Linguistic Research 33(Special Edition), 107-136

A multifactorial analysis of particle placement in Korean EFL learners' writings*

Yong-hun Lee[a] · Ha-Eung Kim[b] · Gyu-Hyeong Lee[b]
(Chungnam National University[a] · Hannam University[b])

Lee, Yong-hun, Ha-Eung Kim, and Gyu-Hyeong Lee. 2016. A multifactorial analysis of particle placement in Korean EFL learners' writings. *Linguistic Research 33*(Special Edition), 107-136. This paper statistically investigated the particle placement in Korean EFL learners' writings, based on the corpus data. Two corpora were selected for the study. One was the Korean component of the TOEFL11 corpus, and the other was the ICE-GB corpus. After all the sentences with particles were extracted from these two corpora, twelve linguistic factors were manually encoded. Then, the data were statistically analyzed with R. Two statistical analyses were adopted. One was logistic regression, and the other was Behavioral Profiles. Through the analysis, the following facts were observed: (i) five linguistic factors were involved in the choice of particle placement, (ii) two linguistic factors were involved in the differences of two groups of speakers, and (iii) one type of construction (with the 'verb + object + particle') showed similar characteristics in both groups but the other type of construction was different in two groups of speakers. **(Chungnam National University · Hannam University)**

Keywords alternation, particle placement, competition model, logistic regression, behavorial profiles

1. Introduction

All the natural languages have alternations, and English is not an exception. English also has many different kinds of alternations, and particle placement is one of them. Let's see the following sentences.[1]

---

* We wish to thank two anonymous reviewers of this journal for their helpful comments and suggestions. All remaining errors, however, are ours.

1 Gries (1999) used the term *particle movement* while Gries (2001) used the term *particle placement*. The former study adopted Chomsky's transformational-generative grammar approach (Chomsky 1957, 1965) and thought that particles moved from one position to another. The latter study did not presuppose such movement analysis. This paper adopted Gries' second approach and called the syntactic phenomena in (1) *particle placement*. This means that this paper did not presuppose the

(1) a. John picked *up the book*.
b. John picked *the book up*.

(1a) consists of a subject, a verb, a particle, and an object, whereas (1b) consists of a subject, a verb, an object, and a particle. (1a) will be called a *po* construction (following the order of the particle and the direct object), and (1a) will be called a *op* construction.

Alternations make the English as a foreign language (EFL) learners difficult to learn the target language (here, English), since they have to acquire the following two knowledge on the construction: Not only do they have to know which verbs can occur in both types of constructions but also they have to study which of the two constructions should be selected if a verb occurs in both constructions. Naturally, such difficulty leads to the tendencies that the EFL learners avoid one construction and prefer the other. It has been frequently observed that L2 learners often took an avoidance strategy when they perceived a construction of target language difficult to produce. Since Korean has no construction which corresponds to particle placement, it will be hard to claim that the L1 transfer effects (Odlin 1989, 2003) are involved in the particle placement by the Korean EFL learners.

However, there is a way to investigate the behaviors of Korean EFL learners and those of English native speakers (English as a Native Language; ENL). It is possible to construct statistical models for two groups of speakers based on the corpus data and to compare the characteristics of these two groups of speakers in using the particles of English. In fact, Lee *et al*. (2015) constructed a statistical model for the particle placement of Korean EFL learners and investigated which linguistic factors involved the choice of alternation. However, in that paper, there was no direct comparison. Though the study succeeded in revealing the characteristics of particle placement of Korean EFL learners, it did not examine which linguistic factors made non-native English by Korean EFL learners.

In order to solve this problem, this paper took the concept of interlanguage (IL) and Bates and MacWhinney's Competition Model as theoretical bases. That is, this paper constructed the statistical models for Korean-English IL and native English, and all the linguistic factors *compete* for the choice of alternations in the same

movement of a particle from one position to another. Instead, this paper supposed that the particle placement was decided by the influences and interactions of many linguistic factors.

sentences. On the other hand, this paper took a corpus-based approach as a methodological basis, and it took two corpora (one for the ENL speakers and the other for the Korean EFL learners) and compared the analysis results of these two corpora. Through the analysis, it would be revealed which linguistic factors behaved differently in the Korean EFL learners' writings.

