	3	4	5	6
低い	やや低い	どちらかといえば低い	どちらかといえば高い	やや高い	高い

（記入例）

単語A	単語B	関連度
hamster (ハムスター)	mouse (ハツカネズミ)	6
hamster (ハムスター)	hippopotamus (カバ)	1

では、以下の単語ペアについて、関連度の評定をお願いいたします。

単語A	単語B	関連度
vulture (ハゲワシ)	eagle (ワシ)	
lighthouse (灯台)	factory (工場)	
mica (雲母)	crystal (水晶)	
zinc (亜鉛)	silver (銀)	
audacity (図太さ)	bravery (勇気)	
hemp (麻)	wool (羊毛)	
mica (雲母)	diamond (ダイヤモンド)	
archipelago (列島)	island (島)	
pancreas (膵臓)	stomach (胃)	

Appendix A-2: Similarity Judgment in Experimental Study 2

語彙学習実験 予備調査

以下の「単語A」と「単語B」について、意味がどれだけ似ているかを「1～6で」評価してください。2語の意味が**近いと思うほど高い数値**をつけてください。
日本語訳はあくまで参考にし、できるだけ英語本来のニュアンスで考えてください。

1	2	3	4	5	6
低い	やや低い	どちらかといえば低い	どちらかといえば高い	やや高い	高い

（記入例）

単語A	対訳	単語B	対訳	類似度
stone	石ころ	pebble	丸石	5
stone	石ころ	gem	宝石	1

では、以下の単語ペアについて、類似度の評定をお願いいたします。

単語A	対訳	単語B	対訳	類似度
upheaval	(混乱、問題となる) 激変	upset	混乱	
inure	(困難に) 慣れさせる	familiarize	慣れさせる、習熟させる	
foible	(性格上の) 弱点、ご愛嬌	limitation	限界	
innuendo	当てこすり	nuance	ニュアンス	
gleam	かすかな光	twinkle	きらめき、輝き	
impart	伝える、知らせる	present	発表する	
gall	苦痛の種、イライラの種	headache	頭痛の種	
facet	一面	viewpoint	観点	

Appendix B-1: Example of Word List in Learning Phase (Categorical Associates)

次に、制限時間 10 分で、次の 30 単語の**意味を**覚えて下さい。

表の右に**「関連語」が提示されている語は、その語と関連づけて**覚えるようにして下さい。

英単語	和訳	関連語	関連語のカテゴリ
avalanche	雪崩	flood（洪水）	災害
sleet	霙	rain（雨）	天気
gimlet	錐	scissors（はさみ）	工具
ardor	熱意	fun（楽しみ）	感情
protractor	分度器	ruler（定規）	文房具
equilibrium	釣り合い	power（力）	物理（力学）
rivulet	小川	pond（池）	地形（水辺）
flesh	肌色	orange（だいだい色）	色
constellation	星座	star（星）	宇宙
bonito	かつお	salmon（鮭）	魚類

Appendix B-2: Example of Word List in Learning Phase (Collocates)

次に、制限時間 5 分で、次の 20 単語の**意味を**覚えて下さい。

このページの単語（単語 1－10）は、右のコロケーション例を**参考にしながら**覚えるようにして下さい。

英単語	和訳	コロケーション例
veterinarian	獣医	chief veterinarian (高位の獣医)
protractor	分度器	clear protractor (透明な分度器)
archipelago	列島	beautiful archipelago (美しい列島)
rivulet	小川	thin rivulet (細い小川)
popsicle	アイスキャンデー	red popsicle (赤いアイスキャンデー)
equilibrium	つり合い	chemical equilibrium (化学的なつり合い)
platypus	鴨の嘴	male platypus (オスの鴨の嘴)
insulator	絶縁体	natural insulator (天然の絶縁体)
mica	雲母	little mica (小さい雲母)
cicada	蝉	adult cicada (成虫の蝉)

