



Korean EFL Learners' Underuse of the English Perfect: A Collostructional Analysis of Written Learner Corpora*

Kang San Noh (Korea University) **Sanghoun Song** (Korea University)



This is an open-access article distributed under the terms of the Creative Commons License, which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

Received: March 24, 2023

Revised: April 13, 2023

Accepted: May 15, 2023

Kang San Noh (1st author)
Graduate Student,
Dept. of Linguistics, Korea Univ.
Tel: +82-2-3290-1648
Email: kasan1998@korea.ac.kr

Sanghoun Song
(corresponding author)
Associate Professor,
Dept. of Linguistics, Korea Univ.
Tel: +82-2-3290-2177
Email: sanghoun@korea.ac.kr

*This work was supported by the Korea University Insung Research Grant.

ABSTRACT

Noh, Kang San and Sanghoun Song. 2023. Korean EFL learners' underuse of the English perfect: A collostructional analysis of written learner. *Korean Journal of English Language and Linguistics* 23, 342-357.

This study examines whether and how Korean EFL learners underuse verbal lexemes as the English perfect using a series of corpus-based methods involving two written learner corpora, which amount to 3,590,194 words in total. It is difficult to detect which forms are significantly underused by language learners because the underused patterns are such that the learners avoid producing them in the data and are therefore less visible, unlike the overused patterns. Accordingly, the current work performs collostructional analysis to measure the association strength between the perfect construction and the specific verbal lexemes, thereby demonstrating that collostructional analysis works effectively to reveal the underused patterns in the language learners' data. The collostructional analysis reveals that 416 verbal lexemes were underused while only 4 verbal lexemes were overused. It is proposed that this underuse behavior of Korean EFL learners might result from the lack of grammatical markers for the perfect in Korean and the use of temporal adverbials. Along this line of research, this paper proposes the effectiveness of collostructional analysis in measuring associations and potential pedagogical implications.

KEYWORDS

association strength, collostructional analysis, English perfect, learner corpora, underuse

1. Introduction

This study aims to detect Korean EFL learners' underuse of verbal lexemes as the English perfect. The underuse phenomenon refers to the avoidance of specific linguistic structures, especially when such structures are considered difficult by L2 learners (Laufer 2000; Smith 1982). As there are no constructions or systems exactly corresponding to the English perfect in Korean, a series of studies have indicated that the English perfect is generally difficult for Korean EFL learners to use (Yoon 2000; Lee and Kang 2013; Shin and Chung 2013; Choi 2018). That is, they are expected to avoid the English perfect and opt for another form (e.g., the simple past), as illustrated in (1) from the Gachon Learner Corpus (GLC).

- (1) a. So, I saw many car accident since I was in here. (GLC)
 b. Because violence by teenagers has intensified since they have been exposed to lots of violent movies. (GLC)

The previous studies on Korean EFL learners' acquisition and use of the perfect imply that they are likely to prefer (1a) over (1b), underusing sentences such as the latter, in which verbal lexemes are used as the perfect. However, the previous studies have focused on case studies of Korean EFL learners specifically selected for each experiment. For a more empirically macroscopic analysis, the present study makes use of the corpus data on a large scale, analyzing the two Korean learner corpora and the Corpus of Contemporary American English (COCA) as a reference. Specifically, the combination of the two Korean corpora amounts to 3,590,194 words in total, and the selected excerpts from the COCA consist of 36,643,094 words.

In addition, unlike the previous studies, the present study uses a new method to detect the underuse phenomenon. The problem that L2 researchers face when they study the underuse phenomenon is that what is being underused may not be transparent. While overuse can be detected quantitatively with relatively high raw frequencies, underuse is more difficult to detect. Taking such difficulties into consideration, the present study aims to compare L2 learners' use of the perfect with native-like use by means of collostructional analysis, which can provide the strength of association between a construction and a word. The target structure in question is the English perfect, which is a complex morpho-syntactic construction consisting of “*have*” followed by a past participle. This study takes into consideration the finite forms of the English perfect (i.e., the present perfect, the past perfect, and the future perfect). We expect to address the nonfinite forms in future research.

2. Theoretical Background

2.1 Underuse Phenomenon

In the field of second-language acquisition, distributional differences in language use—namely overuse and underuse—reflect the degree of use of specific expressions. The overuse phenomenon occurs when the frequency of specific linguistic items is higher than a reference point. In contrast, the underuse phenomenon occurs when specific linguistic items are avoided (Smith 1982). More specifically, the latter refers to L2 learners' decision to choose one form with which they feel safer for delivering the intended meaning over another (Laufer 2000). For instance, Dagut and Laufer (1985) refer to a case in which Hebrew EFL learners underuse English phrasal verbs.

(2) We didn't believe that John could never _____ his friends. (Dagut and Laufer 1985)

Hebrew EFL learners were asked to fill in the blank in (2) using one of the given four verbs (*let down*, *solve*, *disappoint*, *carry on*). Either the single-word verb '*disappoint*' or the phrasal verb '*let down*' was the right answer, and the other two were distractors. Hebrew EFL learners explicitly preferred '*disappoint*' over '*let down*', implying their preference for single-word verbs over phrasal verbs. However, underuse does not simply result from L2 learners' mere ignorance of the target structures but implies that L2 learners have difficulty putting the avoided structures into use; that is, while L2 learners have knowledge of the avoided structures, they may choose not to use them, replacing them with others that feel more familiar.