This paper had the following research questions.

(2) Research Questions
   a. Which factors influenced the choice of particle placement, both in ENL speakers' and Korean EFL learners' writings?
   b. Which factors interacted with the L1 (English vs. Korean)?
   c. Are the syntactic behaviors of particle placement of the Korean EFL learners' writings similar to those of the ENL speakers' or not?

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

This paper is organized as follows. In Section 2, previous studies were reviewed, especially focused on the concept of interlangauge and Bates and MacWhinney's Competition Model were introduced. Section 3 provided explanations on corpus data and research method. The collected corpus data were analyzed with a logistic regression in Section 4, and the same data were analyzed with a BP analysis in Section 5. Section 6 contained discussions on the findings, and Section 7 concluded this paper.

2. Previous studies[2]

2.1 Interlanguage and second language acquisition

In the study of Second Language Acquisition (SLA), the concept of interlaguage is very important. Interlanguage (IL) usually refers to the rule system of the target

[2] The theoretical bases described in Section 2.1 and Section 2.2 also provided in other papers of the first author (Yoon and Lee 2016; Lee *et al*. 2016), even though the papers handled other types of linguistic alternations (modals *can*/*may* and dative alternation). However, the content of Section 2.1 and Section 2.2 were mentioned here since they were theoretical basis of this paper and they were necessary in the discussion section in Section 6.

language which has been developed by the L2 learners who have not yet reached native-like high level of proficiency. Usually, IL means an intermediate language which is constructed between the native language (L1) and the target language (here, English). Bialystok and Sharwood Smith (1985: 116) described IL as follows: IL is "a linguistic system which is unlike that used by the native speaker, but one which is nonetheless systematic in the structural sense." Their study (1985: 101) also clarified that "IL denotes a product: it is the outcome of language use."

However, their study was not the first study on IL. Before their study, Selinker (1969) first mentioned the existence of an IL. According to Selinker (1969: 71), "an IL may 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 is assumed that such behavior is highly structured. In comprehensive language transfer work, it seems to me 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." [our emphasis]

IL is known to be involved not only in the mental representation of systematically organized information about the target language but also in the effective and efficient retrieving of the knowledge in appropriate situations (Bialystok and Sharwood Smith 1985). According to Bialystok and Sharwood Smith (1985: 106), "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."

It is well known that IL is clearly different from L1 and the target language L2 (here, English) and that the differences were originated from the language transfer effects. Language transfer refers to the tendency of the ESL/EFL learners that they apply their L1 knowledge into the grammar of another language (L2) when they acquire the L2. According to Selinker (1969), language transfer refers to "a process occurring from the native to the foreign language." Odlin (1989) mentioned that L2 learners made use of their L1 knowledge during L2 learning process. Odlin (2003) also pointed out that the L1 influence of previous linguistic subsystems had also been shown in almost every subfield of linguistics, including phonetics, phonology, morphology, syntax, and semantics.

According to Adjemian (1976), the permeability of IL grammar could be the essential difference between native and non-native language varieties. Adjemian

(1976) mentioned that IL grammars were interim grammars in nature and that they were not fixed by their very nature. That is, IL grammars change and develop continuously as time goes on.

This characteristic property of IL has opened various ways of research, from experimental and introspective methods to corpus-based and probabilistic approaches. In fact, as Granger (2002) pointed out, previous studies in SLA research mainly focused on experimental and/or introspective methods of data investigation. Granger (2002: 5) pointed out that "[l]earner corpus research provides a way to combine non-experimental and quantitative approaches to learner language. What is more, 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)." Hanks (2000: 211) also mentioned that "what a corpus gives us is the opportunity to study traces and patterns of linguistic behaviour."

