Linguistic Research 35(Special Edition), 83-103
DOI: 10.17250/khisli.35..201809.004

Interlanguage development of young Korean EFL learners' modal usage: A learner corpus study*

Daria Soon-Young Seog · Incheol Choi*****
(Kyungpook National University)

Seog, Daria Soon-Young and Incheol Choi. 2018. Interlanguage development of young Korean EFL learners' modal usage: A learner corpus study. *Linguistic Research* 35(Special Edition), 83-103. The current paper reports on a comparison study conducted between a learner corpus, the KNU English Learner Corpus (KELC), and a native speaker corpus, the Corpus of Contemporary American English (COCA), to investigate in what ways and to what extent the young Korean learners of KELC use English modal verbs. The results reveal that the young Korean learners underused modals *could*, *would*, *should*, *may*, *might* and *must* in general. However, *will* and *can* were used frequently with a statistically significant overuse of *will* occurring after the intermediate level. Furthermore, interlanguage development is observed with nativelike usage frequency of *can* from the earlier stages while occurrences of *should* and *would* emerging at later stages. A closer examination of the observed under and over usages also discloses that the difficulty of combining tense with the modal verbs impedes the learners' preterit form use of the modal verbs. Additionally, the use of epistemic modality is observed much later than the root modality such as possibility, permission, and ability. These findings coincide with the first language modal acquisition order reported in literature (Coates 1983). **(Kyungpook National University)**

Keywords modal verbs, learner corpus, foreign language acquisition, second language acquisition, epistemic modality, root modality

1. Introduction

English modal verbs are nothing but several delimited forms of verbs including *can*, *will*, *may*, *shall*, and *must* as well as their preterit forms such as *could*, *would*, *might*, *should*, and *ought*. Nevertheless, the problems associated with these verbs in the process of second language acquisition are not that simple.

---

* We would like to thank the anonymous reviewers of this journal for their valuable comments.
** First author
*** Corresponding author

Problems occur not only due to grammatical complexity, but also due to semantic fuzziness of the verbs. For example, as suggested by Zadeh (1965) and Coates (1983), modality can be divided into root modality and epistemic modality. Root modality is associated with the semantic concepts of obligation, ability, futurity, and permission whereas epistemic modality is associated with the concepts of possibility and necessity (Jackendoff 1972; Brennan 1993). The problem is that each concept is not unique to a single modal verb. Several modal verbs can be used to express one kind of modality. For instance, *can* occurs in root ability, root permission, and epistemic possibility. However, the fact that *may* also occurs in root permission, and epistemic possibility may confuse learners. Preterit forms of modal verbs may further complicate the problem. The preterit form *could* can occur in past possibility as the past of *can*. Moreover, *could* can also express epistemic possibility and hypothetical possibility. To make matters worse, the distinction of the concepts is not always clear. Therefore, second language learners identify modal verbs as one of the most difficult factors for them to learn and acquire.

Building on Choi (2017), this paper reports on the findings from a corpus-based study of English modal verb use by young Korean learners.[1] According to Hunston (2002), comparing the corpora of non-native speakers and native speakers can identify instances of learners' underuse or overuse of target language factors. These comparisons reveal how and at which points the language learners significantly differ from the native speakers. As for interlanguage development, the observation of interlanguage change was possible since the learner corpus, the KNU English Learner Corpus (KELC), consists of data from six proficiency levels. Analyses of the findings show how the learners deviate from the native speakers' norm in terms of using modal verbs in the earlier beginner levels, and how they become closer to the norm in the later advanced levels. The tendency observed in this study is compared with that evidenced in the literature regarding first language acquisition of English modal

1 The research by Choi (2017) focuses on the comparison of Korean learners' use of English modal verbs and native speakers' use of them without consideration of proficiency levels while limiting the research scope to four modal verbs can, could, may and might. The current study investigates the interlanguage development of young Korean learners' use of ten English modal verbs with a focus on how the ten English modal verb usage differs among the six proficiency levels.

verbs (Coates 1983; Wells 1985; Altman 1984).

The first aim of the current paper is to compare a learner corpus, KELC, with a native speaker corpus, the Corpus of Contemporary American English (COCA), for English modal use to clarify the similarities and differences in the patterns of usage by the young Korean learners compared to the first language learners reported in literature (Wells 1979, 1985; Kuczaj and Daly 1979). The second aim is to capture the developmental characteristics of the Korean learners' usage of English modal verbs. Lastly, this study aims to reveal which semantic factors mainly motivate the young Korean learners' use of the modal verbs. In pursuit of these aims, modal verbs were automatically extracted using the Concordance search function in WordSmith 5.0 from the annotated KELC corpus. KELC was annotated in accordance with the BNC part of speech coding system.

2. Modal verbs in the acquisition literature

The literature dealing with the acquisition of modal verbs reports that the acquisition process is deeply affected by the semantic classification of the verbs. Hofmann (1966) introduced the term root modality to refer to non-epistemic interpretations of modals concerning abilities, permission, and obligation.

