

Features of Lexical Collocations in L2 Writing: A Case of Korean Adult Learners of English

YiBoon Chang

(Seoul National University)

Chang, YiBoon. (2018). Features of lexical collocations in L2 writing: A case of Korean adult learners of English. *English Teaching*, 73(2), 3-36.

Developing small learner and native corpora, this case study examines how Korean L2 learners used six types of lexical collocations in L2 writing to address (a) the frequency and acceptability of learner collocations, (b) problematic constituents of deviant collocations, and (c) possible sources of the learner difficulties. The overall frequency (about 8% of each corpus) and relative frequencies of each collocation type were similar between the learner and native corpora in descending order of adjective-noun, verb-noun, noun-noun, adverb-verb, adverb-adjective, and noun-verb combinations. The average and individual acceptability rates of each collocation type were around 70% and the problematic constituents were found both in nodes and collocates. L2 influence on learner difficulties mostly lied in confusions about synonyms, overuse of delexical verbs, and use of correct collocations in wrong contexts. Relying on L1 semantic representations, the learners produced non-habitual combinations, misrepresented the intended meaning, and paraphrased L2 collocations. Pedagogical implications arose for teaching L2 collocations about the importance of considering the immediate context of L2 writing and taking different approaches to different types of collocations.

Key words: lexical collocation, L2 writing, learner corpus, lexical approach

1. INTRODUCTION

Scholarly interest in naturally co-occurring multiword expressions (Fernando, 1996; Fraser, 1970; Makkai, 1972; Strässler, 1982; Weinreich, 1969) and L2 learners' non-nativeness incurring the misuse of multiword expressions (Conzett, 2000; Howarth, 1998; Lewis, 2000) has motivated comparing authentic language productions of native and non-native speakers (e.g., Granger, 1998; Howarth, 1998). The advent of corpora, computerized

large collections of language, made it easier to disclose the authentic language use (Huston, 2002) and the arbitrary restrictions in multiword combinations. Corpus linguists denied the traditional dichotomy of grammar and vocabulary, and suggested *lexis* as a new unit of language that contains information about the contexts of word use including co-occurring words (Lewis, 2000; Sinclair & Renouf, 1998). This new approach to language is often called *lexical approach* or more teaching-oriented conception of *lexical syllabus*.

Drawn from the lexical approach, collocations, habitually co-occurring multiword expressions in natural texts in statistically significant ways (Lewis, 2000), have generated particular interests from linguists and L2 researchers with their restricted substitutability of the constituents, unlike other multiword expressions such as idioms being lexically and semantically invariable (Bahns, 1993; Benson et al., 1986; Cowie, 1981; Howarth, 1993; Wood, 1981). L2 learners' lack of collocation knowledge has been well-documented and evidenced in their smaller numbers and types of collocations as well as more deviant forms (Bishop, 2004; Granger, 1998; Howarth, 1998; Levitzky-Aviad, & Laufer, 2013; Lewis, 2000; Muhammad Raji & Hussein, 2003; Nesselhauf, 2003, 2005). One striking finding is that even fairly advanced learners struggled with using collocations properly (Henriksen, 2013; Waldman, 2011; Nesselhauf, 2003, 2005).

Empirical research on Korean learners of English has reported the similar learner difficulties, demonstrated in learners' production of deviant combinations on cloze tests (Kim, 2003; Kim & Yoon, 2008; Lee, 2015; Ma & Kim, 2013; Park, 2003) and a limited number or types of collocations in L2 writing (Chon & Shin, 2009; Sung, 2017). These studies, meanwhile, predominantly focused on a particular type of collocations, namely, verb-noun combinations (e.g., Choi et al., 2015; Kim & Yoon, 2008; Sung, 2017) occasionally plus adjective-noun combinations (Chon & Lee, 2015; Kim, 2012; Lee, 2015). Such a keen interest in verb-noun combinations is consistent with the research on other L2 groups (e.g., Bahns & Eldaw, 1993; Cowie, 1997; Howarth, 1998; Laufer & Waldman, 2011; Nesselhauf, 2003; 2005; Szudarski & Conklin, 2014), probably drawn from its highest frequency and salience in sentence construction (Sung, 2017). As possible sources of Korean learners' difficulties in using L2 collocations, L2 proficiency and L1 influence have been most widely documented suggesting lower L2 proficiency and bigger L1-L2 differences as aggravating factors of the learner difficulties (Chon & Shin, 2009; Kim, 2003; Kim & Yoon, 2008; Kim & Cho, 2010; Kim, 2015; Lee, 2015; Park, 2003).

Existing research on Korean learners' use of L2 lexical collocations, despite its significant achievement so far, is still lack of comprehensive understanding about learners' use of various collocation types and related difficulties. The limited interest in certain types of collocations cannot properly serve learners' needs for using lexical collocations, particularly in L2 writing that requires a wide variety of sentence structures and the concomitant use of diverse word combinations. The scanty empirical research on various

types of lexical collocations might have derived from the highly diversifying definitions and classifications of lexical collocations in the scholarship (Kim, 2003) and the resultant challenge in applying the typologies into actual data (Nesselhauf, 2003). Indeed, the existing studies on Korean learners' production of L2 collocations were often in devoid of thorough distinctions or systematic classifications often mixing up grammatical and lexical collocations (e.g., Chon & Shin, 2009), similar to those of other L2 groups (e.g., Jafarpour et al., 2013; Wu et al., 2010).

To fill these gaps in the empirical studies and extend the existing literature, the present study aims to investigate how Korean learners of English use various types of L2 lexical collocations in L2 writing by addressing the following research questions:

1. How frequently and appropriately do Korean adult learners of English use lexical collocations in L2 writing?
2. What are the problematic constituents of deviant collocations?
3. What are the relating factors of the learner difficulties in using lexical collocations?

To thoroughly examine the features of Korean learners' lexical collocations used in L2 writing, clear definitions of each collocation type and strict criteria for accuracy were applied. To better represent learners' use of collocations in L2 writing, a small learner corpus and a comparable native-speaker corpus were developed by implementing untimed writing tasks on the same topics. Untimed writing tasks are believed to better represent the learners' usual practice of L2 writing consulting various L2 reference tools such as dictionaries than the test setting of timed writing tasks (Koo, 2006). Individual interviews with the learners were also conducted to enrich the understanding of their difficulties in using lexical collocations in L2 writing.

2. BACKGROUND OF THE STUDY

2.1. Collocations

2.1.1. Defining collocations

With the growing interests in multiword expressions in L2 learning, collocations in particular, demands for explicit and comprehensive conceptions and terminologies increased. As Grant and Bauer (2004) indicated, the lack of clear agreement on defining multiword expressions in the literature led to problems in research and pedagogical practice. Traditionally, there are two main perspectives on defining collocations: the

frequency-based approach and the phraseological approach. Nesselhauf (2004) thoroughly discussed these two approaches integrating the relevant reference works. In brief, the frequency-based approach is drawn upon the statistically significant degrees of frequency in multiword occurrences, by which it is examined whether certain words occur together frequently enough to predict their co-occurrences across contexts. The phraseological approach, on the other hand, focuses on syntactic and semantic fixedness in word combinations being commonly adopted in lexicography research.

One of the fundamental differences between the two approaches is in the way of analyzing corpora. The frequency-based approach uses corpus-driven, bottom-up approach (Granger, 2005), taking the frequent co-occurrences of multiword expressions as the legitimate evidence of accepting them as collocations. With this inductive nature, the approach does not have strict criteria on the syntactic or semantic relationship of the constituent words with less prerequisite restrictions on delimiting collocations compared with the phraseological approach. Meanwhile, the frequency-based approach involves a criterion for identifying the unit of frequent accounts. One perspective takes a word-form as the criterion (e.g., Hunston, 2002), considering every inflected form as an independent unit. From the perspective, “study”, “studies”, “studied”, and “studying” are conceived of being different under the same base form of “study.” The inflected word-forms often change the word classes, and Hunston (2002) suggested that the inflected forms in different word classes should not belong to the same base form. Another perspective on the measurement unit centers on the meaning of word-forms (e.g., Nesselhauf, 2005). Accordingly, the following multiword expressions belong to the same collocation as they share the same underlying meaning: “a strong argument”, “argued strongly”, “the strength of the argument,” and argument was strengthened”.

The selected units of measurement are used to analyze corpora via electronic tools known as concordance programs. MonoConc, WordSmith Tools, and the web-based Wordbanksonline and The Corpus of Contemporary American English are the examples of a word-based concordance program. Users select words for the search to examine how the selected words (node words) co-occur with other elements (collocates) in near distances. Concordance programs also show the context of co-occurrence in the concordance lines of cut-off sentences from the corpora involving the search words. From the inductive analyses, the evidence of frequent co-occurrences from concordance programs cannot determine whether the collocations are right or wrong, but suggest what will be acceptable or unacceptable in actual communication.

In contrast, the phraseological approach adopts a top-down approach in identifying and categorizing multiword expressions based on the pre-determined traditional linguistic criteria of syntax and semantics (Granger, 2005). With regard to syntactic restrictions, word combinations conform to grammatical structures such as the construction of adjective-noun

or verb-noun (Fernando, 1996). Concerning semantic restrictions, Soviet phraseologists developed a continuum model in which the restrictedness of combining words is conceptualized in degree and determined by the possibility of substitution and the literalness of elements (Howarth, 1998). Fernando (1996) referred to the two determining features as invariance and non-literalness, and Cowie (1997), commutability and transparency, respectively. The former refers to the degree of substituting the existing elements for other words. For example, “grab” cannot replace “catch” in the collocation of “catch a breath” even if the two words share some meaning. The latter refers to the extent in which the elements are interpreted and used in literal senses. The greater non-literal meaning the constituent words represent, the more fixed the collocations are. According to Cowie (1997), the continuum is constituted by three categories according to the degree of fixedness: free combinations, restricted collocations, and idioms.