However, the usefulness of statistically-grounded approaches with a corpus-based investigation of learner language was clearly mentioned in other studies. Jarvis (2000: 252) mentioned 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." [our emphasis]

If we summarize the previous studies, it can be listed as follows: (i) IL has an independent status and must be dealt with as a system, (ii) language transfer may be one source of the IL, and (iii) IL can systematically be studied using the corpus data and statistical approaches to the data.

2.2 Bates and MacWhinney's competition model (1982, 1989)

The Competition Model (CM) is first developed by Elizabeth Bates and Brian MacWhinney as a psycholinguistic theory of both language acquisition and sentence processing. The most essential ideas of the CM are 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 through the competition of basic cognitive mechanisms in the linguistic environments. The CM claims that the competitive cognitive processes occur in three different types of scales and that it allows us to explain the fact that the language acquisition process takes place across a wide

variety of chronological periods.

Originally, the CM was proposed as a theory in psycholinguistics by which the sentence processing could be explained. Bates and MacWhinney claimed that human beings usually interpret the meaning of a sentence by taking into account various linguistic factors which were included in the given context: such as word order, morphology, and semantic characteristics (e.g. animacy), and so on. When people articulate a sentence, they subconsciously calculate probabilities of each interpretation and choose the best one with the highest probabilities. According to this model, the importance of each linguistic factor are learned inductively by the language learners on the basis of a very constrained set of sentence types and the limited predictions of sentence meaning for a language. Since different languages usually adopt different linguistic factors to encode the meanings, the CM claims that the importance will differ among languages and that the users of a language will use the importance to guide their interpretation of sentences. Accordingly, when human beings learn more than one language, they have to learn which linguistic cues are important in which languages in order to successfully interpret the sentence meaning in any language.

Recently, the CM has been developed into a unified theory which covers both the first and second language acquisition in the studies such as Deshors (2010) and Deshors and Gries (2014). Its scope has been expanded and it is now able to provide an account for several psycholinguistic processes involved in language acquisition; such as cues, storage, chunking, codes, and resonance. The expanded version of the CM mentions that each of these cognitive mechanisms usually control the activation of meaning representations in the target language which competes in the mind of the learners during the acquisition of the target language. As in the original version of the model, the importance weights of each factor are calculated and adjusted in real time based on the learner's experience with the target language. Thus, the CM model claims that the learners gain an increasingly complete and nuanced understanding of the meaning of sentences in the target language as the extensiveness of learners' exposure to the target language increases.

2.3 On particle placement

There have been continuous theoretical studies on English Particle Movement in various linguistic fields: traditional grammar, Chomskyan transformational-generative

grammar, cognitive grammar, discourse-functional approaches, psycholinguistically-oriented approaches, and so on. Gries (1999: 33) closely investigated the claims in previous studies and summarized the linguistic factors which decided the alternation as follows.

Table 1. Linguistic factors that govern the particle placement

Value for construction$_0$	Variable	Value for construction$_1$
Long DO	Length of the DO in words (Length W)	
Long DO	Length of the DO in syllables (LengthS)	
Complex	Complexity of the DO (Complex)	
	NP-Type of the DO: semi-pronominal (Type)	pronominal
Indefinite	Determiner of the DO (Det)	definite
No	Previous mention of the DO (Lm)	yes
Low	◀——Times of preceding mention of the DO (Topm)——▶	high
High	◀——Distance to last mention of the DO (Dtlm/ActPC)——▶	low
High	◀——News Value of the DO——▶	low
Yes	(Contrastive) Stress of the DO	
Yes	Subsequent mention of the DO (Nm)	no
High	◀——Times of subsequent mention of the DO (Tosm)——▶	low
How	◀——Distance to next mention of the DO (Dtnm/ClusSC)——▶	high
	Overall frequency of the DO (Om)	
	following directional adverbial (PP)	yes
Yes	Prep of the following PP is identical to the particle (Part = Prep)	
	Register	
Idiomatic	◀——Meaning of the VP (Idiomaticity)——▶	literal
Low	◀——Cognitive Entrenchment of the DO——▶	high
Inanimate	Animacy of the DO (Animacy)	animate
Abstract	Concreteness of the DO (Concreteness)	concrete

Here, *construction$_0$* refers to the sentences with the order of 'verb + particle + object' as in (1a), while *construction$_1$* refers to the sentences with the order of 'verb + object + particle' as in (1b).[3] This table enumerated eighteen different linguistic factors and demonstrated that several different types of factors, not a single factor, actually influenced the choice of the constructions.

