

## ***A Corpus-based Study of To-Infinitive Errors in Korean College Freshmen's Writing***

**Jung Eun Kim**

*Soongmoon High School, Korea*

**Isaiah WonHo Yoo**

*Sogang University, Korea*

This study aims to investigate how accurately *to*-infinitives are used by Korean learners of English who have completed primary and secondary education in Korea. Data were retrieved from a learner corpus consisting of essays written by 815 incoming freshmen at a university in Seoul, Korea. Of the 2,309 tokens of *to*-infinitives from the corpus, only 171 tokens were found to be errors, a finding which suggests that Korean students with high-intermediate proficiency in English possess a firm understanding of the use of *to*-infinitives. In addition to those 171 error tokens, 65 tokens of omission errors were manually identified; of these 236 error tokens, substitution was the most common type of error (103 tokens, 43.6%), which was closely followed by omission errors (99 tokens, 42.0%). A close examination of all the errors revealed that several types of errors derive from the learner's lack of knowledge regarding subcategorization, i.e., information about selecting proper complementation types. Teachers should therefore provide L2 learners with detailed subcategorization information, especially when new verbs are introduced.

**Keywords:** written learner corpus, error analysis, *to*-infinitive, Korean college students

## Introduction

According to the English curriculum released by the Korean Ministry of Education (2011), the goal of English as a school subject is to help students acquire English as a communication tool. In order to help students and teachers of English accomplish that goal, the Korean Ministry of Education (2011) presented a list of English expressions and sentence structures that should be taught in secondary school (pp. 42-54 and pp. 91-103). On this predetermined list of language items, the *to*-infinitive figures prominently as it is frequently used in many of the English expressions and sentence structures to be taught. This should not come as a surprise since the most frequent words in English are in fact function words, which specify grammatical relations with little or no lexical meanings (O'Keeffe, McCarthy, & Carter, 2007, pp. 33-37). A quick corpus frequency list of the Brown Corpus and that of the Frown Corpus generated by *MonoConc Pro 2.2*, a commercially available concordancer, indeed show that the most frequently used word in these two corpora are the definite article, which is followed by *of*, *and*, *to*, and *a* in this exact order in both the corpora.

The complexity of the deceptively simple-looking *to*-infinitive is well known. Most grammar books, as well as English textbooks, published in Korea distinguish three different uses of *to*-infinitives, i.e., nominal, adjectival, and adverbial uses. Most grammars published in the U.S. and Britain, on the other hand, explain in detail the various functions that the *to*-infinitive serves in sentences. For example, Quirk, Greenbaum, Leech, and Svartvik (1985, pp. 1061-1062) discuss the nominal functions of the *to*-infinitive as the subject, direct object, subject complement, appositive, adjectival complementation, and prepositional complement, while Biber, Johansson, Leech, Conrad, and Finegan (1999, pp. 693-783) present a more elaborate scheme for the functions of *to*-infinitives: post-predicate infinitive clauses controlled by verbs, verbs taking extraposed *to*-clauses, subject noun phrases and subject predicative *to*-clauses, and *to*-clauses controlled by adjectives.

In addition to this inherent complexity of the various functions of *to*-infinitives, the fact that the *to*-infinitive looks the same as the preposition *to* poses further difficulty to learners of English. According to Choi and Yoo's (2012, p. 325) corpus-based study on *V-ing* forms, many of the errors that they discovered are rooted in the fact that the students erroneously used the gerund because they mistook the infinitival marker *to* for the preposition *to*, e.g., (1):

- (1) It's too hard to **\*remodeling** our school.

A further complication arises from the fact that a number of verbs can take either a gerund or a *to*-infinitive as complements, which results in little or no difference in meaning for some verbs such as *begin*, *start*, and *continue*, e.g., (2), in which both the sentences mean the same (Cowan, 2008, p. 506):

- (2) a. She continued arguing her client's case.  
b. She continued to argue her client's case.

For other verbs such as *remember*, *forget*, and *try*, however, taking a *to*-infinitive complement does result in a difference in meaning, e.g., (3), in which (3a) means that "he did the task and then remembered doing it" and (3b) means that "he remembered he had the task and then did it" (Cowan, 2008, p. 508):

- (3) a. He remembered mailing it.  
b. He remembered to mail it.

Because of this inseparable relationship between a *to*-infinitive and a gerund, most research studies regarding *to*-infinitives have centered on its comparison with gerund complementation (e.g., Lee & Choe, 2013) or have focused on its syntactic properties (e.g., Hyde, 2000), while relatively a few studies have looked at L2 learners' ability to use *to*-infinitives properly based

on learners' actual writing samples. In an attempt to fill this gap in the current literature, this study aims to investigate how accurately *to*-infinitives are used by Korean L2 learners of English who have completed primary and secondary education in Korea and to identify the frequently occurring error types regarding *to*-infinitives.