(1) a. I *can* lift a car.
b. You *may* leave now.
c. You *must* finish your homework.

The modal verb *can* in (1a) expresses ability which is attributed to the referent of the subject. The verb *may* in (1b) expresses permission by which the listener can leave at the speech time. The verb *must* in (1c) is concerned with imposing an obligation on the listener. These notions are also known as deontic modality.

The On the other hand, the modal verbs such as *must* and *can't* in (2) express the epistemic modality which is concerned with the knowledge speakers have and the conclusion they draw on the basis of that knowledge.

(2) a. John *must* be in his office by now.
b. It *can't* be Jane. She is not in Seoul now.
c. It *may* rain tomorrow.

In (2a) the speaker concludes that *John* is in his office based on certain specific previous knowledge (e.g. the time *John* left his home for the office). Similarly, in (2b) the proposition that the person who rang the bell was not *Jane* came from the logical conclusion considering *Jane's* non-existence in *Seoul*. In (2c) the modal *may* indicates that the speaker made the assumption based on self-assured reasoning process (e.g. it's already quite cloudy at the time of speech).

In the literature, second language acquisition of English modal auxiliaries is seen to follow the first language acquisition pattern. Gibbs (1990) carried out a series of experiments to compare the acquisition of English modality by Panjabi speaking children to the findings in first language acquisition. The results showed that the Panjabi speakers make very few errors in using the modal verbs for the meaning of permission, ability, and possibility. That is, the learners acquired those concepts and used them correctly before two years' exposure in the English-speaking community. In contrast, the Panjabi children exhibited problems with the hypothetical possibility and the epistemic possibility and necessity. One thing to note here is that in this work, the notion of possibility was divided into root possibility and epistemic possibility. Accordingly, the modal verb *can* in (3a), which expresses probability of the realization of the event, belongs to the root modality category.

(3) a. Smoking can damage your lung.
b. What could you do if you had the money?

Hence, the observations in Gibbs's study (1990) indicate that the root possibility as in (3a) is acquired earlier than the epistemic possibility as in (2c) or the hypothetical possibility as in (3b). The results accord with the findings in the first language acquisition literature (Coates 1983). Coates (1988) reported further supporting data collected from 12-year-old and 8-year-old subjects making epistemic or root distinctions; the conclusion was that the acquisition of

the modal system is achieved at a relatively later age. Likewise, Kuczaj's (1977) results suggested that children between 2;6 and 3;6 produce in conversation more utterances with root modals than with epistemic modals when compared to children between the ages of 4;0 and 5;9.

In general, the literature states that first language learners of English even by age 7 still had not acquired the hypothetical or the epistemic possibility (Kuczaj and Daly 1979; Wells 1979, 1985; Perkins 1983). Wells' (1979) study of 60 children stated that by 2;6[2], more than 50% of the children used *can* to convey both ability and permission and used *will* to communicate intention. Furthermore, Wells (1985) reported on a second sample of children revealing that by 3;3, children correctly used all categories of root modality. By contrast, *may* and *might* uses were observed to express possibility by age 3;3 and use of modals conveying certainty emerged much later around 5 years old with epistemic uses of *will* appearing even later than expressions of certainty. The hypothetical modality using *would* or *could* was systematically absent or rarely observed in the sample. Similarly, Perkins (1983) investigated a large corpus consisting of spontaneous conversation among six 12-year-old children and revealed that the children only exhibited full acquisition of the adult modal system, especially the epistemic or hypothetical uses of modals, later in development.

3. Learner corpus: KELC

The current study examines data from KELC which was constructed using the essay writings produced by 232 elementary school students enrolled in the children's English program at Kyungpook National University (KNU) in Daegu, Korea.[3] KELC consisting of 168,526 words from 830 writing files is a collection of eight essay-writing tests that students took during a three-year period. Most of the students had previous exposure in English as a foreign language through

2 The notation refers to years and months.

3 The learner corpus, KELC, was built using the data from the essay writing level tests administered by the KNU language institute staff led by Lee and Bae every three months from year 2006 to 2008.

formal instruction in class at their elementary schools or through private tutoring. After enrolling in the English program at KNU, they received additional 6 hours of English instruction per week. English native speaker instructors conducted the classes.

The carefully designed KELC is a valid learner corpus that complies with Granger's (2013) suggested definition of corpora that avoids including inappropriate data types. Granger's definition of corpora is as follows:[4]

> Computer learner corpora are electronic collections of authentic FL/SL textual data assembled according to explicit design criteria for a particular SLA/FLT purpose. They are encoded in a standardized and homogeneous way and documented as to their origin and provenance (Granger, 2013, pp. 6).