Many researchers tend to confuse the two approaches to identifying and categorizing collocations or integrate them in their studies (Nesselhauf, 2005). Meanwhile, Granger (2005) persuasively advocated the reconciling of the two approaches in that the traditional phraseological approach should complement the units of collocations identified by the frequency-based approach. In a similar vein, the units defined by quantitative methods do not always produce linguistically or pedagogically valuable accounts of multiword expressions, and more linguistically-defined categories from the phraseological approach are expected to compensate for the limitation. The following discussion about defining lexical collocations, thus, is aligned with this integrating approach to defining collocations.

2.1.2. Lexical collocations

As briefly indicated in the introduction, collocations are broadly classified into grammatical collocations and lexical collocations. Grammatical collocations consist of dominant words and grammatical words; the former are content words and the latter do not convey significant meaning but serve grammatical function. On the other hand, lexical collocations are composed of two (or more) content words in different word classes constructing particular syntactic structures in various degrees of semantic fixedness. Semantic restrictedness of lexical elements is normally determined by the degrees of substitution and literalness from the phraseological perspectives. Under the syntactic and semantic restrictions, various syntactic patterns of lexical collocations in English language have been identified from frequency accounts in corpora, as shown in Table 1 (Seretan, 2005).

TABLE 1
Types of Lexical Collocations in English (Seretan, 2005)

Sources	Types
Lexical collocations in BBI dictionary	A-Adv, N-P-N, N-A, N-V, V-Adv, V-N
Hausmann's collocation definition	A-Adv, N-A, N-[P]-N, N-V, V-Adv, V-N
Xtract collocation extraction system	N-A, N-D, N-P, N-V, V-Adv, V-N, V-P, V-V
FipsCo System	N-A, N-N, N-P-N, N-V, V-N, V-P, V-P-N
WordSketch concordance system	A-P, N-A, N-Conj-N, N-N, N-P-N, N-V, V-A, V-N, V-P

Note. Initials of constituent word classes are the followings: A (Adjective), Adv (Adverb), Conj (Conjunction), D (Determiner), N (Noun), P (Preposition), V (Verb). A short dash is used to present the word combination and brackets stand for optional inclusions of the constituents.

The elements of lexical collocations commonly include lexical elements of adjective, adverb, noun, and verb. All classifications also involve grammatical elements such as conjunction, determiner, and preposition, but the kinds and necessity of grammatical elements appear to be different in each classification according to its syntactic pattern.

2.2. Korean Learners' Difficulties in Using L2 Lexical Collocations

Between the two aspects of learners' use of L2 collocations, namely comprehension and production, researchers have taken a more interest in production since the learners more struggle with producing appropriate collocations than understanding them (Brashi, 2006). Given that the semantic properties of lexical collocations are usually derivable from the components, it would be a natural result that L2 learners have relatively higher receptive collocation knowledge than productive knowledge. Production of L2 learners' collocations has been commonly examined in a written mode due to its permanency and the resultant convenience of data collection and analysis. Empirical research on the production of L2 collocations by Korean learners of English has also favored collecting written data by implementing various elicitation methods such as cloze tests, multiple-choice (e.g., Kim, 2003; Ma & Kim, 2013) or open-ended (e.g., Kim & Yoon, 2008; Lee, 2015; Park, 2003), and timed writing tasks (e.g., Chon & Shin, 2009). While multiple-choice cloze tests were occasionally used to measure receptive collocation knowledge (e.g., Chon & Lee, 2015), cloze tests, even with given choices, are inherently to generate, not decipher, intended meaning and thought to be more pertinent to representing productive collocation knowledge in this study. These test-oriented elicitation methods seemingly reflect researchers' preference for measuring the learners' general collocation knowledge/competence to investigating their actual use of collocations. Some recent works have addressed the issue by analyzing the existing large corpora such as Yonsei English Learner Corpus (e.g., Kim, 2015; Sung, 2017). Given that the writing samples in these learner

corpora were usually collected in test environments such as time writing tasks, however, there is still an urgent need for studying authentic L2 writing in a more natural setting to grasp the nature of learners' use of L2 lexical collocations.

Korean learners' lack of productive collocation knowledge has been empirically demonstrated by high error rates or uncertainty in their answers to collocation tests (Choi et al., 2015; Kim & Yoon, 2008; Ma & Kim, 2013; Park, 2003) or over-/under-/misuse of collocations in L2 writing (Kim, 2015; Sung, 2017). With the biased interest in a verb-noun combination in the literature as indicated in Introduction, the empirical evidence is relevant to this particular collocation type in the strict sense. In fact, influences of collocation types on Korean learners' collocation use have been seldom researched, while a few studies reported related results. Kim (2003), in the study about four types of collocations (adjective-noun, verb-noun, preposition-noun, verb-preposition), found that Korean high school students more struggled with verb-noun combinations than adjective-noun combinations. Chon and Lee (2015) reported the similar results about Korean college students producing more errors in verb-noun combinations than adjective-noun combinations at a statistically significant level. The greater difficulty in verb-noun combinations than adjective-noun combinations, however, contradicts with the findings of Park (2003) investigating five collocation types: noun-verb, adjective-noun, adverb-verb, verb-noun, and adverb-adjective combinations. Her study about 133 Korean college students found that noun-verb combinations were the most problematic and adverb-adjective combinations were the least problematic across all proficiency levels. Comparing verb-noun and adjective-noun combinations, only advanced students more struggled with verb-noun combinations, whereas the low and intermediate level students more struggled with adjective-noun combinations. The counterevidence for Korean learners' greater difficulties in adjective-noun combinations was also reported by Kim (2012) even across L2 proficiency levels. In sum, Korean learners' relative difficulties in different collocation types have been underexplored with inconclusive findings.

When it comes to possible sources of Korean learners' difficulties in using L2 lexical collocations, various factors of L1 and L2 influences have been investigated. L1 influence refers to inter-language or crosslinguistic influence from L1 knowledge on L2 collocation knowledge, while L2 influence encompasses any aspects of L2 knowledge interacting with L2 collocation knowledge. Most empirical studies took multiple factors of each influence (e.g., general L2 proficiency and L2 vocabulary knowledge) or both influences (e.g., L1 influence and L2 proficiency) into account.

L1 influence is rooted in L1-L2 similarities and differences in the conceptual and lexical representations, often conceptualized as *congruency* (Choi et al., 2015; Chon & Lee, 2015; Kim, 2012; Lee, 2016) or *predictability* (Kim & Yoon, 2008). When L2 collocations have "word-for-word L1 translation" equivalents (Choi et al., 2015, p. 32), they are considered

to be more congruent or predictable, accordingly easier for L2 learners to learn, than those without L1 equivalence. Comparing the learner performance on congruent and incongruent L2 collocations, empirical research has widely confirmed higher accuracy of congruent collocations (Choi et al., 2015; Chon & Lee, 2015; Kim, 2012; Kim & Yoon, 2008; Lee, 2016) in the interplay with L2 proficiency and collocation types. L2 proficiency seems to play a critical role in learners' coping with incongruent collocations as seen in the greater struggles of low proficiency learners (Choi et al., 2015; Lee, 2016). Influences of collocation types have produced conflicting findings with no significant influences (Kim 2012) versus noticeable influences of adjective-noun collocations on learners' more difficulties in congruent collocations than incongruent ones (Chon & Lee, 2015).

As for L2 influence, L2 proficiency and L2 vocabulary knowledge have earned a focal interest, and L2 proficiency is predominantly incorporated into the empirical research as a base L2-related factor. Statistically significant correlations between L2 proficiency and L2 collocation knowledge have been well documented (Kim, 2003; Kim, 2015; Lee, 2016; Park, 2003), and L2 vocabulary size is also found to interact with L2 collocation knowledge (Kim, 2012; Ma & Kim, 2013; Park, 2003). Among the various aspects of L2 vocabulary knowledge, knowledge in delexical verbs and synonyms has been indicated as noteworthy sources of learner difficulties specifically in using verb-noun and adjective-noun collocations, respectively. Delexical verbs have a wide and common usage in English language, especially in verb-noun collocations in which they can borrow most of semantic properties from the noun collocates. Ma and Kim (2013) indicated Korean high school students' lack of knowledge in delexical verbs (with approximately a 40 percentage of correct responses) despite their high frequency in textbooks. Kim and Yoon (2008) reported Korean learners' overuse of delexical verbs, which was recently confirmed by Kim (2015) specifically about the verb, *make*, with low proficiency learners. Learners' overuse of delexical verbs are thought to relate to their lack of knowledge in appropriate verb collocates of noun nodes and misconception about delexical verbs to be indistinctively applicable to any nouns (Lee, 2015). Korean learners' confusions about L2 synonyms and resultant misuse of L2 collocations have been reported mostly in choosing the adjective collocates in adjective-noun collocations (Chon & Lee, 2015; Kim, 2008). Possible test effects were addressed about these confusions in that the learners are more likely to struggle with choosing the right collocates for congruent collocations when given synonymous distractors (Chon & Lee, 2015; Kim, 2012).