Let's see how these factors can be related with the alternation of Particle Movement. For example, *LengthW* (the first factor in Table 1) refer to the length of object in words. If the object is long, native speakers tend to choose *construction$_0$* rather than *construction$_1$*. If the object is short, the native speakers prefer to use

3 Two constructions (*construction$_0$* and *construction$_1$*) correspond to the *po* constructions and the *op* constructions respectively in this paper.

construction$_1$, rather than *construction*$_0$. The factor *Det*, the fifth factor, refers to the determiner of the object. If the determiner of object is indefinite (such as *a* or *an*), native speakers tend to choose *construction*$_0$ rather than *construction*$_1$. If the determiner is definite (such as the), native speakers prefer *construction*$_1$ rather than *construction*$_0$. Table 1 contains all the related factors which cover most of linguistic fields: phonology, syntax, semantics, pragmatics, and discourse analysis. In order to solve this problem, Gries (2001) pointed out the problems of previous monofactorial analyses and employed a multifactorial analysis, where all the factors in Table 1 were taken into consideration simultaneously. These studies used a Generalized Linear Model (GLM) and statistically analyzed how each factor played a role in the choice of construction. They also took a linear discriminant analysis (LDA) and a classification and regression tree (CART) to analyzed the corpus data.

Based on these studies, Lee *et al*. (2015) extracted all the sentences with English particles from the Korean component of the TOEFL11 corpus (Blanchard *et al*. 2013) and examined the behaviors of particle placement in Korean EFL learners' writings.[4] The sentences with particles were divided into two types of constructions: intransitive vs. transitive. The following plot illustrated the distribution of these two types of constructions in the Korean EFL learners' writings. (Lee *et al*. 2015: 119)

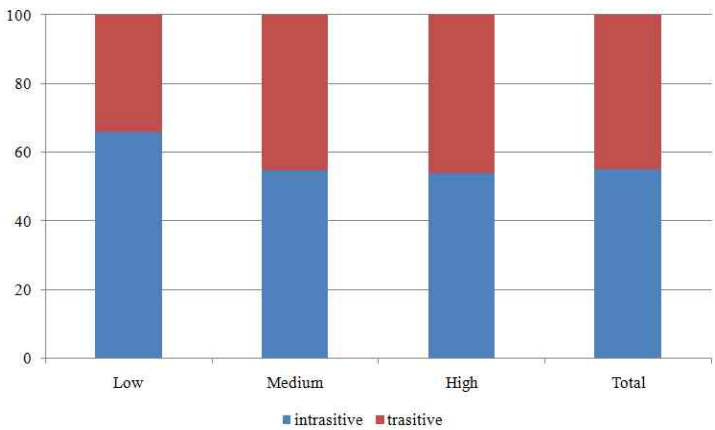

Figure 1. Distributions of intransitive and transitive constructions

As this plot demonstrated, more than half of the constructions with particles were

4 The detailed explanations on the TOEFL11 corpus will be provided in Section 3.1.

intransitive.