One of the strengths of KELC is that it is subdivided into six sub-corpora, which are classified by the proficiency levels of the young Korean learners of English at the time of testing. KELC contains data from Primary (PRI), Basic (BAS), Pre-intermediate (PIN), Intermediate (INT), Advanced (ADV), and Post-advanced (PAV) levels. With respect to the level decisions, native speakers participated as raters. Two raters independently scored the essay writings for coherence, content, and grammar. Another rater scored spelling and text length. Under the direction of the researchers who conducted the project, the raters were trained, and the rating discrepancies were adjusted through statistical componential scoring and fully crossed rater design (Bae and Lee 2012).

4. Results

The present study utilized a text retrieval software, WordSmith Tools version 5.0, to analyze and examine the text in the corpora (Scott 2008).[5] Table 1 presents the descriptive statistics of KELC. As shown, the total token count for KELC is 168,526 with each sub-corpus consisting of more than 2 thousand words

4 Granger (2013) states that his definition of corpora is based on Sinclair's (1996) definition of corpora. See Granger (2013) for detailed explanations of several key notions of the definition.

5 For more information on WordSmith Tools, refer to http://lexically.net/wordsmith/.

except the primary corpus. The level classification in KELC allowed the researchers to study the developmental characteristics of the learners and capture pattern changes in errors correlated with the different proficiency levels.

Table 1. Descriptive statistics of KELC data

Proficiency Level	N of Distinct Words	N of Words	N of Files
Primary (PRI)	289	1,435	23
Basic (BAS)	1,427	24,012	200
Pre-Intermediate (PIN)	2,025	39,454	214
Intermediate (INT)	2,273	41,128	180
Advanced (ADV)	2,263	31,210	113
Post-Advanced (PAV)	2,479	31,287	100
Total	10,756	168,526	830

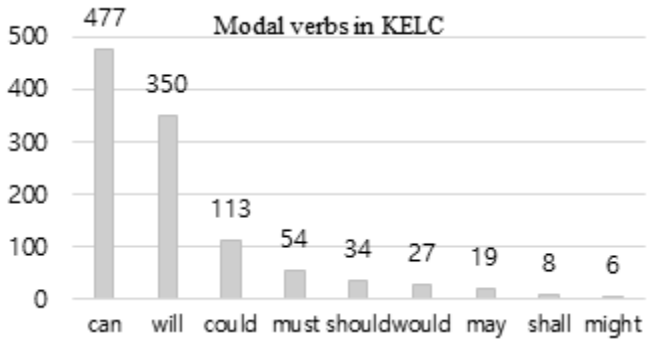

Figure 1. Observed frequencies of models used by young Korean learners of KELC

From the total 168,526 tokens in KELC, WordSmith Tools found 1,088 tokens of modal verbs. Figure 1 shows the number of tokens for each of the identified modal verbs used by the Korean learners. Figure 1 clearly illustrates the usage hierarchy of the modal verbs in KELC. Among the 9 modal verbs, *can* with 477 occurrences is used most frequently. *Will* with 350 occurrences and *could* with 113 occurrences follow *can* in terms of high frequency of use. Conversely, *might* and *shall* are rarely used by the young Korean learners occurring only 6 and 8 times, respectively.

4.1 Overuse and underuse

To uncover the characteristics of the learners' interlanguage, it is necessary to reveal the non-nativelike features in the interlanguage (Granger 1998). Comparing the interlanguage with the native speaker language is one way to achieve the non-nativelike usage distinctions. Correspondingly, the learners' overused and underused modals can be identified via comparison with the native speakers' use of the modal verbs. Table 2 presents the comparison results of the total occurrences of modals in KELC and in the written sections of COCA. The normalized counts as well as the raw frequency of the modal verbs are provided in the table. The normalized counts show that the native speakers in COCA used the modal verbs 1,058 times while the Korean learners in KELC used them 650 times.

Table 2. Comparison of total occurrences of modals: NNS (KELC) vs. NS (COCA) corpora

	KELC	COCA
Observed Frequency	1,088	4,493,985
Normalized Frequency*	650	1,058

*Occurrences are normalized per 100,000 words and rounded to the nearest single digits

When the detection rates of real words and nonce words were compared, in the unviable Dunning (1993) suggested that the comparison of corpora can be effectively carried out through the log-likelihood test. Besides, Rayson and Garside (2000) and Yae (2015) reported that using the log-likelihood test for corpora comparison research is more reliable than the Pearson's chi-squared test. They also stated that the chi-squared value becomes unreliable when dealing with very low and very high frequency words as well as when comparing a relatively small corpus to a much larger one. Correspondingly, log-likelihood calculation is appropriate for the current study investigating high frequency words, modals, by comparing a relatively small learner corpus, KELC, with a much larger native speaker corpus, COCA. Log-likelihood tests were conducted for estimation of underuse and overuse of the total modal verbs and the individual modal verbs. The results for the total number of modal verbs used

showed that the young Korean learners significantly underused modal verbs compared to the native speakers of English (log-likelihood=307.4, Bayes factor=287.53).[6] Table 3 presents each modal verb underused or overused by the Korean learners compared to the native speakers.

