

# The Use of Collocations by Chinese Learners of Korean\*

Hyeson Park  
(Keimyung University)

**Park, Hyeson. 2007. The Use of Collocations by Chinese Learners of Korean.** *The Linguistic Association of Korea Journal*, 15(2), 1-22. This paper examines the use of collocations by Chinese learners of Korean as an L2. Spoken narrative data collected from 45 learners at three proficiency levels and 15 Korean native speakers were analyzed focusing on S+V, O+V, adverb (ideophones)+V, and noun+hata collocations. The analysis reveals the following: 1) A developmental pattern, though weak, was observed, indicating a positive relationship between the learners' general L2 proficiency and collocation competence, a result inconsistent with what was observed by Howarth (1998) and Bonk (2001). 2) The learners produced S+V collocations more than O+V collocations, while the Korean speakers produced twice as many O+V collocations as S+V collocations. This implies that high frequency of a target structure in the input does not automatically render the target to be noticed and acquired by a learner. 3) The adverb (ideophones) +verb collocations, highly frequent in the Korean language, but rare in Chinese, showed low frequency in the learner data, a result confirming the general observation in L2 research of L1 influence in L2 learning. More research with diverse data collection methods is suggested to better understand how collocations are learned, stored, and used in L2.

**Key Words:** collocations, L2 learning, Chinese learners, Korean as an L2

## 1. Introduction

---

\* An earlier version of this paper was presented at the 2006 Winter Conference of the Society of Modern Grammar held at Pukyong National University. I am grateful to the anonymous reviewers for their comments and suggestions. This work was supported by the Keimyung Bisa Research Grant in 2005.

## 2 Hyeson Park

Prefabricated or multiword units prevalent in natural language have drawn researchers' attention for a long time, and recent progress in corpus linguistics and computational tools has accelerated research on multiword units in both theoretical and applied linguistics (Sinclair, 1991; Nattinger & DeCarrico, 1992; Ellis, 1997; Cowie, 1998; Granger, 1998; Howarth, 1998; Mel'čuk, 1998; Jackendoff, 2002; Wray, 2002; Nesselhauf, 2003). A general consensus emerging on the nature of word combinations out of the recent research is that knowledge of word combinations is part of a native speaker's linguistic competence together with knowledge of syntactic rules, and that word combinations can no longer be set aside as an idiosyncratic phenomenon; the way speakers acquire, store, and use them deserves more attention. This claim has a significant theoretical implication in that it contrasts with the tradition of mainstream generative grammar, which places a great emphasis on syntax as a rule-governed productive system, while ignoring the lexicon as a storage of idiosyncratic lexical information (Pawley & Syder, 1983; Jackendoff, 2002; Wray, 2002).

Prefabricated units do not form a single homogeneous category; rather, they refer to a continuum of categories with free combinations at one end and pure idioms at the other. The focus of the current study is the word combinations which lie in the fuzzy zone between free combinations and true idioms, which are called collocations (Cowie, 1998). Specifically, this study examines the use of collocations in second language (L2) learning, which has been recognized as one of the most important factors for successful L2 learning. The use of incorrect word combinations, rather than incorrect grammatical usage, was found to be the main source of a foreign 'accent' even in advanced L2 learners' language (Pawley & Syder, 1983).

Though there has been constant interest in collocations in L2 learning, our understanding of L2 collocations is not as rich as it should be, the main reason being the limitations in research methods and target languages studied. That is, the majority of L2 collocation studies have examined English collocations using written data collected through controlled or semi-controlled methods, such as questionnaires or in-class

compositions (Granger, 1998; Howarth, 1998; Altenberg & Granger, 2001; Nesselhauf, 2003). This paper attempts to broaden our understanding of L2 collocations by examining the use of Korean collocations by Chinese learners of Korean in their spoken narratives. It is hoped that an analysis of spoken data produced by Chinese learners of L2 Korean will help us better understand how L2 learners learn and store collocations in the process of L2 learning.

This paper is organized as follows: in section 2, we examine definitions of collocations proposed by previous research, which is followed by section 3 with a review of research on collocations in L2 learning. Section 4 provides a description of the data collection method and results of the current study. A brief summary in section 5 concludes the paper.