In order to achieve these objectives, the present study will analyze relevant data retrieved from a learner corpus consisting of essays written by incoming freshmen at a university in Korea and, by doing so, will address the following two research questions:

- 1) How accurately do learners of English who have completed primary and secondary education in Korea use *to*-infinitives?
- 2) What are the types of errors involving *to*-infinitives and the possible causes of such errors?

Descriptively accurate answers to these questions drawn from actual data produced by students may provide both the learners and the teachers of English with linguistic and pedagogical applications that they can utilize.

## Literature Review

There have been attempts to discuss the use of *to*-infinitives based on a corpus-driven analysis, especially with regard to written texts. Goh and Kim (2009), for example, investigated the uses of *to*-infinitives in English newspapers published in Korea and compared them with those of *to*-infinitives in American and British newspapers. To analyze the use of *to*-infinitives in Korean English newspapers, they compiled the Korean English Newspaper Corpus (the KENC), which consists of approximately 100,000 words from the editorials of three leading English newspapers in Korea, i.e., *The Korea Herald*, *The Korea Times*, and *JoongAng Daily*. Goh and Kim compared the patterns of *to*-infinitives used in the KENC with those of the

Frown Corpus (Freiburg-Brown Corpus) and the FLOB Corpus (Freiburg-LOB Corpus), which are based on written texts of American English and British English, respectively.

According to Goh and Kim (2009), English newspapers published in Korea have a tendency to overuse *to*-infinitive structures in comparison with American and British newspapers, a result which was found to be statistically significant. One of the reasons for this tendency is that Korean English newspapers used “marginal auxiliaries” such as *have to* and *need to* (Quirk et al., 1985, p. 236), which contain the infinitival marker *to*, much more frequently than did American and British newspapers. Another reason a *to*-infinitive has a higher likelihood of appearing in Korean English newspapers than in American and British newspapers is that Korean English newspapers used extraposition much more frequently (Goh & Kim, 2009):

(4) It’s up to you to make an appointment.

In (4), *to make an appointment* has been “generated in subject position and then moved, or extraposed, to the end of the sentence” and the empty subject position has been subsequently filled with the non-referential subject *it* (Celce-Murcia & Larsen-Freeman, 1999, p. 669)

There also have been a few studies that dealt with general patterns of errors found in a learner corpus consisting of Korean university students’ writing samples (e.g., Cha, 2004), as well as those that focused on a specific aspect of the errors found in a learner corpus (e.g., Choi & Yoo, 2012). However, no attempts have been made to identify specific error types regarding *to*-infinitives found in a learner corpus consisting of English produced by students in Korea.

In their study of the *V-ing* forms, i.e., gerunds and present participles, found in Korean university student’s essays, Choi and Yoo (2012, p. 318) found that the students rarely made errors when using *V-ing* forms: only 292 of all the 3,843 tokens of the *V-ing* forms in the students’ essays (fewer than eight percent) were found to be errors, of which 138 were gerund-related.

What is interesting, however, is the fact that the incorrect substitution of a gerund for a *to*-infinitive, e.g., (5), accounted for about a fourth of all the gerund-related errors (34 of the 138 tokens, 24.6%) (Choi & Yoo, 2012, pp. 325-326):

- (5) a. I want **\*talking** about my problem.  
b. I hope **\*making** real friends.

Given the fact that Korean students learn early on in their English education that *want* and *hope* are two of the verbs that can only take a *to*-infinitive complement, it would be interesting to ascertain whether students also make errors in which they substitute a *to*-infinitive for a gerund with verbs such as *enjoy*, *finish*, and *avoid*, which cannot take a *to*-infinitive complement.

With respect to the possible sources of errors involving *to*-infinitive structures, Dulay, Burt, and Krashen (1982) state that because three different forms – i.e., *that*-clauses, *to*-infinitives, and gerunds – can be used as a complement, errors can occur due to the inappropriate selection of one of these complements. Dulay et al. (1982, p. 195) also point out that learners often fail to understand that the “implied subject” of a *to*-infinitive may be omitted only when it is the same as the subject of the main clause, e.g., (6), and sometimes make the mistake of omitting the subject of an infinitive when it is required, e.g., (7):

- (6) We plan to go to New York next week.  
(7) Mother has a lot of work. Daddy expects to stay at her office late.  
(Intended ... Daddy expects her to stay at her office late)

Dulay et al. (1982) serve as a good point of departure for analyzing errors involving *to*-infinitives, as they offer a major classificatory system by which most learner errors can be analyzed, as well as list several possible situations in which *to*-infinitive errors can occur, but they fall short of offering an array

of specific subcategories of errors by which all the errors involving *to*-infinitives can be analyzed. Needless to say, identifying such specific subcategories will require a close examination of an ample amount of data retrieved from a learner corpus.