Traditionally, two strands of studies have focused on the underuse phenomenon. On the one hand, it is claimed that L2 learners commit avoidance due to the structural differences between L1 and L2 (Kleinmann 1977). Kleinmann reports that Hebrew-speaking students of English avoided English phrasal verbs since Hebrew lacks a structure equivalent to English phrasal verbs. On the other hand, it is also argued that underuse can result from structural or semantic identity between L1 and L2; for example, Dutch ESL learners avoided idiomatic English phrasal verbs that highly resembled the corresponding Dutch phrasal verbs (Hulstijn and Marchena 1989). The present study takes the approach that L2 learners' underuse results from the structural differences between their L1 and the target language (L2).

As it has been pointed out that L1 could influence L2 learners' acquisition of grammatical morphemes (Murakami 2013), it is also likely that Korean EFL learners' acquisition of the perfect is influenced by Korean. The fact that Korean (L1) lacks the structure equivalent to the perfect in English (L2) is expected to contribute to Korean EFL learners' underuse behavior. In regard to their underuse behavior, a series of studies have focused on the underuse of specific constructions such as verb-particle constructions (Kweon 2006; Sung 2020). Korean EFL learners' underuse of the perfect was also considered by several studies (Yoon 2000; Lee and Kang 2013; Shin and Chung 2013; Choi 2018), but they were focused on experimental tests or pedagogical implications rather than on underuse behavior itself. With this theoretical background in mind, this study focuses on Korean EFL learners' underuse of verbal lexemes as the perfect by taking a macroscopic analysis of the two Korean EFL written learner corpora.

2.2 The Perfect

2.2.1. Definition of the perfect

The perfect is a complex morpho-syntactic construction comprising "*have*" followed by a past participle (Ritz 2012). A series of arguments have been made on whether the perfect is a tense, an aspect, or both. Those who argue that the perfect is closer to a tense claim that the interpretation of "completion" or "result" is not intrinsic to the perfect itself (McCoard 1978). In contrast, Huddleston (1988) argues that the perfect falls under the category of an aspect: even though it resembles the past tense in terms of "completion," the perfect focuses on the current or resultant state. Thus, whether the perfect is a tense or an aspect depends on how one defines the meaning of an aspect and how one thinks of the perfect semantically. Following Huddleston (1988), this study considers the perfect as an aspect. Specifically, the perfect is seen as a grammatical marker for expressing temporality.

2.2.2. Present perfect, past perfect, and future perfect

Jespersen (1924) claims that the English present perfect is more concerned with the present results of past events rather than the past events themselves. Let us consider the following instance given by Jespersen (1924):

(3) Now, I have eaten enough. (Jespersen 1924)

In (3), the present adverbial '*now*' must be applied to the present result and not to the past event. That is, '*now*' is applied to the present moment of having finished the action of eating rather than to the action of eating, which took place in the past. Thus, Jespersen (1924) concludes that the present perfect is distinct from the preterite (the simple past tense).

Ritz (2012) argues that the past and future perfect are distinguished from the present perfect in that the latter is a canonical perfect; specifically, the past and future perfect can modify the event by using a past temporal adverbial as in (4).

(4) a. Dean had/will have arrived in Sydney the day before we left/leave for Paris/on the first of December. (Ritz 2012)

b. Dean had/will have got up at seven. After that, he had/will have eaten breakfast, then had/will have taken the bus to go to his office....

In (4), the past temporal adverbial '*on the first of December*' and '*at seven*' can be used in either the past perfect or the future perfect. However, if a past temporal adverbial were to be used in the present perfect, it would result in ungrammatical sentences, as in (5).

(5) a. *Martin has left two days ago. (Ritz 2012)

b. *Martin has left on the first of December.

In (5a), the past temporal '*two days ago*' is not compatible with the present perfect. The past temporal adverbial '*on the first of December*', which was used in (4a), results in ungrammatical sentence (5b); thus, (4) and (5) show an asymmetry between the past and future perfect on the one hand and the present perfect on the other.

2.2.3. Constructionist approach to the perfect

Binnick (1991) takes a compositional approach toward the perfect and states that the perfect seems to be compatible with every tense inflection. Klein (1992) argues that the meanings of the perfect system are compositionally derived. From the standpoint of this compositional approach, the tense of the auxiliary is combined with the participle, which denotes the anteriority relation. However, Michaelis (1993) argues that the three perfect types—namely, the present perfect, past perfect, and future perfect—represent distinct grammatical constructions. For instance, the present perfect and the past perfect are distinguished in that the former is subject to some parochial constraints, as in (6).

(6) a. Our committee chair has (?? angrily) tendered his resignation. (Michaelis 1993)

b. Our committee chair has angrily tendered his resignation every time we have asked him to take a

controversial stand on something.

- c. [Pushed to take a stance], our committee chair had angrily tendered his resignation, [and we were hard pressed to find a replacement].

As seen from (6a), the manner adverb '*angrily*' is not permitted in the case of the resultative present perfect. Conversely, the existential perfect is compatible with the manner adverb as in (6b); that is, the constraint regarding the exclusion of the manner adverb is characteristic of the resultative present perfect. (6c) shows that the past perfect expression delivering the same interpretation is not limited by such a constraint.