## 2. Defining collocations

Firth (1957, cited in Wray, 2002) succinctly summarized the significance of word combinations by saying that one knows a word from 'the company that it keeps'. The relationship between a word and its company, however, is not uniform but varies, resulting in several subcategories of word combinations. It is still being debated how to classify and define the subcategories, but a widely recognized classification is found in Howarth (1998). His classification, influenced by the tradition of Russian phraseological studies, puts an emphasis on the gradient rather than categorical nature of word combinations:

Free combinations: consist of elements that are used in their literal senses and freely substitutable (e.g. carry a trumpet, want a car, on the top of the table)

(Restricted) collocations: consist of one element that is used in a restricted sense and another used in a nonrestricted sense (e.g. take a picture, blow a fuse, under attack)

Figurative idioms: have a metaphorical meaning as well as a

#### 4 Hyeson Park

literal interpretation (e.g. blow your trumpet, under the microscope)

Pure idioms: non-compositional and the most opaque and fixed category (e.g. sweeten the pill, blow the gaff, under the weather)

A problem with the above classification, which relies on such descriptors as restricted vs. nonrestricted senses or substitutability, is that there are many cases where the boundaries between free combinations and collocations, or between collocations and idioms are not so clear. This led some researchers to utilize statistical measures in identifying collocations so that two or more items were judged to be collocations when the frequency of their co-occurrence exceeded a pre-fixed threshold. Granting its problems, however, this paper adopts the phraseological or linguistic sense of collocations presented by Howarth (1998) rather than the frequency-based classification since statistical extraction of collocations has often produced word combinations that cannot be considered genuine collocations (cf. Hong et al, 2001).

The importance of collocations among the subcategories of word combinations was noted by Mel'čuk (1998), who stated that collocations constitute the absolute majority of word combinations and should be paid special attention. Some characteristics of collocations summarized by Mel'čuk include the following interrelated properties: arbitrariness, language specificity, and unidirectional selection. The arbitrary nature of collocations explains the restricted substitutability. For example, 'strong' in 'strong coffee' cannot be replaced by a synonym 'powerful', and a 'heavy smoker' cannot be a 'weighty smoker'. Substitution of 'do' with 'give' in the expression 'do (someone)a favor' results in a different meaning from its original meaning ('to do something that will have a good effect on someone'). In English you like 'strong coffee' not 'thick coffee'; however, in Korean you like 'thick coffee'-- an example of language-specificity. In English you 'wear' shirts, glasses, caps, and shoes. However, Korean uses different verbs for each of these items:

*ipta*, *kkita*, *ssuta*, and *sinta*, respectively. Another characteristic of collocations is that one element selects another element it wants to keep company with. In 'do a favor', 'favor' selects 'do'; 'The horse' chooses 'neighs' but not 'barks'. In Korean *kechitmal* 'lie' chooses *ppalkan* 'red' to express an 'utter lie', and *nat* 'face' selects *ikta* 'ripe' to express *nat-i ikta* 'a familiar face'. All these examples show that arbitrariness, language specificity and unidirectional selection are not independent, but rather interrelated properties in forming collocations.

### 3. Previous research on collocations in L2

Pawley and Syder(1983) noted that what is intriguing about native speakers' linguistic competence is their ability to make the distinction between natural vs. unnatural word combinations, and that this is what sets a native speaker apart from a non-native speaker. The arbitrariness and language specificity of collocations appear to be what make them a serious obstacle for a successful L2 learning.

Studies on word combinations in L2 have mostly centered on the learning of true idioms rather than collocations, which are more frequent and productive than true idioms. The main variables in the few studies that examined learning of collocations include L1 influence on the acquisition of L2 collocations, the relationship between collocation competence and general L2 proficiency, and differences between native and non-native collocation competence. Nesselhauf (2003) examined writings by German learners of English, focusing on verb+object collocations (e.g. make an effort, do dishes, reach a conclusion etc.). His analysis of learner errors revealed that even advanced learners had considerable difficulty in producing correct collocations, mainly due to L1 influence. The German learners made fewer errors when the L1 and L2 collocations were congruent in that word for word translation of an L1 collocation rendered a natural L2 collocation. Nesselhauf argues that the role of L1 in L2 learning, which has been downplayed by those researchers who put a great emphasis on the universal aspects of language learning, should be paid more attention.

Howarth (1998) examined collocations in English academic writings by native and non-native writers, focusing also on the verb+object collocations. Howarth found that collocations constituted 38% of the verb+object combinations produced by the native speakers in the study, while in the non-native writing the percentage was 25%. His examination of the relationship between the use of collocations and general L2 proficiency revealed that the two were not strongly correlated, which led him to suggest that the use of collocations might depend more on an individual's stylistic choice rather than on her general L2 proficiency and that one might not be able to predict a learner's L2 proficiency based on her use of collocations.