In sum, the properties of collocations are decided largely from a frequency-based approach or a phraseological approach and integration of the two is known to be desirable. Lexical collocations, distinguished from grammatical collocations in the lack of grammatical words, have various types depending on the constituent parts of speech. Research on Korean learners' use of lexical collocations has demonstrated their lack of

collocation knowledge, mostly in verb-noun and adjective-noun combinations, and suggested diverse L1 and L2 influences on their struggles with L2 collocations.

3. METHODOLOGY

3.1. Data Collection

3.1.1. Participants

To compare lexical collocation use of native speakers and Korean speakers of English in writing in terms of frequency and distribution of collocation types, eight Korean adult learners and eight adult native speakers of English were recruited. The Korean speakers consisted of two females and six males aged between 23 and 30, all of whom had a Bachelor's degree with at least 10-year experiences of learning English in a public or private educational system. They had different academic backgrounds in undergraduate studies including statistics, police science, electrical engineering, business administration, political science, and English education. Recruiting participants from various academic backgrounds was to prevent the influence of background knowledge in their L2 writing about generic topics. In the lack of information about their English proficiency, the first writing drafts were evaluated by a Korean-Canadian English instructor with a five-year teaching experience in a prestigious national university in Korea. Since the essay topic was from the international English language testing system (IELTS) task 2 writing as described in 3.1.2, the evaluation used the IELTS task 2 writing band descriptors (public version) as rubrics. The writing proficiency of the learners was around advanced or upper intermediate scoring 6.9 an average on the 0 to 9 band spectrum, ranging from the lowest 6 to the highest 8.

The native speakers were aged from 24 to 40, half of whom were from Canada, Australia, and England, and the other half were Korean-Americans. The Korean-American speakers were considered to have retained their native intuition since they lived in English-speaking countries over 20 years and stayed there during the puberty. All of the native speakers had a Bachelor's degree and some of them were in graduate studies. Given these backgrounds, it is assumed that the native speaker participants are assumed to have nativeness and sufficient collocation knowledge, and have demonstrated the knowledge in their L1 writing tasks in this study. Their use of lexical collocations in L1 writing was not assessed for accuracy in this regard, which conforms to the common practice in L2 research.

The Korean learners were regular members of an online study group from different

backgrounds; some were students from different colleges, and others worked for different organizations. All of them were highly motivated to learn English, especially English writing, believing that a high English proficiency would be essential in their academic and occupational success. They were aware of that their use of English was not native-like despite its grammatical accuracy. Some of them had already utilized reference materials from news articles or Google webpages to search for conventional native-like expressions for L2 writing. While they had not been familiar with the notion of collocations, they acknowledged the existence and nature of arbitrary co-occurring multi-words in English language after being informed of the concept from the researcher.

Two raters, one from Canada and the other from England, were recruited to judge the acceptability of lexical collocations produced by Korean participants. The Canadian rater was working as an instructor and coordinator of a college foreign language center, with a Master's degree in teaching English and completed a doctoral coursework. The English rater stayed in Korea over 10 years teaching English with a Bachelor's degree in theology. Both of them were familiar with Korean adult learners' use of English from various teaching experiences. They were to judge the acceptability of lexical collocations identified by the researcher and to provide corrections or alternatives for deviant collocations.

3.1.2. Materials

To introduce lexical collocations to the Korean participants, an introductory lecture was provided by the researcher. The lecture offered examples of lexical collocations with a fill-in-the blank task and a translation task. Six types of lexical collocations excluding grammatical elements were selected from the types in Table 1: adjective and noun (A-N), adverb and adjective (Ad-A), adverb and verb (Ad-V), noun and noun (N-N), noun and verb (N-V), and verb and noun (V-N). The fill-in-the blank task asked the participants to choose appropriate collocates for the given head words. By discussing the proper collocates, the participants came to be aware of the restrictions on combining words for collocations. The translation task required the participants to use particular head words to translate Korean sentences into English sentences. The participants' translations were compared with the original English sentences later. Then, Collins Wordbanks Online web concordance was introduced by the researcher with a short lecture on how to use the program to search out proper collocates for head words.

Writing tasks for data collection were to write argumentative essays about six general issues from the topics of IELTS task 2 writing (e.g., *Some people think that university should not provide theoretical knowledge, but to give practical training to benefit society. What is your opinion?*). There was no restriction on the content of essays under the topics. One requirement was that the Korean participants should involve at least one lexical

collocation in each type by using Collins Wordbanks Online web concordancer. They were also allowed to use any other references of their choice such as monolingual/bilingual dictionaries or online dictionaries. Each essay should have 500 to 700 words and the tasks were implemented in untimed manners.

The requirement of searching lexical collocations from the web concordancer was to ensure the Korean participants' use of lexical collocations and produce rich data, considering the widely-reported underuse of collocations from non-native speakers (Granger, 1998). It was also to motivate the Korean participants to engage in the study by serving their needs for improving English. The Collins Wordbank Online English corpus in the web concordance was composed of 56 million words of contemporary written and spoken texts from books, ephemera, radio, newspapers, magazines, and transcribed speech. The query system allowed users to search any patterns from a single word to parts of speech. Users could search out how frequently certain words co-occur in authentic language data either in the whole corpora or in each subcorpus. The researcher subscribed to Wordbanksonline for multiple accesses and had the Korean participants join the subscription.

3.1.3. Data collection procedures

To effectively collect the writing data from the participants, the researcher created a temporary online website giving exclusive access to the Korean participants. All the materials for data collection were shared via the website for the participants' references. The participants posted their essays on the writing board and their works were accessible to one another. The posted writing should include their results of searching lexical collocations on the concordance with their intended meaning to deliver. The reports of search results involved the participants' trials and errors including their reluctant use of alternatives, which helped the researcher to judge the acceptability of collocations. For example, one participant failed to find the exact match between his or her intended meaning and the concordancer search-out. The participant finally reported two collocations, one from the concordancer as an alternative collocation and the other from Google search engine better matched with the intended meaning. Meanwhile, the one selected for the writing was only accounted as the participant's use of lexical collocations in data analyses. Native speakers' essays were collected by email in a Microsoft word format upon the request from the researcher.

Before the recruitment, the researcher joined the online study group where the Korean participants had attended offline meetings on a regular basis, and introduced the research project. With the approval of the head member, the researcher joined the meeting to give the one-time introductory lecture on lexical collocations and a training session for using the

concordance were offered by the researcher. Those who consented were offered the access to the online website to complete writing tasks within six weeks submitting one essay per week on average. The researcher took a rather flexible stance so that the participants would not be too much burdened, and their submissions of six essays completed within 12 weeks. The researcher identified lexical collocations from the essay drafts, and had the native raters judge their acceptability and provide alternatives or corrections, if any. The native raters' comments and some corrections on the electronic documents were delivered to each participant through the online website. The raters' reviews were usually given to the participants within two weeks for each essay. Any further questions or comments from the participants could be posted and shared via the online website. It is plausible that the participants developed their L2 collocation and writing knowledge in the data collection process, the possible learning effects were not counted into the present study because the study aimed to explore a relatively natural setting of learners using collocations in L2 writing. The adult learners were not oriented to particular instructional goals as in the school setting being exposed to any sources of learning in their daily lives. There would have been a faint possibility of dramatic changes in their knowledge to seriously affect the study results in this regard; even if it were so, that would be a part of their 'natural' setting of using (and learning) L2 collocations to be explored.

For data collection from native speakers, the researcher recruited them by personal contacts and on/off-line help-wanted advertisements. They were informed of the writing tasks and compensations for their contributions. Once they consented on the participation, the writing topics were sent to each participant by email. No requirements or limitations except for avoiding plagiarism were imposed on them. They were asked to submit two essays per week, but the schedule was also flexible. The native participants completed the writing tasks within three to four weeks on average.

3.2. Data Analysis

Since the data were collected from a limited number of participants with particular writing tasks, this study adopted a case study method of a holistic single case with embedded units (Baxter & Jack, 2008); the eight Korean and native-speaker participants were situated in the same context of constructing each corpus. Features of each corpus and learner difficulties were identified by cross-unit analyses, and interview analyses of the individual units were used for data triangulation. In this regard, the findings from these small corpora are confined to the given case as preliminary research, not being generalized to other cases or the entire population of Korean adult learners of English.

To overview the data analysis process, the researcher identified lexical collocations from each draft of the Korean and native participants, and then counted the occurrences in each

type. The collocations of Korean learners were further analyzed with regard to their acceptability and the sources of deviant use. In the process of data collection and analysis, the researcher conducted individual interviews with all eight Korean participants centering on their deviant use of lexical collocations to corroborate the textual analyses. Each interview lasted for about one and a half an hour online via the website in Korean. The researcher had prepared for some interview questions driven from the data analysis and any emerging questions were also addressed during the interviews.

3.2.1. Identifying lexical collocations

Identifying collocations from large corpora usually relies on collocation extraction tools or systems such as lemmatization and chunking or parsing to identify collocation candidates prior to statistical analyses (Seretan, 2005; Seretan & Wehrli, 2006). However, using those tools often fails to incorporate contextual meaning in extracting collocations. Considering that the current study pursues an in-depth study on a small scale learner corpus, hand-extracting approach was thought to be more suitable. To incorporate a systematic and thorough method, the study adopts the approach of Nesselhauf (2005), which integrated the phraseological and frequency-based approaches being appreciated for its detailed and rigorous hand-extracting process (Cobb, 2006).