## Methodology

### Data Collection

In order to evaluate Korean students' ability to use *to*-infinitives properly, this study collected its data from a learner corpus consisting of writing samples produced by incoming freshmen students at a university in Seoul, Korea, in February 2010. The decision to use students' writing samples rather than speaking samples to investigate their ability to use *to*-infinitives was based on the fact that unlike *that*-clauses and *wh*-clauses, both of which are most common in conversation, "infinitive clauses are considerably more common in the written registers than conversation" (Biber et al., 1999, p. 699).

The learner corpus consists of 815 essays written by students majoring in the humanities or engineering, who were asked to write an essay as part of the placement test for their General English requirement. The students majoring in the humanities were asked to write on the following essay topic: "If you could change one important thing about your hometown, what would you change? Use reasons and specific examples to support your answer." The students majoring in engineering were also given a similar topic to write on: "If you could make one important change in a school that you attended, what change would you make? Use reasons and specific examples to support your answer." Both groups were given 50 minutes to write their essays in computer labs on campus, and they were not allowed to use any dictionaries or reference books.

The students who wrote the essays can be assumed to be high-intermediate or advanced learners of English, as the vast majority of them placed at or

above the 98 percentile of the English subject on the Korean College Scholastic Ability Test. The essays were collected by using the *Criterion*® Online Evaluation Service, which offers a variety of essay topics and also has the ability to rate essays using the automatic assessment system *e-rator* (Kim, 2009).

### Procedures

Using the concordancing program *MonoConc Pro 2.2.*, we were able to retrieve all the 2,928 tokens of *to* and 68 tokens of *to* from the learner corpus, which contained 201,382 words. Of these tokens, 667 were instances of the preposition *to*, e.g., (8a), and 20 accounted for the non-interpretable items designated as “Others” in Table 1 below, e.g., (8b), leaving the total number of tokens of *to*-infinitives at 2,309:

- (8) a. There are about 5 **to** 6 schools inside my town.  
 b. When he is teaching, he was often used **\*to** picture and shape.

TABLE 1  
*Distribution of the Tokens of To in the Learner Corpus*

	Tokens	Percentage
<i>To</i> -infinitive	2,309	77.1%
Preposition <i>to</i>	667	22.2%
Others	20	0.7%
Total	2,996	100%

According to Norrish (1983), there are two main approaches to analyzing errors: one is using pre-determined categories, and the other is using post-determined categories in accordance with the materials used for the analysis. The former, which is more common, is to conduct an error analysis based on a set of pre-determined categories of errors that are commonly produced by the learners, and those categories are usually established by previous studies or researchers. The latter, on the other hand, provides a more tailored way of

sorting errors since the errors themselves determine the categories of errors, but using this approach takes more time as it requires multiple sorting and re-sorting processes.

The present study uses post-determined categories in combination with pre-determined categories. Using Dulay, Burt, and Krashen's (1982, p. 150) "surface strategy taxonomy" of learner errors, this study classified all the errors involving the *to*-infinitive into the following four major categories: substitution, omission, addition, and others. Subsequently, all the errors in these major categories have been re-sorted into the subcategories found to be relevant to all the errors involving *to*-infinitives, which are shown in Table 2 below. It should also be noted that some obvious spelling mistakes and grammatical errors irrelevant to the questions at hand have been corrected in the examples provided in this paper so as not to unnecessarily impede the reader's understanding of the examples.

TABLE 2  
*Error Types Involving To-Infinitives*

Error Type	Example
<b>Substitution</b>	
<i>To</i> + gerund	I prefer to <b>*taking</b> a bus.
<i>To</i> + noun	Most of buildings need to <b>*renovation</b> .
Semantically wrong verb	I don't have chance to <b>*hear</b> the professor's great literature lesson.
<i>To</i> -infinitive for gerund	Students feel like <b>*to study</b> .
<i>To</i> -infinitive for prepositional phrase	It is mind to love school that is capable <b>*to change</b> university.
Wrong voice	My thought to make one important change in our school is too ideal to <b>*realize</b> .
<b>Omission</b>	
Leaving out <i>to</i>	I want <b>*(to)</b> read many books.
Leaving out verb after <i>to</i>	I want to <b>*(have)</b> more train stations.