Bonk (2001), based on the results of collocation and general English proficiency tests administered to 98 adult learners of English, cautiously agreed with Howarth's suggestion that L2 collocation competence did not develop in proportion to general L2 proficiency. He observed that lack of correlation between general English proficiency and collocation competence was more evident among the learners in the intermediate level than those in the advanced or beginning level.

Granger (1998) tested the hypothesis that non-native speakers might underuse collocations compared to native speakers by examining written production data collected from advanced French learners of English and native English speakers. She focused on one category of adverbs, that is, amplifiers/modifiers ending in *-ly*. She found that there was a clear difference between the native and non-native groups in their use of these adverbs. The non-native speakers overused some amplifiers (e.g. *completely* and *totally*), while underusing others (e.g. *highly*) compared to the native speakers. Granger at first hypothesized L1 influence as a possible cause for the different patterns between the two groups. However, through an independent experiment eliciting native and non-native speakers' judgments on possible word combinations, she reached a different conclusion: the poor collocation performance of the non-native speakers resulted from the learners' underdeveloped sense of significant collocations; that is, while the native speakers recognized the amplifiers as part of collocations, they were treated as separate items

by the non-native speakers. A similar observation is found in Wray (2002), where he states:

Because native speakers start with big units and analyze them only as necessary, their treatment of collocations can be framed not so much in terms of loose associations per se, but fully formulaic pairings which have become loosened.... In contrast, the adult learner's collocations are to be seen as separate items which become paired. It is this pairing, and particularly, the establishment of the strength of the association, which causes the characteristic difficulties.... (p. 211)

Thus, according to Wray and Granger, the processes involved in the storage and retrieval of collocations are not identical between native and non-native speakers.

Wray (2002) provides an extensive review of previous studies on multiword units in L2 learning. Of the 20 studies he reviewed, English was the target language in 12 studies, French and Swedish being the other two major languages studied. Hence not much is known about the acquisition of prefabricated units including collocations in other languages, especially in East Asian languages. Regarding acquisition of Korean as an L2, few empirical studies have been conducted on the acquisition of collocations. There are a few studies that mention the role of collocations in learning L2 Korean, but the main focus of these studies is either learner errors of all types or vocabulary learning in general. For example, Lee (2002) studied errors made by learners of Korean as an L2. She examined writings (i.e. free composition, diary, in-class composition) of 260 learners from 37 language backgrounds, focusing on grammatical errors (e.g. case particles, tense, and connective suffixes), lexical errors, and spelling errors. She did not pay special attention to collocation errors, but noted in passing that lexical errors by the advanced learners were as frequent as those by beginning learners, and that one way to help learners avoid lexical errors is to emphasize the co-occurrence relationship among words in language

lessons. The lack of correlation between general L2 proficiency and use of collocations observed by Lee appears to be consistent with what Howarth (1998) and Bonk (2001) noticed in L2 English. Han and Kang (2004), noting the importance of vocabulary, especially of collocations in achieving high proficiency in L2 Korean, presented useful teaching methods for collocations utilizing large corpora. Their proposal, however, was not accompanied by empirical data except for a few examples of collocation errors.

Yang (2004) is one of the few studies to focus on the acquisition of collocations in L2 Korean. The data Yang analyzed consisted of 244 controlled or semi-controlled compositions by Russian learners enrolled in a Korean language program. The combinations she examined were subject+verb and object+verb. The errors incurred by a wrong choice of the noun or verb constituted 41.7% of the total errors, the majority of which involved a wrong choice of the verb (36.8% of the total errors). Her detailed analysis of the errors revealed that the main causes of the errors were L1 influence, L2 influence, and application of wrong strategies. A problem of Yang's study is that the learners were not equally distributed across the proficiency levels; the advanced learners constituted more than two thirds of the participants, who produced the majority of errors observed. Hence, it is not possible to examine the relationship between general L2 proficiency and the use of collocations.

The above review of previous studies on collocations in L2 shows that most of the studies have focused on English as an L2 and analyzed controlled or semi-controlled written production data, without paying much attention to the use of collocations in spoken learner language. This study examines and analyzes spoken narrative data by Chinese learners of L2 Korean focusing on the following points: 1) the relationship between general L2 proficiency and collocation use, 2) the difference between native and non-native speakers in their use of collocations, and 3) other factors involved in L2 collocation use (e.g. the role of input, L1 influence, data collection method etc.). In the next section, we present the procedures and results of our study.