Appendix C-1: Example of Cued-Recall Test (Categorical Associates)

次に、共に提示した関連語をヒントに、覚えた単語の意味をご記入ください。

不確かな記憶、断片的な記憶でも、できるだけ書き出すようにして下さい。

（例）eel（ウナギ）→ 「細長いもの」

英単語	和訳	関連語
bonito		salmon（鮭）
sleet		rain（雨）
archipelago		forest（森）
avalanche		flood（洪水）
popsicle		chocolate（チョコレート）
insulator		motor（モーター）
equilibrium		power（力）
vulture		penguin（ペンギン）
ardor		fun（楽しみ）
platypus		frog（蛙）
audacity		kindness（優しさ）
lighthouse		factory（工場）
navel		knee（ひざ）
protractor		ruler（定規）
constellation		star（星）
saliva		tear（涙）
sheath		bomb（爆弾）
rivulet		pond（池）
flesh		orange（だいだい色）
gimlet		scissors（はさみ）

Appendix C-2: Example of Cued-Recall Test (Collocates)

次に、以下の表現の意味をご記入ください。

不確かな記憶、断片的な記憶でも、できるだけ書き出すようにして下さい。

（例）eel（ウナギ）→ 「細長いもの」

英語表現	意味
professional prosecutor	
plastic sheath	
sudden sleet	
southern constellation	
green persimmon	
wild mane	
huge avalanche	
rose petal	
human saliva	
healthy pancreas	

Appendix D: Example of Multiple-Choice Test

最後に、覚えた単語の意味を選択肢の中から選んでください。

lagoon	池	潟	水溜り	湖	湾
innuendo	含蓄、暗示	言及	当てこすり	内緒話	ニュアンス
facet	観点	一面	側面	特色	部分
elation	楽しみ	意気揚々	歓喜	幸せ	興奮
calibre	力量、特長	力、強さ	能力	優越、優勢	才能
spout	洪水	流れ	しぶき	泉、噴水	噴出、噴水
fervency	欲求	熱望	貪欲さ	熱烈さ	熱狂、熱中
edification	情報	教化、啓発	教育	案内	指導

Appendix E: Target Words, Related Words, Mean Relatedness Ratings, and Categories in Experimental Study 1 (With the Translation in Parentheses)

	High relatedness		Low relatedness		
Target words	L2 form	RV	L2 form	RV	Category
constellation	star	5.92	gravity	1.85	space
(星座)	(星)		(重力)		
protractor	ruler	5.31	ink	2.85	stationery
(分度器)	(定規)		(インク)		
sleet	rain	5.15	wind	3.15	weather
(霙)	(雨)		(風)		
bonito	salmon	4.69	octopus	3.15	life in the sea
(かつお)	(鮭)		(タコ)		
equilibrium	power	4.69	speed	1.85	dynamics
(釣り合い)	(力)		(速さ)		
rivulet	pond	4.38	beach	2.85	aquatic landform
(小川)	(池)		(砂浜)		
flesh	orange	4.31	green	2.92	color
(肌色)	(橙色)		(緑)		
avalanche	flood	3.92	storm	2.85	disaster
(雪崩)	(洪水)		(嵐)		
ardor	fun	3.92	fear	2.31	emotion
(熱意)	(楽しみ)		(恐怖)		
gimlet	scissors	3.62	glue	2.77	tool
(錐)	(はさみ)		(糊)		

