

A Multifactorial Analysis of English Dative Alternations in Korean EFL Learners' Writings

Yong-Hun Lee*

(Hannam University)

Cheongmin Yook

(Hallym University)

Bomi Lee, Yeonkyung Park

(Hannam University)

Lee, Yong-Hun, Yook, Cheongmin, Lee, Bomi, & Park, Yeonkyung. (2016). A multifactorial analysis of English dative alternations in Korean EFL learners' writings. *English Teaching*, 71(3), 145-168.

This paper adopts a multi-factorial analysis and investigates English dative alternations in Korean EFL learners' writings, by directly comparing dative alternations of native speakers and those of Korean EFL learners' writings. Two corpora were chosen for the comparison: the Switchboard corpus (written sections) and the Korean component of the TOEFL11 corpus. After all the sentences with dative alternations were extracted from these two corpora, and seventeen linguistic factors were manually encoded, the data were statistically analyzed in R. This paper was theoretically based on the concept of interlanguage in the Second Language Acquisition and Competition Model of Bates and MacWhinney (1982, 1989). Through the analysis, the following facts were observed: (i) the Korean EFL learners used ditransitive constructions more frequently than prepositional constructions, (ii) eight main factors and four interactions with the L1 were statistically significant, and (iii) the ditransitive constructions which the Korean EFL learners used were closer to the prepositional constructions rather than the ditransitive constructions in Korean.

Key words: dative alternation, interlanguage, Competition Model, logistic regression, Behavioral Profile

* Yong-Hun Lee: First author; Cheongmin Yook: Corresponding author

1. INTRODUCTION

Natural languages may have different kinds of syntactic alternations, and the alternations have been the target of linguistic studies. English also has different types of alternations, and dative alternation is one of them. Two different types of constructions are allowed in the dative alternations in English, and they are illustrated in (1) (Lee, Yook, Lee, & Park, 2015).

- (1) a. Mary gave *Tom a book*.
b. Mary gave *a book to Tom*.

(1a) consists of an agent, a verb, a recipient, and a theme; and it is called a *ditransitive* construction. On the other hand, (1b) consists of an agent, a verb, a theme, and a recipient with the insertion of a preposition (here *to*) between two NPs; and this construction is called a *prepositional dative* construction.¹

This kind of alternation poses difficulty for English as a foreign language (EFL) learners, since they have to know (1) which verbs can occur in both constructions, and (2) which of the two constructions should be selected if a verb occurs in both constructions. Such difficulty leads to avoidance of one construction and preference for the other. It has been found that L2 learners often turn to an avoidance strategy when they perceive a word or a structure of target language difficult to produce. Korean has a construction which can correspond to the English prepositional dative constructions, but many scholars have claimed that the language has no construction which exactly matches the English ditransitive constructions. Consequently, as Yook (2012, 2013) pointed out, it is difficult for Korean EFL learners learn English dative alternations.

Based on the previous studies on dative alternations, Lee et al. (2015) investigated the English dative alternations in Korean EFL learners' writings, with corpus data and statistical analyses. Though the study observed many interesting properties of the dative alternations by the Korean EFL learners, it contained no direct comparison with a Native Language speaker (ENL; a native speaker) data.² Accordingly, there are some limitations

¹ In Lee et al. (2015), (1a) was called a double object construction (DC) and (1b) was an adpositional construction (AC). However, this paper follows the terminology in Gries and Deshors (2015) and uses the terms *ditransitive* construction for (1a) and *prepositional dative* construction for (1b), respectively.

² This does not mean that the study did not include any comparison with dative alternations by ENL speakers. The study also contained a comparison of the results with those of Bresnan, Cueni, Nikitani, and Baayen (2007). However, the comparison was indirect. That is, Lee et al. (2015) used the analysis results of Bresnan et al. (2007), not the real corpus data.

in investigating the exact properties of the English dative alternations in Korean EFL learners.

In order to address this problem, this paper took real corpus data of (American) ENL speakers and directly compared the dative alternations of Korean EFL learners with those of ENL speakers. Through the direct comparison, the following research questions will be closely investigated.³

1. Which factors interacted with the L1 differences (English vs. Korean)?
2. Which factors influenced the choice of dative alternations, both ENL speakers and Korean EFL learners?
3. Are the ditransitive constructions in the Korean EFL learners' writings close to the ditransitive constructions or the prepositional transitive constructions in English?

Among these three questions, the first two questions would be answered with a logistic regression analysis, and the last one would be answered with a Behavioral Profiles (BP) analysis.

2. LITERATURE REVIEW⁴

2.1. Interlanguage and Second Language Acquisition

In the literature of Second Language Acquisition (SLA), interlanguage (IL) refers to the rule system that has been developed by L2 learners who have not yet reached native-like high level of proficiency. Usually, IL is an intermediate language which is constructed between the native language (L1) and the target language (here, English). According to Bialystok and Sharwood Smith (1985), IL is a linguistic system which is unlike that used by the native speaker, but one which is nonetheless systematic in the structural sense. They also claim that IL denotes a product and that it is the outcome of language use.

In fact, before their study, Selinker (1969) first mentioned that recognition of the existence of an IL could not be avoided and must be dealt with as a system. According to

³ One might add more research question(s) on the distributions of ditransitive constructions and prepositional dative constructions in two different groups. However, the distributions were shown during the analysis process (Figure 1). This paper focused on which factors and interactions resulted in the distributions, not the distributions themselves.

⁴ Lee et al. (2015) included a review on the previous studies on dative alternations in Korean EFL learners' English. Accordingly, they were not included in this paper.

Selinker (1969), an IL can be linguistically described using as data the observable output resulting from a speaker's attempt to produce a foreign norm, i.e., both his errors and non-errors. It can be assumed that such behavior is highly structured. In comprehensive language transfer work, it seems that recognition of the existence of an IL cannot be avoided and that it must be dealt with as a system, not as an isolated collection of errors.

IL is known to be involved both in the mental representation of systematic organized information about the target language and in the procedures for effectively and efficiently retrieving that knowledge in appropriate situations (Bialystok & Sharwood Smith, 1985). Bialystok and Sharwood Smith (1985) mentions that it is more as a system than as a product that IL has triggered most interest, since IL is concerned with the outcome of mental functioning which attributes to the learner specific limitations in two aspects of mental processing.

IL is clearly different from L1 and the target language L2 (here, English). The source of the differences is known to be language transfer, which refers to the tendency of ESL/EFL learners to apply their L1 knowledge to another language (L2) when learning the L2, i.e., a process occurring from the native to the foreign language, according to Selinker (1969). Odlin (1989) pointed out that L2 learners bring the knowledge of their L1 grammar to the L2 learning process. Odlin (2003) also noted that the L1 influence of previous linguistic subsystems has also been shown, including phonetics, phonology, morphology, syntax, and semantics.