Table 3. Comparison of each modal: NNS (KELC) vs. NS (COCA) corpora

	Observed Frequencies		Normalized Frequencies*		Over/Under-Use	
	KELC	COCA	KELC	COCA	Log-Likelihood	Bayes Factor
Can	477	1,103,476	283	260	+ 3.3	-16.5
Will	350	737,279	208	174	+ 10.5	-9.3
Could	113	663,022	67	156	- 109.4	89.5
Must	54	187,160	32	44	- 6.1	-13.7
Should	34	303,104	20	71	- 86.7	66.9
Would	27	905,351	16	213	- 525.1	505.2
May	19	339,874	11	80	- 157.4	137.5
Shall	8	16,358	5	4	+ 0.3	-19.5
Might	6	225,147	4	53	- 134.4	114.5
Ought	0	13,214	0	3	- 10.5	- 9.4

* Occurrences are normalized per 100,000 words and rounded to the nearest single digits.
** The log-likelihood value is always a positive number. The UCREL log-likelihood wizard by Rayson inserts '+' for overuse and '-' for underuse of corpus 1 (KELC) relative to corpus 2 (COCA).

The modals in Table 3 are listed in order from the most to the least observed occurrences in KELC. The Korean learners' modal verb usage hierarchy is as follows: *can* > *will* > *could* > *must* > *should* > *would* > *may* > *shall* > *might* > *ought*. Although *ought* is included in the hierarchical order, not even a single occurrence was observed in KELC. On the other hand, the native speakers used *can* most frequently with *would*, *will*, and *could* following in order.

Figure 2 displays the normalized frequencies of the modal verbs in KELC and COCA with the log-likelihood values for each comparison. From Figure 2, the overuse to underuse hierarchy of the modal verb usage by the Korean learners in comparison with the native speakers are identified as follows: *will* >

6 When the log-likelihood value is higher than 10, the difference is significant at the level of $p<0.001$. If the Bayes factor is higher than 10, it indicates a very strong evidence against the null hypothesis.

can > *shall* > *must* > *ought* > *should* > *could* > *might* > *may* > *would*. In general, *will* is overused and *should*, *could*, *might*, *may*, and *would* are significantly underused while usage frequencies of *can*, *shall*, and *must* do not significantly differ in comparison to the native speakers' use of the modal verbs.

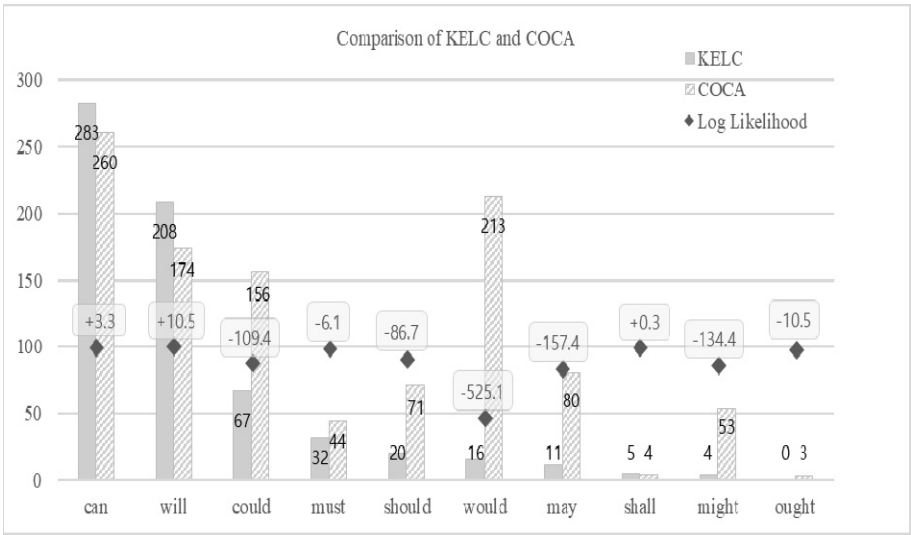

Figure 2. Comparison of modal verb occurrences in KELC and COCA normalized per 100,000 words

When the occurrences of the modal verbs from the two corpora were compared by the log-likelihood test, the analysis revealed that the Korean learners significantly underused *should*, *could*, *might*, *may*, and *would* (log-likelihood values are 86.7, 109.4, 134.4, 157.4, and 525.1, respectively). In contrast, the learners significantly overused only one modal, *will* (log-likelihood=10.5). Given that the young Korean learners generally underused the modal verbs, the overuse of *will* is noteworthy. The Korean learners in KELC also tended to use *can* slightly more than the native speakers in COCA did (log-likelihood=3.3). Furthermore, the results reveal a notable difference concerning the use of the preterit modal verbs; compared to the native speakers, the young Korean learners significantly underused the preterit modal verbs. Thus, occurrences of *could*, *would, should,* and *might* were rarer in KELC than in COCA.

4.2 Use of modals by proficiency levels

As presented in Table 4, the young Korean learners of KELC used *can* earlier than any other modal with 8 observances even at the primary level.