vulture	eagle	5.92	penguin	2.85	bird
(禿げ鷲)	(鷲)		(ペンギン)		
archipelago	island	5.31	forest	1.85	terrestrial landform
(列島)	(島)		(森)		
sheath	sword	5.08	bomb	1.23	weapon
(鞘)	(剣)		(爆弾)		
insulator	electricity	4.23	motor	3.23	electricity
(絶縁体)	(電気)		(モーター)		
lighthouse	port	4.54	factory	1.77	facility
(灯台)	(港)		(工場)		
audacity	bravery	3.69	kindness	2.77	personality
(図太さ)	(勇気)		(優しさ)		
platypus	duck	4.31	frog	2.77	animal around river
(鴨の嘴)	(鴨)		(蛙)		
popsicle	candy	3.92	chocolate	3.00	confectionery
(ｱｲｽｷｬﾝﾃﾞｰ)	(飴)		(ﾁｮｺﾚｰﾄ)		
navel	chest	3.85	knee	3.08	body
(へそ)	(胸)		(ひざ)		
saliva	sweat	3.69	tear	3.08	body fluid
(唾液)	(汗)		(涙)		
veterinarian	doctor	5.46	singer	2.46	profession
(獣医)	(医者)		(歌手)		
eyelash	eye	5.38	tooth	3.08	face
(まつ毛)	(目)		(歯)		

pancreas	stomach	4.85	tongue	2.77	organ
(すい臓)	(胃)		(舌)		
homoiotherm	mammal	4.77	virus	1.62	classification of life
(恒温動物)	(哺乳類)		(ウイルス)		
bib	apron	4.46	tie	2.46	clothes
(よだれかけ)	(エプロン)		(ネクタイ)		
breaststroke	diving	4.46	soccer	2.00	sport
(平泳ぎ)	(ダイビング)		(サッカー)		
porcupine	hedgehog	4.77	monkey	2.62	animal on land
(山荒らし)	(ハリネズミ)		(猿)		
mane	fur	4.00	horn	3.15	body of animal
(たてがみ)	(毛)		(角)		
blender	oven	3.77	camera	2.23	home appliance
(ミキサー)	(オーブン)		(カメラ)		
gingko	pine	4.46	rose	2.85	plant
(いちょう)	(松)		(バラ)		

Note. RV stands for relatedness value.

Appendix F: Target Words, Synonyms, Mean Similarity Ratings in Experimental Study 2 (With the Translation in Parentheses)

Target words	Rank of similarity	Synonyms	Similarity values
abate	S1	moderate	
(減らす、和らげる)		(和らげる、控えめにする)	4.20
	S2	soothe	
		(和らげる、楽にする)	4.00
	S3	reduce	
		(減少させる)	3.70
	S4	calm	
		(静める)	3.30
abbreviate	S1	shorten	
([言葉などを]短縮する)		(縮める)	4.50
	S2	shrink	
		(縮ませる)	4.10
	S3	reduce	
		(減少させる)	2.70
	S4	edit	
		(編集する)	2.50
aberration	S1	abnormality	
(逸脱)		(異常、変則)	3.90
	S2	irregularity	
		(不規則、変則)	3.10
	S3	variation	
		(変動)	2.70
	S4	rarity	
		(まれなもの)	2.40
annihilate	S1	extinguish	
(全滅させる)		(消す、消滅させる)	5.00
	S2	destroy	
		(破壊する、撲滅する)	4.60

	S3	kill	
		(殺す)	3.90
	S4	erase	
		(削除する)	3.20
batter	S1	beat	
(乱打する、痛めつける)		(叩く、殴る)	4.60
	S2	abuse	
		(虐待する)	3.90
	S3	attack	
		(攻撃する)	3.70
	S4	assault	
		(激しく襲撃する)	3.50
beget	S1	generate	
([よくないものを]生む)		(発生させる、生み出す)	3.50
	S2	produce	
		(引き起こす、もたらす)	3.30
	S3	trigger	
		(誘発する、引金となる)	2.90
	S4	promote	
		(促進する)	1.90
calamity	S1	disaster	
(大災害)		(災害)	5.10
	S2	crisis	
		(危機)	4.10
	S3	tragedy	
		(悲劇、悲劇的事件)	3.50
	S4	accident	
		(事故)	2.70
calibre	S1	ability	
(力量、特長)		(能力)	4.30
	S2	talent	
		(才能)	3.80
	S3	strength	3.40