Taking this kind of asymmetry into consideration, the present study takes a constructionist approach to the English perfect. The traditional compositional approach claims that what differentiates the present perfect from the past perfect is the temporal reference point, that is, the tense of the auxiliary combines with the anteriority relation, which the following participle denotes. If the present tense combines with the anteriority relation, then the corresponding interpretation will be the present perfect. In the same way, the past tense combines with the anteriority relation to yield the past perfect interpretation. However, if the perfect is simply the combination of the tense of the auxiliary and the anteriority relation, then why does the past perfect permit the application of the manner adverb, while the present perfect does not do so? This kind of asymmetry between the present perfect and the past and future perfect indicates that the compositional account of the perfect is not adequate. In sum, the English perfect is now viewed as the English perfect construction. As will be discussed in Section 3.3, taking a constructionist approach to the perfect makes it possible to analyze it by performing collocation analysis.

3. Present Study

The main research question is whether collocation analysis can show underuse of the perfect aspect. We will address this question by comparing the use of the perfect by Korean EFL learners with that of native English speakers. Then, if Korean EFL learners underuse certain verbal lexemes as the English perfect, we will discuss linguistic factors that could potentially result in their underuse behavior.

3.1 Learner Corpus

The present study used both a Korean EFL learner corpus and a reference corpus on a large scale, as macroscopic analysis can be effective for detecting the overuse or underuse phenomenon. For instance, Zhang (2014) extracts the overused and underused English concluding connectives by Chinese EFL learners using the BNC (British National Corpus) as a reference corpus and SWECCL (Spoken and Written English Corpus of Chinese Learners: 4,560,029 words) as a learner corpus. In a similar way, we used two Korean EFL learners' corpora, which amounted to 3,590,194 words in total: the Gachon Learner Corpus (GLC, Carlstrom and Price 2012) and the Yonsei English Learner Corpus (YELC, Rhee and Jung 2012). The GLC consists of 2,507,899 words, and the YELC consists of 1,082,295 words. Both the GLC and YELC consist of the written data produced by Korean students in their undergraduate program. In regard to the GLC, 2,500 Korean EFL learners at Gachon University were given 20 questions and asked to write a response that ranged from 100 to 150 words. The 20 selected questions concerned personal experiences or opinions (e.g., *Have you ever done something that was really thrilling or frightening? What did you do?*) and their writings were collected via Google Forms. The YELC consists of essays written by first-year undergraduate students at Yonsei University. The collected essays were the results of the diagnostic tests

for measuring their English proficiency.

For the L1 data, which are geared toward the comparison of distribution, we used the Corpus of Contemporary American English (COCA, Davies 2008) as the source of a reference corpus. Specifically, we extracted excerpts from COCA consisting of the following five genres from 1990 to 2014: “spoken,” “fiction,” “magazine,” “newspaper,” and “academic.” In sum, 125 sections were extracted (i.e., the abovementioned 5 genres * 25 years), and 12 sections were randomly selected. The selected excerpts amounted to 36,643,094 words in total, which is approximately 10 times as large as the two Korean EFL learner corpora combined.¹

3.2 Corpus Annotation Process

The present study duplicates the corpus annotation process by Song and Oh (2021). The annotation process mainly consists of two parts: the first step is automatic annotation, and the following step is manual correction performed by human annotators. For the process of automatic annotation, Full Forest Treebanking (Packard 2015) was used to generate a large number of syntactic trees. Since the generation of a large set of syntactic trees allows the extraction of potentially meaningful linguistic information, Full Forest Treebanking served as the basis of the annotation process.

The overall process of automatic annotation was conducted using English Resource Grammar (ERG, Flickinger 2000), which is an open-source grammar implementation for English based on Head-driven Phrase Structure Grammar (HPSG). The automatic analyzer identified the finite forms of the perfect (*have/has/had* + a past participle), but it may not work perfectly. Thus, the process of automatic annotation was followed by manual correction by human annotators, for which an online workbench was used as the basis. Ten human annotators (graduate and undergraduate Korean students) conducted manual correction and cross-checking. Each data point was cross-checked by four human annotators to minimize potential errors that could deteriorate the process of annotation.²

3.3 Collostructional Analysis

From the standpoint of construction grammar (CxG), all levels of grammatical analysis inevitably involve constructions (Goldberg 2006). Based theoretically on construction grammar, collostructional analysis is an extension of collocational analysis that targets a particular construction and shows which lexemes are either attracted or repelled in it (Stefanowitsch and Gries 2003). These lexemes are called *collexemes* (i.e., occur more frequently or less frequently), while the target construction is called a *collostruct*. The combination of the two is a *collostruction*. Collostructional analysis is applied to extract the collexemes of a collostruct; doing so provides the strength of association or the *collostructional strength* of the collexemes. We perform collexeme analysis, one of the three kinds of collostructional analysis (the others being distinctive collexeme analysis and covarying collexeme analysis), which measures the degree of attraction/repulsion of a lexeme to a slot in the target construction (Stefanowitsch and Gries 2003).

Collostructional strength can be calculated by using the following measures: the *t* test, *z* test, G-test, and Fisher's exact test. Among them, Fisher's exact test is the most commonly used, as it is computed on the basis of the

¹ Rather than using the whole COCA, we used the selected excerpts from COCA as it is not always better to use extremely large reference corpora. For more information concerning the size of reference corpora, see Sardinha (2000).