The six types of lexical collocations were determined by their syntactic structures. Thus, the adjacency, order, or overlapping of elements did not influence extracting the components. For example, two verb-object constructions, “make profit” and “spend cost”, were identified in the following example even if their components did not appear in adjacency and in verb-object order: “However, it is only a logical deduction and the real *profit* increase working women can actually *make* has not proven yet, while the **costs** government *spends* is a concrete fact” (Subject 2, Topic 4). When multiple elements were coordinated in a single construction, multiple combinations were identified. For instance, two verb-noun combinations (“esteem fame”, “esteem riches”) and two adjective-noun combinations (“worldwide fame” and “worldwide riches”) were identified from the following example: “For I *esteem their worldwide fame and riches* as that of the country” (Subject 4, Topic 1).

The components were limited into two elements excluding function words such as preposition, infinitives, or clauses. Identifying the components of collocation was based on the meaning of word-forms; thus, “develop an idea” and “developed ideas” were regarded as same V-N collocations, but “development of ideas” was excluded because of its different syntactic structure. In V-N collocations, only VO constructions (verb - direct object) and VOO constructions (verb - indirect object - direct object) were included. Meanwhile, noun-of-noun combinations were not included in this study following Benson

et al. (1986) and Nesselhauf (2005).

Delimiting lexical collocations from other free multiword combinations was based on the reference to Oxford Collocations Dictionary for Students of English (OCD), BBI Combinatory Dictionary of English (BBI), Collins COBUILD English Dictionary (CCED), and Collins Wordbankonline program (WB). Above all, combinations indicated in OCD and BBI were regarded as collocations. However, if the combinations were easily delimited without substitutes when their elements were used in literal meaning, they were classified as free combinations (e.g., “wear clothes”, “earn money” etc.). Since one of the core features of collocations is their random restrictions, the co-occurrences with the lack of arbitrary restrictions were excluded.

The combinations not appearing in OCD or BBI were examined by CCED. The researcher checked the definitions of each element in CCED to see whether the element had any restrictions on combining other elements. If CCED said the elements combined with “something” or “somebody”, they were regarded as free combinations. If the element appeared to combine with certain (groups of) words, they were regarded as collocations. Even the same element had different restrictions on combinations depending on each meaning (e.g., “use something” vs. “use drugs”).

When it was difficult to examine the arbitrary restrictions in CCED, the researcher tried to hand-search the combinations on WB. The researcher randomly chose four synonyms of the collocate, and then searched for the four different combinations in WB to see if there were any restrictions on co-occurring. When two or more combinations had more than five occurrences in WB, it meant that the collocate had two or more substitutes for itself, indicating the combinations were free collocations. If no or only one substitute was found for the collocate, the node should only combine with the collocate and the combination was judged a collocation.

3.2.2. Determining acceptability of lexical collocations

Acceptability of lexical collocations used by the Korean participants was judged by four sources on the three scales: acceptable, questionable, and unacceptable. The first two sources were two native raters from England and Canada who marked the acceptability of each collocation on the essay drafts. ‘Acceptable’ means that the combinations were existing collocations and used in appropriate contexts. ‘Questionable’ indicated that the raters could not understand the meaning or were not sure about the acceptability. ‘Unacceptable’ meant that the combinations were hard for the raters to understand and not used by ordinary native speakers. For those judged ‘unacceptable’, the raters were asked to suggest appropriate expressions.

The third source of determining acceptability was OCD and BBI. If there were the same

expressions in either OCD or BBI, the combinations were categorized into acceptable collocations. When they did not appear on the dictionaries, it was regarded as questionable collocations. Only when one of the synonyms of the collocates was found with the node in the dictionaries, the combinations were judged unacceptable. The fact that other combinations with a synonymous collocate deliver the same meaning indicates that the node exclusively combines with the synonymous collocate.

The final source for acceptability was WB. The combinations judged ‘questionable’ or ‘unacceptable’ based on OCD or BBI were searched in WB. When the search results showed five or more occurrences, the combinations were judged acceptable by WB. From one to four occurrences were considered to be questionable, and no occurrence, unacceptable.

These four judgment results of an English rater, a Canadian rater, collocation dictionaries (OCD and BBI), and a concordance program (WB) were integrated to make the ultimate judgment. Table 3.1 shows a scheme for the judgment suggested by Nesselhauf (2005) with the only difference in the symbols of the degrees of acceptability.

TABLE 2
Acceptability Judgment

Judgment 1	Judgment 2	Judgment 3	Judgment 4	Ultimate Judgment
A	A	A	A	A
A	A	A	Q	A
A	A	A	U	(A)
A	A	Q	Q	(A)
A	A	Q	U	(A)
A	A	U	U	Q
A	Q	Q	Q	Q
A	Q	Q	U	Q
Q	Q	Q	Q	Q
Q	Q	Q	U	Q
A	Q	U	U	(U)
A	U	U	U	(U)
Q	Q	U	U	(U)
Q	U	U	U	U
U	U	U	U	U

Note. A stands for acceptable, (A), largely acceptable, Q, questionable, (U), largely unacceptable, U, clearly unacceptable, respectively.

The order of judgment did not matter because each judgment was equally influential in determining the ultimate acceptability. The determining process was a bit modified from that of Nesselhauf (2005) in which the raters’ judgments were taken into consideration only when the combinations could not be judged in collocation dictionaries and a corpus program. The present study included raters’ judgment in all cases, since the dictionaries and the concordance program had a limitation when it comes to examining whether the

combinations were used in proper contexts or not. In fact, the combinations judged acceptable from the dictionaries and the program turned out to be deviant in the ultimate judgment at times. It indicates that the modified process was more suitable for determining the acceptability of collocations used in writing.

4. RESULTS AND DISCUSSION

4.1. Distribution and Acceptability of Lexical Collocations

Section 4.1 addresses the Research Question 1 about how frequently and appropriately Korean adult learners of English used lexical collocations in their writing. Distribution and acceptability of lexical collocation use are reported in subsection 4.1.1 and 4.1.2, respectively.

4.1.1. Distribution of lexical collocations

The learner corpus of Korean adult learners of English consisted of 25,663 words from 48 essays with approximately 535 words for each essay sample on average. The corpus contained 2,263 lexical collocation occurrences (8.8% of the total word counts) producing 1,974 different combinations. The native speaker corpus consisted of 14,822 words from 24 essays with approximately 618 words for each essay sample on average. The corpus contained 1,197 lexical collocation occurrences (8.1% of the total word accounts) producing 1,060 different combinations. The collocation occurrences were comparable between the learner and native speaker corpora with 8.8% and 8.1%, respectively.

Whereas non-native speakers were reported to produce collocations much less frequently than native speakers due to their lack of collocational competence (Granger 1998), this study found that Korean learners produced almost the same amount of lexical collocations with 8.8% out of total words in comparison to that of native speakers with 8.1%. This contradictory finding may be related to the fact that the present study analyzed a smaller scale of data to identify six types of lexical collocations while Granger (1998) investigated a significantly larger scale of data (234,514 words) addressing only one type of adverb-adjective collocations. Hence, Granger (1998) sorted out possible collocations using text-retrieval software while this study did handwork to delimit lexical collocations, which might have influenced the results of identifying collocations. The research design of the current study, asking Korean learners to use at least one collocation for each type and allowing them to access various sources including the web concordancer while writing, would have also influenced their frequent use of lexical collocations comparable to the use

of native speakers. These considerable differences in research methods and design are expected to contribute to the conflicting findings between the two studies.

In the learner corpus, the largest occurrence of collocations in one essay was 86 (72 different types) from 639 words, and the smallest, 29 (28 different types) from 523 words. The average number of different combinations produced per essay was about 41. In the study of Nesselhauf (2005), about ten verb-noun collocations were produced per essay with the average length of 500 words. Considering that the present study included total six collocation types, five more than her study, the overall occurrences of lexical collocations in this study are less than those of her study. This difference might have been because each collocation type was not produced with equal frequency in the learner corpus.

TABLE 3
Distribution of Lexical Collocation Types of Korean Learners

A-N	V-N	N-N	Ad-V	Ad-A	N-V	Total
823	707	225	108	61	50	1974
41.7%	35.8%	11.4%	5.5%	3.1%	2.5%	100%

TABLE 4
Distribution of Lexical Collocation Types of Native Speakers

A-N	V-N	N-N	Ad-V	Ad-A	N-V	Total
454	363	153	40	26	24	1060
42.8%	34.2%	14.4%	3.8%	2.5%	2.3%	100%

Table 3 shows the disproportionate distribution of lexical collocation types from Korean learners. Adjective-noun and verb-noun collocations were the most commonly produced by Korean learners occupying almost three-fourths of the total lexical collocations together. Adverb-adjective and noun-verb combinations appeared to be the least common taking less than five percent of total collocation use, respectively. The finding is analogous to that of Wu et al. (2010) reporting the highest frequency in verb-noun and adjective-noun collocations in L2 student essays written without using resources. Although the classifications of collocation types were different between the two studies, dominant production of adjective-noun and verb-noun collocations can be confirmed, justifying the exclusive scholarly interest in these two types of collocations in the literature of Korean learners' use of lexical collocations.