		(力、強さ)	
	S4	superiority	
		(優越、優勢)	2.50
candour	S1	honesty	
(誠実さ、包み隠さぬさま)		(正直、誠実)	4.70
	S2	openness	
		(率直さ、開放的さ)	4.10
	S3	directness	
		(単刀直入)	3.80
	S4	frankness	
		(ざっくばらんな)	3.50
clasp	S1	hug	
(抱きしめる)		([愛情を持って]抱く)	4.70
	S2	fold	
		(抱える、包む)	3.80
	S3	envelop	
		(包む、覆う)	2.90
	S4	squeeze	
		(絞る、締め付ける)	2.10
concordat	S1	contract	
(協定)		(契約)	4.70
	S2	understanding	
		(合意、取り決め)	4.10
	S3	negotiation	
		(交渉)	3.50
	S4	sale	
		(販売)	2.00
consternation	S1	surprise	
(非常な驚き、狼狽)		(驚き)	4.30
	S2	shock	
		(動揺、ショック)	4.10
	S3	anxiety	
		(不安)	3.20

	S4	fear (恐怖、不安)	3.20
contraption (珍妙な機械、からくり)	S1	device (道具、装置)	3.70
	S2	mechanism ([小さな]機械)	3.60
	S3	machine (機械)	3.30
	S4	invention (発明)	2.60
credence ([根拠のない]信頼)	S1	trust ([直感的な]信用)	3.90
	S2	confidence ([根拠のある]信用)	3.60
	S3	reliance (信頼、依存)	3.10
	S4	acceptance (受容、容認)	2.20
upheaval ([混乱、問題となる]激変)	S1	revolution (変革)	4.10
	S2	disturbance (混乱、騒動)	3.70
	S3	trouble (騒ぎ、いざこざ)	3.30
	S4	upset (混乱)	3.20
defile (けがす、冒涜する)	S1	spoil (台無しにする、損なう)	4.20
	S2	harm (害する、傷つける)	3.90
	S3	ruin (めちゃくちゃにする)	3.30
	S4	destroy	3.00

		(破壊する)	
detritus	S1	waste	
(残骸)		(廃棄物)	3.80
	S2	litter	
		([散らかった]ごみ)	3.60
	S3	trash	
		(がらくた、くず)	3.20
	S4	scrap	
		([金属の]くず)	2.70
edification	S1	instruction	
(教化、啓発)		(指導)	4.50
	S2	education	
		(教育)	4.20
	S3	guidance	
		(案内)	2.60
	S4	information	
		(情報)	2.00
elation	S1	excitement	
(意気揚々)		(興奮)	4.10
	S2	joy	
		(歓喜)	4.10
	S3	delight	
		(楽しみ)	3.70
	S4	happiness	
		(幸せ)	2.80
emend	S1	correct	
([印刷・出版前]校訂する)		(訂正する)	5.10
	S2	revise	
		(改訂する)	4.70
	S3	check	
		(確かめる)	3.50
	S4	polish	
		(磨きをかける)	3.00

envisage	S1	expect	
([将来を]心に描く)		(予期する、期待する)	3.50
	S2	predict	
		(予測する)	3.50
	S3	intend	
		(予定する)	3.40
	S4	forecast	
		(予言する、予報する)	2.90
expedite	S1	accelerate	
(迅速化する)		(加速する)	4.40
	S2	hurry	
		(急ぐ、せきたてる)	3.80
	S3	aid	
		(助ける、手伝う)	2.50
	S4	stimulate	
		(激励する)	2.20
facet	S1	part	
(一面)		(部分)	4.10
	S2	side	
		(側面)	4.00
	S3	viewpoint	
		(観点)	2.90
	S4	characteristic	
		(特色)	2.70
fancier	S1	fan	
(愛好家)		(ファン)	4.60
	S2	addict	
		(依存者、大ファン)	4.10
	S3	enthusiast	
		(熱狂者)	3.80
	S4	expert	
		(名人、専門家)	2.20
fathom	S1	interpret	3.70