Adjemian (1976) pointed out that the permeability of IL grammar could be the crucial difference between native and non-native language varieties. According to Adjemian, IL grammars (unlike the L1 grammars) are interim grammars in nature, and they are not fixed by their very nature. Instead, they develop and change continuously.

This variable nature of IL has opened a various way from experimental and introspective methods to corpus-based and probabilistic approaches. Previous studies on SLA research mainly adopted experimental and introspective methods of data investigation, as Granger (2002) points out. Granger (2002) states that learner corpus research provides a means to combine non-experimental and quantitative approaches to learner language. Furthermore, "a corpus-based approach to learner language allows the researcher to identify the characteristics of particular IL varieties (i.e., the interactions of particular L1s and L2s)" (p. 5). Hanks (2000) also notes that what a corpus gives us is the opportunity to study traces and patterns of linguistic behavior.

Along with these lines of thoughts, the usefulness of statistically grounded approaches for a corpus based investigation of learner language is clearly discussed in Jarvis (2000). Jarvis (2000) claims that L1 influence refers to any instance of learner data where a statistically significant correlation (or probability-based relation) is shown to exist between some features of learners' IL performance and their L1 background.

As an interim summary, it can be said that (i) IL has an independent status and must be dealt with as a system, (ii) language transfer effects are the origin of the IL, and (iii) IL can systematically be studied using corpus data and statistical approaches to data.

2.2. Bates and MacWhinney's Competition Model (1982, 1989)

The Competition Model (CM) was developed by Elizabeth Bates and Brian MacWhinney as a psycholinguistic theory of both language acquisition and sentence processing. The crucial idea of the CM is that the meaning of a language must be interpreted through the comparison of *a number of (linguistic) factors* within a sentence, and that a language is acquired/learned through the *competition* of basic cognitive mechanisms in the linguistic environment. According to the CM, competitive cognitive processes occur in three different types of scales, and the CM enables us to explain the fact that language acquisition takes place across a wide variety of chronological periods.

Originally, the CM was proposed as a theory of sentence processing in psycholinguistics. Bates and MacWhinney stated that human beings tend to interpret the meaning of a sentence by taking into account various linguistic factors included in the given context, such as word order, morphology, and semantic characteristics (e.g., animacy), and so on. When people formulate a sentence, they calculate the probabilities of each interpretation with unconscious mechanisms and choose the best with the highest probabilities. According to the model, the weight or importance of each linguistic factor is learned inductively on the basis of a constrained set of sentence types and the limited predictions of sentence meaning for a language. Because different languages adopt different linguistic factors to signal meaning, the CM claims that the weights will differ between languages and that the users of a given language will use the weights or importance to guide their interpretation of sentences. Thus, when human beings learn more than one language, they must learn which linguistic cues are important in which languages in order to successfully interpret the sentence meaning.

Recently, the CM has been developed into a unified theory which covers not only first but also second language acquisition (see Deshors, 2010; Deshors & Gries, 2014). Its scope has been expanded and it is now able to explain several psycholinguistic processes involved in language acquisition; such as cues, storage, chunking, codes, and resonance. The expanded version of the CM claims that each of these cognitive mechanisms controls the activation of meaning representations in the target language which competes in the mind of the learners during their acquisition of the language. As in the original version of the model, the weights of each linguistic factor are computed and adjusted in real time based on the learner's experience with the target language. Thus, the CM model claims that as the extensiveness of learners' exposure to the target language increases, they will gain an

increasingly complete and nuanced understanding of the meaning of sentences in the target language.

2.3. A Corpus-based Approach: Lee et al. (2015)

Although there have been several previous studies on the English dative alternations by Korean EFL learners (e.g., see Lee, 1997; Oh, 2010; Yook, 2012; Yook, 2013), their studies were experimental and introspective in nature, as Granger (2002) described. Unlike these previous studies, Lee et al. (2015) employed a corpus-based approach and statistically analyzed the corpus data.

Lee et al. (2015) used the Korean component of the TOEFL11 corpus to examine which linguistic factors influenced the choice of dative alternations in Korean EFL learners' writings. From this corpus, all the sentences were extracted, which included the English dative alternations. Then nine linguistic factors were manually encoded. Finally, the collected data were statistically analyzed with a generalized linear mixed-effects, model following Zuur, Ieno, Walker, Saveliev, and Smith (2009).

From the analysis, the following facts were observed: (1) Korean EFL learners' choice of dative alternation was influenced by the following five factors: animacy of theme, definiteness of theme, pronominality of recipient, length difference between theme and recipient, and individual verbs; (2) despite some similarities in choice of alternation between Korean EFL learners and ENL speakers, the two groups also exhibited differences, and (3) these differences originated from the different effects of these factors.

After the statistical analysis of the collected data, the results were indirectly compared with the analysis results in Bresnan et al. (2007). Table 1 illustrates the comparison results (Lee et al., 2015, p. 148).

TABLE 1
Comparison with Bresnan et al. (2007)

ENL Speakers	Korean EFL Learners
Accessibility of Recipient	
Accessibility of Theme	
Pronominality of Recipient	Pronominality of Recipient
Pronominality of Theme	
Definiteness of Recipient	
Definiteness of Theme	Definiteness of Theme
Animacy of Recipient	
	Animacy of Theme
Length Differences	Length Differences
Verb	Verb

As this table shows, nine factors are involved in the choice of dative alternations in the

ENL speakers, while five linguistic factors influenced the choice of dative alternations in the Korean EFL learners' writings.

Although Lee et al. (2015) succeeded in identifying which linguistic factors influenced the choice of dative alternations in Korean EFL learners' writings, a direct comparison with the ENL data was impossible, because the comparison was conducted with the analysis results of Bresnan et al. (2007), not with real corpus data. This paper was designed to solve this problem.

3. RESEARCH METHOD

3.1. Research Procedure

Our research in this paper proceeded as follows. First, two corpora were selected. One is the written part of the Switchboard corpus (LDC Catalo No.: LDC97S62; Godfrey, Holliman, & McDaniel, 1992) containing English sentences of ENL speakers, and the other is the Korean component of the TOEFL corpus (LDC Catalo No.: LDC2014T06; Blanchard, Tetreault, Higgins, Cahill, & Chodorow, 2013), which includes English sentences of Korean EFL learners.⁵ Next, all sentences with dative alternations were extracted from both corpora, using NLP Tools (Lee, 2007). A total of forty verbs were used in the extraction.⁶ Then seventeen different linguistic factors were manually encoded, following Deshors (2010) and Deshors and Gries (2014). Lastly, a statistical analysis of the corpus data was done with the help of R (R Core Team, 2016), including a logistic regression (a multi-factorial analysis) and a Behavioral Profile (BP) analysis.