As seen in Table 4, the distribution of lexical collocation types in native speakers also accorded with that of Korean learners as in the overall collocation occurrences. The hierarchy of frequently-used lexical collocations was the same: adjective-noun > verb-noun > noun-noun > adverb-noun > adverb-adjective > noun-verb collocations (“>” stands for “more than”). Apart from slightly more use of noun-noun combinations in native speaker

corpus, no noticeable differences were observed between the two corpora. The dominant use of adjective-noun and verb-noun collocations in the current native speaker data can be ascribed to their major role in constructing subjects, objects, and predicate structures, as indicated in the literature (Choi et al., 2015; Sung, 2017). On the other hand, adverb-verb and adverb-adjective collocations predominantly modify other elements in clauses and sentences to emphasize or intensify meaning, playing a relatively less essential role as “optional” elements to add details (Lewis, 1997 as cited in Kim, 2003, p. 239) and accordingly being used less frequently.

Interview results confirmed these different roles and importance of each collocation type in sentence construction. Two Korean learners reported that they did not place much importance on adverbs or adjective in writing because they do not influence meaning making as significantly as nouns or verbs do as seen in the following excerpt (a).

Interview excerpt (a):

Adverbs or adjectives are easy to be ignored because they are not the central components of sentences. They express the least important meaning and I just skipped them, not searching in dictionaries or a concordance when I do not know how to use them. (Subject 2)

The other learner attributed his insufficient knowledge of adjectives or adverbs to their relatively insignificant roles in sentence construction as shown in the interview excerpt (b). Lacking the “repertoire” for these modification elements, he could not use them to the desired or intended extent.

Interview excerpt (b):

I seem to know only 20 adverbs in the whole. When asked to list adverbs, I can only come up with ‘very’ or ‘really.’ I don’t know many kinds of adverbs, and I also cannot search for the appropriate expressions. It was hard to find the expressions exactly delivering the suitable meanings. (Subject 5)

One interesting result is that noun-verb collocations, which can serve both subject and predicate in sentences, were used the least frequently. Cognizant of their remarkable underuse of the combination, the learners assumed that it was substantially influenced by the topics or genres of writing tasks. The learners were aware of the possible integral functions of noun-verb combinations in sentence construction, but they pointed out that noun constituents, the node of the combination, are far more specific than other nodes. As there are only a limited number of nouns to form noun-verb collocations (e.g., bee, bomb, research etc.), the learners did not have many chances to use those specific noun

constituents to write about general topics. One learner indicated that she could have used more noun-verb combinations if the topics had required professional knowledge or certain conventions such as reporting research results or explaining graphs or tables. To corroborate this finding, further studies are needed to examine the use of noun-verb collocations in various genres of writing.

The similarities in the overall occurrences of collocations and distribution of each type between the Korean learners and native speakers seem not to coincide with the well-documented L2 learners' lack of collocation knowledge. However, the current findings are not indicative of the same level of collocational competence in both groups. Above all, the result from Korean learners included deviant collocations as well as correct ones, and their use of collocations reflected the raw frequency not the frequency of appropriate collocations. Thus, a direct comparison of collocation use between the two groups does not demonstrate their collocational competence. In addition, the similarities between the two corpora would have resulted from the distinctive task instructions; this study controlled the writing topics of the both corpora and required the Korean learners integrating at least one collocation for each type. Under this setting of encouraging the learners to use collocations in L2 writing, they could have shown comparable performance with the native speakers.

4.1.2. Acceptability of lexical collocations

To determine whether Korean learners had difficulties in using lexical collocations in L2 writing, the overall acceptability of L2 lexical collocations was examined. The result of determining the acceptability with a 5-point scale confirmed that lexical collocations were difficult for Korean learners to use as shown in Table 5. Approximately two-thirds of the lexical collocations produce were acceptable (judged A; clearly acceptable, or (A); largely acceptable). The rest one-third (33.1%) were classified into deviant expressions when questionable collocations were included, and about one-fifth of total collocations were unacceptable (judged U; clearly unacceptable, or (U); largely unacceptable).

TABLE 5
Acceptability of Lexical Collocations Produced by Korean Learners

clearly acceptable	largely acceptable	questionable	largely unacceptable	clearly unacceptable	Total
790	531	274	236	143	1974
40%	26.9%	13.9%	12%	7.2%	100%

The approximately 33% error rate in this study is substantially lower than the error rates reported from the cloze test results: 70% (Kim, 2003), 44% (Park, 2003), 50.7% (Kim & Yoon, 2008), and 62% (Ma & Kim, 2013). This conflicting finding would have derived

from the task differences; learners could avoid uncertain collocations in the elicitation tasks with concomitant lower error rates than those of cloze tests in which they cannot use the avoidance strategy. Furthermore, the current untimed L2 writing tasks allowed the learners to check the appropriateness of candidate collocations from various sources including the concordancer. To compare the finding with that of a similar elicitation task, the one-third of inappropriate collocations used by Korean learners corresponds that of Nesselhauf (2005) about the verb-noun collocation use of advanced German-speaking learners. Nesselhauf (2005) suggested that about 20-30% of verb-noun collocations produced in the writing of advanced learners were deviant regardless of their L1 backgrounds. Given the advanced or upper intermediate writing proficiency of the Korean learners, the similar amount of overall deviant lexical collocations in this study is comparable to that of Nesselhauf's study.

To closely examine the deviant collocations, the error rate of each collocation type was analyzed as seen in Table 6. Approximately one third of each collocation type was considered questionable or unacceptable with the exception of noun-noun type. About 36% of adjective-noun collocations were deviant, slightly higher than the average deviation rate (33.1%) presumably attributed to their largest occurrences. Verb-noun and noun-verb collocations produced nearly 34% deviations, respectively, which is almost similar to the average deviance rate. Adverb-verb and adverb-adjective collocations were deviated by 30.6% and 31.1%, respectively, which is lower than the average deviance ratio.

TABLE 6

Deviant Collocations in Each Type Produced by Korean Learners

	A-N	V-N	N-N	Ad-V	Ad-A	N-V	Total
produced	823	707	225	108	61	50	1974
deviant	297	243	44	33	19	17	653
percentage	36.1%	34.3%	19.6%	30.6%	31.1%	34%	33.1%

On the other hand, noun-noun collocations produced exclusively less deviant forms. Asked about this noteworthy result during the interviews, the learners pointed out the ease of learning and retrieving noun-noun collocations from the distinctive process of combining the elements of noun-noun combinations. The learners found it easier to retain the noun-noun combinations as a whole chunk when encountering them in reading, perceiving that the both constituents are equally important in representing the meaning. By the same token, the learners had to simultaneously come up with the two noun elements in producing the noun-noun collocations, whereas they usually selected the node first and then tried to match appropriate collocates for the node in producing the other collocations. Since the noun-noun collocations allow less diverse combinations with less substitute collocates for one node, the learners were less likely to produce deviant combinations.

Under the general assumption that the collocations with more deviant forms would have

been more difficult for the learners to produce, the current finding of similar student difficulties across collocation types contradicts the dissimilar student difficulties reported by Park (2003) based on the responses to cloze tests. According to Park (2003), Korean college learners of English had different collocational competence in the following order: noun-verb < adjective-noun < adverb-verb < verb-noun < adverb-adjective (“<” stands for “less difficult”). Allowing for the lack of noun-noun combinations in Park’s analysis and the different task type, the hierarchical student difficulties for each collocation type do not match with the similar deviation ratios in all the collocation types in this study.

Park (2003) attributed the hierarchy of Korean learners’ difficulties to the different L2 collocation input or exposure; more common combinations such as noun-verb and adjective-noun would have been easier for the learners, whereas less common ones such as adverb-adjective combinations, more difficult to get familiar to leading to the lower scores in the cloze test. While the powerful impact of input on Korean L2 learners’ use of collocation has been reported in the literature (e.g., Choi et al., 2015; Sung, 2017), a counter evidence was also presented about delexical verbs about verb-noun collocations (Ma & Kim, 2013). Furthermore, native speakers did not frequently use noun-verb collocation in this study, possibly refuting its more frequent input. The contradictory findings should be addressed with further research on Korean learners’ difficulties in relation to the types of lexical collocations.

4.2. Problematic Constituents of Deviant Lexical Collocations

This section addresses the Research Question 2 about the problematic constituents leading to the deviant use of lexical collocations so as to specify the difficulties of Korean adult learners’ use of lexical collocations in L2 writing. To explore the nature of learners’ inappropriate use of lexical collocations, the elements of deviance were analyzed as seen in Table 7. Some deviations were from one deviant constituent, while others, from both constituents deviant. Deviances in one constituent were found both in the node and the collocate, but deviant collocates were more common than deviant nodes, especially in the most dominant combinations of adjective-noun and verb-noun, as Table 7 shows. More errors in collocates are highly predictable given that L2 learners usually match collocate candidates with the chosen node when they are lack of the knowledge about proper collocations as a whole chunk. This collocation-generating process is reflected on the cloze tests created in the literature which ask the learners to come up with appropriate collocates for the given nodes and context.

On the other hand, the current data from writing tasks present a new aspect of L2 learners’ difficulties in choosing the nodes and both elements. Table 7 demonstrates that the errors in both constituents were comparable to those in the collocates (114 vs. 110 in

TABLE 7
Problematic Constituents of Deviant Collocations

Collocation Types	Problematic Constituents			Total
	Adjective	Noun	Both	
Adjective-Noun	110	73	114	297
Verb-Noun	Verb	Noun	Both	Total
	97	30	116	243
Noun-Noun	Noun 1	Noun 2	Both	Total
	7	12	25	44
Adverb-Verb	Adjective	Verb	Both	Total
	8	10	15	33
Adverb-Adjective	Adverb	Adjective	Both	Total
	6	2	11	19
Noun-Verb	Noun	Verb	Both	Total
	0	10	7	17

Note. The node of each combination was marked in bold.

adjective-noun, and 97 vs. 116 in verb-noun) and the error rates on the nodes were not negligible ranging from 10% (adverb-adjective) to 30% (adverb-verb). The following six examples from the learner corpus display different deviant types according to the problematic constituents: a node, a collocate, and both.