(推察する)		(解釈する)	
	S2	perceive	
		(わかる、悟る)	3.10
	S3	understand	
		(理解する)	2.60
	S4	absorb	
		(取り入れる)	1.90
fervency	S1	eagerness	
(熱烈さ)		(熱望)	4.60
	S2	enthusiasm	
		(熱狂、熱中)	4.30
	S3	greed	
		(貪欲さ)	2.90
	S4	appetite	
		(欲求)	2.70
foible	S1	weakness	
([性格上の]弱点、ご愛嬌)		(弱点)	4.20
	S2	defect	
		(欠点)	3.50
	S3	lack	
		(不足)	2.40
	S4	limitation	
		(限界)	1.70
frolic	S1	jump	
(はしゃぐ)		(飛び跳ねる)	3.60
	S2	leap	
		(跳ぶ、跳躍する)	3.30
	S3	hop	
		(ぴょんと跳ぶ)	3.10
	S4	dance	
		(踊る、跳ね回る)	2.90
gall	S1	bother	
(苦痛の種、イライラの種)		(面倒、悩みの種)	4.60

	S2	nuisance (迷惑、迷惑なもの)	4.00
	S3	pain (不愉快なもの)	3.90
	S4	headache (頭痛の種)	3.20
galvanize (衝撃を与える、刺激する)	S1	excite (興奮させる)	3.80
	S2	shock (ぎょっとさせる)	3.50
	S3	awaken (目覚めさせる、喚起する)	3.00
	S4	urge (かりたてる、強く迫る)	2.90
gamut (全領域、全範囲)	S1	breadth (範囲、広さ)	3.40
	S2	area (領域、分野)	3.40
	S3	sequence (順序、連続体)	2.50
	S4	variety (変化、多様性)	1.70
gleam (かすかな光)	S1	shine (光、輝き)	3.70
	S2	twinkle (きらめき、輝き)	3.20
	S3	flash (閃光)	2.80
	S4	beam (光線)	2.80
guzzle (がつがつ食う)	S1	eat (食べる)	4.10
	S2	chew	2.60

		(噛む)	
	S3	consume	
		(消費する)	2.60
	S4	swallow	
		(飲み込む)	2.40
halloo	S1	shout	
(大声で呼びかける)		(叫ぶ、大声で言う)	4.30
	S2	yell	
		(大声をあげる)	4.00
	S3	scream	
		(絶叫する)	3.30
	S4	roar	
		(大声で言う、怒鳴る)	2.90
hoard	S1	stock	
([金品の]貯蔵、蓄え)		(蓄え、在庫)	4.30
	S2	treasure	
		(財宝、財産)	3.40
	S3	heap	
		(堆積)	2.50
	S4	mass	
		(かたまり、集積)	2.10
hurtle	S1	dash	
(突進する、爆走する)		(突進する)	4.10
	S2	rush	
		(急ぐ)	3.90
	S3	run	
		(走る)	3.60
	S4	race	
		(競走する、急ぐ)	2.90
immerse	S1	sink	
(浸す、沈める)		(沈める)	4.00
	S2	soak	
		(濡らす)	3.20

	S3	drench	
		(ずぶ濡れにする)	3.00
	S4	dip	
		(ちょっと浸す、つける)	2.90
impair	S1	damage	
(減じる、損なう)		(損害を与える)	3.90
	S2	reduce	
		(減少させる)	3.80
	S3	spoil	
		(台無しにする、損なう)	3.80
	S4	compromise	
		(妥協する)	2.40
impart	S1	announce	
(伝える、知らせる)		(知らせる、公表する)	4.60
	S2	report	
		(知らせる、報告する)	4.50
	S3	circulate	
		(広める、流通させる)	3.30
	S4	present	
		(発表する)	3.00
implementation	S1	fulfillment	
(実行、実施)		(履行、達成)	4.00
	S2	achievement	
		(達成)	3.50
	S3	realization	
		(実現)	3.10
	S4	application	
		(応用)	2.70
implore	S1	beg	
(哀願する)		(請う)	4.30
	S2	request	
		(要請する、懇願する)	4.00
	S3	ask	3.40

