Relationship Between Depth of Collocation Knowledge and L2 Proficiency Using the Depth Test of Collocation Knowledge

Kiwamu KASAHARA  
Hokkaido University of Education

Rie KOIZUMI  
Tokiwa University

Abstract

The purpose of this study is to examine how L2 learners’ collocation knowledge can reflect their proficiency of a target language. The experiment was conducted with 159 first-year university students, who took (a) an English proficiency test called the Computerized Assessment System for English Communication (CASEC) and (b) a collocation test named the Depth Test of Collocation Knowledge (DTCK), which was developed to measure the depth of participants’ collocation knowledge. DTCK includes two types of test items: One asks test-takers to choose a word that collocates with the most frequent meaning of a target word, and the other asks them to choose a word that collocates with another less frequent meaning of the target word. We found the following: (a) There was a high positive correlation between CASEC and DTCK, which indicates that learners’ collocation knowledge can be a strong indicator of their L2 proficiency; (b) the higher L2 proficiency was, the deeper L2 collocation knowledge he or she showed; and (c) the participants had more knowledge of the collocates combined with the frequent meaning of the target words than of the collocates combined with less frequent meaning of the target words, regardless of their proficiency.

1. Introduction

1.1 Measuring Vocabulary Knowledge

It is widely believed that vocabulary knowledge strongly reflects L2 proficiency, so it can be a reliable indicator of one’s L2 proficiency. A large number of studies have shown strong correlations between vocabulary knowledge and four skills, especially reading comprehension (Alderson, 2000; Anderson & Freebody, 1981; Coady, Magoto, Hubbard, Grane, & Mokhtari, 1993; Koda, 1989; Laufer & Sim, 1985a, 1985b). This is the primary reason why many vocabulary researchers have engaged in measuring vocabulary knowledge in the past few decades (Milton, 2009; Read, 2000; Schmitt, 2010).
However, measuring vocabulary knowledge is far from easy because it consists of various aspects. Nation (2001) proposed a famous categorization in which he divided vocabulary knowledge into three major categories—form, meaning, and use—each of which has three subcategories with the receptive and productive distinctions. From the perspective of measurement, vocabulary knowledge is often discussed in the context of three dimensions: lexical breadth, lexical depth, and fluency (Daller, Milton, & Treffers-Daller, 2007). Lexical breadth reflects the size of someone’s vocabulary, or the number of words for which a learner knows the primary meaning. Lexical depth measures the depth of the learner’s knowledge about the words for which he or she knows the primary meaning. This concept includes not only meanings but also associations, grammatical functions, collocations, and constraints on use. Fluency measures how quickly the learner is able to draw from his or her mental lexicon when he or she uses words or information about words.

Of the three dimensions, measuring vocabulary size (or lexical breadth) has been explored and developed the furthest. Vocabulary size can often be a reliable indicator of a learners’ L2 proficiency because receptive understanding of a word’s meaning is the first necessary step of vocabulary acquisition (Milton, 2009). In terms of novice or intermediate L2 learners, vocabulary size can serve as the most crucial indicator of their proficiency because they are not expected to have acquired various types of vocabulary knowledge other than connections between form and meaning, which are acquired first (Meara, 1996). In addition, vocabulary size is relatively easy to measure. A vocabulary size test can include numerous items and be conducted in a short period, regardless of whether it is a checklist type of test (Meara & Jones, 1990) or a multiple-choice test (Mochizuki, 1998; Nation, 1990, 2001).

Measuring lexical depth and fluency is far more complex. As stated above, lexical depth includes various aspects such as associations, grammatical functions, collocations, and constraints on use. These are so closely interrelated that it is hard to identify and measure certain specific aspects (Read, 2000). Lexical fluency is even more difficult to measure: Measuring how fast a learner can use his or her vocabulary knowledge requires a very sensitive method. Measuring fluency is a new area of research on applied linguistics, and few studies have been done on it (Hase et al., 2011; Iso, Aizawa, & Tagashira, 2011; Mochizuki, 2010). Therefore, this study focuses on measuring lexical depth, which has a longer history of research.