Leaving out infinitival subject	I will not permit <b>*(NP)</b> to build easily.
<b>Addition</b>	
Unnecessary <i>to</i>	I strongly <b>*to</b> say that.
Unnecessary verb after <i>to</i>	We tend to <b>*apply give</b> good score professor.
Unnecessary tense-marking	Everyone has to <b>*had</b> their own part.
<b>Others</b>	
	There are some methods <b>*to down</b> fee.
	Universities seem to <b>*don't</b> care about that.

---

## Results and Discussion

A close examination of the 2,309 tokens of *to*-infinitives retrieved from the learner corpus revealed that only 171 tokens were used erroneously. In addition to those 171 tokens of incorrect uses of *to*-infinitives, 65 other tokens of omission errors, in which *to* was omitted as in (9), were manually identified:

- (9) a. If I could make one important thing, I want **\*(to)** build big camp in Seoul.  
b. I think school must be friendly to students and make students want **\*(to)** go school.

Even with these 65 manually identified omission errors, the errors still account for less than 10 percent of all the *to*-infinitive tokens, as shown in Table 3 below. The finding that over 90 percent of all the tokens of *to*-infinitives were used correctly (2,138 of the 2,374 tokens) does not seem surprising in light of the fact that Choi and Yoo (2012, p. 318) also found that the vast majority of *V-ing* forms were used correctly by the same group of students (3,551 of the 3,843 tokens of *V-ing* forms, 92.4%).

TABLE 3  
*Distribution of Correct and Incorrect Uses of To-Infinitives*

	Tokens	Percentage
Correct	2,138	90.1%
Incorrect	171	7.2%
Omitted <i>to</i>	65	2.7%
Total	2,374	100%

All the 236 tokens of errors involving the *to*-infinitive were subsequently classified into the following four major categories: substitution, omission, addition, and others. As Table 4 shows below, substitution was the type of error that was most frequently committed, accounting for 103 tokens (43.6%), which is closely followed by the omission errors (99 tokens, 42.0%). The addition errors, on the other hand, accounted for only 29 tokens (11.6%). The remaining five tokens could not be classified as any of the three aforementioned categories and were consequently grouped together as others (2.1%).

TABLE 4  
*Distribution of the To-Infinitive Errors*

	Tokens	Percentage
Substitution	103	43.6%
Omission	99	42.0%
Addition	29	12.3%
Others	5	2.1%
Total	236	100%

### **Substitution Errors**

A closer examination of all the 103 tokens of substitution errors revealed that the substitution errors can be re-sorted into the following subcategories: (a) *to* + gerund, (b) *to* + noun, (c) semantically wrong verb, (d) *to*-infinitive

for gerund, (e) *to*-infinitive for prepositional phrase, and (f) wrong voice. Of these six subcategories, substituting a gerund for the dictionary form of a verb after *to* was the most commonly occurring type of error, e.g., (10), accounting for over a third of all the substitution errors (36 tokens, 35.0%), as shown in Table 5 below:

- (10) a. I prefer to **\*taking** a bus.  
 b. I was about to **\*playing**.  
 c. I know it is not easy to **\*changing** one's character.

TABLE 5  
*Subcategories of the Substitution Errors*

	Tokens	Percentage
<i>To</i> + gerund	36	34.9%
<i>To</i> + noun	15	14.6%
Semantically wrong verb	25	24.3%
<i>To</i> -infinitive for gerund	12	11.6%
<i>To</i> -infinitive for prepositional phrase	14	13.6%
Wrong voice	1	1.0%
Total	103	100%

(10c) is particularly interesting in that unlike the other two examples, (10c) could also be categorized as an instance of an addition of unnecessary *to*, as it could be considered an instance of extraposition, in which a clause in subject position moves to the end of a sentence and the subject position is consequently filled by a non-referential *it*. Although “[g]erund clauses generally resist extraposition,” e.g., (11), there are exceptions to this rule, e.g., (12), (Cowan, 2008, p. 476):

- (11) a. Understanding this lesson was easy.  
 b. \*It was easy understanding this lesson.  
 c. Herb's winning the gold medal came as a complete surprise to his coach.

- d. \*It came as a complete surprise to his coach Herb's winning the gold medal.
- (12) a. *Protesting the new policies* would be no use.  
b. It would be no use *protesting the new policies*.