The TOEFL11 corpus was a learner corpus which was released by the English Testing Service (ETS) in 2014, and consists of essays written during the TOEFL iBT® tests in 2006-2007. It contains 1,100 essays for each of the 11 native languages (Arabic, Chinese, French, German, Hindi, Italian, Japanese, Korean, Spanish, Telugu, and Turkish), totaling 12,100 essays. All essays were taken from the TOEFL independent task, where test-takers

⁵The reasons why the Switchboard corpus was chosen were (1) that the corpus data were also used in the analysis of Bresnan et al. (2007), and (2) that the corpus represented American English, which Korean English is known to be based on. However, there was a notable difference between Bresnan et al. (2007) and this study. Though Bresnan et al. (2007) used both spoken and written components of the corpus data, this paper used only the written component of the data. This strategy was chosen, since the Korean EFL learner data (the Korean component of the TOEFL11 corpus) were written naturally. Accordingly, the results of the Switchboard corpus in this paper might be (or must be) different from the results in Bresnan et al. (2007).

⁶ For the forty verbs used in the study, see Lee et al. (2015).

were asked to write an essay in response to a brief writing topic. The essays were sampled as evenly as possible from eight different topics. The corpus also provides the score levels (low/medium/high) for each essay. The Korean component of the TOEFL11 corpus contains 95,066 word tokens for the low level, 202,531 word for the medium level, and 30,787 tokens for the high level. A total of 328,384 word tokens were included in the Korean component of the TOEFL11 corpus.

3.2. Encoding of Variables

After all the sentences were extracted from two types of corpora, seventeen linguistic factors were manually encoded for the extracted data, following Deshors (2010) and Deshors and Gries (2014).⁷ Table 2 shows the encoded factors.⁸

TABLE 2
Encoded Factors

ID Tag Type	ID Tag	ID Tag Levels
Data	L1	English, Korean
Length	RELENGTH	
	PATLENGTH	
	LENGTHDIFF	
Syntax	VOICE	active, passive
	REALIZATION	ditransitive, prepositional transitive
	RECPRONOMINALITY	nonpronoun, pronoun
	PATPRONOMINALITY	nonpronoun, pronoun
Semantics	VERBSEMANTICS	abstract, communication, future, prevention, transfer
	RECANIM	animate, inanimate
	PATANIM	animate, inanimate
	RECDEF	definite, indefinite
	PATDEF	definite, indefinite
	RECSEMANTICS	abstract, concrete, human, informational
	PATSEMANTICS	abstract, concrete, human, informational
Discourse	REACCESSIBILITY	accessible, given, new
	PATAACCESSIBILITY	accessible, given, new

Following the study in Atkins (1987), each linguistic factor and its level were called ID tag

⁷ This process is called *operationalization*, as in Deshors (2010) and Deshors and Gries (2014).

⁸ Note that nine linguistic factors were used in Lee et al. (2015), while seventeen factors were encoded in this paper. Since more linguistic factors were involved in the analysis, more accurate prediction would be possible for the dative alternations (both for the ENL speakers and the Korean EFL learners). Also note that seventeen factors were hierarchically organized in Table 2.

and ID tag levels, respectively. These variables were used in the statistical analysis.⁹

3.3. Statistical Analysis

Two types of statistical analyses can be applied to the collected corpus data: a monofactorial analysis and a multifactorial analysis, depending on how many factors or variables are involved in the analysis. In the former type, only the effects or influences of each factor/variable are statistically examined. In the latter type, however, not only the effects/influences of each factor/variable but also their interactions among the factors/variables are statistically analyzed. Though both kinds of analyses are available, multifactorial analyses are known to be more appropriate for linguistic data, since many linguistic factors/variables and their interactions are involved in linguistic phenomena.¹⁰ The multifactorial analyses of linguistic data are supported by the studies in Cognitive Linguistics. For example, Langacker (2000) points out that the goal of multifactorial analysis is to conceive of [linguistic] entities in connection with one another (e.g., for the sake of comparison, or to assess their relative position), not just as separate, isolated experiences. This is linguistically important because relationships figure in the meaning of almost all expressions, many of which (e.g., verb, adjectives, prepositions) actually designate relationships. Gries (2003) also adopted a multi-factorial analysis and investigated the distributions of particle placement in ENL speakers' English. Deshors (2014) also mentioned, in her studies on *can* and *may*, that the multifactorial approach also helped the authors make a connection between degrees of grammatical complexity of speakers' utterances and learners' lexical choices during second language production. For instance, they observed that *can* rather than *may* was more frequently used by French English learners (compared to native speakers) in more complex grammatical environments such as negated or subordinated linguistic contexts.

This paper also took a multi-factorial approach and used a Generalized Linear Model (GLM) with logistic regression in the statistical analysis, since it is one of the simplest and most widely adopted analyses. For regression analysis, Deshors (2014) described binary

⁹ In Table 2, REC and PAT refer to the recipient and patient, respectively. The classifications of the factor VERBSEMANTICS followed the traditions in Levin (1993) and Bresnan et al. (2007): the 'transfer' of possession of an object (e.g., *give a book to someone*); the 'future transfer' of objects (e.g., *owe money to someone*); the 'communication' of information (e.g., *tell a story to someone*); the 'prevention of possession' (e.g., *deny food to someone*); and 'abstract' (e.g., *allow someone a privilege*), which contains nearly every other remaining sense.

¹⁰ This does not imply that monofactorial analyses are meaningless in the investigations of linguistic phenomena. The monofactorial analyses play an important role when we examine how much each linguistic factor influences the linguistic phenomena.

logistic regression as a confirmatory statistical technique that allowed the analyst to identify possible correlations between the dependent and the independent variables. Ultimately, this statistical approach allowed us to see what factors influence learners' choices of *may* and *can*.

During the analysis process, a stepwise model selection procedure was applied as follows, which was similar to the model selection process of mixed models.¹¹ First, an initial model was constructed with all of the factors and their interactions with the L1. Second, a new model was constructed where one factor or one interaction was deleted from the previous model. Third, the newly constructed model was compared with the previous model with an ANOVA. Fourth, an optimal model was chosen according to some criteria such as significance testing (with p -values) or information criteria: if a model m_1 contained a factor f or an interaction i but a model m_2 did not contain f or i , (i) when the p -value of the ANOVA test was significant ($p < .05$), this implied that factor f or interaction i must not be deleted from the model, and model m_1 was selected, and (ii) when the p -value of the ANOVA was not significant ($p > .05$), this implied that factor f or interaction i can be safely deleted from the model, and model m_2 was selected. The processes continued until all the factors and their interactions were scrutinized.