1.2 Measuring Lexical Depth

Mochizuki (2010) argued that L2 vocabulary size could predict a large portion of learners’ L2 proficiency. However, measuring lexical depth has a role in clarifying the process of L2 vocabulary acquisition. L2 learners cannot obtain various aspects of vocabulary knowledge at one time. Encountering the same words repeatedly in different contexts gradually and incrementally deepens a learner’s knowledge of the words (Nation, 2001, 2008). Measuring various aspects of the vocabulary knowledge of L2 learners in different developmental stages can contribute to
elucidating the mechanisms of complex L2 vocabulary learning and to developing an effective methodology for teaching vocabulary.

According to Read (2000) and Schmitt (2010), there are two approaches to measuring lexical depth: One is to ask learners how well they know specific words, and the other is to test some aspects of their vocabulary knowledge. The former approach takes the form of a questionnaire and asks learners how far they think they have progressed along a continuum that ranges from receptive to productive mastery of specific words. The Vocabulary Knowledge Scale (Paribakht & Wesche, 1997; Wesche & Paribakht, 1996) is a typical example of this type of questionnaire. The latter approach includes various types of tests that measure specific aspects of vocabulary knowledge. One well-known test is the Word Associates Test (Read, 1993, 1998, 2000), which measures learners’ knowledge of paradigmatic or syntagmatic links to specific words. Another famous test is the Test of English paradigmatic and syntagmatic Derivatives (Schmitt & Zimmerman, 2002), which measures learners’ knowledge of derivative forms of words.

1.3 Measuring Collocation Knowledge

Of the tests that measure some components of vocabulary knowledge, a promising one that can reflect learners’ L2 proficiency is a collocation test. Schmitt (2010) insisted that collocation is “one of the most important types of ‘contextualized’ word knowledge” and is “a good candidate for a depth-of-knowledge test” (p. 229). A collocation is a combination of two or more words that happens more frequently than would happen by chance (e.g., strong coffee). Collocations are so widespread in English that a mastery of them can discriminate native English speakers from those learning English (Durrant & Schmitt, 2009; Siyanova & Schmitt, 2007). Moreover, the depth of one’s collocation knowledge can also distinguish advanced learners of English from novice ones (Hsu, 2007; Schmitt, 2010). Acquiring collocation knowledge is a long and time-consuming process for L2 learners. Clarifying the process can shed light on the mechanisms of L2 vocabulary acquisition.

One problem with early collocation tests was that they used a small number of items, regardless of whether they took the form of L1-L2 translation tests (Bahns & Eldaw, 1993, used 15 items) or fill-in-the-blank tests (Farghal & Obiedat, 1995, used 22 items). This problem was overcome by some studies that used larger numbers of items, which were tested out of context (Barfield, 2003 [used 120 items], 2009 [used a maximum of 90 items]; Gyllstad, 2005 [used more than 100 items]). Such studies showed that advanced learners tend to comprehend and produce more collocations than novice learners. However, there is another serious problem—previous studies have not considered the quality of the collocations tested. Thus, they dealt with the size, not the depth, of collocation knowledge. Researchers in previous studies selected collocations without considering the developmental stages of learners’ collocation knowledge. A high-frequency word with multiple meanings has a large number of collocations because each meaning of the word has several typical collocations. For example, the word black has the
frequent collocation *black hair* (i.e., dark hair) and the less frequent collocation *a black look* (i.e., an angry expression). It seems natural that L2 learners would first learn the collocates that combine with the most frequent meaning of the word (*black hair*) and later learn the collocates that combine with a less frequent meaning of the word (*a black look*). Thus, advanced learners may know more of the latter type. However, this prediction needs to be tested because of the lack of studies examining the quality of collocation knowledge. A collocation test that includes these two types of collocations would be useful for research and pedagogy in distinguishing L2 learners in different developmental stages.