² For more information concerning the procedure of annotation/retrieval, see Flickinger and Yu (2013) and Suppes et al. (2014).

hypergeometric distribution (Gries and Ellis 2015), and it is not based on any distributional assumptions, such as normality. For this reason, even though Fisher's exact test requires a highly intensive computational procedure, it is performed for the sake of precision. Collocation analysis requires a two-by-two contingency table of a particular lexeme, as shown in Table 1.

Table 1. The Two-by-Two Contingency Table of *avoid*

	The Perfect	¬ The Perfect	Sum
Learners	2	2,362	2,364
COCA	90	987	1,077
Sum	92	3,349	3,441

Table 1 shows the distribution of '*avoid*' in perfect construction. Using a two-by-two contingency table and the programming language R, Fisher's exact test was performed (R Core Team 2019).³ The calculation process quantifies the degree of attraction/repulsion or the degree of association, and as a result, it indicates in which of the two cases (the learner corpora or the English native corpora [COCA]) is more underused as the perfect. The following code is used to perform Fisher's exact test. `fisher.test(testor<-rbind(c(2, 2362), c(90, 987)), alternative = "less")$p.value` Using this R script, the collocation strength of '*avoid*' can be calculated, as shown in Table 2.

Table 2. The Collocation Strength of *avoid*

Lexeme	p.value (Collocation Strength)
<i>avoid</i>	$5.87 * 10^{-44}$

In collocation analysis, the p value based on the contingency table represents the degree of association or the collocation strength (Stefanowitsch and Gries 2003). In Table 2, the collocation strength is $5.87 * 10^{-44}$. Following this process, we calculated the ratio of the perfect in two cases: the Korean EFL learners (i.e., GLC and YELC) and the native English speakers (i.e., the COCA excerpts). The results show that the ratio is higher in the latter case, indicating that '*avoid*' is repelled or underused when it is used as the perfect by Korean EFL learners. Repeating the computing process of the collocation strength of each lexeme, Section 4 provides the list of the collexemes most strongly attracted to or repelled by the English perfect construction by Korean EFL learners.

³ All statistics reported in this study were computed with the 2019 version of the R package.

4. Results

4.1 Distribution of the Perfect

From the two annotated Korean EFL learner corpora (i.e., GLC and YELC), all the cases of the three perfect constructions are counted, and the distribution within the two corpora is provided (Table 3).

Table 3. The Distribution of Perfect Constructions in Korean EFL Learner Corpora

	Present Perfect	Past Perfect	Future Perfect	Sum
YELC	1,470 (80.64%)	349 (19.14%)	4 (0.22%)	1,823
GLC	5,632 (88.61%)	710 (11.17%)	14 (0.22%)	6,356
Sum	7,102	1,059	18	8,179

Following the constructionist approach to the perfect (Michaelis 1993), the present study measured the proportions of each discrete perfect construction. In YELC, the proportion of the present perfect amounts to 80.64%, which is roughly four times higher than that of the past perfect (19.14%). In GLC, the proportion of the present perfect is 88.61%, which is nearly eight times higher than that of the past perfect (11.17%). In both GLC and YELC, the proportion of the future perfect was marginal, being 0.22% in both cases.

4.2 Collexeme Analysis

The following is the result of the collexeme analysis, which produced both lists of overused and underused collexemes. The results show that only 4 collexemes were attracted to the perfect. In contrast, 416 collexemes turned out to be repelled, indicating the underuse of the perfect construction.

4.2.1. The attracted collexemes

Table 4. List of Attracted Collexemes

Rank	Lexeme (n)	p.value (Collostructional Strength)
1	see (1447)	$1.87 * 10^{-23}$
2	worry (100)	$1.92 * 10^{-07}$
5	report (15)	$9.77 * 10^{-05}$
7	ride (84)	0.003816542

Table 4 lists the collexemes that turned out to be attracted to the perfect. The numbers in parentheses refer to the raw frequency of the verbal lexemes in their perfect form. From 20 collexemes that turned out to be attracted, those that occurred fewer than 10 times were filtered out, leaving four collexemes in Table 4. Regarding why certain verbal lexemes were overused as perfect, it could be that Korean EFL learners were induced to overuse specific verbal lexemes. Since both GLC and YELC consist of essays written as an answer to a series of questions, it is likely that certain verbal lexemes within the given questions were frequently used. For instance, when the following question from the GLC “*Do you ever worry about using the internet? Why or why not?*” is provided, learners are likely to use the verbal lexeme ‘worry’ in their answers, eventually resulting in its overuse.

4.2.2. The repelled collexemes

The results show that 416 collexemes were repelled or underused as perfect in GLC and YELC. For simplicity, Table 5 presents the top 30 repelled collexemes.