This paper adopted another multifactorial analysis, a Behavioral Profile (BP) analysis. BP analysis is a statistical method that closely examines the behavioral properties of each linguistic factor. The analysis represents the similarity or dissimilarity of the components with a dendrogram (which was created by hierarchical agglomerative cluster analysis). It was developed by Gries and Otami (2010) and Gries (2010a). This statistical method was originally used to analyze synonymy and/or antonymy in lexical semantics. However, the method can also be used here, since the use of the modal constructions in the EFL learners' writings can be classified based on the behavioral properties (similarity or dissimilarity) of linguistic factors.

4. REGRESSION ANALYSIS

4.1. Logistic Regression with GLM

Since the dependent variable REALIZATION has two values (*ditransitive* and *prepositional dative*), a binary logistic regression is necessary. The first step for the logistic regression is to set up the initial model. Then model selection procedures were applied (cf.

¹¹ The model selection process described here also appears in other papers of the first author in this paper. However, it was mentioned again here for readers' convenience.

Section 3.3) and the final (optimal) model was selected. Table 3 shows an initial model of our study, and Table 4 the final model.

TABLE 3

Initial model

Model	Specification
model01	REALIZATION~L1+LENGTHDIFF+PRONREC+PRONTH+VERBSEM+ANIMREC+ANIMTH+DEFREC+DEFTH+ACCESSREC+ACcessth+L1:LENGTHDIFF+L1:PRONREC+L1:PRONTH+L1:VERBSEM+L1:ANIMREC+L1:ANIMTH+L1:DEFREC+L1:DEFTH+L1:ACCESSREC+L1:ACcessth

TABLE 4

Final model

Model	Specification
model02	REALIZATION~L1+LENGTHDIFF+PRONREC+VERBSEM+ANIMREC+ANIMTH+DEFTH+ACCESSREC+L1:LENGTHDIFF+L1:VERBSEM+L1:ANIMTH+L1:ACcessth

As one can observe in Table 3 and Table 4, the eight main factors and four interactions with L1 survived in the final model.

4.2. Analysis Results

After we obtained the final model, all the main factors/variables and their interactions with L1 were statistically analyzed, as in Table 5 and Table 6.

TABLE 5

Analysis Results (Main Factors)

	<i>df</i>	Deviance	AIC	LRT	<i>p</i>
<none>	832.12	886.12			
L1	1	970.95	998.95	69.18	.001
LENGTHDIFF	1	1133.82	1161.82	232.05	.001
PRONREC	1	937.05	965.05	35.28	.001
PRONTH	1	901.88	929.88	.11	.741
VERBSEM	4	1001.26	1023.26	99.49	.001
ANIMREC	1	908.77	936.77	7.01	.008
ANIMTH	1	901.89	929.89	.12	.724
DEFREC	1	905.19	933.19	3.42	.064
DEFTH	1	918.44	946.44	16.67	.001
ACCESSREC	1	925.51	953.51	23.74	.001
ACcessth	1	902.11	930.11	.34	.559

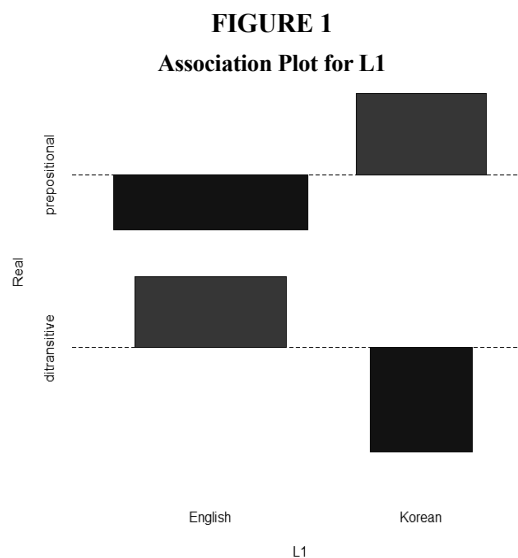
TABLE 6

Analysis Results (Interactions)					
	<i>df</i>	Deviance	AIC	LRT	<i>p</i>
L1:LENGTHDIFF	1	838.04	890.04	5.92	.015
L1:PRONREC	1	832.15	884.15	.03	.858
L1:PRONTH	1	832.58	884.58	.47	.494
L1:VERBSEM	3	858.19	906.19	26.08	.001
L1:ANIMREC	1	834.22	886.22	2.10	.147
L1:ANIMTH	1	838.99	890.99	6.88	.009
L1:DEFREC	1	833.42	885.42	1.30	.254
L1:DEFTH	1	832.34	884.34	.23	.632
L1:ACCESSREC	1	848.46	900.46	16.34	.001
L1:ACCESSTH	1	832.28	884.28	.16	.685

These two tables show that seven main factors and four interactions with L1 were statistically significant in the model. These tables also show that the factor ANIMTH survives in the final model because of its interactions with the factor L1.

Since the final model was obtained, it was possible to investigate how the Korean EFL learners' choices of dative alternation differed from the ENL counterparts. In what follows, we will illustrate some of the main findings with two types of graphic representations.

Among the main factors, one factor (i.e., L1) was chosen and examined with a graphic tool as an example. Figure 1 shows the association plot for L1.

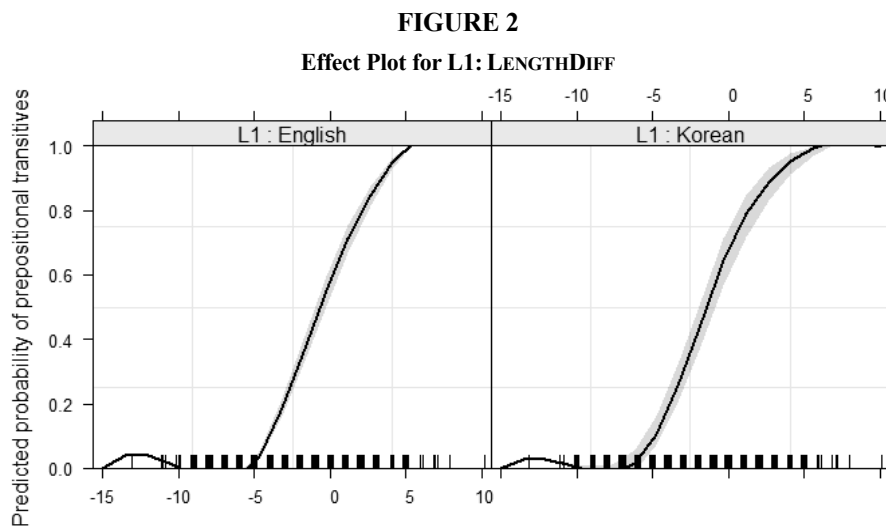


In the association plot, the effects of the factor are represented by the baseline (the dotted line) and rectangles above and below the baseline. Here, the baseline (the dotted line)

represents the expected frequency of each value for a given factor. The width of the rectangle is proportional to the square root of the expected frequency, and the height of the rectangle is proportional to the standardized residual.