The purpose of this study is to investigate whether L2 learners’ depth of collocation knowledge reflects their L2 proficiency. This study considers two-word combinations that occur together as collocations. The present study developed a collocation test called the Depth Test of Collocation Knowledge (DTCK), which adopted two types of collocations: One consists of a target word and another word combined with the target word’s most frequent meaning, and the other consists of a target word and another word combined with a less frequent meaning of the target word. All the collocations in DTCK were either combinations of (a) an adjective and a noun (e.g., *wild animal*) or (b) two nouns (e.g., *tree branch*) because two-word combinations are the most basic collocations (Simpson-Vlach & Ellis, 2010), and adjective/noun and noun/noun combinations are two of the most basic two-word combinations (Aitchison, 2003; Barfield, 2009).

2. Method

2.1 Research Questions

The present study addressed the following three research questions:

(1) Is there a positive correlation between L2 learners’ collocation knowledge and their L2 proficiency?

(2) Do learners with higher L2 proficiencies have a deeper knowledge of L2 collocations than those with lower proficiencies?

(3) Are collocates combined with the most frequent meaning of a target word more familiar to learners than collocates combined with a less frequent meaning of the target word, regardless of the learner’s L2 proficiency?

Although previous studies have asked questions similar to Research Question 1 and suggested positive correlations between collocation knowledge and the proficiency of L2 learners (Hsu, 2007; Schmitt, 2010), this study advances previous studies in two ways. First, we examine the quality of collocation knowledge (differences in collocation types) in Research Questions 2 and 3. Second, the test used for this study has a sufficient number of test items (50 items). These characteristics make the current study unique, opening a new research avenue.
2.2 Participants

The participants in the present study were 159 first-year university students, all of whom were Japanese learners of English with more than six years of English learning experience. Originally, 160 students joined the experiment, but one student who received no points on DTCK was excluded.

2.3 Materials

2.3.1 Depth Test of Collocation Knowledge (DTCK)

DTCK was developed for this study to measure participants’ English collocation knowledge. The test-takers were asked to choose two words from four options that collocate syntagmatically with a target word. One collocate combines with the most frequent meaning of the target word (high-frequency collocates, or HFCs), and the other combines with a less frequent meaning of the target word (low-frequency collocates, or LFCs).

First, 25 target words were chosen from *JACET List of 8000 Basic Words* (JACET 8000; Japan Association of College English Teachers Basic Word Revision Committee, 2003), a word list developed specifically for Japanese learners of English that includes 8,000 words selected based on frequency. All of the target words were high-frequency adjectives or nouns with multiple meanings that belong to the list’s 1,000- or 2,000-word level. The authors referred to one of the widespread English-English monolingual dictionaries for English as a second/foreign language learners, *Longman Dictionary of Contemporary English* (LDCE; 5th ed., Pearson Longman Limited, 2009), and decided that the first definition listed for each target word was the most frequent meaning. The other entries were then categorized into less frequent meanings of the target word. Then, the two types of collocates (HFCs and LFCs) were selected by referring to *LDCE* and *Longman Collocations Dictionary for Students of English* (Oxford University Press, 2002). All of the collocates were also high-frequency nouns or adjectives from *JACET 8000’s* 1,000- or 2,000-word level. Third, two distractors for each test item were chosen from *JACET 8000*. The authors selected distractors that did not create a collocation when combined with the target word, which the authors determined using the two dictionaries mentioned above.

To ensure the validity of correct and incorrect options in DTCK, the study employed three experienced English teachers at a university: one native English speaker from Canada and two native Japanese speakers. They were instructed to check the following two points: (a) whether the HFCs were combined with the most frequent meaning of the target words and the LFCs were combined with a less frequent meaning of the target words and (b) that the distractors did not co-occur with the target words. Any collocates or distractors to which one or more of the teachers objected were replaced by other candidates until there was no disagreement. The final items chosen are listed in the Appendix.
One point was given for each correct answer; no points were given for a wrong answer. The highest possible score for each subtest of collocate types was 25 points; the highest possible score for the entire test was 50 points.