Since particle placement occurred in the transitive constructions, all the sentences with the transitive constructions were extracted from the corpus data. Then, the distributions of two types of constructions with particle placement were examined. The following plot illustrated the distribution of two types of constructions (*op* vs. *po*) in the Korean EFL learners' writings. In this plot, the lower parts of the bars corresponded to the *po* constructions and the upper parts to the *op* constructions. (Lee *et al*. 2015: 120)

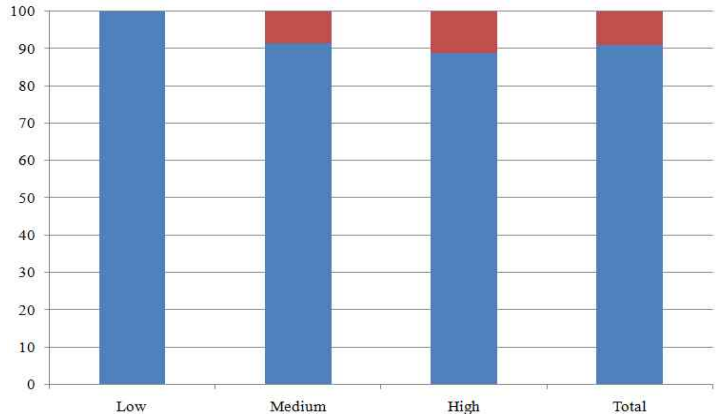

Figure 2. Distributions of *op* and *po* constructions

As this plot illustrated, the *po* constructions occupied more than 80% of the sentences with particles. That is, the Korean EFL learners were reluctant to separate the particles from the verbs. However, the ratio of the *op* constructions increased, as students' levels of proficiency went up.

Since the *op* constructions were also observed in the Korean EFL learners' writings, Lee *et al*. (2015: 119) manually encoded eight (linguistic) factors to the extracted sentences: *Level* (of proficiency), *Complexity*, *Animacy*, *Definiteness*, *Pronominality*, *Idiomacity*, *Concrete*, and *Length* (in words).[5] After the encoding process, the study employed a Generalized Linear (Regression) Model with a logistic regression and statistically examined which linguistic factors were involved in the choice of alternations between the *op* constructions and the *po* constructions.

5 These linguistic factors were also included in Table 1 (Section 3.2).

Through the analysis, it was observed that Korean EFL learners employed a different strategy in the particle placement and that only some factors were used for the selection of constructions. Unlike the ENL speakers, four linguistic factors were statistically significant in Korean EFL learners' writings (*Animacy*, *Pronominality*, *Concreteness*, and *Length*). That is, these four linguistic factors played a crucial role in the determination of choice between the *op* constructions and the *po* constructions. It was also found that there were some differences in the ratio of two constructions (*po* vs. *op*) as the level of proficiency went up. However, the differences were not statistically significant.

3. Research method

3.1 Research procedure

The research procedure in this paper was as follows. First, two corpora were selected. One was the British component of the International Corpus of English (ICE-GB; Nelson *et al*. 2002) and the other was the Korean component of TOEFL corpus (LDC Catalo No.: LDC2014T06; Blanchard *et al*. 2013). The former was for the English sentences of the ENL speakers, and the latter was for the English sentences of the Korean EFL learners. Among the texts in the ICE-GB corpus, the written parts were taken.[6] Then, since the latter corpus did not tag information in the text, all the sentences in this corpus were tagged with a CLAWS tagger.[7] In the next step, all the sentences with the particles were extracted from both corpora, using ICE-CUP and NLPTools (Lee 2007) respectively.[8] Then, twelve different linguistic factors were manually encoded, following the studies in Deshors (2010) and Deshors

---

6 Strictly speaking, since the TOEFL11 corpus included students' essays, the texts which correspond to this genre was from W1A-001 to W1A-010 in the ICE-GB corpus, which contained untimed student essays. However, the volume of the texts were too small to be compared with the Korean component of the TOEFL11 corpus. Accordingly, we extended the scope of the study to the whole part of the written texts, since we thought that it would also be meaningful to compare the tendencies in the Korean EFL learners' writings against the general tendencies in the written part of the ICE-GB corpus.

7 http://ucrel.lancs.ac.uk/claws/trial.html

8 Particles were tagged with PRTCL in the ICE-GB corpus, and they were tagged with RP in the tagged version of the TOEFL11 corpus.