As this association plot indicates, the ENL speakers (labeled as English) use ditransitive constructions more frequently but prepositional transitive constructions less often than the Korean EFL learners. That is, the association plot shows that the Korean EFL speakers use ditransitive constructions less frequently and prepositional dative constructions more often than the ENL speakers.

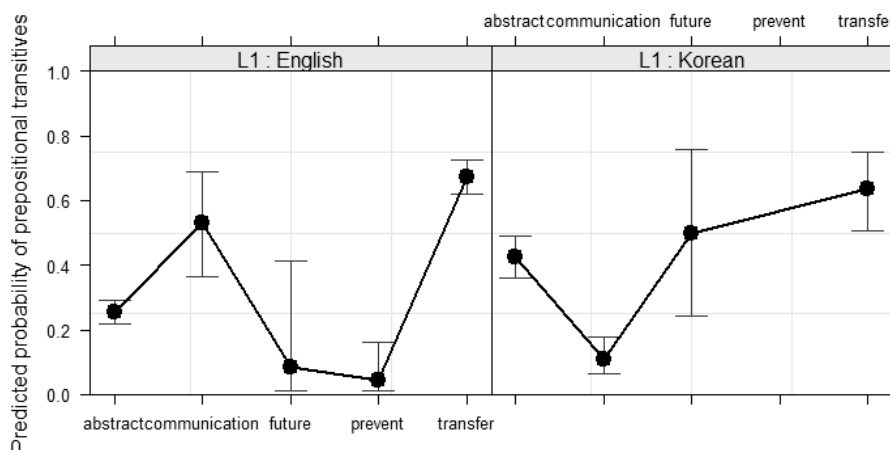
For the interaction with L1 by effect plots, Figure 2 illustrates the effect plot for L1: LENGTHDIFF.



As seen in this effect plot, a similar tendency appeared both in the English ENL speakers and the Korean EFL learners. That is, as the LENGTHDIFF value (RECLength-PATLength) increased, both types of speakers used the prepositional dative constructions more frequently than the ditransitive constructions. However, the range of LENGTHDIFF in Korean was slightly wider than for English, even though the range of both varieties covered from -5 to 5 (word length). Also note that the 95% bands for the Korean EFL learners were slight wider than the bands for the ENL speakers.

Figure 3 shows the effect plot for L1: VERBSEM.

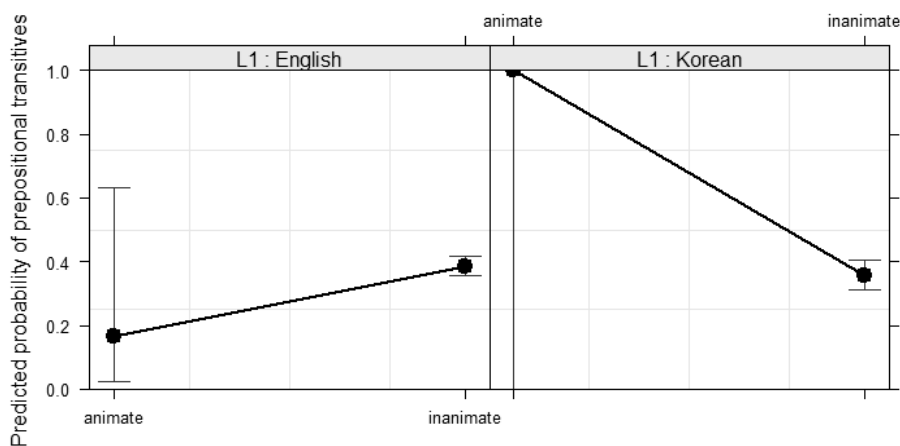
FIGURE 3
Effect Plot for L1: VERBSEM



As seen above, the overall tendencies of both types of speakers differed, except for the verb ‘transfer’. For ‘abstract’, the Korean EFL learners used prepositional dative constructions more frequently. For ‘communication’, the ENL speakers used more prepositional dative constructions. For ‘future’, the Korean EFL learners used prepositional dative constructions more frequently. For ‘prevent’, the Korean EFL learners used neither the prepositional dative constructions nor the ditransitive constructions.

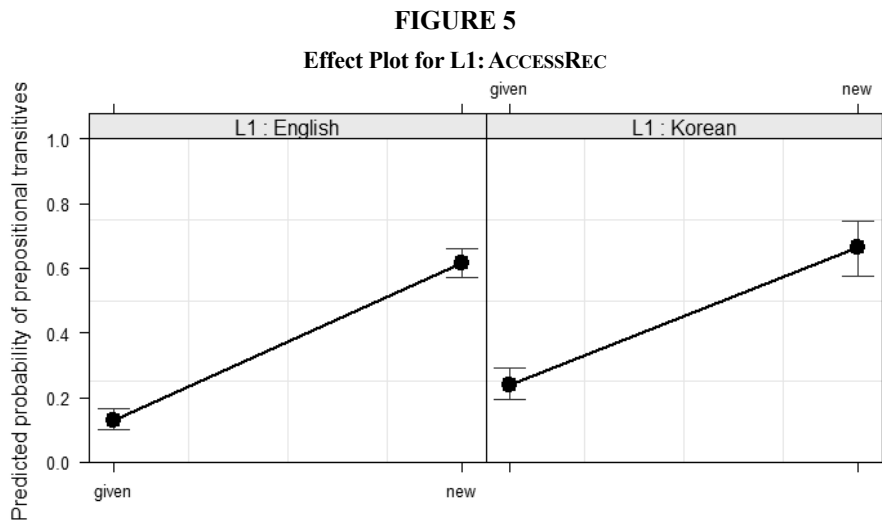
Figure 4 shows the effect plot for L1: ANIMTh.

FIGURE 4
Effect Plot for L1: ANIMTh



As seen above, when animacy of theme (ANIMTH) referred to an inanimate entity, both types of speakers demonstrated a similar tendency. When it referred to an animate entity, the ENL speakers used ditransitive constructions more frequently, while the Korean EFL learners used prepositional dative constructions more frequently.

Figure 5 shows the effect plot for L1: ACCESSREC.



The overall tendencies were similar in both types of speakers. However, the Korean EFL learners used more prepositional dative constructions in both 'given' and 'new' contexts. Also note that the two lines were not parallel, as the slope of the Korean data was slightly steeper than for the English data. This implied that there was an interaction between the two factors L1 and ACCESSREC.