2.3.2 Computerized Assessment System for English Communication (CASEC)

To measure the participants’ English proficiency, we used CASEC, which was developed by the Japan Institute for Educational Measurement, Inc. (JIEM). According to JIEM (n.d._a), there is a high positive correlation (r = .84) between CASEC and Test of English for International Communication (TOEIC) scores. CASEC is a computer adaptive test for evaluating test-takers’ ability to communicate in English. Employing Item Response Theory enables CASEC to present each test-taker with questions suitable to his or her level of English proficiency. The test is taken on a computer, and test-takers’ answering patterns tell the CASEC system which level of items is appropriate for exactly measuring their proficiency. The test, which takes 40-50 minutes to complete, consists of four sections: vocabulary, expressions, listening to main ideas, and listening to detailed information. Scores vary from zero to 1,000. Based on their CASEC scores, the participants were divided into three groups: the Upper Group (n = 53), the Middle Group (n = 53), and the Lower Group (n = 53). A one-way ANOVA revealed that there was a significant difference between the groups in terms of their CASEC scores, F(2, 156) = 374.47, p < .0001. Post hoc Tukey tests showed that the Upper Group had significantly better scores than the Middle and Lower Groups, and that the Middle Group had significantly higher scores than the Lower Group.

2.4 Procedure

First, the participants were placed in a computer room and given 50 minutes to complete CASEC. Then, they were given 10 minutes to take DTCK.

2.5 Data Analysis

To examine Research Question 1, the authors used the Pearson product-moment correlation coefficients between the total DTCK scores, the HFC and LFC scores, and the total CASEC scores. We also estimated the actual degrees of correlation when test reliabilities are perfect (e.g., Glass & Hopkins, 1996); this estimation is called correction for attenuation, and corrected correlation coefficients are derived after the correction is made.

To investigate Research Question 2, a two-way ANOVA (3 groups [between-subjects] x 2 collocate types [within-subjects]) was conducted to investigate whether there was an interaction between the groups and the types of collocations. When there was a statistically significant interaction, one-way ANOVAs were employed to see the differences between the groups in terms of the HFC and LFC scores.

To examine Research Question 3, paired-sample t tests were carried out between the HFC
and LFC scores in each group. According to Keppel and Wickens (2004), ANOVAs makes the following assumptions: (a) “The observations from any subject must be independent of those from every other subject, and the distribution of observations is the same for every subject” (for between-subjects analyses; p. 442) and (b) “the variances of the different treatment scores were all the same and the correlations among them were identical” (i.e., sphericity assumption for within-subjects analyses; p. 443). We checked the first assumption and found that it was satisfied. For the second assumption, Mauchly’s Test of Sphericity showed a significant result, and we used the Greenhouse-Geisser corrected result.

3. Results

Table 1 shows the means, standard deviations, and minimum and maximum values of all participants and those in each subgroup in CASEC and DTCK (HFCs, LFCs, and the total). The table also includes mean percentages converted using the means and top scores to allow for