Table 5. List of the Most Repelled Collexemes

Rank	Lexeme (n)	p.value (Collostructional Strength)
1	make (120)	0
2	have (387)	4.41 * 10 ⁻¹⁸⁸
3	get (157)	2.96 * 10 ⁻¹⁷⁴
4	give (70)	2.56 * 10 ⁻¹³²
5	change (137)	7.65 * 10 ⁻¹²⁰
6	use (233)	8.59 * 10 ⁻¹²⁰
7	become (156)	6.97 * 10 ⁻¹⁰⁶
8	do (373)	8.53 * 10 ⁻⁹¹
9	go (229)	7.67 * 10 ⁻⁸⁶
10	learn (120)	5.39 * 10 ⁻⁶⁷
11	lose (29)	8.70 * 10 ⁻⁶⁷
12	think (94)	1.99 * 10 ⁻⁶⁶
13	take (101)	4.79 * 10 ⁻⁶¹
14	help (10)	3.59 * 10 ⁻⁵⁵
15	allow (14)	3.79 * 10 ⁻⁵⁵
16	choose (14)	3.61 * 10 ⁻⁵⁰
17	find (44)	3.93 * 10 ⁻⁴⁹
18	study (50)	7.26 * 10 ⁻⁴⁹
19	cause (15)	6.07 * 10 ⁻⁴⁸
20	develop (69)	8.11 * 10 ⁻⁴⁸
21	ban (11)	6.39 * 10 ⁻⁴⁷
22	spend (27)	2.84 * 10 ⁻⁴⁵
23	avoid (2)	5.87 * 10 ⁻⁴⁴
24	grow (75)	1.07 * 10 ⁻⁴³
25	come (50)	3.29 * 10 ⁻⁴³
26	improve (20)	2.10 * 10 ⁻³⁹
27	affect (19)	2.48 * 10 ⁻³⁹
28	talk (12)	4.29 * 10 ⁻³⁹
29	agree (7)	8.60 * 10 ⁻³⁹
30	struggle (6)	3.52 * 10 ⁻³⁸

Table 5 shows that it is difficult to provide a generalization that can classify the top 30 repelled collexemes as a single group with semantic coherence. For instance, while 'make', 'get', and 'give' are typically used in the ditransitive construction (Goldberg 2006), the remaining 27 collexemes are not particularly associated with the ditransitive construction. Even when one uses another criterion such as the lexical aspect (Vendler 1957), it still seems difficult to group all the repelled collexemes as a whole. While some of the underused collexemes ('have', 'think', and 'agree') could be classified as stative verbs, other collexemes could be accomplishment verbs ('develop' and 'grow') or achievement verbs ('find' and 'ban'). Considering that the aspect is a relatively abstract

construction (Stefanowitsch and Gries 2003), there is no strong reason to expect that the perfect will be related to specific types of collexemes. Thus, while it is difficult to say that all the underused verbal lexemes are semantically congruent, it can be said that various verbal lexemes were underused as the perfect regardless of their semantic classes.

4.3 Tupleization

Applying collostructional analysis instead of traditional collocational analyses can be effective for the detection of the underuse phenomenon. In contrast to traditional collocational analyses, which mainly focus on the relationship between words, collostructional analysis aims to determine the association strength between a construction and a word. Collostructional analysis is also distinguished from previous collocational analysis, which focuses on the raw frequencies of occurrence. Raw frequencies can be ineffective in measuring the association strength of collocates; for instance, in Zipfian distributions, where a small number of function words cover a large proportion of the data, the relative significance of content words is marginalized. This can be a nuisance for those who want to take subtle semantic characteristics into account (Manning and Schütze 2000).

To separate the effect of raw frequencies from the measurement of association, the process of *tupleization*—a multidimensional method that keeps relevant corpus-linguistic dimensions of information separate—can be implemented (Gries 2019). Relevant corpus-linguistic dimensions of information refer to a series of characteristics such as frequency, the strength of association/contingency, and dispersion. Thus, association measures such as Fisher's exact test (including other measures such as the G-test or *t* test) need to be compared to each word's tuple {pure frequency, pure association} (Gries 2022). Specifically, the two tuples in which the two variables (e.g., frequency and association/strength) are separated from each other are as follows: if the overused and underused collexemes are shown simultaneously, then separating the strength of association from raw frequencies can result in a clear contrast between the overused and underused collexemes.

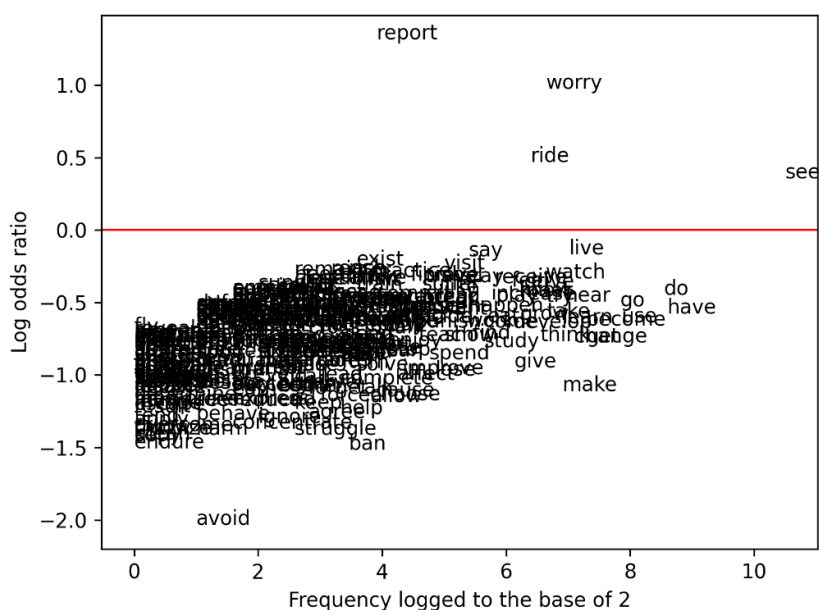


Figure 1. 2-tuples: Separating Frequency (x-axis) and Association (y-axis)