An optimal model was constructed for the data, and goodness of fit for the obtained model was tested. Since a logistic regression was used in the analysis (cf. Section 4.1), *C*-statistics was used for comparison.¹²

The *C*-statistics for Table 4 was .93445. For the *C*-values, Harrell (2001) stated that "*C*-values range from .5 to 1 and the higher the value, the better a regression model is at classifying or predicting the dependent variable; *C*-values $\geq .8$ are generally considered

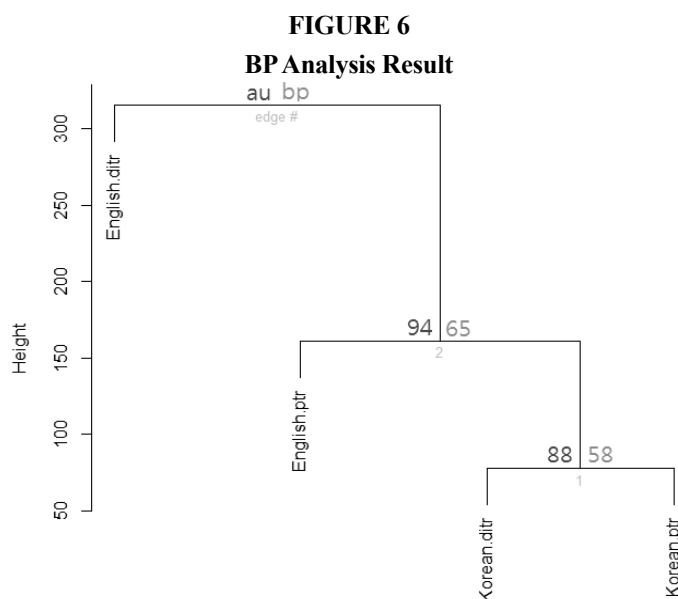
¹² The *C*-statistic refers to the probability that predicting the outcome is better than chance, and it is commonly used to compare goodness of fit of logistic regression models. The values for this measure range from .5 to 1.0. A value of .5 indicates that the model is no better than chance, and a value of 1.0 indicates that the model perfectly identifies which fits the given data and which does not. A model is typically considered reasonable when the *C*-statistic exceeds .7 and strong when it exceeds .8. (Hosmer, Lemeshow, & Sturdivant, 2000).

good” (p. 248). Note that the *C*-value for our final model is .93445. This suggests that our statistical model is very good for explaining the differences between the ENL speakers and the EFL learners.

5. The BP Analysis

As Table 5 and Table 6 indicated, these two tables show that eight main factors and four interactions with L1 were statistically significant in the model ($p < .05$). These results showed that dative alternations of Korean EFL learners are differed with those of the English ENL speakers. Then the question was how much this tendency among the Korean EFL learners differed from the English ENL speakers. To answer this question, a BP analysis was performed.¹³

Among the factors in Table 2, the combination of L1 and REALIZATION were chosen as a dependent variable and the other factors plus VERB were used as independent variables. Figure 6 shows the dendrogram which resulted from the analysis (multiscale bootstrap resampling clustering). In this figure, *ditr* refers to ditransitive constructions and *ptr* to prepositional dative constructions.



¹³ We used Gries (20101b) in the actual analysis.

Here, the horizontal lines indicate which component(s) can be grouped with which component(s), and the vertical lines indicate the distance between the two groups. Two numeric values in the dendrogram refer to the AU (approximately unbiased) *p*-value and BP (bootstrap probability) value for each cluster.

As shown in Figure 6, Korean.ditr was combined with Korean.ptr first, which can be represented as {Korean.ditr, Korean.ptr}. This implies that Korean.ditr was closer to Korean.ptr. If the same reasoning had been applied, English.ditr would have been grouped with English.ptr, which can be represented as {English.ditr, English.ptr}. However, the result was that English.ptr was patterned with {Korean.ditr, Korean.ptr}, which can be represented as {English.ptr, {Korean.ditr, Korean.ptr}}. On the other hand, English.ditr formed another independent group by itself. This implies that the behaviors of Korean.ditr were closer to English.ptr or Korean.ptr, not to English.ditr.

6. DISCUSSION

This paper used two types of analyses to compare dative alternations of Korean EFL learners with those of the (American) ENL speaker counterparts: a regression analysis (GLM) and a BP analysis.

In the GLM regression analysis in Section 4, it was observed (i) that the ENL speakers use ditransitive constructions more frequently than the Korean EFL learners, and (ii) that seven main factors and four interactions with L1 were statistically significant in the model. These results showed that dative alternations in the Korean EFL learners' writings differed from those of the (American) ENL speakers.

Among the seventeen linguistic factors, eight main factors (L1, LENGTHDIFF, PRONREC, VERBSEM, ANIMREC, ANIMTH, DEFTH, and ACCESSREC) and four interactions with L1 (L1:LENGTHDIFF, L1:VERBSEM, L1:ANIMTH, and L1:ACCESSREC) were significant. These results differed from those of Lee et al. (2015).

As Table 1 showed, only four main factors played significant roles in the the dative alternations in Lee et al. (2015): Pronominality of Recipient (PRONREC), Definiteness of Theme (DEFTH), Animacy of Theme (ANIMTH), and Length Differences (LENGTHDIFF). If this result is compared with the four interactions with L1 in this study, only two factors overlap: ANIMTH and LENGTHDIFF,¹⁴ excluding the factor VERBSEM from the discussion,

¹⁴ It would be meaningless if the four linguistic factors in Lee et al. (2015) were compared with eight main factors in this study. The reason is that these eight main factors in this study play the same important role both for the (American) ENL speakers and the Korean EFL learners. These factors are involved in the determination of dative alternations, whose construction is chosen in

since this factor was added in this paper. The factor Accessibility of Recipient (ACCESSREC) was not significant in Lee et al. (2015), but it was statistically significant in this study. The factors Pronominality of Recipient (PRONREC) and Definiteness of Theme (DEFTH) were statistically significant in Lee et al. (2015), but were statistically insignificant in this study ($p = .858$ and $p = .633$, respectively). However, these two factors were statistically significant in Table 5. This fact implied (1) that these two factors certainly played important roles in the choice of dative alternations, (2) that their effects were similarly crucial among both types of speakers, and (3) that they could not be used as factors to distinguish the Korean EFL learners from the (American) ENL speakers.

There may be two reasons for these discrepancies. The first reason arose from the direct comparisons with the ENL corpus data. As mentioned in Section 2.3, the comparison in Lee et al. (2015) was conducted with the analysis results of Bresnan et al. (2007), not with real corpus data. On the other hand, the comparison in this paper was performed with real corpus data (the Switchboard corpus). Accordingly, the results in this paper might reveal more facts about the dative alternation of two different groups. Second, as mentioned in Section 3.1, only the written component of the corpus data was selected in this study, whereas both spoken and written components were involved in the study of Bresnan et al. (2007). This difference might also cause the discrepancies in the results.