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 is a kind of learner corpus which was released by the English Testing Service (ETS) in 2014.[9] The corpus contains the essays written by the EFL learners during the TOEFL iBT® tests in 2006-2007. This corpus includes a total of 1,100 essays per each of the 11 native languages (Arabic, Chinese, French, German, Hindi, Italian, Japanese, Korean, Spanish, Telugu, and Turkish). Therefore, the corpus consists of a total of 12,100 essays. All of the essays were taken from the TOEFL independent task, in which the test-takers were asked to write an essay in response to a brief writing topic. All the essays were sampled as evenly as possible from eight different topics. In addition, the corpus also provides the score levels (low/medium/high) for each essay. Among the essay texts in the corpus, 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 are included in the Korean component of the TOEFL11 corpus. These sentences were the targets of the investigation.

3.2 Encoding variables

After all the sentences with particles were extracted from two types of corpora, twelve linguistic factors were manually encoded to the extracted sentences, following Deshors (2010) and Deshors and Gries (2014). Table 2 shows the encoded linguistic factors.

Table 2. Encoded linguistic factors

ID Tag Type	ID Tag	ID Tag Levels
Length	LengthS	Length in Syllables
	LengthW	Length in Words
Morphology	Voice	active, passive
Syntax	Const	op, po

9 Strictly speaking, the TOEFL11 corpus is an archive, rather than a corpus, since no tagging and parsing information was encoded in the texts. However, it will be called a corpus in this paper, since an archive is also a corpus in a broad sense.

	NPType	lexical, pronominal, semi-pronominal
	Definite	definite, indefinite
	Complex	complex, intermediate, simple,
	PP	no, yes (following directional adverbials)
	Part=PP	no, yes
Semantics	Animacy	animate, inanimate
	Idiomacity	idiomatic, literal
	Concreteness	abstract, concrete

Following the study in Atkins (1987), each linguistic factor and its level were called ID tag and ID tag levels respectively. These variables were used in the statistical analysis.[10]

3.3 Statistical analysis

This paper took two types of statistical analyses. One was a Generalized Linear (Regression) Model (GLM) with logistic regression, and the other was Gries' BP analysis. These two methods were multifactorial in nature. There are several studies that mentioned the necessity of the multifactorial analyses. Among them, Langacker (2000: 3) mentioned that "to conceive of [linguistic] entities in connection with one

10 Among the linguistic factors included in Gries (2003), the following factors were not included in the o*perationalization* process of this study: *Register* (spoken vs. written), *Om* (overall frequency of the direct object), *Lm* (previous mention of the direct object), *Nm* (subsequent mention of the direct object), *Topm* (times of preceeding mention of the direct object), *Tosm* (times of subsequent mention of the direct object), *Dtlm*/*ActPC* (distance to last mention of the direct object), *Dtnm*/*ActSC* (distance to next mention of the direct object), *CohPC* (cohesiveness of the direct object to the preceeding discourse), and *CohSC* (cohesiveness of the direct object to the subsequent discourse). There were three reasons why these factors were not included in the study of this paper. First, the factor *Register* was excluded in this study, since the Korean component of the TOEFL11 corpus included only the written texts. Second, all the other linguistic factors except *Register* were the factors which were defined within the discourse contexts. Gries (2013) calculated the value of these factors within the boundary of 10 sentences before and after the target sentences. In sum, these discourse-related factors were defined within the 21 sentences. The problem was that the texts in the Korean component of the TOEFL11 corpus were too short to get reliable values from the essays. (This was the major reason why those linguistic factors were not included in this study.) Third, we thought that the sentence-internal linguistic factors in Table 1 would play important roles, in addition to the discourse-related factors. These three reasons were why the above (discourse-related) factors were not included in the analysis.

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."

This paper also took a multi-factorial approach and used a GLM in the statistical analysis, because it is one of the simplest and most widely-adopted analyses. For regression analysis, Deshors (2014: 11) mentioned that "[b]inary logistic regression is a confirmatory statistical technique that allows the analyst to identify possible correlations between the dependent and the independent factor/variables. Ultimately, this statistical approach allows us to see what factors influence learners' choices of alternation."