Examples (1) and (2) illustrate the deviations in the nodes of collocations. Example (1) was an adverb-verb collocation, and the node verb, *mix*, did not properly deliver the intended meaning of human relationship. Example (2) was an adjective-noun collocation, an existing correct collocation, but the choice of node noun, *time*, was judged inappropriate for the given context by the native raters.

(1) *socially **mix** → socially **interact** (deviance: node)

As society becomes more complicated and as more people ***are socially mixed**, relationship among people becomes more superficial. (S6T3: Subject 6, Topic 3)
→ ...as more people **socially interact**, relationships among people...

(2) *great **time** → great **opportunity** (deviance: node)

Most of all, this precious one year is a ***great time** for speculation. (S5T2)
→ Most of all, this precious one year is a **great opportunity** for speculation.

Examples (3) and (4) show how collocates were wrongly chosen in deviant collocations. Example (3) was an adjective-noun collocation and the chosen collocate, *steady*, does not co-occur with the node, *water*, even if it is a synonym of *still*, the proper collocate. Example (4) also demonstrates how a synonymous collocate led to wrong collocations. The chosen collocate for the verb-noun collocation, *squander*, shares the meaning of the

right collocate, *waste*. However, *squander* combines with money or talent, not time, so that it was not appropriate for the node, *time*.

(3) ***steady** water → **still** water (deviance: collocate)

‘The ***steady water** will rot’ is a Korean proverb. (S6T4)

→ ‘**Still waters** stagnate’ is a Korean proverb.

(4) ***squander** time → **waste** time (deviance: collocate)

The problem is that many of the students tend to have little or no idea about the necessity and purpose of their proceeding to university largely because of lack of social experience and deliberation on one’s future, which in many cases steers them to fail in adapting to university life or ***squander priceless time** with worrying about what they want and what they should do. (S2T2)

→ ... fail in adapting to university life or **waste precious time** ...

Examples (5) and (6) are the cases when the learners produced completely inappropriate collocations. In the noun-verb combination in example (5), the wrong expression, *worry has evaporated*, sounds like direct translation from learners’ L1 or a rather figurative expression. Anyhow, such a combination neither exists nor delivers its intended meaning. Example (6) was an adverb-verb collocation and the node, *say*, was not appropriate in the context and the collocate, *resolutely*, did not make a chunk with the node, either.

(5) ***worry evaporates** → **concern lessens** (deviance: both)

Now that ***worry has evaporated**, and we are facing the opposite fear. (S5T4)

→ Now that **concern has lessened**, and we are facing the opposite fear.

(6) ***resolutely say** → **strongly feel** (deviance: both)

I ***resolutely say** that there are much more opportunities in the campus to support what students need and qualified systems to pick up the goal than in the off-campus life. (S3T1)

→ I **strongly feel** that there are much more...

These findings indicate that the learners had trouble choosing appropriate nodes as well as searching collocates, which significantly contributes to their inappropriate use of lexical collocations in L2 writing. Deviances in nodes are more serious than in collocates since nodes are the bases in constructing collocations and representing the intended meaning. It is indeed hardly possible for L2 learners to appropriately use collocations with inappropriate nodes. When learners are able to choose the proper nodes for the context,

they only need to learn which collocates are combined with the nodes. As Lewis (2000) suggested, in this case, learners' difficulties in choosing correct collocates can be lessened by leading them to acquire the node and collocate together as a single chunk from the beginning of their learning vocabulary. However, difficulties in choosing proper nodes cannot be resolved by simply raising learners' awareness on lexical collocations, but requires significant overall improvement of L2 proficiency to properly represent their intended meaning. In this regard, Korean learners' difficulties in using lexical collocations in L2 writing are fundamentally rooted in their general L2 writing abilities not limited to their knowledge in L2 collocations.

4.3. Factors of Difficulties in the Use of Lexical Collocations

Deviant elements in the learner corpus provided clues about what would have led the learners to produce inappropriate lexical collocations in their L2 writing. In this section, possible sources of the deviations are examined to address the Research Question 3.

4.3.1. Intra-language factors (L2 influence)

The learners were found to build wrong combinations by failing to distinguish the subtle differences in synonyms or near-synonyms in L2, namely, English language in this study. Synonyms originally refer to the words meaning the same as other words or expressions, but they are broadly defined as formally or semantically related words in this context. The learners were often confused about the subtle semantic differences of synonyms in the lack of knowledge in English vocabulary, thereby choosing inappropriate constituent words. Three cases were identified regarding the learners' confusion in selecting the right words among the synonyms. First, they chose formally related but semantically unrelated synonyms as shown in Example (7).

- (7) ***Humane imagination** and creativity are essential for our future vision, and these things have been sustaining our society. (S5T6)

The deviant adjective-noun collocation, *humane imagination*, was created by the learner's confusion between *human* and *humane*. *Human* and *humane* look similar to each other in their forms, but the former means "relating to people", and the latter, "kind or compassionate". The learner should have used *human* to deliver the intended meaning in the context.

Second, learners' use of formally unrelated but semantically related synonyms gave rise to deviant collocations, which was the most common case of misusing synonyms.

Examples (8) and (9) illustrate the problems.

(8) These days, it is frequent to meet ***unknown people** because of the improvement of transportation and division of labor. (S1T3).

(9) In addition, students can broaden their perspective on life while traveling by meeting and seeing ***unacquainted people**, places, and culture to learn the fact that the world they had known was not all, as well as having time to reflect on themselves seriously. (S2T2).

The combinations marked in bold in Examples (8) and (9) should be replaced by *new people*. The adjectives, *unknown* and *unacquainted* are semantically related with the word, *new*, but they do not convey the same meaning with *new people* as an adjective-noun collocation when combined with the noun, *people*. According to CCED, *new people* refers to those whom one encounters for the first time having not been previously aware of their existence. On the other hand, *unknown person* means someone whose name you do not know, or whose character you do not know anything about. In this sense, *unknown people* are semantically comparable to “unidentified” people, the existence of whom one is aware of but has not met. The combination of *unacquainted people* is an awkward combination because *unacquainted* is not used as an adjective to modify the following nouns but usually used in the form of *be unacquainted with*.

Third, formally and semantically related words made it difficult for the learners to choose appropriate elements for lexical collocations. Example (10) displays a case in which a learner misused *progression* for *progress*.

(10) ***Great progression** is largely based on “know why”, a theoretical knowledge, and university is an academic place where people learn “know why” rather than “know how”. (S1T5)

CCED regards *progression* as a synonym of *progress*, but their meanings are subtly different from each other. *A progression* refers to a gradual development from one state to another, while *progress* is the process of gradually improving or drawing nearer to achieving or completing something. *Progress* expresses more positive changes than *progression* in that it implies an improvement. Furthermore, *progression* refers to the “result” of development whereas *progress* focuses on the “process” of improvement. In this regard, *progress* is more appropriate than *progression* in this context.

Learners’ inappropriate use of synonyms in producing L2 collocations shows that they had insufficient knowledge in the meanings and usage of each word. Words categorized as

synonyms in monolingual and bilingual dictionaries share meaning with one another in some aspects, but they are neither used in the same contexts nor combined with the same words. However, the learners tended to combine the elements based on the literal or salient meanings out of context drawn from their own knowledge or dictionaries, producing wrong collocations.

Another noteworthy intra-lingual factor for deviant collocations was related to so-called, ‘stretched-verb construction’ (Nesselhauf, 2005) in verb-noun combinations. Stretched verbs refer to verbs which cannot deliver meanings in verb-noun combinations, which are more commonly called ‘light verbs’ or ‘delexical verbs’ in linguistics. In the verb-noun combinations with light verbs, noun elements mainly deliver meanings and verb elements contain little semantic content. Typical examples of light verbs include *take*, *make*, *give*, *have* etc. Deviant collocations in relation to these light verbs are exemplified in Table 8.

Some deviations were caused by misusing other light verbs. In *make effect*, for example, the learner should have used the verb, *have*, as a light verb to be combined with the noun, *effect*. Other deviations were caused by misusing light verbs when heavy verbs should have been used in the given contexts as seen in the examples, *make a balance* (for *achieve balance*) and *give image* (for *project image*). In these combinations, the verb elements should have contained more specific meanings as in *achieve* and *project*, respectively. By the same token, deviations were also made when the nouns take particular non-light verbs to deliver specific meanings but the learners did not have the knowledge about it. For example, the learners misused the light verb, *get*, without knowing that these nouns combine with particular verbs such as *take/join* (for *class*) and *earn* (for *income*), as shown in Table 8. These findings correspond to the well-documented overuse of light verbs by Korean learners of English to compensate for their lack of collocation knowledge (Kim & Yoon, 2008; Kim, 2015; Lee, 2015).