The regression analysis results also demonstrated that Bates and MacWhinney's CM (1982, 1989) could be successfully applied to model the IL, which was located between the Korean (L1) and the target language (English). In Section 4, we used a statistical method (specifically, a binary logistic analysis) to construct two kinds of statistical models, one constructed from the English corpus data, and the other from the Korean EFL learner data. The former was a statistical model for the target language (English) and the latter was a statistical model for the Korean-English IL.¹⁵ In these statistical models, many linguistic factors competed as choices in the given contexts. This competition reflected the claim of the CM that each cognitive mechanism controls the activation of meaning representations in the target language and that these cognitive mechanisms compete in the mind of the learners during their acquisition of the language. The statistical models also reflect the claim of the CM that human beings tended to interpret the meaning of a sentence by taking into account various linguistic factors included in the given context.

If the CM were applied to the dative alternations between the two groups, they could be explained as follows. First, there was an environment where dative alternation was available. Then the speakers had to decide which construction to use between the

the given contexts.

¹⁵ Note that the second statistical model was not the model for L1, since the corpus data contained English sentences, not Korean sentences.

ditransitive construction and prepositional dative construction, i.e., two constructions competed for the same linguistic environment. To decide which construction could be used, the speakers would make use of various linguistic factors which were helpful for the decision. Some factors might be advantageous for the ditransitive constructions and others might be for the prepositional dative constructions. The final choice of the construction depended on the strengths of each factor.

The strengths of each linguistic factor in the alternations were highly influenced by the linguistic structure of the L1 and the target language. Thus, three different models could be constructed for a single syntactic alternation. For dative alternations of Korean EFL learners, for example, three different models could be constructed: one for the dative alternations in Korean (if possible), another for those in the Korean-English IL, and another for those in English (by ENL speakers). By comparing these three models, it was possible to uncover which factors made the EFL learners' English more non-native-like and how the L1 influenced the choice of alternations. Among these three models, the second and the third models were constructed for this paper. Accordingly, it was possible to examine which linguistic factors made the EFL learners' writings non-native-like. However, the model for the dative alternations in Korean was not constructed in this paper. Therefore, further studies are necessary, where the Korean dative alternation model is constructed and the model compared with the model for Korean-English IL. Then it would be possible to investigate if and how the L1 influenced the choice of dative alternations by the Korean EFL learners.

Some might ask if such complex process (competition of alternate constructions) really occurs in the human brain. Bates and MacWhinney (1982, 1989) presented several pieces of evidence which demonstrated that such a process really takes place, and Bates and MacWhinney (1989) also explain how the CM might be applicable to the study of SLA. Consequently, it would be natural to apply the CM to dative alternations by Korean EFL learners, which was done in this paper.

Through the comparisons of two models, it was systematically shown that the dative alternations in the (American) ENL speakers differed from those of the Korean EFL learners. The results statistically attested Adjemian's (1976) claim that the permeability of IL grammar could be the crucial difference between native and non-native language varieties.

This procedure indirectly attested Jarvis' (2000) claim that L1 influence refers to any instance of learner data where a statistically significant correlation (or probabilitybased relation) is shown to exist between some features of learners' IL performance and their L1 background. This also indirectly confirmed Selinker's (1969) claim that recognition of the existence of an IL cannot be avoided and must be dealt with as a system.

In the BP analysis in Section 5, however, it was observed that Korean.ditr patterned with

Korean.ptr first, which implies that Korean.ditr was closer to Korean.ptr. It was also found that English.ditr formed another independent group by itself. This implies that the behaviors of Korean.ditr were closer to English.ptr or Korean.ptr, not to English.ditr. The possible reason for these differences might be an L1 transfer effect, as discussed in Selinker (1969) and Odlin (1989, 2003). However, it would be premature to claim that the result originated from L1 transfer effects. In order to prove L1 transfer effects, it is necessary to perform the same kind of analysis for the corpus data in Korean (such as the Sejong Corpus). Then the results have to be compared with those in this paper. This paper did not provide such an analysis. In addition, it is doubtful that Korean really contains constructions corresponding to English ditransitive constructions. Accordingly, it is impossible to claim that the tendency for Korean.ditr to be closer to English.ptr or Korean.ptr originated from L1 transfer effects.

From the results, however, it could be concluded that the differences between two groups of speakers (ENL speakers vs. Korean EFL speakers) were much stronger than the distinctions between the ditransitive and prepositional dative. If the constructional distinction had been much stronger than the group differences, Korean.ditr would have patterned with English.ditr and Korean.ptr with English.ptr, which could be represented as $\{\{\text{Korean.ditr, English.ditr}\}, \{\text{Korean.ptr, English.ptr}\}\}$. The fact that Korean.ditr combined with Korean.ptr could be interpreted as evidence that the group differences were much stronger than the distinctions between the ditransitive and prepositional dative.

Returning to the research questions raised in Section 1, how can these questions be answered based on the results? For the first question, Table 5 provided the answers. Seven linguistic factors influenced the choice of dative alternations, both for the (American) ENL speakers and the Korean EFL learners: L1, LENGTHDIFF, PRONREC, VERBSEM, ANIMREC, DEPTH, and ACCESSREC. For the second questions, Table 6 provided the answers. Four main factors interacted with the L1: L1:LENGTHDIFF, L1:VERBSEM, L1:ANIMTH, and L1:ACCESSREC. Accordingly, those interactions could distinguish the (American) ENL speakers from the Korean EFL learners. For the third question, Figure 6 provided the answers. The ditransitive constructions in the Korean EFL learners' writings were much closer to the prepositional dative constructions in Korean or the constructions in English.

7. CONCLUSION

This paper statistically investigated how various linguistic factors influenced the choice of dative alternations in Korean EFL learners' English and American ENL speaker counterparts. Two corpora (the written component of the Switchboard corpus and the Korean component of the TOEFL11 corpus) were selected in this study. After all the

sentences with dative alternations were extracted from the corpora, seventeen linguistic factors were manually encoded for each sentence. Then a logistic regression model and a BP analysis were applied to statistically analyze the factors playing a role in the selection of constructions and how they affected the selection among the two different groups of speakers. How various linguistic factors influenced the choice of dative alternation was closely examined through effects plots (Figure 2, Figure 3, Figure 4, and Figure 5).

Through the analysis, the following facts were observed: (i) the Korean EFL learners used prepositional dative constructions more frequently than ditransitive constructions, (ii) eight main factors and four interactions with the L1 were statistically significant, and (iii) the ditransitive constructions which the Korean EFL learners used in their writings were close to the prepositional dative constructions rather than the ditransitive constructions in Korean.