		(頼む)	
	S4	encourage	
		(励ます、仕向ける)	2.20
inauguration	S1	start	
(幕開け、開業)		(開始)	4.40
	S2	birth	
		(誕生、出現)	3.70
	S3	origin	
		(発端、始まり)	3.50
	S4	appearance	
		(発生、登場)	3.30
innuendo	S1	implication	
(当てこすり)		(含蓄、暗示)	3.10
	S2	whisper	
		(内緒話)	2.90
	S3	reference	
		(言及)	2.00
	S4	nuance	
		(ニュアンス)	1.70
inure	S1	adjust	
([困難に]慣れさせる)		(適合させる)	3.90
	S2	accustom	
		(慣れさせる、順応させる)	3.80
	S3	familiarize	
		(慣れさせる、習熟させる)	3.40
	S4	harden	
		(強固にする)	2.30
jaunt	S1	trip	
(小旅行、気晴らし旅行)		(旅行)	4.50
	S2	excursion	
		([団体の]小旅行、遠足)	4.10
	S3	cruise	
		(船旅)	3.00

	S4	drive (ドライブ)	2.90
jilt ([恋人を]捨てる)	S1	dump (捨てる)	4.30
	S2	betray (裏切る)	3.50
	S3	reject (拒絶する)	3.00
	S4	leave (去る)	3.00
kink (よじれ、湾曲)	S1	distortion (ゆがみ、ねじれ)	4.30
	S2	curl (ねじれ、カール)	3.40
	S3	knot (結び目、こぶ)	2.20
	S4	loop (輪)	2.10
knack (こつ、技巧)	S1	method (方法)	3.50
	S2	approach (手がかり、アプローチ)	3.00
	S3	ability (能力、実力)	2.60
	S4	genius (天分)	2.50
knead (練る、こねる)	S1	mix (混ぜる)	3.50
	S2	twist (よじる、ひねる)	3.30
	S3	shape (形作る)	2.50
	S4	crush	2.30

		(押しつぶす)	
lagoon	S1	bay	
(潟)		(湾)	4.00
	S2	lake	
		(湖)	2.70
	S3	pond	
		(池)	2.70
	S4	pool	
		(水溜り)	2.60
leeway	S1	freedom	
(ゆとり、裁量)		(自由)	3.50
	S2	space	
		(余地、余白)	3.50
	S3	flexibility	
		(融通)	3.30
	S4	independence	
		(独立、自立)	2.00
luminary	S1	star	
(有名人、すぐれた人)		(花形、スター)	4.40
	S2	master	
		(名人、達人)	4.20
	S3	leader	
		(第一人者)	3.20
	S4	expert	
		(専門家、権威)	3.00
meekness	S1	obedience	
(おとなしさ、従順)		(従順、服従)	4.00
	S2	modesty	
		(謙遜)	3.90
	S3	gentleness	
		(穏やかさ、親切さ)	3.70
	S4	patience	
		(辛抱強さ)	2.60

ramble	S1	wander	
(散策する)		(ぶらつく、放浪する)	4.30
	S2	trek	
		(徒歩旅行する)	3.80
	S3	drift	
		(さまよう、漂流する)	3.50
	S4	hike	
		(ハイキングする)	3.20
regalia	S1	dress	
([要職を示す]礼服)		(礼装、正装)	4.50
	S2	clothes	
		(衣服)	3.90
	S3	uniform	
		(制服)	3.40
	S4	costume	
		(衣装、扮装)	3.40
relish	S1	excitement	
(強い喜悦、興奮)		(興奮)	4.80
	S2	pleasure	
		(喜び)	4.00
	S3	satisfaction	
		(満足)	2.90
	S4	liking	
		(愛好)	2.60
spout	S1	fountain	
(噴出、噴水)		(泉、噴水)	4.60
	S2	spray	
		(しぶき)	4.10
	S3	flood	
		(洪水)	2.90
	S4	stream	
		(流れ)	2.50
tariff	S1	tax	4.80