## 4. The present study

### 4.1. Methods

*Participants:* 45 Chinese students enrolled in a Korean language program at a Korean university participated in the study: 13 advanced, 15 intermediate, and 17 beginning learners. The levels of the students were determined according to the level of the class they were attending at the time of the data collection. The advanced group had studied Korean about 23 months on average, the intermediate group 12 months, and the beginning group seven months. 15 Korean university students participated as a control group.

*Procedures:* The Chinese students were recruited from those enrolled in a Korean language program in the fall of 2005. The students were paid \$5 each for their participation. The Korean university students were volunteers from the researcher's class in the fall semester of 2004.

The data were collected through an elicitation task using the 24-page wordless picture book *Frog, where are you?* (Mayer, 1969). The subjects were interviewed individually by the researcher. After spending some time talking with the participants to help them feel comfortable, the researcher showed them the picture book and asked them to look at the pictures one by one trying to understand what the story was about. Then, the participants were asked to narrate the story while looking at the pictures again. Two tape-recorders were placed on the table and recorded the participants' narration. The task took about 15 minutes for the Chinese learners and less than 10 minutes for the Korean students. The recorded narrations were then transcribed and analyzed, focusing on the following categories of collocations: subject+verb, object+verb, adverb+verb, and the *-hata* ('do') verb construction.

### 4.2 Results and discussion

Table 1 presents the number of types of collocations for each category produced by each group. The results will be discussed according to each category.

*Subject + verb collocations:* The Chinese students did not produce diverse types of S+V collocations. The learners at the beginning level produced four types of S+V collocations, which increased to eight at the intermediate level, and nine at the advanced level. The native Korean adults produced 15 types of S+V collocations. The difference between the native speakers and the non-native speakers was not statistically significant ( $\chi^2 = 6.88$ ,  $p > .05$ ). Only one collocation in this category was produced by all the participants: *hwa-ka nata*<sup>1)</sup> (anger-NOM arise) 'get angry'.

Table 1. Number of types of collocations produced by the participants

	subject+ verb	object+ verb	adverb+ verb	-hata (do)	etc.	total
Korean univ.(15)	15	28	36	20	11	110
Chinese advanced (13)	9	8	12 (chacta:10)	18	3	50(41)*
intermediate (15)	8	4	6 (chacta:5)	13	3	34(30)
beginning (17)	4	2	9 (chacta: 7)	13		28(22)

\* The number in the parenthesis is obtained when all the instances of adverb+*chacta* (search) is treated as a single item.

There was one collocation produced by all the Chinese learners, but not by the Korean students: *kipwun-i cohta* (mood-NOM good) 'be in a good mood'. The beginning learners produced three incorrect

---

1) The Korean data in this paper are transcribed using the Yale romanization system (Martin 1992).

collocations.

- (1) a. \*sori-ka issta => sori-ka nata  
 sound-NOM exist sound-NOM arise  
 'There is sound.'
- b. \*sayngkak issta => sayngkak hata  
 thinking exist thinking do  
 'Think.'
- c. #sikan-i kellita => sikan-i cinakata  
 time-NOM take time-NOM pass  
 'Time passes.'

In (1a) and (1b), the beginning learners overgeneralized *issta* (exist/have). It may be the case that they were utilizing *issta* as a default verb when they did not have the correct target verb in the lexicon. (1c) is not an incorrect collocation by itself, but it is incorrect in the context it was used.

The intermediate learners added six more S+V collocations to the two mentioned above, some of which are presented in (2).

- (2) a. sikan-i kata  
 time-NOM go  
 'Time passes.'
- b. sori-ka sikkulepta  
 sound-NOM noisy  
 'It is noisy.'
- c. hokisim-i issta/manhta  
 curiosity-NOM have/a lot  
 'be curious'

They made three collocation errors with this category:

- (3) a. \*pyeng-i kocangnata => pyeng-i kkaycita  
 bottle-NOM out of order bottle-NOM break

12 Hyeson Park

- 'The bottle is broken.'
- b. \*pel-i kay-lul ttaylita => pel-i kay-lul sso-ta  
bee-NOM dog-ACC beat bee-NOM dog-ACC sting  
'The bees sting the dog.'
- c. \*hwa-ka nayta => hwa-lul nayta  
anger-NOM arise anger-ACC arise  
'get angry'

In (3a) and (3b) the learners did not know the verbs that the nouns select and chose incorrect verbs. The sentence in (3c) means the same as *hwa-ka nata*, but since the verb in (3c) is a transitive verb, the accusative case particle *-lul* should have been used instead of the nominative particle *-ka*.