The procedure and results demonstrated that it was possible to construct a model of IL (here, English-Korean IL) using statistical methods, based on Bates and MacWhinney's CM. The results also showed that it is possible to statistically model and analyze linguistic behaviors with different L1 backgrounds, with this type of statistical model. We hope that developments and statistical tools like those above will ultimately place findings within our discipline on a more solid foundation.

REFERENCES

- Adjemian, C. (1976). On the nature of Interlanguage systems. *Language Learning*, 26(2), 297-320.
- Atkins, B. (1987). Semantic ID tags: Corpus evidence for dictionary senses. In *Proceedings of the third annual conference of the UW centre for the new Oxford English dictionary*. 17-36.
- Bates, E., & MacWhinney, B. (1982). Functionalist approaches to grammar. In E. Wanner & L. Gleitman (Eds.), *Language acquisition: The state of the art* (pp. 73-218). Cambridge: Cambridge University Press.
- Bates, E., & MacWhinney, B. (1989). Functionalism and the competition model. In B. MacWhinney & E. Bates (Eds.), *The cross-linguistic study of sentence processing* (pp. 3-73). Cambridge: Cambridge University Press.
- Bialystok, E. & Sharwood Smith, M. (1985). Interlanguage is not a state of mind: An evaluation of the construct for second language acquisition. *Applied Linguistics*, 6, 101-117.
- Blanchard, D., Tetreault, J., Higgins, D., Cahill, A., & Chodorow, M. (2013). *TOEFL11: A corpus of non-native English*. ETS RR-13-24. Princeton, NJ: Educational Testing

Service.

- Bresnan, J., Cueni, A., Nikitina, T., & Baayen, R. (2007). Predicting the dative alternation. In G. Bouma, I. Krämer & J. Zwarts (Eds.), *Cognitive foundations of interpretation* (pp. 69-94). Amsterdam: Royal Netherlands Academy of Arts and Sciences.
- Deshors, S. (2010). *Multifactorial study of the uses of may and can in French-English interlanguage*. Unpublished doctoral dissertation, University of Sussex, Sussex.
- Deshors, S. (2014). Constructing meaning in L2 discourse: The case of modal verbs and sequential dependencies. In G. Dylan & M. Sjölin (Eds.), *Subjectivity and epistemicity: Stance strategies in discourse and narration* (pp. 329-348). Lund: Lund University Press.
- Deshors, S., & St. Greis. (2014). A case for the multifactorial assessment of learner language: The uses of *may* and *can* in French-English interlanguage. In D. Glynn & J. Robinson (Eds.), *Corpus methods for semantics: Quantitative studies in polysemy and synonymy* (pp. 179-204). Amsterdam: John Benjamins.
- Godfrey, J., Holliman, E., & McDaniel, J. (1992). SWITCHBOARD: Telephone speech corpus for research and development. *Proceedings of the acoustic, speech, and signal processing* (pp. 517-520). San Francisco, CA: IEEE.
- 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. 3-33). Amsterdam: John Benjamins.
- Gries, S. T. (2003). *Multifactorial analysis in corpus linguistics: A study of particle placement*. New York: Continuum.
- Gries, S. T., & Deshors, S. (2015). EFL and/vs. ESL?: A multi-level regression modeling perspective on bridging the paradigm gap. *International Journal of Learner Corpus Research*, 1(1), 130-159.
- Gries, S. T., & Otani, N. (2010). Behavioral Profiles: A corpus-based perspective on synonymy and antonymy. *ICAME Journal*, 34, 121-150.
- Gries, S. T., (2010a). Behavioral Profiles: A fine-grained and quantitative approach in corpus-based lexical semantics. *The Mental Lexicon*, 5(3), 323-346.
- Gries, S. T., (2010b). Behavioral profiles 1.01: A program for R 2.7.1 and higher.
- Hanks, P. (2000). Do word meanings exist? *Computers and the Humanities*, 34(1/2), 205-215.
- Harrell, F. (2001). *Regression modeling strategies with applications to linear models, logistic regression, and survival analysis*. Berlin: Springer.
- Hosmer, D., Lemeshow, S., & Sturdivant, R. (2013). *Applied logistic regression* (3rd ed.). Hoboken, NJ: Wiley.
- Jarvis, S. (2000). Methodological rigor in the study of transfer: Identifying L1 influence in the Interlanguage lexicon. *Language Learning*, 50(2), 245-309.

- Langacker, R. (2000). *Grammar and conceptualization*. Berlin: Mouton.
- Lee, D. (1997). Acquisition of dative alternation in English by second language learners. Unpublished doctoral dissertation, University of South Carolina, South Carolina.
- Lee, Y. (2007). *Corpus analysis using NLPTools and their applications: Applications to linguistic research, English education, and textbook evaluation*. Seoul: Cambridge University Press.
- Lee, Y., Yook, C., Lee, B., & Park, Y. (2015). Factors involved in Korean EFL learners' choice of English dative constructions. *English Teaching*, 70(4), 139-159.
- Levin, B. (1993). *English verb classes and alternations*. Chicago: University of Chicago Press.
- Odlin, T. (1989). *Language transfer*. Cambridge: Cambridge University Press.
- Odlin, T. (2003). Cross-linguistic influence. In C. Doughty & M. Long (Eds.), *The handbook of second language acquisition* (pp. 436-386). New York: Blackwell.
- Oh, E. (2010). Recovery from first-language transfer: The second language acquisition of English double objects by Korean speakers. *Second Language Research*, 26(3), 407-439.
- R Core Team. (2016). *R: A language and environment for statistical computing*. Vienna: R Foundation for Statistical Computing.
- Selinker, L. (1969). Language transfer. *General Linguistics*, 9(2), 67-92.
- Yook, C. (2012). L1 influence on ESL learners' acquisition of English dative constructions. *English Teaching*, 67(2), 27-50.
- Yook, C. (2013). The role of frequency in Korean learners' acquisition of English dative construction. *English Teaching*, 68(1), 179-200.
- Zuur, A., Ieno, E., Walker, N., Saveliev, A., & Smith, G. (2009). *Mixed effects models and extensions in ecology with R*. Berlin and New York: Springer.

Applicable levels: Secondary, tertiary

Yong-Hun Lee, Bomi Lee, Yeonkyung Park
Department of English Language & Literature
College of Liberal Arts, Hannam University
70 Hannam-ro, Daedeok-gu
Daejeon, Republic of Korea
Phone: 042-629-7320
E-mail: yleuiuc@hanmail.net, lbmloveyo@hanmail.net, withbyk@gmail.com

CheongminYook
Department of English
College of Humanities, Hallym University
Administration Building 2626
1 Hallymdaehak-gil, Chuncheon
Gangwon-do
Phone: 033-248-1533
E-mail: cmYook@hallym.ac.kr

Received on June 1, 2016

Reviewed on July 15, 2016

Revised version received on August 15, 2016