Figure 1 represents the separation of association from raw frequencies in both overused and underused collexemes. What is provided in Figure 1 is congruent with the results of collexeme analysis. It shows that only 4 verbal lexemes ('see', 'worry', 'report', and 'ride') were overused as the perfect. In contrast, 416 verbal lexemes were underused, resulting in a striking contrast. While all the attracted collexemes are located above 0 in terms of the y-axis, the repelled ones are located below 0; that is, the attracted collexemes have a meaningfully higher degree of association compared to the repelled collexemes. Compared to the y-axis, no significant disparity is found between the scope of the x-axis of the overused and underused collexemes. Considering that the x-axis represents the raw frequencies of the overused and underused collexemes, this implies that if raw frequency alone had been considered, it would have been difficult to detect the strength of association; accordingly, the detection of the underuse phenomenon would have been infeasible. However, collostructional analysis and the process of tuplezation allow us to measure the association strength, thus paving the way for the detection of the underuse phenomenon.

5. Discussion

5.1 The Perfect Aspect in Korean

It has been argued that Korean lacks an aspect system corresponding to the English perfect. From this traditional perspective, Korean EFL learners' underuse of the English perfect is understandable. As the perfect aspect does not exist in Korean, Korean EFL learners are highly likely to have difficulties learning or using the English perfect. In contrast to the common claim that the perfect aspect does not exist in Korean, several studies have pointed out that Korean does have the perfect aspect (Yang 2017; Kim 1999). For instance, the following Korean sentences are cited as an example of the perfect aspect.

(8) a. Changswu-nun cikum-kkaci cemsim-ul mek-ess-ta.

Changswu-Nom now-until lunch-Acc eat-Pst-Decl

'Changswu has eaten lunch until now.'

b. Seoul-hayng KTX-ka pangkum tte-nass-ta.

Seoul-bound for KTX-Nom just now leave-Pst-Decl

'KTX bound for Seoul has left just now.'

(Yang 2017)

As seen in (8), the past temporal marker 'ess/ass' is used to denote the perfect aspect. This makes the perfect in Korean less apparent compared to its counterpart in English. That is, Korean lacks distinctive grammatical markers for the perfect, which could make it difficult to distinguish the perfect from the simple past, as in (9).

(9) Jihye-nun cip-ey kass-ta.

Jihye-Nom home-to go-Pst-Decl

'Jihye went (has gone) home.'

(Yang 2017)

The case in (9) shows that the same sentence can be simultaneously interpreted as the past tense and the perfect.

Yang (2017) points out that the lack of distinctive grammatical markers for the perfect seems to have misled previous researchers into the conclusion that the perfect aspect does not exist in Korean.

What we suggest here is that Korean EFL learners' acquisition of the English perfect could be difficult regardless of whether one accepts the existence of the perfect in Korean or not. If one accepts the traditional claim, Korean EFL learners are likely to face difficulties acquiring the perfect, which is entirely a new concept. On the other hand, if one accepts the claim that Korean does have the perfect aspect, it is likely that they will have a hard time acquiring the English perfect due to the structural differences between the two languages. That is, the lack of distinctive grammatical markers for the perfect could result in Korean EFL learners' underuse of the English perfect. As the perfect in Korean lacks its own grammatical markers and is not clearly distinguishable from the past tense, Korean EFL learners could have difficulties understanding or using the English perfect, which has its own grammatical markers. That is, Korean EFL learners might end up using the past tense in certain situations where native English speakers would have typically used the perfect.

Furthermore, temporal adverbials could be an additional factor for Korean EFL learners' underuse of the English perfect. A series of studies are supportive of the argument that grammatical markers for tense and aspect can be redundant or unnecessary for L2 learners. It has been pointed out that the L2 expression of temporal reference develops when a reference is established by solely using temporal adverbials (Meisel 1987; Bardovi-Harling 1992). If L2 learners make use of such adverbials, this could deter their application of grammatical markers for tense and aspect. Similarly, Schumann (1987) notes that L2 temporal reference can be made by solely using temporal adverbials (adverbs such as *now*, *always*, or prepositional phrases), a serialization of events, a calendric reference (dates), or an implicit temporal reference that can be inferred from a specific context. Thus, L2 learners' acquisition of grammatical markers for tense and aspect can be blocked by the availability of other means of delivering temporality.

The same process can be observed in Korean EFL learners' acquisition of the English perfect. Korean EFL learners' avoidance of the perfect becomes more prominent in regard to the perfect of persistent situations and the perfect of the recent past. Consider the following cases in (10) and (11), from COCA and GLC/YELC, respectively:

- (10) a. He has spent nearly **25 years** studying whales. (COCA 1990 SPOK)
 b. I think I have **just** found my new best friend. (COCA 2001 NEWS)
 (11) a. He is an ethic teacher and spent **three years** with us. (YELC)
 b. And I **just** found a solution. (GLC)

(10a) and (11a) both involve the persistence of time. However, while the meaning of persistence is conveyed by the perfect in (10a), the very same meaning of persistence is conveyed by the simple past tense in (11a). (10b) and (11b) both convey the same message of recency, which is implied by the adverb *just*. Nonetheless, (10b) and (11b) are different in that only (10b) delivers the message of recency with the perfect. What (10) and (11) show is that Korean EFL learners in GLC and YELC prioritize delivering the semantic/pragmatic side of the perfect. Making full use of the morphological/grammatical side of the perfect may be unnecessary since some of the meanings of the perfect can be alternatively conveyed by temporal expressions, as in (10b) and (11b). This can result in Korean EFL learners' underuse of verbal lexemes as the perfect.