The advanced learners produced seven more S+V collocations besides the two produced by all the learners. They include the following:

- (4) a. kaykwuri-ka sayngkita  
frog-NOM get  
'obtain/get a frog'
- b. sok-i sanghata  
inside-NOM rotten  
'get one's feeling hurt'
- c. hokisim-i issta/manhta/nata  
curiosity-NOM have/a lot/arise  
'be curious'

Two types of errors were made by the advanced learners:

- (5) a. \*pel-i kay-lul mwulta => pel-i kay-lul ssota  
bee-NOM dog-ACC bite bee-NOM dog-ACC sting  
'The bees sting the dog.'
- b. \*sori-ka issta/naota/nakata => sori-ka nata  
sound-NOM exist/come out/go out sound-NOM arise  
'There is sound.'

Like the lower level learners, the advanced learners did not seem to have learned the verbs that collocate with the nouns 'bee' and 'sound'. Overall, compared to the other types of collocations, the S+V collocations produced by both the native and non-native speakers were not diverse, possibly due to their infrequent use in the target language.

*Object+verb collocations:* The native speakers produced twice as diverse O+V collocations as the S+V collocations, which is consistent with the general observation that the verb and the object form a tighter unit than the subject and the verb, resulting in more O+V idioms than S+V idioms. However, the non-native speakers produced the same or fewer types of O+V collocations than the S+V collocations. The difference between the native and non-native performance was statistically significant ( $\chi^2=37.13$ ,  $p<.05$ ). The O+V type collocations did not seem to have drawn much attention from the non-native speakers despite their frequent presence in the language input the learners were exposed to. The O+V collocations produced by all the participants are the following:

- (6) a. *cam-ul cata*  
 sleep-ACC sleep  
 'sleep a sleep'
- b. *hwa-lul nayta*  
 anger-ACC arise  
 'get angry'

The beginning level learners produced the above two collocations and one incorrect collocation. Instead of *cam-ul cata*, one learner uttered '*cam-ul issta*' (sleep-ACC exist/have). As is the case with the S+V collocations, this error shows that the learners used *issta* (exist/have) as a default verb when they did not have the correct target verb in their lexicon.

The intermediate learners used two more collocations besides the two in (6):

- (7) a. tomang-(ul) kata  
 escape-(ACC) go  
 'run away'  
 b. sinpal-ul sinta  
 shoes-ACC put on  
 'put on shoes'

In Korean, different verbs are used to express the event of 'putting something on one's body'. For example, *ipta* is used for wearing clothes, *sinta* for shoes, and *kkita* for glasses. The intermediate learners chose the correct verb for shoes in (7b), showing that they learned the distinction at least for shoes. One collocation error was made by an intermediate learner. He uttered "*\*changmwun-ul kyeta*" (window-ACC turn on) when the target verb was *yelta* (open).

The advanced learners produced nine types of O+V collocations. In addition to the collocations produced by the lower level learners, the following were also produced by the advanced learners.

- (8) a. kaykwuri-lul khiwuta/kiruta  
 frog-ACC raise  
 'raise a frog'  
 b. sori-(lul) chita  
 sound-ACC do hard  
 'shout'  
 c. cangnan-(ul) chita  
 mischief-ACC do hard  
 'play a trick on'  
 d. os-ul ipta  
 clothes-ACC put on  
 'put on clothes'

-*Chita* is a verb which combines with diverse nouns or verbs. The advanced learners began to use *chita* as part of a collocation, though

not without errors. One instance of error was found where *sori* (sound) was combined with *cicta* (bark) instead of *chita* (shout). Some learners used the verb *khiwuta/kiruta* (raise) correctly, but some made an error with this verb; *cikhita* (guard) was used instead of the correct verb by one learner.