Table 1

Descriptive Statistics for CASEC and DTCK Scores

<table>
<thead>
<tr>
<th>Test</th>
<th>Mean</th>
<th>SD</th>
<th>Minimum</th>
<th>Maximum</th>
</tr>
</thead>
<tbody>
<tr>
<td>All participants (N = 159)</td>
<td>418.77</td>
<td>139.04</td>
<td>95</td>
<td>876</td>
</tr>
<tr>
<td>CASEC</td>
<td>418.77</td>
<td>139.04</td>
<td>95</td>
<td>876</td>
</tr>
<tr>
<td>DTCK (HFC)</td>
<td>21.01</td>
<td>3.00</td>
<td>5</td>
<td>25</td>
</tr>
<tr>
<td>DTCK (LFC)</td>
<td>13.96</td>
<td>2.59</td>
<td>4</td>
<td>20</td>
</tr>
<tr>
<td>DTCK (Total)</td>
<td>34.97</td>
<td>4.57</td>
<td>9</td>
<td>44</td>
</tr>
<tr>
<td>The Upper Group (n = 53)</td>
<td>575.72</td>
<td>73.51</td>
<td>486</td>
<td>876</td>
</tr>
<tr>
<td>CASEC</td>
<td>575.72</td>
<td>73.51</td>
<td>486</td>
<td>876</td>
</tr>
<tr>
<td>DTCK (HFC)</td>
<td>23.17</td>
<td>1.76</td>
<td>17</td>
<td>25</td>
</tr>
<tr>
<td>DTCK (LFC)</td>
<td>15.32</td>
<td>2.59</td>
<td>9</td>
<td>20</td>
</tr>
<tr>
<td>DTCK (Total)</td>
<td>38.49</td>
<td>3.61</td>
<td>9</td>
<td>44</td>
</tr>
<tr>
<td>The Middle Group (n = 53)</td>
<td>413.60</td>
<td>44.01</td>
<td>334</td>
<td>481</td>
</tr>
<tr>
<td>CASEC</td>
<td>413.60</td>
<td>44.01</td>
<td>334</td>
<td>481</td>
</tr>
<tr>
<td>DTCK (HFC)</td>
<td>20.87</td>
<td>2.82</td>
<td>5</td>
<td>24</td>
</tr>
<tr>
<td>DTCK (LFC)</td>
<td>16.35</td>
<td>2.52</td>
<td>4</td>
<td>20</td>
</tr>
<tr>
<td>DTCK (Total)</td>
<td>35.47</td>
<td>4.22</td>
<td>9</td>
<td>41</td>
</tr>
<tr>
<td>The Lower Group (n = 53)</td>
<td>266.98</td>
<td>52.78</td>
<td>95</td>
<td>333</td>
</tr>
<tr>
<td>CASEC</td>
<td>266.98</td>
<td>52.78</td>
<td>95</td>
<td>333</td>
</tr>
<tr>
<td>DTCK (HFC)</td>
<td>19.00</td>
<td>2.73</td>
<td>11</td>
<td>24</td>
</tr>
<tr>
<td>DTCK (LFC)</td>
<td>12.87</td>
<td>2.19</td>
<td>9</td>
<td>18</td>
</tr>
<tr>
<td>DTCK (Total)</td>
<td>31.87</td>
<td>3.15</td>
<td>25</td>
<td>40</td>
</tr>
</tbody>
</table>

Note. Top scores = 25 for the HFCs and LFCs and 50 for the whole test. Raw scores were used for statistical tests.
comparisons between the three types of DTCK scores. The mean percentages for the whole group were 84.04% for HFCs, 55.84% for LFCs, and 69.94% for the whole test, suggesting that the difficulties of the three types of DTCK were within an appropriate range. The reliability (Cronbach's alpha) for the total DTCK scores was not high ($\alpha = .61$); reliability for the HFC scores was higher ($\alpha = .70$), but reliability for the LFC scores was low ($\alpha = .20$), probably due to small standard deviations. To compensate for the low reliabilities, corrections for attenuations were employed to estimate the actual degree of correlations when test reliabilities were 1.0 (Glass & Hopkins, 1996), as explained in section 2.5 above.

Table 1 shows the CASEC average score was 418.77 out of 1,000, among the whole participant group, 575.72 in the Upper Group, 413.60 in the Middle Group, and 266.98 in the Lower Group. Using the conversion table provided by JIEM (n.d.), each CASEC mean roughly corresponds to 315-355 in the TOEIC scores for the whole group, 465-545 for the Upper Group, 315-355 for the Middle Group, and 205-240 for the Lower Group. This suggests that the study participants generally had beginning or intermediate levels of English proficiency.

Table 2

*Pearson Correlation Coefficients (in the Upper Right Diagonal) and Corrected Correlation Coefficients (in the Left Diagonal) Between CASEC and the Three Types of DTCK (N = 159)*

<table>
<thead>
<tr>
<th>Test</th>
<th>CASEC</th>
<th>DTCK (HFC)</th>
<th>DTCK(LFC)</th>
<th>DTCK (Total)</th>
<th>Reliability</th>
</tr>
</thead>
<tbody>
<tr>
<td>CASEC</td>
<td>—</td>
<td>.60**</td>
<td>.45**</td>
<td>.65**</td>
<td>.93*</td>
</tr>
<tr>
<td>DTCK (HFC)</td>
<td>.74</td>
<td>—</td>
<td>.33**</td>
<td>.85**</td>
<td>.70</td>
</tr>
<tr>
<td>DTCK (LFC)</td>
<td>1.00*</td>
<td>.88</td>
<td>—</td>
<td>.78**</td>
<td>.20</td>
</tr>
<tr>
<td>DTCK (Total)</td>
<td>.86</td>
<td>1.00*</td>
<td>1.00*</td>
<td>—</td>
<td>.61</td>
</tr>
</tbody>
</table>