The fact that some constructions do allow the extraposition of a gerund clause may explain why students make mistakes like (10c). Be that as it may, (10c) has been categorized as an instance of a substitution error since the construction *It is not easy* does not allow the extraposition of a gerund, as shown in (11b).

The erroneous substitution of a noun for the dictionary form of a verb, e.g., (13), or that of a *to*-infinitive for a gerund, e.g., (14), on the other hand, was not as common, as there were only 15 (14.6%) and 12 (11.7%) such tokens, respectively:

- (13) a. Most of buildings need to **\*renovation**.  
b. I am proud of my school and happy to **\*entrance** my school.
- (14) a. So we were upset and stopped **\*to do** our work.  
b. After **\*to do** five missions, finally, I can make my life better.

In fact, the second most commonly occurring subcategory of substitution was using semantically incorrect verbs, e.g., (15), accounting for 25 tokens (24.3%):

- (15) a. I don't have chance to **\*hear** the professor's great literature lesson.  
b. There are many middle-school and high-school students to **\*trip** the *Ojukheon*.

One unexpected finding was that there were 14 tokens of substituting a *to*-infinitive for a prepositional phrase, e.g., (16), in which *by taking* should have been used instead of *to take* and *of changing* instead of *to change*:

- (16) a. Every morning going to work, people can exercise **\*to take** bicycles.  
b. It is mind to love school that is capable **\*to change** University.

The remaining one token was an instance of using the wrong voice, more specifically substituting the active voice for the passive:

- (17) My thought to make one important change in our school is too ideal to **\*realize**.

### Omission Errors

The 99 tokens of the omission errors are further classified into the following three subcategories: (a) leaving out *to*, (b) leaving out the verb after *to*, and (c) leaving out the subject of the *to*-infinitive. As shown in Table 6 below, almost two-thirds of all the omission errors involve leaving out *to*, e.g., (18), accounting for 65 of the 99 tokens (65.6%):

- (18) a. I want **\*(to)** read many books.  
b. I want my hometown **\*(to)** have a good environment.

As Cowan (2008, p. 500) points out, there are only a few verbs in English that allow a bare-infinitive complement, namely “causative verbs” such as *make*, *let*, and *have* and “perception verbs” such as *hear*, *observe*, *see*, and *watch*. Thus, *read* in (18a) should be preceded by *to* in order to function as a complement of the main verb *want*. The majority of these errors occurred in the form of (18a), where two verbs are juxtaposed with nothing between them; there were only four errors such as (18b), where two verbs are juxtaposed with an NP between them.

TABLE 6  
*Subcategories of the Omission Errors*

	Tokens	Percentage
Leaving out <i>to</i>	65	65.6%
Leaving out verb after <i>to</i>	26	26.3%
Leaving out infinitival subject	8	8.1%
Total	99	100%

The other two subcategories of omission involve leaving out items other than *to*, namely, the verb after *to* and the subject of the *to*-infinitive. Of the two, leaving out the verb occurred much more frequently, as there were 26 such tokens (26.3%), e.g., (19):

- (19) a. So, we have to **\*(be)** careful too much.  
 b. But in Deajeon, Eunhangdong is only one place to play that is not enough to join thousands people. And that one place is just nice for buying things like clothes, shoes and accessory that people who want to see movie or dance or eat various food are hardly go there. And that place is so boomed with young people like student of high school, middle school and elementary school that adult and old people are not so easy to **\*(get)** along with other people there.

In (19a), the reader does not have to consider the context to determine what the omitted verb is: the meaning of the sentence is clear even without the verb *be*, which is required only because *careful* is an adjective and *have to* needs to be followed by a verb. The fact that *be* is a linking verb without any semantic content seems to be the reason why the learner omitted the verb. In fact, directly translated into Korean, (19a) makes perfect sense without the linking verb. Many other instances of leaving out the verb after *to*, however, required taking the context into consideration in order for the reader to determine the missing verb, as is the case with (19b).

The least commonly occurring subcategory of omission errors was leaving out the subject of the *to*-infinitive, e.g., (20), accounting for the remaining eight tokens (8.1%):

- (20) One more thing that I was disappointed with my school cafeteria is its interior. Its walls were white and there was no decoration. Chairs and tables are obsolete. Cafeteria was so old fashioned that many young fashionable freshmen did not like to sit there and have the lunch. They will have their lunch or dinner at neat restaurant outside the campus and it is waste of time and money. For these reasons, I strongly urge **\*(NP) to change** the school cafeteria.