The native speakers' inventory of O+V collocations totalled 28, doubling the number of S+V collocations. Many of the collocations are specific to the Korean language and difficult to translate to English. Some examples include:

- (9) a. swum-ul    cwukita  
       breath-ACC    kill  
       'hold one's breath'
- b. aykyo-lul    ttelta  
       charms-ACC    act  
       'be all smiles'
- c. sinkyeng-ul    ssuta  
       nerves-ACC    use  
       'pay attention'
- d. cengsin-ul        chalita  
       mind/sprit-ACC    set up  
       'recover one' senses'
- e. hotulkap-ul        ttelta  
       exuberance-ACC    act  
       'be over-excited'

These expressions are usually used in very informal contexts and not so uncommon in native speakers' everyday language. However, none of these collocations were produced by the non-native speakers, even by the advanced learners. This indicates that building up collocations in an L2 is not an easy task, and that high frequency in the input does not guarantee automatic acquisition.

*Adverb+ verb*: Talmy (1985) classified languages into two groups based on the way motion events are expressed: verb-framed vs.

satellite-framed languages. In a satellite-framed language, the main verb conflates the manner and motion of an event, and the path is expressed separately by particles. On the other hand, in a verb-framed language, the path and motion of an event are conflated in the main verb, and the manner of the event is expressed by separate structures such as adverbials or gerundivals. According to Talmy's typology, English and Chinese are satellite-framed languages and Korean a verb-framed language. According to Slobin (2003), another strategy many verb-framed languages use to describe manner is ideophones or memetics, and Korean is well known for its rich lexicon of ideophones (Sohn,1994). This characteristic was clearly seen in the native speakers' narrative data, which contained 36 adverb+verb collocations, many of which involved ideophones.

(10) chembeng	ppacita	(splashing fall)
pelttek	ilenata	(with a jerk stand up)
oswun	doswun cal	salta (harmoniously well live)
hekepcikep	tomangchita	(all flustered run away)
kuwsek	kuwsok	chacta (everywhere search)
ontey	kantey	epta (no where found)
coyong	coyong	kata (silent, silent go)
twuripen	twuripen	salphita. (look around wonderingly)
pwulsswuk	naota	(suddenly come out)
kkamccak	nollata	(all of sudden be surprised)

The adverbials in these expressions describe spatial or acoustic dimensions of manner of motion. The absence of the adverbials in these expressions does not render the expressions ungrammatical, but they provide a more precise or detailed description of the manner with which the action occurs.

The use of ideophones was very rare in the non-native speakers' data. The beginning learners produced nine adverb+verb collocations, but seven of them had *chacta* (search) as the main verb, with adverbials meaning 'everywhere' combined with it (e.g. *yeoki ceki*, *etitunci*,



*yeorekaci, yeokikkaci*). Some of them are unnatural combinations. A learner tried to express "*kkamccak nollata*" but in an incorrect form (*camkkan nollata*). The intermediate learners produced six adverb+verb collocations, fewer than the beginning learners. Five of them had the 'everywhere + *chacta* (search)' structure. The one expression without *chacta* had an ideophone functioning as a main verb (*pelcip ttwuk* "beehouse with a thump (fall)). The advanced learners were not much different from the two lower groups with regard to the use of adverb+verb collocations. They produced 12 collocations, but 10 of these had *chacta* (search) as the verb. Many of the 'adverb+*chacta*' combinations were unnatural (e.g. *amuwteyna, amwudunci, amwukenato*). The other two expressions with a verb other than *chacta* were:

- (11) *kkamccak nollata* (all of a sudden be surprised)  
*phwudeng ttelecita* (with splash fall)

According to Talmy's typology, Chinese belongs to the satellite-framed language group like English, though Slobin (2003) puts Chinese in a new group called serial verb language. Whether Chinese is categorized as a satellite-framed language or serial verb language, a characteristic of Chinese is that it expresses manner of motion with the verb or serial verbs, and does not have a rich inventory of ideophones. This must have contributed to the dearth of adverb+verb collocations in the non-native speakers' narratives. The adverb+verb collocations are common especially in story telling in Korean, and the absence of these expressions in the Chinese students' narratives has resulted in the 'foreign accent' in their narratives.