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 twelve linguistic factors and their interactions with L1. Second, a new model was constructed where one factor or one of the interactions was dropped from the previous model. Third, the new model was compared with the previous model using an ANalysis Of VAriance (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 ($p<.05$), it implied that the factor *f* or an interaction *i* must NOT be deleted from the model and the model m_1 was selected in this case, and (ii) when the *p*-value of the ANOVA was NOT significant ($p>.05$), it implied that the factor *f* or an interaction *i* can be safely deleted from the model and the model m_2 was selected in this case. The processes continued until all the factors and their interactions were exhausted.

This paper also used another multifactorial analysis, a Behavioral Profile (BP) analysis. The BP analysis is a statistical method which closely investigates the behavioral characteristics of each linguistic factor. As an analysis result, the analysis represents the similarity/dissimilarity of the components with a dendrogram (created by the hierarchical agglomerative cluster analysis). It was developed by Gries and Otami (2010) and Gries (2010a). This analysis method was originally invented to analyze the synonymy or the antonymy in lexical semantics. However, the method

11 The stepwise model selection procedure was included in other papers of the first author. However, it was mentioned here again for readers' convenience.

can also be adopted in this study, because the particle placement in the ENL speakers' writings and the Korean EFL learners' ones 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 *Const*(*ruction*) has one of the two values (*op* or *po*), a binary logistic regression is necessary. The first step for the logistic regression is to set up the initial model. After that, model selection procedure was applied (cf. 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 which was obtained after the model selection procedure.

Table 3. Initial model

Const~L1+LengthS+LengthW+Voice+NPType+Definite+Complex+PP+PartPP+Animacy+Idiomacity+Concreteness+L1:LengthS+L1:LengthW+L1:Voice+L1:NPType+L1:Definite+L1:Complex+L1:PP+L1:PartPP+L1:Animacy+L1:Idiomacity+L1:Concreteness

Table 4. Final model

Const~L1+LengthS+NPType+PP+Idiomacity+Concreteness+L1:NPType+L1:Idiomacity

As you can observe in Table 3 and Table 4, the six main factors and two interactions with L1 survived in the final model.

After the final model was obtained, 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)

	df	Deviance	AIC	LRT	*p*	
<none>		427.29	459.29			
L1	1	460.67	490.67	33.383	7.569e-09	***
LengthS	1	440.89	470.89	13.598	0.0002265	***
LengthW	1	429.22	459.22	1.929	0.1648431	
NPType	2	499.75	527.75	72.460	<2.2e-16	***
Definite	2	430.72	458.72	3.425	0.1804020	
Complex	2	430.49	458.49	3.197	0.2022482	
PP	1	462.89	492.89	35.597	2.426e-09	***
Part=PP	1	427.32	457.32	0.028	0.8670574	
Animacy	1	427.32	457.32	0.033	0.8556356	
Idiomacity	1	428.22	458.22	0.928	0.3355077	
Concreteness	1	440.84	470.84	13.550	0.0002323	***

Table 6. Analysis results (interactions)

	df	Deviance	AIC	LRT	*p*	
L1:LengthS	1	407.10	465.10	2.0580	0.15141	
L1:LengthW	1	405.07	463.07	0.0199	0.88793	
L1:NPType	2	411.64	467.64	6.5907	0.03705	*
L1:Definite	2	408.38	464.38	3.3307	0.18912	
L1:Complex	2	405.49	461.49	0.4403	0.80241	
L1:PP	1	406.67	464.67	1.6202	0.20306	
L1:Part=PP	1	405.05	463.05	0.0000	0.99998	
L1:Animacy	1	406.75	464.75	1.7012	0.19214	
L1:Idiomacity	1	409.56	467.56	4.5191	0.03352	*
L1:Concreteness	1	405.69	463.69	0.6415	0.42317	

Here, ' ' (not significant) is used when $p>0.1$; '.' (marginally significant) when $p<0.1$; '*' (significant) when $p<0.05$; '**' (very significant) when $p<0.01$; and '***' (highly significant) when $p<0.001$.