*Note. The values corrected for attenuation were computed using the following formula (Glass & Hopkins, 1996, p. 126): (Pearson $r$) / $\sqrt{(\text{reliability of one test} \times \text{reliability of another test})}$; for example, the corrected coefficient between CASEC and DTCK (HFC) is derived from $$.60 / \sqrt{(.93 \times .70)} = .74.$$*

*Reported by JIEM (n.d._b). *The actual computed value was beyond 1.00.

** $p < .01$.

Table 2 presents the Pearson correlation coefficients and coefficients corrected for attenuation between the tests. The Pearson correlation coefficients show significant positive moderate correlations between CASEC and the total scores of DTCK ($r = .65$) and between CASEC and the HFC scores ($r = .60$); there was a less moderate correlation between CASEC and the LFC scores ($r = .45$). However, the corrected coefficients suggest that such correlations were underestimated due to low reliabilities and that all relationships could be high (.74 to 1.00). Furthermore, the Pearson correlations suggest positive high correlations between the total and each type of collocate ($r = .85$ and .78) and a low correlation between the two types of collocates
(r = .33). However, again, the corrected coefficients indicate that all three associations are strong (.88 to 1.00). It should be noted that the corrected coefficients are estimates, which can be overestimated and sometimes go beyond 1.00.

The result of the two-way ANOVA showed that there was a significant interaction between the three groups and the two types of collocates, F(2, 156) = 3.88, p < .05, with a small effect size, η² = .01. In terms of the HFC scores, the one-way ANOVA revealed that there was a significant difference between the three groups, F(2, 156) = 37.51, p < .0001, η² = .32. Post hoc Tukey tests showed that the Upper Group received significantly higher scores than the Middle and Lower Groups and that the Middle Group received significantly higher scores than the Lower Group. The results were a little different for the LFC scores. According to the one-way ANOVA, there was a significant difference between the groups, F(2, 156) = 14.38, p < .0001, η² = .16. Post hoc Tukey tests showed that the Upper Group was significantly superior to the Middle and Lower Groups. However, there was no significant difference between the Middle and Lower Groups. Finally, paired-sample t tests were conducted within each group to examine whether there was a difference between the two types of collocation knowledge. The results showed that the HFC scores were significantly higher than the LFC scores in all of the groups: for the Upper Group, t(52) = 24.77, p < .0001, r = .96; for the Middle Group, t(52) = 15.94, p < .0001, r = .91; and for the Lower Group, t(52) = 11.68, p < .0001, r = .85.