*Hata construction*: The question of whether the noun+ *hata* construction should be treated as a collocation or a compound verb is debatable. In their extensive discussion of collocations, Im (2002) and Kim (2000) included a few examples of the *hata* construction. Considering its resemblance to the light verb + object construction in English (e.g. do a favor, make a decision, give permission), it would not be unreasonable to classify the *hata* construction as a collocation. The

non-native learners produced noun+ *hata* collocations more than other types of collocations: 13 by the beginning learners, 13 by the intermediate learners, 18 by the advanced learners, and 20 by the native speakers. There was no difference between the native and non-native performance ( $\chi^2=2.37$ ,  $p > .05$ ). The noun+*hata* collocations produced by all the groups include:

- |                               |                         |
|-------------------------------|-------------------------|
| (12) sayngkak hata (think),   | kyelhon hata (marry)    |
| hayngpok hata (be happy)      | palkyen hata (discover) |
| sayngwhal hata (live)         | insa hata (greet)       |
| sinki hata (be curious)       | date hata (date)        |
| kosayng hata (suffer)         | pikon hata (be tired)   |
| dangwhang hata (be at a loss) |                         |

As can be seen from the examples, the noun+ *hata* collocation is very productive in the learners' lexicon. One of the reasons for the relative easiness the learners have for this type of collocation may be that many of the nouns combined with *-hata* are sino-Korean words with corresponding Chinese characters; that is, similarity between L1 and L2 might have resulted in the active use of the construction. Another possible reason may be that the noun+ *hata* combination is not a collocation, but a compound verb, which is stored and retrieved as a unit in the L2 mental lexicon while genuine collocations are stored as separate units, as suggested by Wray (2002) and Granger (1998).

Four errors were found with the noun+*hata* collocations. The intermediate learners produced the following two errors: \**wihem issta* (*wihem hata*) 'be dangerous', \**sayngwhal cinayta* (*sayngwhal hata*) 'live'. The advanced learners made the following error: \**kyelsim ceonghata* (*kyelsim hata*) 'make a decision'.

## 5. Summary and conclusion

In this paper, we examined the use of collocations by Chinese learners of Korean as an L2. Narrative data collected from the learners

at three proficiency levels were analyzed focusing on the S+V, O+V, adverb+V, and noun+ *hata* collocations. The picture emerging from the analysis can be summarized as follows: 1) Overall, the learners did not produce many collocations. The semi-spontaneous spoken production task of story telling may have increased processing load, leading the learners to choose an avoidance strategy. A comprehension task targeting the collocations produced by the native speakers may help reveal whether these collocations are part of the learners' lexicon or not. 2) A very weak developmental pattern was observed; the total number of types of collocations increased from 29 at the beginning level, to 35 at the intermediate level, and 51 at the advanced level. Even though the gap between the levels is not big, the advanced students produced almost twice as many diverse types of collocations as the beginning students, and the difference among the groups was statistically significant ( $\chi^2=5.68$ ,  $p=.058$ ). This result indicates that the learners' general L2 proficiency and collocation competence may be positively related, contrary to what Howarth (1998) and Bonk (2001) observed and led them to suggest that use of collocations may be more strongly related to learners' stylistic choice than their L2 proficiency. We conjecture that the different modes of data collection might be responsible for the different results between the current study and the previous ones. That is, writing tasks do not involve as heavy a processing load as spontaneous speech, and allow the learners to apply conscious editing strategies. 3) The learners' production of S+V and O+V collocations did not show much difference in the number of types. This diverges from the native speakers' performance and the frequency pattern found in the language input they are exposed to; in L1 data O+V collocations outnumber S+V collocations, a consequence of the tighter relation between the object and the verb than the subject and the verb in natural language. This indicates that a high frequency of a target structure in the input does not automatically lead to noticing and acquisition of the target. As Wary (2002) suggests, for some reason collocations in the target language input seem to escape learners' attention. 4) The adverb+verb collocations have a high frequency in the

native speakers' data, but are very rare in the learners'. Since Chinese does not have a rich inventory of ideophones in its lexicon, the adverb+verb collocations in Korean may not be familiar to them and hence are difficult to learn.

In conclusion, the result of this study seems to be consistent with the results of previous research with regard to the role of L1 in learning L2 collocations; the absence of a congruent structure between L1 and L2 may hinder learning of the target structure in L2. The relationship between general L2 proficiency and collocation competence, however, reveals a mixed picture. More research with diverse data collection methods is needed to provide a conclusive answer to this issue.