TABLE 8

Examples of Deviations Caused by Using Inappropriate Stretched Verbs

Deviation	Stretched Verb	Appropriate Combination
make effect	make	have effect
make a balance	make	achieve a balance
give image	give	project image
have brooch	have	wear brooch
get a class	get	take/join a class
get income	get	earn income

One interesting intra-lingual factor for the deviations was the cases when the L2 combinations themselves were correct as existing collocations, but used in wrong contexts and ultimately judged deviant due to the misrepresentation of intended meaning. In this case, the learners have the knowledge of habitually co-occurring combinations, but are lack

of the knowledge about their exact meanings or situations to be used. One example is the misuse of *catch a breath* when *take a breath* should have been used. According to CCED, meanings of the two collocations are subtly different as follows:

take a breath:

When you take a breath, you breathe in once.

catch a breath:

1. When you catch your breath while you are doing something energetic, you stop for a short time so that you can start breathing normally again.
2. If something makes you catch your breath, it makes you take a short breath of air, usually because it shocks you.

While *catch a breath* shares some meaning of *take a breath* in that the person breathe in, the situations of using *catch a breath* are rather purposeful and specific unlike those of *take a breath*. Without the knowledge about these differences between the two, the learners would have had difficulties in using a suitable collocation in the given context. Learners' misuse of existing collocations in the given contexts, which are acceptable in other contexts, is a serious issue since they can hardly be aware of the misuses without external judgment. Furthermore, it was reported that raters of L2 writing foreground the appropriateness of writers' messages when they evaluate L2 collocations used (Chon & Shin, 2009). Without particular attention to the usage of existing collocations in terms of their meanings and the contexts of use, in this vein, this type of misuse of collocations should be extremely challenging for L2 learners to deal with.

4.3.2. Inter-language factors (L1 influence)

Direct L1 translation as an interfering L1 influence on producing L2 collocations was a typical source of deviant collocations. Meanwhile, the influence of L1 cannot be verified in this study which did neither scrutinize the process of learners producing L2 collocations nor thoroughly analyze linguistic similarities between L1 and L2. L1 influence on L2 collocation use is thus speculated based on the semantic similarity between L1 and L2, which "was considered an indication that influence was likely (Nesselhauf, 2003, p. 234). In this study, the researcher, a native speaker of Korean and fluent speaker of English, judged whether the deviant combinations derived from L1 translations as exemplified in Table 9.

Some deviances on L1-translated L2 collocations demonstrate the learners' lack of knowledge in habitually co-occurring elements in L2 collocations as seen in *big popularity*, *dig up gold*, *do a role*, and *handle equipment*. In these adjective-noun and verb-noun

TABLE 9
Examples of Deviations Caused by Korean Equivalent Combinations

Deviation	Korean Equivalent	Appropriate Combination
big popularity	keun ingi	wide popularity
dig up gold	geumeul kaenaeda	mine gold
do a role	yeokaleul hada	play a role
real society	hyeonsil sahoe	real world
uncivilized society	migae sahoe	primitive society
handle equipment	jangbireul daruda	operate equipment
cultural superiority	munhwajeok uwi	cultural dominance
feel culture	munhwareul neukkida	experience culture
know world	sesangeul alda	experience world
lead a trend	yuhaengeul judohada	set trends
catch stars	seutareul japda	sign up/recruit stars

combinations, the node nouns habitually occur with particular collocates in L2 collocations to represent the common semantic expressions about the nodes. In the lack of knowledge in the arbitrary restrictions, however, the learners translated the L1 equivalents based on the literal meaning and generated those deviant combinations. In the cases of *real society* and *uncivilized society*, L2 collocations have common counterparts for the intended meaning of L1-translated combinations, *real world* and *primitive society*, respectively. Although the synonyms, *society* and *uncivilized*, adopted in the combinations substantially share the semantic properties of *world* and *primitive*, they do not co-occur with the chosen collocate and node.

L1-influenced deviant L2 combinations also derived from the mismatches between the intended meanings constructed in L1 and the semantic representation of L1-translated L2 combinations. In other words, the Korean equivalents convey the intended meanings, but L1-translated English combinations do not properly represent the meanings as seen in *cultural superiority*, *feel culture*, *know world*, *lead a trend*, and *catch a star*. In these deviant combinations, *feel culture* and *know world*, the Korean equivalent for “feel”, *neukkida* and that of “know”, *alda*, entail the meaning of “experience”. However, their English translations, *feel* and *know*, do not deliver the meaning of “experience” combined with the node nouns, *culture* and *world*, respectively. In *lead a trend*, the Korean equivalent refers to achieving dominance or popularity by introducing innovation into the field of interest. In this sense, establishing a new movement or development in the current tendency is properly represented by *set trends* in L2 collocations, but not by the verb collocate *lead*. By the same token, *catch a star* was supposed to refer to hiring celebrities, which is clearly represented in L2 collocation, *sign up/recruit a star*, but the L1-translated verb collocate *catch* does not deliver such a meaning.

Another interesting aspect of L1 influence was observed in learners’ paraphrasing the intended meaning based on L1 semantic representations, often in an unnecessarily

elongated or elaborated fashion, to compensate for their incapability to come up with proper L2 collocations. Examples (11), (12), and (13) show how the learners paraphrased their intended meanings based on L1 instead of using relevant L2 collocations.

- (11) ***The difference of fashion from today and the past is in shift time.**
 → **Fashion has shifted over time.**
- (12) We are easily showed ***items famous fashion designer made.**
 → We have been often shown **famous designer clothes.**
- (13) ...some students would possibly try to ***use this time without good reason...**
 → ...some students would possibly try to **waste their time...**

Example (11) shows how simply the idea can be represented with an appropriate noun and a verb not in the lengthy sentence. Example (12) illustrates the learner's lack of knowledge in the noun-noun collocation, *designer clothes*. While Example (13) does not necessarily show the learner's lack of knowledge in the collocation, *waste time*, the suggested collocation definitely delivers the intended meaning in a more economic manner. These unnecessarily complicated or elaborated expressions might have derived from learners' resort to L1 conceptions about their intended meaning in the lack of knowledge about proper collocations to represent the meaning. Brevity is not a requisite for good writing, but unnecessarily complicated expressions easily distract readers' attention and even blur or misrepresent the ideas at times. While the adherence to L1 resulted from the lack of L2 collocation knowledge, the inappropriate products were rooted in L1 influence. In this regard, learners' strategic L1-based representation of the intended meaning in L2 writing can be resolved by extending their L2 collocation knowledge.

5. CONCLUSION AND IMPLICATIONS

The present study explored how Korean adult learners of English used lexical collocations in L2 writing, addressing (a) the frequency and acceptability of learner collocations, (b) problematic constituents of deviant collocations, and (c) possible sources of the learner difficulties. Six types of lexical collocations were examined in comparison with that of native speakers: adjective-noun, verb-noun, noun-verb, noun-noun, adverb-verb, and adverb-adjective combinations. The overall occurrences of learner collocations constituted nearly 8% of the learner corpus from six 500-700 word argumentative essays about general topics, and the most frequent types of collocations were adjective-noun and verb-noun combinations taking nearly 76% of the overall occurrences. The overall frequency of collocations and distributions of frequencies in each type were almost the

same between the learner and native speaker corpus. Overall acceptability of learner collocations was around 70%, and the acceptability of each collocation type was also around 70% except for 80% acceptability of noun-noun collocations. Learners' difficulties in using lexical collocations, demonstrated in approximately 30% of deviant collocations, were reflected on their inappropriate choices of nodes, collocates, and the both. Comparable error rates in the both elements of deviant combinations indicate learners' difficulties in choosing the nodes as well as the collocates. As possible sources for the learner difficulties, intra- and inter-language factors were identified. Intra-language factors derived from learners' lack of knowledge in L2 synonyms and light verbs, and inappropriate use of correct collocations in wrong contexts. Inter-language factors from L1 influence included learners' combining non-habitually co-occurring elements and misrepresentation of the intended meaning based on L1 semantic representations. Learners' L1-driven elaboration of intended meaning, which could have clearly represented by suitable L2 collocations, was also found as L1 reliance to compensate for the lack of L2 collocation knowledge.

The current findings about various collocation types from the authentic learner production data provide important implications for serving Korean learners' difficulties in using collocations in L2 writing. First and foremost, instructions on producing L2 lexical collocations should be embedded in the immediate context of learners' L2 writing, through which their intended meaning can be properly represented. The learners' frequent errors on the both elements and misuse of right collocations in wrong contexts indicate that L2 learners' collocation knowledge is not limited to the arbitrary rules of combining multi-words. Using L2 collocations reside in the whole ecology of representing learners' intended meaning in L2 writing ranging from choosing the proper head nodes to organizing the ideas in the sentence or paragraph level. Thus, teaching L2 collocations should address how to construct meaning by using collocations in harmony with the overall sentence and paragraph structures and organizations. Instructional focus on the representation of intended meaning corresponds to the essence of learning L2 collocations that is to "store a new collocation in the L2 lexicon and link it to an existing concept" (Yamashita & Jiang, 2010, p. 652).

This study also suggests the need for different instructional approach to different collocation types based on the disproportionate productions. The most frequent types, adjective-noun and verb-noun collocations, require consistent and long-term approach to enlarge learners' repertoire of these widely-applicable L2 collocations, considering the slow progress of L2 collocation knowledge (Levitzky-Aviad & Laufer, 2013). In that process, developing L2 vocabulary depth should be one of the focal instructional goals given that the learners often chose wrong constituents of L2 collocations from their confusions about L2 synonyms. Instructions on less frequent types such as adverb-

adjective or adverb-verb collocations can also benefit the learners, even more instantly, enabling them to use the collocations in L2 writing instead of avoiding them due to the lack of collocation knowledge. The suggested genre-specific usage of noun-verb collocations can be applied to instructional practice by linking specific genres of L2 writing with teaching the combinations.