Table 4. Observed modal frequencies at each developmental stage of KELC

	PRI	BAS	PIN	INT	ADV	PAD	Total
Token (Total)	1,435	24,012	39,454	41,128	31,210	31,287	168,526
Can	8	31	99	136	108	95	477
Will	0	11	65	117	70	87	350
Could	0	1	5	25	20	62	113
Must	0	0	10	18	15	11	54
Should	0	1	6	4	6	17	34
Would	0	0	1	4	4	18	27
May	0	1	4	6	3	5	19
Shall	0	1	0	0	7	0	8
Might	0	0	0	3	1	2	6
Ought	0	0	0	0	0	0	0
Sum (Modals)	8	46	190	313	234	297	1088

Note: The labels refer to proficiency levels of the young Korean learners of KELC. PRI: Primary, BAS: Basic, PIN: Pre-intermediate, INT: Intermediate, ADV: Advanced, PAV: Post-advance

As Table 5 represents the normalized frequencies of the observed modal verb occurrences in KELC and COCA. As seen in Table 5, the normalized count per 100,000 words for *can* use at the primary level of 557 is much higher than the normalized counts of the other levels or that of COCA. Since the total number of tokens at the primary level was too small (1,435 tokens) to draw any conclusive results, the high normalized value does not indicate that the learners actually use *can* more frequently in the earlier levels. However, the findings clearly show that the learners used *can* as frequently as the native speakers did from the earlier stages.

Table 5. Normalized modal frequencies for each developmental stage of KELC and COCA (normalized per 100,000 words)

	KELC							COCA
	PRI	BAS	PIN	INT	ADV	PAD	Mean	
Can	557	129	251	331	346	304	283	260
Will	0	46	165	284	224	278	208	174
Could	0	4	13	61	64	198	67	156
Must	0	0	25	44	48	35	32	44
Should	0	4	15	10	19	54	20	71
Would	0	0	3	10	13	58	16	213
May	0	4	10	15	10	16	11	80
Shall	0	4	0	0	22	0	5	4
Might	0	0	0	7	3	6	4	53
Ought	0	0	0	0	0	0	0	3

Note: The labels refer to proficiency levels of the young Korean learners of KELC. PRI: Primary, BAS: Basic, PIN: Pre-intermediate, INT: Intermediate, ADV: Advanced, PAV: Post-advance

Additionally, an examination of the normalized values for *will* at each developmental stage of KELC indicates that the use of *will* by the Korean learners did not stay static throughout the different levels. The learners rarely used *will* at the primary or basic level. Nonetheless, at the pre-intermediate level, the observed frequency of *will* was 65, which calculates to 165 when normalized per 100,000 words. The comparison of the normalized value of 165 at the pre-intermediate level in KELC with the normalized value of 174 for *will* in COCA reveals that at the pre-intermediate level, the learners used *will* almost as frequently as the native speakers. From the intermediate level to the post-advanced level, the learners used *will* more frequently than the native speakers in COCA. As a whole, the overuse of the modal *will* is not very prominent, but the underuse of the modal in the earlier two proficiency levels clearly indicate the overuse of the modal *will* at the more advanced levels.

In contrast, *would*, which is considered as the preterit form of *will*, is rarely used throughout the levels. The total observed occurrences of *would* is only 27 and when normalized per 100,000 words, the value is 16. Given that the native speakers' normalized value is 213, it is certain that *would* is the most underused of the modal verbs investigated. Like *would*, *could* and *may* were also underused by the Korean learners. However, in the case of *could*, the tendency to underuse

suddenly disappears at the post advanced level. Table 5 shows that at the post-advanced level, the normalized value is 198 which is more than three times the normalized values of the intermediate or advanced level, 61 and 64, respectively.

4.3 Use of modals by advanced learners

The learners at the post-advanced level of KELC did not reach the level of mastering English. However, they rarely made mistakes regarding subject verb agreement and frequently used the past tense forms of verbs. They also quite often used complex sentences including complement clauses and relative clauses. Thus, it will be interesting to examine how the relatively advanced learners used the modal verbs.

Table 6. Use of modal verbs at the post-advanced level

	Observed Frequencies		Normalized Frequencies*		Log Likelihood**	Bayes Factor
	KELC (PAD)	COCA	KELC (PAD)	COCA		
Can	95	1103476	304	260	+ 2.2	-17.7
Will	87	737279	278	174	+ 16.6	-3.3
Could	62	663022	198	156	+ 3.2	-16.6
Would	18	905351	58	213	- 50.3	30.4
Should	17	303104	54	71	- 1.4	-18.5
Must	11	187160	35	44	- 0.6	-19.3
May	5	339874	16	80	- 24.0	4.1
Might	2	225147	6	53	- 20.7	0.9
Shall	0	16358	0	4	- 2.4	-17.5
Ought	0	13214	0	3	- 1.9	-17.9

* Occurrences are normalized per 100,000 words and rounded to the nearest single digits.
** The log-likelihood value is always a positive number. The UCREL log-likelihood wizard by Rayson inserts '+' for overuse and '-' for underuse of corpus 1 (KELC) relative to corpus 2 (COCA).