When there is no overt infinitival subject in the sentence, it is assumed that the implied infinitival subject is equivalent to “the subject of the main clause” (Dulay et al., 1982, p. 195), as in (20), where the subject of *to change* is interpreted to be *I*. However, it is clear that the intended subject of *to change* is the university that the learner is attending, and the reason for omitting this infinitival subject seems to be the fact that the intended subject of *to change* is so evident that the learner did not think it necessary to overtly mention it.

### **Addition Errors**

As with the omission errors, all the addition errors proved instances of three subcategories: (a) unnecessary *to*, (b) unnecessary verbs after *to*, and (c) unnecessary tense-marking on the verb. Of these three subcategories, adding an unnecessary *to* was by far the most commonly occurring error, e.g., (21), as it accounted for over half of the 29 addition errors (17 tokens, 58.6%):

- (21) a. First, I want to set a new school zone to make student \***to** study in clean neighborhood.
- b. Teachers just thought (that) students \***to** go to good university and they didn't care about what students like, (and) what students is good at.
- c. I strongly \***to** say that some of them are worth trying to the school.

(21a) is an example of a phenomenon called “overgeneralization” (or regularization) (Dulay et al., 1982, p. 157), in which the learner used the more common *to*-infinitive complement rather than the required bare infinitive for the causative verb *make*. In (21b), *to* is unnecessary as the verb *think*, unlike *want*, can take a *that*-clause complement but not an NP + *to*-infinitive complement. Thus, (21b) can also be considered an instance of overgeneralization. It is unclear, however, why the learner added the unnecessary *to* before *say* in (21c), and we cannot rule out the possibility that the student might have accidentally left out *have* before *to*. Table 7 below lists the three subcategories of addition errors and shows the frequency for each subcategory.

TABLE 7  
*Subcategories of the Addition Errors*

	Tokens	Percentage
Unnecessary <i>to</i>	17	58.6%
Unnecessary verb after <i>to</i>	2	6.9%
Unnecessary tense-marking	10	34.5%
Total	29	100%

Unlike adding an unnecessary *to*, adding an unnecessary verb after *to* was found to be quite rare, as there were only two such tokens:

- (22) a. All I have to **\*do memorize** is to keep my response, and not to give any harm to others.  
b. We tend to **\*apply give** good score professor.

As can be seen in the two sentences above, it is unclear as to why the learners added the extra verbs *do* and *apply*, since correcting the grammatical mistakes will not resolve the semantic incongruity: *All I have to do to memorize is to keep my response* and *We tend to apply giving a good score to professors* still do not make any sense. It may thus be safe to argue that these two instances of adding an unnecessary verb after *to* seem to be instances of performance mistakes.

Unnecessary tense-marking, on the other hand, was found in ten different tokens (34.5%), and both the third-person singular present, e.g., (23a), and the simple past, e.g., (23b), were found to be marked on the verbs that should have been used as dictionary forms:

- (23) a. So I want to **\*makes** friends.  
b. So, if I can, I would change my hometown's road system as like an ancient roman's, clear and accurate, one. If it is possible, we don't have to **\*lost** the way, and don't have to be angry in a traffic jam, it means we can reduce the time to our destination, and the new road system will make the view of the city better. This changes make my hometown citizen feel that we live in a clear environment and advanced quality of life.

Again, it is unclear why errors like these appear in the learner corpus. If the subject in (23a) were a third-person singular, using *makes* might make sense, but the subject *I* does not even require inflecting the main verb for person, let alone the *to*-infinitive. In (23b), since all the other sentences are in the present tense, there is no reason to use the past-tense verb *lost* after the *to*-infinitive. Thus, unnecessary tense-marking also seems to a performance mistake.

## Others

Of the 236 tokens of errors involving *to*-infinitives, five tokens could not be classified as substitution, omission, or addition errors. Two of them were instances of using an adjective or an adverb after *to*:

- (24) a. One of ways to **\*cheaper** the price, school will manage their finance perfectly.
- b. There are some methods **\*to down** fee.

In (24), the learners seem to have used “an alternative term expressing the meaning of the target lexical item as closely as possible” (Brown, 2007, p. 138). In (24a), the learner used the comparative form of an adjective, *cheaper*, which is grammatically wrong, yet we can assume that he or she might have meant to use the verb *lower*, which also looks like the comparative form of the adjective *low*. Therefore, the learner seems to have used the “approximation” strategy by choosing a word in his or her interlanguage lexicon (Brown, 2007, p. 138). In (24b), the correct verb also seems to be *lower*, but the adverb *down* was used instead. The adverb *down* can actually be used as a verb, but when it is used as a verb, it is not synonymous with *lower*. Thus, (24b) also seems to be an instance of using the approximation strategy.