REFERENCES

- Bahns, J. (1993). Lexical collocations: A contrastive view. *ELT Journal*, 47(1), 56-63.
- Bahns, J., & Eldaw, M. (1993). Should we teach EFL students collocation? *System*, 21(1), 101-114.
- Baxter, P., & Jack, S. (2008). Qualitative case study methodology: Study design and implementation for novice researchers. *The Qualitative Report*, 13(4), 544-559.
- Benson, M., Benson, E., & Ilson, R. (1986). *The BBI dictionary of English word combinations*. Amsterdam: Benjamins.
- Bishop, H. (2004). The effect of typographic salience on the look up and comprehension of unknown formulaic sequence. In N. Schmidt (Ed.), *Formulaic sequences: acquisition, processing, and use* (pp. 227-244). Philadelphia, PA: John Benjamins Publishing Company.
- Brashi, A. (2006). Collocability as a problem in L2 production. *Reflection on English Language Teaching*, 8 (1), 21-34.
- Choi, Y., Chon, Y., & Han, M. (2015). L2 learners' knowledge of verb-noun collocations: Congruency, L2 proficiency and learning strategies. *Korean Journal of Applied Linguistics*, 31(3), 31-63.
- Chon, Y., & Shin, D. (2009). Collocations in L2 writing and rater's perceived writing proficiency. *Korean Journal of Applied Linguistics*, 25(1), 101-129.
- Chon, Y., & Lee, S. (2015). L2 collocational knowledge by congruency and writing proficiency of EFL university learners. *English Language & Literature Teaching*, 21(1), 49-74.
- Cobb, T. (2006). Review of Nadja Nesselhauf (2005): "Collocations in a learner corpus". *The Canadian Modern Language Review*, 63(2), 293-295.
- Cowie, A. (1981). The treatment of collocations and idioms in learners' dictionaries. *Applied Linguistics*, 11(3), 223-325.
- Cowie, A. (1997). Phraseology in formal academic prose. In J. Aarts, I. de Mönnink, & H. Wekker (Eds.), *Studies in English and language teaching* (pp. 43-56). Amsterdam and Atlanta: Rodopi.
- Fernando, C. (1996). *Idioms and idiomaticity*. Oxford: Oxford University Press.

- Fraser, B. (1970). Idioms within a transformational grammar. *Foundations of Language*, 6, 22-42.
- Granger, S. (1998). Prefabricated patterns in advanced EFL writing: Collocations and formulae. In A. Cowie (Ed.), *Phraseology: theory, analysis and application* (pp. 145-160). Oxford: Oxford University Press.
- Granger, S. (2005). Pushing back the limits of phraseology: How far can we go? In C. Cosme, C. Gouverneur, F. Meunier, & M. Paquot (Eds.), *Proceedings of phraseology 2005 conference* (pp. 165-168). Louvain-la-Neuve, Oct. 13-15, 2005.
- Grant, L., & Bauer, L. (2004). Criteria for re-defining idioms: Are we barking up the wrong tree? *Applied Linguistics*, 25(1), 38-61.
- Henriksen, B. (2013). Research on L2 learners' collocational competence and development – a progress report. In C. Bardel, C. Lindquist, & B. Laufer (Eds.), *L2 vocabulary acquisition, knowledge, and use: New perspectives on assessment and corpus analysis* (pp. 29-56). European Second Language Association 2013.
- Howarth, P. (1993). A phraseological approach to academic writing. In G. Blue (Ed.), *Language learning and success: Studying through English* (pp. 58-69). London: Macmillan.
- Howarth, P. (1998). The phraseology of learners' academic writing. In A. Cowie (Ed.), *Phraseology: theory, analysis, and application* (pp. 161-186). Oxford: Oxford University Press.
- Hunston, S. (2002). *Corpora in applied linguistics*. Cambridge: Cambridge University Press.
- Jafarpour, A., Hashemian, M., & Alipour, S. (2013). A corpus-based approach toward teaching collocation of synonyms. *Theory and Practice in Language Studies*, 3(1), 51-60.
- Kim, D. (2008). An analysis of Korean university EFL learners' adjective-noun collocational competence. *Studies on English Language & Literature*, 34(4), 261-282.
- Kim, H., & Yoon, H. (2008). Effects of predictability in L1 on the use of L2 verb-noun collocations. *English Teaching*, 63(2), 237-259.
- Kim, J., & Cho, Y. (2010). A study of the relationship between collocational knowledge and lexical knowledge/general English proficiency: Focus on synonyms of verbs and adjectives. *Korean Journal of English Language and Linguistics*, 10(3), 447-471.
- Kim, K. (2012). L1 influence on L2 collocational knowledge: By level of receptive vocabulary knowledge and general language proficiency. *Korean Journal of Applied Linguistics*, 28(3), 1-38.
- Kim, N. (2003). An investigation into the collocational competence of Korean high school

- EFL learners. *English Teaching*, 58(4), 225-248.
- Kim, S. (2015). The effect of learner proficiency and L1 transfer on the use of Make by Korean EFL learners of English. *Language Research*, 51(1), 139-166.
- Koo, K. (2006). *Effects of using corpora and online reference tools on foreign language writing: A study of Korean learners of English as a second language*. Unpublished doctoral dissertation. The University of Iowa. U.S.A.
- Laufer B., & Waldman, T. (2011). Verb-noun collocations in L2 writing: A corpus analysis of learners' English. *Language Learning*, 61(2), 647-672.
- Lee, S. (2015). Korean EFL learners' productive collocation knowledge and its relationship to their writing proficiency. *English Language and Literature Teaching*, 21(4), 59-84.
- Lee, S. (2016). L1 influence on the process of L2 collocation: An experimental study of Korean EFL learners. *Linguistic Research*, 33, 137-163.
- Levitzky-Aviad, T., Laufer, B. (2013). Lexical properties in the writing of foreign language learners over eight years of study: single words and collocations. In C. Bardel, C. Lindquist, & B. Laufer (Eds.), *L2 vocabulary acquisition, knowledge, and use: New perspectives on assessment and corpus analysis* (pp. 127-148). European Second Language Association 2013.
- Lewis, M. (2000). Language in the lexical approach. In M. Lewis (Ed.), *Teaching collocation* (pp. 126-154). England: Language Teaching Publications.
- Ma, J. H., & Kim Y. (2013). Korean high school English learners' knowledge of collocations: Focusing on delexical verbs: *make, get, and take*. *Language Research* 49(1), 45-71.
- Makkai, A. (1972). *Idiom structure in English*. The Hague: Mouton.
- Muhammad Raji, Z. & Hussein, A. F. (2003). Translational collocational strategies of Arab learners of English: A study in lexical semantics. *Babel*, 49(1), 59-81.
- Nesselhauf, N. (2003). The use of collocations by advanced learners of English and some implications for teaching. *Applied Linguistics*, 24(2), 223-242.
- Nesselhauf, N. (2005). *Collocations in a learner corpus*. Amsterdam & Philadelphia: Benjamins.
- Park, S. (2003). Lexical collocation use by Korean EFL college learners. *SNU Working Papers in English Linguistics and Language* 2, 19-40.
- Seretan, V. (2005). Induction to syntactic collocation patterns from generic syntactic relations. In *proceedings of the 19th International Joint Conference on Artificial Intelligence* (pp. 40-49). Sydney, Australia.
- Seretan, V., & Wehrli, E. (2006). Accurate collocation extraction using a multilingual parser. In *Proceedings of the 21st International Conference on Computational Linguistics and 44th Meeting of the Association for Computational Linguistics* (pp.

- 953-960). Sydney, Australia.
- Sinclair, J., & Renouf, A. (1998). A lexical syllabus for language learning. In R. Carter & M. McCarthy (Eds.), *Vocabulary and language teaching* (pp. 140-158). Harlow: Longman.
- Sung, M. (2017). Verb-noun collocations in EFL writings: Crosslinguistic influence and input frequency. *Korean Journal of Applied Linguistics*, 33(3), 91-115.
- Strässler, J. (1982). *Idioms in English: A pragmatic analysis*. Tübingen: Gunter Narr Verlag.
- Szudarski, P., & Conklin, K. (2014). Short- and long-term effects of rote rehearsal on ESL learners' processing of L2 collocations. *TESOL Quarterly*, 48(4), 833-842.
- Weinreich, U. (1969). Problems in the analysis of idiom. In J. Puhvel (Ed.), *The substance and structure of language* (pp. 23-81). Berkeley: University of California Press.
- Wood, M. (1981). *A definition of idiom*. Manchester, England: Center for Computational Linguistics, University of Manchester.
- Wu, S., Witten, I. H., & Franken, M. (2010). Utilizing lexical data from a Web-derived corpus to expand productive collocation knowledge. *ReCall*, 22(1), 83-102.
- 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(4), 647-668.

Applicable levels: Tertiary

YiBoon Chang
Foreign Language Education Center
Language Education Institute, Seoul National University
1 Gwanak-ro, Gwanak-gu, Seoul 08826, South Korea
Phone: 02-880-8560
Email: yiboon22@gmail.com

Received on March 1, 2018

Reviewed on April 15, 2018

Revised version received on May 15, 2018