5.2 Pedagogical Implications

As discussed in Section 5.1, Korean does not have distinctive grammatical markers for the perfect, which makes

the distinction between the past tense and the perfect less prominent. Taking this into consideration, we suggest the explicit teaching of the context in which the English perfect is used. By providing EFL learners with exemplary usages and contexts, it is expected that their understanding of the target construction will be improved (Gilquin 2015). For instance, Park and Choe (2013) point out that Korean EFL learners properly used the present perfect when they were provided with explicit lexical cues (the context in which the present perfect is recommended or should be used). Thus, it is expected that more usage-based teaching materials that do not end up simply providing abstract grammatical rules about the perfect could be helpful for the acquisition of the perfect. We believe that the underuse by Korean EFL learners in the present study shows the need to provide them with appropriate usages and contexts of the perfect.

6. Conclusion

This study performed collocation analysis to measure the degree of association between the English perfect and specific verbal lexemes. This is because collocation analysis can function as an effective method of detecting the underuse phenomenon. By separating raw frequency on the one hand and the association as an effect size on the other, the impact of association/contingency can be considered. In contrast to simply considering the raw frequencies of collexemes, contrasting the degree of association between collexemes makes the detection of the underuse phenomenon possible. Collocation analysis using two Korean EFL learner corpora (GLC and YELC) shows that Korean EFL learners are inclined to underuse specific verbal lexemes as perfect. Specifically, 416 verbal lexemes were found to be underused as perfect in GLC and YELC; in contrast, only 4 verbal lexemes were overused as perfect by Korean EFL learners. We claim that such a disparity between the number of overused verbal lexemes and that of underused lexemes indicates the underuse of the English perfect by Korean EFL learners. The analysis of the overused verbal lexemes shows that it is likely that Korean EFL learners were induced to overuse specific words due to the questions given to them. On the other hand, the analysis of the underused verbal lexemes reveals that Korean EFL learners underused various kinds of verbs as perfect.

Below, we discuss the perfect as a grammatical marker and its impact on Korean EFL learners and the implications of this study. Concerning the perfect's use as a grammatical marker, it is posited that L2 learners can find such grammatical markers redundant or insignificant because Korean lacks the distinctive grammatical markers for the perfect and alternative options (i.e., temporal adverbials) are available. As a result, Korean EFL learners could be induced to generally underuse the English perfect.

In future studies, Korean EFL learner corpora other than GLC and YELC need to be considered for a more rigorous generalization regarding Korean EFL learners' underuse of the English perfect construction. In addition, potential criticisms might require more detailed analyses of the corpus data. In our defense, the present study mainly focused on the application and analyses of learner corpus data on a large scale, which differentiates it from previous studies on the underuse phenomenon. Although a detailed typological or semantic analyses of the underused collexemes may be needed, such processes are expected to be conducted in future studies. While our study needs further clarifications or modifications, we expect this paper to show that applying learner corpus data to the field of L2 studies has potential benefits. At the same time, we hope that this paper can serve as an exemplar for those who want to apply collocation analysis to research on the underuse phenomenon.

References

- Bardovi-Harlig, K. 1992. The use of adverbials and natural order in the development of temporal expression. *International Review of Applied Linguistics in Language Teaching* 30(4), 299–320.
- Binnick, R. 1991. *Time and the Verb*. Oxford, UK: Oxford University Press.
- Carlstrom, B and P. Nathan. 2012. *The Gachon Learner Corpus*. Retrieved from <http://koreanlearnercorpusblog.blogspot.kr/p/corpus.html>
- Choi, J. 2018. Use of the present perfect by Korean learners of English: A descriptive case study. *Foreign Languages Education* 25(3), 23–51.
- Dagut, M. and B. Laufer. 1985. Avoidance of phrasal verbs – A case for contrastive analysis. *Studies in Second Language Acquisition* 7(1), 73–79.
- Davies, M. 2008. The Corpus of Contemporary American English (COCA): 560 million words, 1990-Present. Retrieved from <https://corpus.byu.edu/coca/>.
- Flickinger, D. 2000. On building a more efficient grammar by exploiting types. *Natural Language Engineering* 6(1), 15–28.
- Flickinger, D. and J. Yu. 2013. Toward more precision in correction of grammatical errors. Proceedings of the Seventeenth Conference on Computational Natural Language Learning: Shared Task, 68-73.
- Gilquin, G. 2015. The use of phrasal verbs by French-speaking EFL learners. A constructional and collocation-based approach. *Corpus Linguistics and Linguistic Theory* 11(1), 51-88.
- Goldberg, A. E. 2006. *Constructions at Work: The Nature of Generalization in Language*. Oxford, UK: Oxford University Press.
- Gries, S. T. 2019. 15 years of collocations: Some long overdue additions/corrections (To/of actually all sorts of corpus-linguistics measures). *International Journal of Corpus Linguistics* 24(3), 385–412.
- Gries, S. T. 2022. What so (some of) our association measures measure (most)? Association? *Journal of Second Language Studies* 5(1), 1–33.
- Gries, S. T. and N. C. Ellis. 2015. Statistical measures for usage-based linguistics. *Language Learning* 65(Supplement 1), 1–28.
- Huddleston, R. 1988. *English Grammar: An Outline*. Cambridge, UK: Cambridge University Press.
- Hulstijn, J. H. and E. Marchena. 1989. Avoidance: grammatical or semantic causes? *Studies in Second Language Acquisition* 11(3), 241–255.
- Jespersen, O. 1924. *The Philosophy of Grammar* (Reprint). New York: Norton.
- Kim, C. G. 1999. *Tense Structure & Aspect Recognition of Korean*. Seoul: Taehaksa.
- Klein, W. 1992. The present-perfect puzzle. *Language* 68(3), 525–552.
- Kleinmann, H. H. 1977. Avoidance behavior in adult second language acquisition. *Language Learning* 27(1), 93–107.
- Kweon, S. O. 2006. Avoidance of phrasal verbs by Korean EFL learners. *Studies in Modern Grammar* 44, 175-197.
- Laufer, B. 2000. Avoidance of idioms in a second language: The effect of L1-L2 degree of similarity. *Studia Linguistica* 54(2), 186–196.
- Lee, J. and M. K. Kang. 2013. An English tense and aspect comprehension for effective communicative improvement focused on present perfect aspect and past tense. *The Korean Association of Language Studies* 18(1), 163–180.
- Manning, C. and H. Schutze. 2000. *Foundations of Statistical Natural Language Processing*. Cambridge, MA:

MIT Press.

- McCoard, R. W. 1978. *The English Perfect: Tense Choice and Pragmatic Inferences*. North-Holland: Amsterdam.
- Meisel, J. 1987. Reference to past events and actions in the development of natural second language acquisition. In B. MacWhinney, ed., *First and Second Language Acquisition Processes*, 206-224. New York, NY: Newbury House.
- Michaelis, L. A. 1993. *Toward a Grammar of Aspect: The Case of the English Perfect Construction*. Doctoral dissertation, University of California at Berkeley, CA, USA.
- Murakami, A. 2013. *Individual variation and the role of L1 in the L2 development of English grammatical morphemes: Insights from learner corpora*. Doctoral dissertation, University of Cambridge, Cambridge, United Kingdom.
- Packard, W. 2015. *Full forest treebanking*. Master's thesis, University of Washington, WA, USA.
- Park, M. and M. H. Choe. 2013. Advanced Korean EFL learners' use of the English simple past and present perfect in controlled contexts. *The Linguistic Association of Korea Journal* 21(4), 119-145.
- R Core Team, 2019. *R: A language and environment for statistical computing*. R Foundation for Statistical Computing: Vienna.
- Rhee, S. C. and C. K. Jung. 2012. Yonsei English Learner Corpus (YELC). *Proceedings of the First Yonsei English Corpus Symposium*, 26-36.
- Ritz, M. E. 2012. Perfect tense and aspect. In R. I. Binnick, ed., *The Oxford Handbook of Tense and Aspect*, 881-907. Oxford, UK: Oxford University Press.
- Sardinha, T. B. 2000. Comparing corpora with WordSmith Tools: How large must the reference corpus be? *Proceedings of the Workshop on Comparing Corpora*, 7-13.
- Schumann, J. H. 1987. The expression of temporality in Basilang speech. *Studies in Second Language Acquisition* 9(1), 21-41.
- Shin, H. S. and S. Y. Chung. 2013. A study of English teaching methods on present perfect tense through comparison with past tense. *Journal of British & American Language and Literature* 29(1), 49-69.
- Smith, K. L. 1982. Avoidance, overuse, and misuse: Three trial and error learning strategies of second language learners. *Hispania* 65(4), 605-609.
- Song, S. and E. Oh. 2021. A multimethod approach to Korean speakers' acquisition of unaccusativity in English. *Journal of Cognitive Science* 22(1), 135-186.
- Stefanowitsch, A. and S. T. Gries. 2003. Collocations: Investigating the interaction of words and constructions. *International Journal of Corpus Linguistics* 8(2), 209-243.
- Sung, M. C. 2020. Underuse of English verb-particle constructions in an L2 learner corpus: Focus on structural patterns and one-word preference. *Corpus Linguistics and Linguistic Theory* 16(1), 189-214.
- Suppes, P., T. Liang, E. E. Macken and D. P. Flickinger. 2014. Positive technological and negative pre-test-score effects in a four-year assessment of low socioeconomic status k-8 student learning in computer-based math and language arts courses. *Computers & Education* 71, 23-32.
- Vendler, Z. 1957. Verbs and times. *The Philosophical Review* 66(2), 143-160.
- Yang, Y. J. 2017. A syntactic & semantic analysis on perfective aspect of English & Korean. *The Mirae Journal of English Language and Literature* 22(2), 107-127.
- Yoon, W. J. 2000. A study on the function and use of English perfect tense in discourse. *Journal on English Language Teaching* 12(1), 295-314.
- Zhang, R. 2014. Overuse and underuse of English concluding connectives: A corpus study. *Journal of Language Teaching and Research* 5(1), 121-126.

Examples in: English

Applicable Languages: English

Applicable Level: Tertiary