As shown in Table 6, the learners at the post-advanced level significantly overused *will* (log-likelihood=16.6). They overused *will* since the intermediate level and kept overusing it in the advanced and post-advanced levels. Similarly, at the post-advanced level, the learners overused *can* and *could*. However, the log likelihood test did not strongly support the overuse of these two modal verbs

(Log likelihood=2.2 and 3.2, respectively). The results indicate that the learners' use of these two modal verbs was not significantly different from the native speakers in COCA. One thing to note is that *could* was being underused before the post-advanced level. The occurrences of *could* increased approximately three times in the post-advanced level compared to the occurrences in the intermediate or advanced levels. In addition, the normalized value at the post-advanced level in KELC slightly exceeds that in COCA. Consequently, the acquisition of *could* is considered a recent development for the advanced learners.

On the other hand, the learners underused *would, shall, should, may, might*, and *must* at the post-advance level. Among them, the log likelihood test indicates that *would*, *may*, and *might* were significantly underused. Remember that Table 3 confirmed that 5 modal verbs (i.e. *could, would, should*, *may*, and *might*) were significantly underused when looking at the whole KELC corpus without any consideration of the proficiency levels.[7] Accordingly, the frequencies of *should* and *must* evidently increased, and their values are no longer significantly different from the values in a native speaker corpus (i.e. COCA).

5. Analyses and Discussion

The corpus research conducted in this study shows that the Korean learners exhibit nativelike usage frequencies of *can* and *will* from the earlier stages of acquisition. In addition, although they have some delay in acquiring *could, should* and *must*, they use them quite well at the later stages. However, the learners underused the other modal verbs even at the quite advanced stages. The issue that needs to be considered is whether this result accords with the observations in the literature, namely, the acquisition process/sequence reported in Gibbs (1990) and other previous studies.

5.1 *Can* and *could*

The learners in KELC used *can* as frequently as did the native speakers in

7 The modal verb *shall* was not used at all. Thus, the log likelihood test did not statistically support the underuse. However, it is certain that the learners were still underusing the modal at the level.

COCA even from the very beginning stages. This is in accord with the observation in Gibbs' study (1990). The modal verb *can* expresses ability, permission, and root possibility. As also suggested by Coates (1983), Gibbs (1990) classified the modal verb *can* as a modal delivering root possibility different from the modal verb *may*. This is because the modal verb *can* does not contain the meaning of inference to the extent that the modal verb *may* does. Since root possibility, ability, and permission are acquired in the earlier stages of acquisition, the learners in KELC were able to use the modal verb *can* frequently even at the primary level.

Another reason for the frequent use of the modal verb *can* can be found from the misuse illustrated in (4).

(4) After I wroted all, I *can* sleep well. (KELC: Post-Advanced)

In general, the learners at the post-advanced level used tense quite consistently. However, some of them still had difficulty combining the information of tense with the modal verbs. As a result, some learners used *can* when *could* or *was able to* was necessary. It is natural that this tendency of wrong simplification becomes more prevalent when the proficiency levels are lower. In fact, there was only one case where *BE able to* was used in KELC. Therefore, we suggest that the simplification misuse made by the learners explains the overuse of *can*, in addition to the prevalence of the root modality.

Overall, the learners used *could* less frequently than native speakers. However, such a tendency dramatically changed at the post-advanced level. The reason for the underuse may be attributed to the difficulty of using tense for modal verbs at the earlier levels. However, the learners already frequently used *can* to express root possibility, ability, and permission. Therefore, when they were capable of combining the tense information with the modal verbs, they were able to use *could* frequently for the same reasons as the modal verb *can*. This development may explain the sharp increase of the use of *could* at the post-advanced level.

The learners' misuse of *can* also explains the overuse of *could* as illustrated in (5).

(5) a. I *could* cross the river without a boat, but I didn't even go out. (Hoffman 1993)

b. So both of them *could* eat. And they went to home. (KELC: Post-Advanced)

When *could* is used to deliver the meaning of past possibility, it does not suggest that the event described by the verb phrase comes true, viz., the action is not likely carried out as illustrated in (5a). Therefore, when the event is actually completed, *was able to* is required instead. In sum, the frequent use of *could* resulted from the learners' misuse which was attributable to the rare use of *was able to*, as well as the wide demand of the past root possibility.

5.2 *Will* and *would*

In general, the modal verb *will* expresses futurity, prediction, or generic modality. The young Korean learners of this study mostly used *will* for futurity and prediction as given in (6).context.

(6) a. So I *will* introduce my family. (KELC: Advanced)
b. If the car comes, then you *will* have to go to hospital. (KELC: Advanced)

For the learners in KELC, *will* was the second most frequently used modal following *can*. The young Korean learners used *will* as frequently as the native speakers at the pre-intermediate level and used it more than the native speakers from the intermediate level. This overall overuse may be partially due to the learners' simplification strategy. Just as in the overuse of *can*, the learners may have used *will* even in the circumstances where the native speakers would use the present tense, the progressive aspect, or other modal verbs such as *may* or *would*. This is observed from the KELC data in (7).