The other three tokens are all instances of misordering, as shown below in (25), the first two of which can be corrected simply by reordering words:

- (25) a. So I think to change those things is the first thing of all **\*to do**.
- b. I have to make a rule **\*to not repeat** the regretting.
- c. But universities seem to **\*don't** care about that.

In (25a), *to do* should be placed before *of all*; in (25b), *not* should be placed before *to*. In (25c), however, a simple reordering of *to* and *don't* does not make the sentence grammatical, as *don't* has to be replaced with *not* first.

As with *to repeat* in (25b), *to care* in (25c) should take “secondary verb negation,” i.e., “the use of the negative element *not* to negate a clause that has a verb in one of its *secondary forms* – that is, a verb in its infinitive, bare infinitive, present participle, or past participle form” (Cowan, 2008, p. 93).

Cowan (2008, p. 104) also points out that learners with advanced proficiency still make errors with secondary verb negation and that “[o]ne common error type is the use of *don't* when *not* is required,” as in (26):

(26) Nevertheless, the practice of sports along the campus installations can help foreign students *to don't become segregate*.

As the example in (26) appears in an essay written by an undergraduate Spanish speaker, making an error with secondary verb negation such as the one in (25c) must not have stemmed from negative transfer from the learner's mother tongue.

## Conclusion

This study has investigated the errors involving *to*-infinitives found in a learner corpus consisting of 815 essays written by incoming freshmen at a university in Seoul, Korea. Of the 2,996 tokens of *to* retrieved by using a concordancer, 2,309 were found to be tokens of *to*-infinitives, and only 171 of these 2,309 tokens (7.4%) were found to be errors, a finding which suggests that Korean students with high-intermediate proficiency in English possess a firm understanding of the use of the *to*-infinitive. Also noteworthy is the fact that this distribution of correct and incorrect uses of *to*-infinitives corresponds almost exactly to that of *V-ing* forms reported in Choi and Yoo (2012), i.e., 292 of the 3,843 tokens of *V-ing* forms (7.6%) were used incorrectly.

In addition, to those 171 tokens of errors involving *to*-infinitives, 65 instances where the *to*-infinitive should have been used, i.e., *to* omission

errors, were manually identified as the corpus had not been tagged for errors and a concordancer could not have been used to retrieve tokens of *to*-infinitives that students mistakenly left out. Of the three major categories of errors, substitution was the most commonly occurring type of errors (103 tokens, 43.6%), which was closely followed by omission errors (99 tokens, 42.0%). Addition errors, on the other hand, accounted for only 12.3% (29 tokens), and the remaining five tokens were classified as “Others,” as they did not belong to any of the subcategories of the three major categories.

Of the 103 tokens of substitution errors, using a gerund instead of the dictionary form of a verb after *to* was the most commonly occurring subcategory, accounting for 36 tokens (35.0%). Among the omission errors, leaving out *to* was by far the most frequent subcategory, as it accounted for almost two-thirds of all the omission errors (65 of the 99 tokens, 65.6%). Interestingly, the most commonly occurring subcategory of addition errors was the direct opposite of leaving out *to*: adding an unnecessary *to* before a verb (17 of the 29 tokens, 58.6%).

A close examination of the errors involving *to*-infinitives indicates that several types of errors derive from the learner’s lack of knowledge regarding C-selection, i.e., “[t]he information about the complement types selected by particular verbs and other lexical items” (Fromkin, Rodman, & Hyams, 2011, p. 104). Therefore, it is suggested that the teacher provide L2 learners with detailed C-selection, or subcategorization, information, especially when new verbs are introduced. Needless to say, L2 learners should also be provided with S-selection (S for semantics) included in the lexical entry of verbs, i.e., “a specification of certain intrinsic semantic properties of their subjects and complements” (Fromkin et al., 2011, p. 104).

Overall, the finding that the vast majority of the tokens of *to*-infinitives retrieved from the learner corpus were used properly is indeed encouraging. This finding should, however, be taken with a grain of salt in light of the fact that Koreans tend to overuse *to*-infinitives (Goh & Kim, 2009). According to Schachter and Celce-Murcia (1977), L2 learners may overuse a certain grammatical structure because they avoid producing other grammatical

structures “which they find difficult both in terms of the actual formation of such structures and the conditions for their use” (p. 447). For EFL/ESL teachers, it is important to ascertain what the learners “won’t do, and why” as well as what they will do, and why (Schachter & Celce-Murcia, 1977, p. 447). Therefore, teachers should keep in mind that Koreans’ tendency to overuse *to*-infinitives and their ability to use them relatively well may well mean that they are in fact avoiding producing other grammatical structures.