(関税)		(税)	
	S2	toll	
		(通行料、使用料)	3.30
	S3	charge	
		(料金)	2.40
	S4	fee	
		(授業料、謝礼)	1.70
vivacity	S1	energy	
(快活さ)		(活力)	4.50
	S2	vitality	
		(元気、生命力)	4.30
	S3	vividness	
		(鮮やかさ、明瞭さ)	2.90
	S4	happiness	
		(幸せ)	2.70
vocation	S1	profession	
(職業、適職)		([知的、専門的]職業)	5.00
	S2	job	
		(仕事、作業)	4.60
	S3	business	
		(職業、業務)	4.10
	S4	mission	
		(使命、天職)	4.00

Note. S1–4 stand for the synonyms with the highest, second highest, second lowest, and the lowest similarity, respectively.

Appendix G: Target Words, Collocates, Categorical Associates, Mean Relatedness Ratings, and Categories in Experimental Studies 3 and 4 (With the Translation in Parentheses)

		CAs		
Target words	Collocates	L2 form	RV	Category
constellation	southern constellation	star	5.92	space
(星座)	(南の星座)	(星)		
sheath	plastic sheath	sword	5.08	weapon
(鞘)	(プラスチックの鞘)	(剣)		
sleet	sudden sleet	rain	5.15	weather
(霙)	(突然の霙)	(雨)		
pancreas	healthy pancreas	stomach	4.85	organ
(すい臓)	(健康的なすい臓)	(胃)		
petal	rose peta.	leaf	5.15	body of plant
(花びら)	(バラの花びら)	(葉)		
prosecutor	professional prosecutor	court	5.23	judgement
(検事)	(プロの検事)	(裁判所)		
mane	wild mane	fur	4.00	body of animal
(たてがみ)	(野性的なたてがみ)	(毛)		
avalanche	huge avalanche	flood	3.92	disaster
(雪崩)	(巨大な雪崩)	(洪水)		
saliva	human saliva	sweat	3.69	body fluid
(唾液)	(ヒトの唾液)	(汗)		

persimmon	green persimmon	apple	4.38	fruit
(柿)	(緑色の柿)	(りんご)		
veterinarian	chief veterinarian	doctor	5.46	profession
(獣医)	(高位の獣医)	(医者)		
archipelago	beautiful archipelago	island	5.31	terrestrial landform
(列島)	(美しい列島)	(島)		
mica	little mica	glass	3.85	mineral
(雲母)	(小さな雲母)	(ガラス)		
insulator	natural insulator	electricity	4.23	electricity
(絶縁体)	(天然の絶縁体)	(電気)		
protractor	clear protractor	ruler	5.31	stationery
(分度器)	(透明な分度器)	(定規)		
equilibrium	chemical equilibrium	power	4.69	dynamics
(釣り合い)	(化学的なつり合い)	(力)		
platypus	male platypus	duck	4.31	animal around river
(鴨の嘴)	(オスの鴨の嘴)	(鴨)		
popsicle	red popsicle	candy	3.92	confectionery
(ｱｲｽｷｬﾝﾃﾞｰ)	(赤いｱｲｽｷｬﾝﾃﾞｰ)	(飴)		
rivulet	thin rivulet	pond	4.38	aquatic landform
(小川)	(細い小川)	(池)		
cicada	adult cicada	butterfly	3.69	bugs
(蝉)	(成虫の蝉)	(蝶)		

Note. RV stands for relatedness value, CA stands for categorical associates.