(7) a. Teacher told we *will* go to garden. (KELC: Intermediate)
b. I want we *will* go to the beach one more time. (KELC: Pre-Intermediate)

However, we also suggest that the sudden increase of the use of *will* at the

intermediate level needs to be ascribed to the necessity of the use of futurity rather than any other modality. This assumption can be supported by the use of *BE going to* by the learners. The learners in KELC used *BE going to* 56 times. In other words, when the native speakers use *BE going to* 15 times in writing, the learners use it 34 times. Consequently, both the overuse of *will* and that of *BE going to* can be explained by the necessity of frequent use and the subsequent early adoption of futurity by the learners.[8] Additionally, the modal *will* was used most frequently in the intermediate level while the usage slightly decreased in frequency in the advanced and the post-advanced levels. This decrease may be explained by the adoption of other ways of expressing futurity. For example, the learners used *BE going to* in these levels 34 times. In comparison, *BE going to* was used once while *will* was used 5 times. Before these levels, *BE going to* occurred much less. Specifically, the lower level learners used *BE going to* once when they use *will* 25 times.

The modal verb *would* expresses past generic modality and is used as an indirect form in reported speech. It is also used in hypothetical conditional sentences. Examples of *would* usage from KELC are shown in (8a) and (8b).

(8) a. *Would* you be happy if I cut your hair? (KELC: Post-Advanced)
b. My father said we *would* find him. (KELC: Post-Advanced)

Among the 18 tokens of *would* used by the post-advanced learners in KELC, 3 cases conveyed the meaning of hypothetical possibility as shown in (8a). Otherwise, *would* is used as a kind of indirect form in reported speech as shown in (8b). In our corpus investigation, *would* is the most significantly underused modal. We first suggest that it is attributed to the impossibility of the simple combination between futurity and past tense. Second, the past generic modality and indirect past form are not as prevalent as the root possibility or futurity. Finally, the rarity of *would* may also be explained by the late acquisition of the hypothetical conditional use as suggested in Gibbs (1990).

---

8 This overuse can also be explained by the distinction between the written and spoken dichotomy. The native speakers in COCA rarely used BE going to in formal written documents. However, the young learners do not in general clearly distinguish the writing from the speech forms.

5.3 *Should* and *must*

The modal verb *should* is used as an indirect past form of *shall* or used as a deontic modal expressing obligation. In KELC, *should* was used mostly to express the deontic obligation or necessity as illustrated in (9a) and to describe past event as in (9b).

(9) a. Now I 'm too tired and I think I *should* go a sleep. (KELC: Post-Advanced)
b. When the teacher made the groups, the group members *should* finish the wonderful picture of the park. (KELC: Post-Advanced)

The modal verb *must* expresses epistemic necessity or deontic obligation. For the post-advanced learners in KELC, among the 11 tokens of *must*, 10 cases were used as a deontic modal as in (10a). There was only one case where *must* was used as an epistemic modal as shown in (10b). The rare use of epistemic modality in KELC is in accord with the observation in Gibbs (1990).

(10) a. There's only three chairs in there, so we *must* bring another chair. (KELC: Post-Advanced)
b. In restaurant or any food corner. It *must* have fast food corner. (KELC: Post-Advanced)

Although the learners underused modals in general, the use of *should* and *must* were not significantly infrequent. This may be due to the wide demand of deontic modality and necessity in writing. Notably, the learners used *must* more than *should* in contrast with the frequency of the two verb uses in COCA. This may be due to the difficulty combining tense with the modality. As a result, the use of *should* tends to be concentrated in the post-advanced level whereas that of *must* is quite consistent across all the levels except the first two lowest levels.

6. Conclusion

The The young Korean learners of English observed by the current learner corpus investigation underused modal auxiliaries in general. Predominantly, they underused *could, would, should, may, might* and *must*. Midst this poverty of modals, *will* and *can* were used frequently. Particularly, the overuse of *will* was statistically significant (log-likelihood=10.5) and that tendency became conspicuous after the intermediate level.

The findings reveal that the young Korean leaners of KELC used the modal *can* as frequently as the native speakers did from the earlier stages. In addition, the learners are able to use *can* and *will* to express possibility, permission, and futurity from the early stages of acquisition. The results also show that the learners rarely used *must* or *should* as epistemic modals. In most cases, they used them to express deontic obligation. These findings indicate that the second language acquisition of modal verbs by young Korean learners of this study also follows the order of first language acquisition, with the earlier emergence of root modality functions such as ability, permission, and possibility and the later acquisition of epistemic and hypothetical possibility (Coates, 1983). Moreover, the observed underuse of preterit forms of modals seems to be affected by the difficulty of combining tense with modality. Hence, the learners' use of preterit forms of modals takes place at the later stages of development, mostly at the post-advanced level. Consequently, the use of *could* remarkably increased at the post-advanced level. However, if a type of a modal verb is rarely required linguistically and is acquired with difficulty, the modal may not be observed even at the higher stages of development. This explains why the learners used *would* so rarely in KELC.