Table 3
Top 10 Difficult and Easy HFC and LFC Items

<table>
<thead>
<tr>
<th>Most difficult HFC items</th>
<th>Easiest HFC items</th>
</tr>
</thead>
<tbody>
<tr>
<td>No. 2 (rough ground), No. 22 (newspaper article), No. 25 (odd story), No. 13 (tree branch), No. 15 (dry towel), No. 24 (sharp knife), No. 7 (rich family), No. 20 (strong muscle), No. 18 (flat area), No. 3 (late arrival)</td>
<td>No. 21 (hot day), No. 16 (easy question), No. 12 (black hair), No. 19 (clean room), No. 9 (long river), No. 10 (new address), No. 6 (green vegetable), No. 1 (heavy bag), No. 17 (soft bed), No. 5 (wild animal)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Most difficult LFC items</th>
<th>Easiest LFC items</th>
</tr>
</thead>
<tbody>
<tr>
<td>No. 18 (flat beer), No. 17 (soft job), No. 7 (rich blue), No. 12 (black look), No. 22 (household article), No. 24 (sharp increase), No. 25 (odd shoe), No. 15 (dry wine), No. 4 (loud suit), No. 23 (monthly interest)</td>
<td>No. 9 (long memory), No. 8 (busy street), No. 10 (opening address), No. 1 (heavy traffic), No. 2 (rough idea), No. 3 (late summer), No. 21 (hot issue), No. 14 (bright girl), No. 16 (easy day), No. 20 (strong coffee)</td>
</tr>
</tbody>
</table>
Although not directly related to the research questions, mean percentages were observed in order to gain further insights into learners' collocation knowledge. Thus, it was revealed that the Upper Group had very high percentages of correct answers (92.68%) but that even in this group, some participants failed to correctly identify some of the HFCs (four items on average). Table 3 presents the 10 most difficult and easiest HFC and LFC items. This list indicates that some collocations that look like they would be easy to acquire are not in fact easy for L2 learners to acquire (e.g., dry towel, rich family, and flat area are among the difficult HFC items). Meanwhile, other collocations that seem like they would be difficult to acquire are apparently not very difficult for L2 learners to acquire (e.g., long memory, opening address, and late summer are among the easy LFC items).

4. Discussion

The answer to Research Question 1 was affirmative: This study showed that there was a high positive correlation between the participants’ English proficiency and the depth of their English collocation knowledge. Several studies have revealed that vocabulary knowledge plays a crucial role in learners’ L2 proficiency. The correlation between the two tests shown in this study supports the results of Hsu (2007) and Schmitt (2010): L2 learners’ collocation knowledge can be a reliable indicator of their L2 proficiency.

Research Question 2 concerns whether learners with greater English proficiency have a deeper knowledge of English collocations. The present study indicated that this is the case. The scoring patterns for the two types of collocations, however, turned out to be different. In terms of learners’ knowledge about HFCs, the Upper Group was superior to the Middle Group, which was superior to the Lower Group. This indicates that learners with higher proficiency tend to know more words that combine with the most frequent meanings of the target words.

As for LFCs, the Upper Group had significantly more knowledge about LFCs than the other groups. However, there was no significant difference between the Middle and Lower Groups. Thus, compared to the Lower Group, the Middle Group likely has more knowledge about HFCs, but it has a similar degree of knowledge about LFCs. This implies that knowledge about LFCs is more advanced and is acquired in a later stage of L2 vocabulary acquisition than knowledge about HFCs. This may be because L2 learners tend to have more opportunities to see HFCs than LFCs because HFCs combine with the most frequently used meanings of the target words.

Research Question 3 investigates whether L2 learners are more familiar with collocates that use the most frequent meaning of a target word or collocates that use a less frequent meaning of the target word, regardless of their L2 proficiency. The answer was confirmed by the results of the paired-sample $t$ tests: All of the participants had significantly higher scores on the HFC subtest than on the LFC subtest. On the whole, any spoken or written text has a greater chance of including the former type of collocation than the latter type. Novice and intermediate learners
remember frequently used collocations first, and advanced learners consolidate their knowledge of these collocations while increasing their knowledge of less frequently used collocations. This study supports the assumption that L2 learners have more knowledge about frequently used collocations than less frequently used ones, regardless of their proficiency level.

Finally, this study does have two limitations. First, DTCK had a low reliability (e.g., $\alpha = .61$ for the whole test). One way of improving its reliability is to use Item Response Theory to identify any items that do not fit and then to replace them. Second, it is necessary to confirm the frequencies of the collocations in the test. The present study defines the two frequency-based collocation types according to the order in which their definitions appeared in one monolingual dictionary. However, the actual frequency of each collocation should be corroborated using reliable corpus data. Future studies should replace any target words whose two collocates do not have a significant difference in frequency.

5. Conclusion

This study has clarified three points: First, the collocation knowledge of L2 learners can be a reliable indicator of their L2 proficiency. Second, advanced L2 learners have deeper L2 collocation knowledge than novice and intermediate L2 learners. Third, L2 learners know collocates that use the frequent meanings of target words at an earlier stage of their L2 learning process than collocates that use the less frequent meanings of the target words. This suggests that the first type of collocate is likely to be acquired earlier than the second type. These findings were verified thanks to the development of DTCK and the classification of different types of collocation knowledge. Future studies that further scrutinize the quality of L2 collocation knowledge can shed light on the process of L2 vocabulary acquisition.