Lastly, it should be noted that the findings presented in this study may only be manifestations of the learners’ performance and not of their competence. Nesselhauf (2005) was indeed correct in pointing out that “confronted only with learner corpus data, it cannot be determined if a given feature is a part of the learner’s competence or whether the learner merely uses the feature as a communication strategy” (p. 42). Granger (2002), however, was also correct in pointing out that research studies using data from learner corpora, such as the present study, provide us with “improved descriptions of learner language which can be used for a wide range of purposes in EFL/ESL language acquisition research” (p. 4).

### **The Authors**

*Jung Eun Kim* received an M.A. in English Education from Sogang University. She is currently an English teacher at Soongmoon High School in Seoul, Korea.

Soongmoon High school  
99, Sungmun-gil, Mapo-gu  
Seoul, 04126, Korea  
Tel: +82 27162980  
Fax: +80 27112980  
Email: dhfl2311@naver.com

*Isaiah WonHo Yoo* is Professor in the Department of English at Sogang University. After an undergraduate psychology education at UC Berkeley, he earned a Ph.D. in applied linguistics from UCLA and taught EAP for three years at MIT before coming to Sogang University. His primary research focuses on how corpus linguistics informs language pedagogy. His recent publications have appeared in *Applied Linguistics*, the *Journal of Pragmatics*, the *International Journal of Corpus Linguistics*, and the *Journal of English for Academic Purposes*.

Department of English  
Sogang University  
35 Baekbeom-ro, Mapo-gu  
Seoul, 04107, Korea  
Tel: +82 27058340  
Fax: +82 27150705  
Email: iyoo@sogang.ac.kr

## References

- Biber, D., Johansson, S., Leech, G., Conrad, S., & Finegan, E. (1999). *Longman grammar of spoken and written English*. Harlow, UK: Pearson Education.
- Brown, D. (2007). *Principles of language learning and teaching* (5<sup>th</sup>ed.). New York: Pearson Education.
- Celce-Murcia, M., & Larsen-Freeman, D. (1999). *The grammar book: An ESL/EFL teacher's course* (2<sup>nd</sup>ed.). Boston: Heinle & Heinle.
- Cha, K. (2004). An error analysis of Korean university students' writing. *Foreign Languages Education*, 11(2), 149-169.
- Choi, Y. J., & Yoo, I. W. (2012). A corpus-based error analysis of V-ing forms in Korean college freshmen's essays. *English Teaching*, 67(4), 311-335.

- Cowan, R. (2008). *The teacher's grammar of English*. Cambridge, UK: Cambridge University Press.
- Dulay, H., Burt, M., & Krashen, S. (1982). *Language two*. Oxford, UK: Oxford University Press.
- Fromkin, V., Rodman, R., & Hyams, N. (2011). *An introduction to language* (9<sup>th</sup> ed.). Boston: Wadsworth, Cengage Learning.
- Goh, G. Y., & Kim, N. Y. (2009). The use of the *to*-infinitive in the language of Korean English newspapers: A corpus-based analysis. *Comparative Korean Studies*, 17(1), 33-56.
- Granger, S. (2002). A bird's-eye view of learner corpus research. In S. Granger, J. Hung, & S. Petch-Tyson (Eds.), *Computer learner corpora, second language acquisition, and foreign language teaching* (pp. 4-28). Philadelphia: John Benjamins.
- Hyde, B. (2000). The structures of the *to*-infinitive. *Lingua*, 110(1), 27-58.
- Kim, P. (2009). *Criterion*® program summary. *KATE Forum*, 34(3), 20-24.
- Korean Ministry of Education. (2011). *Yeong-eo-gwa gyo-yug-gwa-jeong* (Publication No. 2011-361). Seoul, South Korea: Ministry of Education.
- Lee, W. H., & Choe, M. H. (2013). A cognitive linguistic approach to teaching gerund and *to*-infinitive complements. *Modern English Education*, 14(2), 101-121.
- Nesselhauf, N. (2005). *Collocations in a learner corpus*. Amsterdam: John Benjamins.
- Norrish, J. (1983). *Language learners and their errors*. London: Macmillan.
- O'Keeffe, A., McCarthy, M., & Carter, R. (2007). *From corpus to classroom*. Cambridge, UK: Cambridge University Press.
- Schachter, J., & Celce-Murcia, M. (1977). Some reservations concerning error analysis. *TESOL Quarterly*, 11(4), 441-451.
- Quirk, R., Greenbaum, S., Leech, G., & Svartvik, J. (1985). *A comprehensive grammar of the English language*. London: Longman.