Finally, we suggest that the learners' strategy plays a role in the skewed frequency of modal verbs. That is, the learners tend to express certain types of modality with a one-word modal verb rather than a two or three-word verb phrase. Although much of this simplification process leads to wrong choices of modals semantically, the literature reports that such process is not only existent, but also helps learners construct grammatical generalization in the early stages of learning (Goldberg 2006). This type of learning tendency explains why the learners overused *will* and *can* excessively.

References

Altman, Roann. 1984. *Assessing modal proficiency in English as a second language*. PhD Dissertation. University of Southern California.

Bae, Jungok and Yae-Sheik Lee. 2012. Evaluating the development of children's writing ability in an EFL context. *Language Assessment Quarterly* 9: 348-374.

Boersma, Paul and David Weenink. 2012. *Praat: Doing phonetics by computer* (version 5.3.23)

Brennan, Virginia. 1993. *Root and epistemic modal auxiliary verbs*. PhD Dissertation. University of Massachusetts, Amherst.

Choi, Incheol. 2017. How do Korean EFL learners acquire the English modal auxiliaries *can*, *could, may,* and *might*?: A learner corpus study. *Studies in Modern Grammar* 92: 233-248.

Coates, Jennifer. 1983. *The semantics of the modal auxiliaries.* London: Croom Helm.

Gibbs, Dorothy. 1990. Acquisition of English modal auxiliaries can, could, may, and might. *Applied Linguistics* 11: 297-314.

Goldberg, Adele. 2006. *Constructions at work: The nature of generalization in language*. New York: Oxford University Press.

Granger, Sylviane. 1998. The computer learner corpus: A versatile new source of data for SLA research. In Sylviane Granger (ed.), *Learner English on computer,* 3-18. London: Longman.

Granger, Sylviane. 2002. A bird's eye view of learner corpus research. In Sylviane Granger, Joseph Hung, and Stephanie Petch-Tyson (eds.), *Computer learner corpora, second language acquisition and foreign language teaching*, 3-33. Philadelphia, PA: John Benjamins.

Granger, Sylviane. 2003: The international corpus of learner English: A new resource for foreign language learning and teaching and second language acquisition. *TESOL Quarterly* 37: 538-546.

Hoffmann, Ronald. 1966. Past tense replacement and the modal system. In Anthony Oettinger (ed.), *Mathematical linguistics and automatic translation*. Cambridge, MA: Harvard University, Harvard Computational Laboratory. Report NSF-17.

Hofmann, Thomas. 1993. *Realms of meaning: An introduction to semantics*. London and New York: Longman.

Hunston, Susan. 2002: *Corpora in applied linguistics*. Cambridge, MA: Cambridge University Press.

Jackendoff, Ray. 1972. *Semantic interpretation in generative grammar*. Cambridge, MA: MIT Press.

Kuczaj, Stan and Mary Daly. 1979. The development of hypothetical reference in the speech of young children. *Journal of Child Language* 6: 563-579.

Kuczaj, Stan. 1977. The acquisition of regular and irregular past tense forms. *Journal of*

Verbal Learning & Verbal Behavior 16: 589-600.
Perkins, Michael. 1983. *Modal expressions in English.* New York: Frances Pinter.
Rayson, Paul and Roger Garside. 2000. Comparing corpora using frequency profiling. In *Proceedings of the Workshop on Comparing Corpora, held in conjunction with the 38th annual meeting of the Association for Computational Linguistics (ACL 2000)*, 1-6.
Scott, Mike. 2008. *WordSmith tools*. Birmingham: Lexical Analysis Software.
Wells, Gordon. 1979. Learning and using the auxiliary verb in English. In Victor Lee (ed.), *Language development.* London: Croom Helm.
Wells, Gordon. 1985. *Language development in the pre-school years*. Cambridge, MA: Cambridge University Press.
Yae, Sunhee. 2015. Grammaticalization of the marginal modal verb *ought to*: A copus-based approach. *Linguistic Research* 32: 773-793.
Zadeh, Lotfi A. 1965. Fuzzy sets. *Information and Control* 8: 338-353.

Daria Soon-Young Seog
Assistant Professor
Dept. of English Education, Teachers College
Kyungpook National University
80, Daehak-ro, Buk-gu,
Daegu, 41566, Rep. of Korea
E-mail: dariadks@knu.ac.kr

Incheol Choi
Professor
Dept. of English Education, Teachers College
Kyungpook National University
80, Daehak-ro, Buk-gu,
Daegu, 41566, Rep. of Korea
E-mail: incheol@knu.ac.kr

Received: 2018. 05. 08.
Revised: 2018. 07. 06.
Accepted: 2018. 07. 06.