One practical implication for classroom teaching derived from the current study is the need for teacher intervention in students’ learning of collocation knowledge. As stated in the Introduction, collocations are usually acquired when learners encounter word combinations multiple times in reading and listening. However, Table 3 suggests that some collocations, even frequently used ones, are not learned through such processes. For such collocations, teachers may sometimes need to explicitly direct students’ attention to collocation links, for example, by assigning vocabulary tasks. Teachers can also consider including some tasks aimed at raising students’ consciousness about collocations in their lessons to boost students’ sensitivity to noticing collocations in future learning.

Acknowledgements

This study was supported by the Ministry of Education, Culture, Sports, Science and Technology in Japan, Grant-in Aid for Scientific Research (B) (No. 22320110). The study was presented at
JACET’s 50th Commemorative International Convention held at Seinan Gakuen University in September 2011. We would like to express our sincere gratitude to Professor Masamichi Mochizuki, Toshihiko Uemura, Kazumi Aizawa, Naoki Sugimori, Shin’ichiro Ishikawa, Tatsuo Iso, Kenji Tagashira, and Shin’ichi Shimizu, who gave us thoughtful and insightful advice.

References


192-209.


University Press.

Appendix

Depth Test of Collocation Knowledge (DTCK) コロケーション知識の深さテスト

*「英語として」下の各語に結びついて意味の通る語を、選択肢から2語ずつ選びなさい。ただし、選択肢の語は（ ）のある位置に来るものとします。解答はすべて解答用紙に記号で記入しなさい。（From the four options, select the two English words that are meaningfully related to the presented word. The selected words will appear before or after the presented words, which are indicated by parentheses. Write ア, イ, ウ, or エ on the answer sheet.）

例. short (    )
ア salt イ shop ウ time エ supply
答. ウ (short time 「短い時間」) エ (short supply 「不足している供給」)
[Answers: ウ (for HFC); エ (for LFC)]
1. heavy (  )
   ア bag イ mathematics ウ traffic エ sunshine [ア; ウ]

2. rough (  )
   ア ground イ fun ウ lunch エ idea [ア; エ]

3. late (  )
   ア aim イ arrival ウ computer エ summer [イ; エ]

4. loud (  )
   ア suit イ coffee ウ silence エ noise [エ; ア]

5. wild (  )
   ア animal イ guess ウ pencil エ pocket [ア; イ]

6. green (  )
   ア air イ kick ウ politics エ vegetable [エ; ウ]

7. rich (  )
   ア blue イ clock ウ family エ laundry [ウ; ア]

8. busy (  )
   ア tree イ people ウ sugar エ street [イ; エ]

9. long (  )
   ア memory イ money ウ river エ tea [ウ; ア]

10. (  ) address
    ア sky イ new ウ opening エ blue [イ; ウ]

11. tall (  )
    ア winter イ boy ウ order エ paper [イ; ウ]

12. black (  )
    ア hair イ cycling ウ flight エ look [ア; エ]

13. (  ) branch
    ア aunt イ domestic ウ tree エ water [ウ; イ]
14. bright (アガス イガル ウスター エ ストーム)
15. dry (アチェア イジュース ウタオル エワイン)
16. easy (アデイ イアイコン ウクィーン エキュークション)
17. soft (アディスタンス イストーム ウジョブ エベッド)
18. flat (アビール イトワラー ウレイン エアリヤ)
19. clean (アレコード イリガイ ウレフューズ エルーム)
20. strong (アコーヒーエイカウアンド ウマスキュール エデイト)
21. hot (アデイアイスイク ウスノウ エダークネス)
22. (アマシンエイリズム ウニューポーパー エハサホール)
23. (アレッド イグレイト ウチェープ エモノユーシュ)
24. sharp (アスポンジ イサーク ウアインCREASE エナイフ)
25. odd (アホエニスティーウエイシューエエトニティ)