### References

- Altenberg, B. & Granger, S. (2001). The grammatical and lexical patterning of make in native and non-native student writing. *Applied Linguistics* 22, 173-195.
- Benson, M., Benson, E., & Ilson, R. (1997). *The BBI dictionary of English word combinations*. Amsterdam: John Benjamins.
- Bonk, W. J. (2001). Testing ESL learners' knowledge of collocations. In T. Hudson & J. D. Brown (Eds.), *A focus on language test development: Expanding the language proficiency construct across a variety of tests*. (pp. 113-142). Second Language Teaching & Curriculum Center. University of Hawaii.
- Cowie, A. P. (1998). *Phraseology: Theory, analysis, and applications*. Oxford: Clarendon Press.
- Ellis, N. (1997). Vocabulary acquisition: Word structure, collocation, word- class, and meaning. In N. Schmitt & M. McCarthy (Eds.), *Vocabulary: description, acquisition and pedagogy* (pp. 122- 139). Cambridge: Cambridge University Press.
- Granger, S. (1998). Prefabricated patterns in advanced EFL writing: Collocations and formulae. In A. P. Cowie (Ed.), *Phraseology:*

- Theory, analysis, and applications* (pp. 145-160). Oxford: Clarendon Press.
- Han, S-H. & Kang, H-H. (2004). Teaching vocabulary using collocation. *Journal of Korean Education* 15, 295-318.
- Howarth, P. (1998). Phraseology and second language proficiency. *Applied Linguistics* 19, 24-44.
- Hong, J-S., Kang, -M., & Choi, H-C. (2001). *Hankwuke Yene Kwankyey Yenkwu*. ('Study of Korean collocations'). Seoul: Weolin.
- Im, H-B. (2002). Kwuke yene-uy kaynyem-kwa ku tongsa/uymicek sengkyek ('On the nature of collocation in Korean and its syntactic and semantic properties'). *Kwukehak* 39, 279-410.
- Jackendoff, R. (2002). *Foundations of language: Brain, meaning, grammar, evolution*. Oxford: Oxford University Press.
- Kim, J-H. (2000). *Yene yenkwu* ('Study of collocations'). Seoul: Hankwuk Mwunhwasa.
- Lee, J-H. (2002). *Hankwuke Haksupca-uy Pyohyon Olyu Yenkwu* ('A research study on language production errors of Korean language learners'). Unpublished doctoral dissertation. Kyunghee University
- Martin, S. (1992). *A reference grammar of Korean*. Rutland, Vermont, and Tokyo, Japan: Charles E. Tuttle Company .
- Mel'čuk, I. (1998). Collocations and lexical functions. In A. P. Cowie (Ed.), *Phraseology: Theory, analysis, and applications* (pp. 23-53). Oxford: Clarendon Press.
- Nattinger, J. R. & DeCarrico, J. S. (1992). *Lexical phrases and language teaching*. Oxford: Oxford University Press.
- Nesselhauf, N. (2003). The use of collocations by advanced learners of English and some implications for teaching. *Applied Linguistics* 24, 223-242.
- Pawley, A. & Syder, F. H. (1983). Two puzzles for linguistic theory: Nativelike selection and nativelike fluency. In J. C. Richards & R. S. Schmidt (Eds.), *Language and communication* (pp. 191-226). London: Longman.
- Sinclair, J. (1991). *Corpus, concordance, collocation*. Oxford: Oxford

University Press.

- Slobin, D. (2003). The many ways to search for a frog: Linguistic typology and the expression of motion event. In S. Stromqvist & L. Verhoeven (Eds.), *Relatin events in narrative: Typological and contextua perspectives* (pp. 219-257). Mahwah, NJ: Lawrence Erlbaum Associates.
- Sohn, H-M. (1994). *Korean*. London: Routledge.
- Talmy, L. (1985). Lexicalization patterns: Semantic structure in lexical forms. In T. Shopen (Ed.), *Language typology and lexical description: Vol. 3. Grammatical categories and the lexicon* (pp. 36-149). Cambridge: Cambridge University Press.
- Wray, A. (2002). *Formulaic language and the lexicon*. Cambridge: Cambridge University Press.
- Yang, S-H. (2004). *A lexical pedagogy through collocation error analysis: Focusing on errors by Russian learners of Korean*. MA Thesis, Yonsei University.

Hyeson Park

Department of English Language and Literature

Keimyung University

1000 Sindang-dong, Dalseo-gu

Daegu, 704-701, Korea

Phone: 82-53-580-5091

Email: [parkhy@kmu.ac.kr](mailto:parkhy@kmu.ac.kr)

Received: 30 March, 2007

Revised: 30 May, 2007

Accepted: 6 June, 2007