These two tables demonstrate that six main factors and two interactions with *L1* were statistically significant in the model. These tables also illustrate that the factor *Idiomacity* survives in the final model because of their interactions with the factor *L1*.

4.2 Linguistic factors with similar tendencies: Research question 1

Since the final model was obtained, it was necessary to investigate how each linguistic factor influenced the choice of alternation both in the Korean EFL learners' writings and the ENL speakers' counterparts. In what follows, we will examine some major findings with two kinds of graphic representations. The analysis in this section was based on the analysis results in Table 5.

The first linguistic factor which was examined was *L1*. Figure 1 shows us the association plot for *L1*.

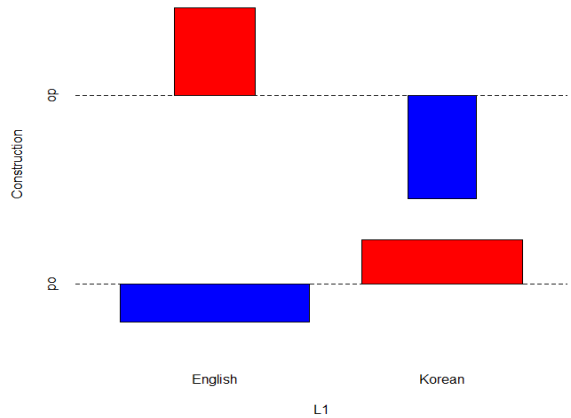

Figure 3. Association plot for *L1*

In the association plot, the effects of *L1* were 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 (here, *L1*). In addition, the width of the rectangle was proportional to the square root of the expected frequency, and the height of the rectangle is proportional to the standarized residual. As this association plot indicates, the ENL speakers (which are labeled as 'English') used more *op* constructions than the Korean EFL learners. In other words, this association plot illustrates that the Korean EFL learners (which are labeled as 'Korean') used less *op* constructions than the (British) ENL speakers.

All the other factors except *L1* were the linguistic factors that were significantly involved in the choice of the *op* constructions and the *po* constructions. Their tendencies were similar in both groups. That is, the tendencies were observed not

only in the (British) ENL speakers' writings but also in the Korean EFL learners'.

The first linguistic factor which showed the tendency was *LengthS*, which was the length of the direct object in syllable.

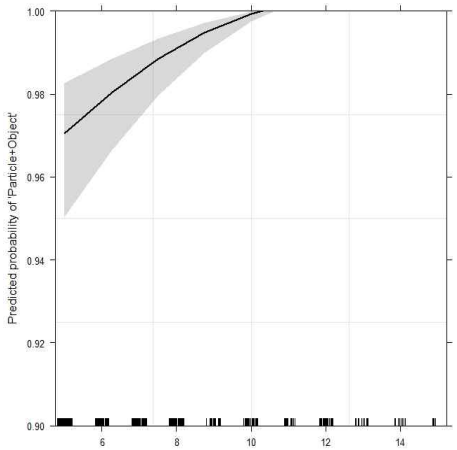

Figure 4. Effect plot for *LengthS*

As you can see in this effect plot, as the *LengthS* value increased, both types of speakers used the *po* constructions more frequently than the *op* constructions. For example, when the syllable length of the direct object was '5', the proportion of the *op* constructions was 0.97. Since the sum of these two constructions (*op* + *po*) must be 1.00 (100%), the proportion of the *po* constructions would be 0.03. However, as the *LengthS* value increased, the proportion of the *op* constructions decreased and the proportion of the *po* constructions increased.

The next linguistic factor which had to be examined was *NPType*, which was the type of the direct object. Three different kinds of NPs were considered here: lexical, pronominal, and semi-pronominal. If the direct object was a pronoun (including reflexives and reciprocals), the NP was encoded as 'pronominal.' If it was not a pronoun, the NP became a candidate of 'lexical.' If the NP contained a indefinite item (*someone* or *something*), it was encoded as 'semi-pronominal.' Figure 5 was the effect plot for this factor.

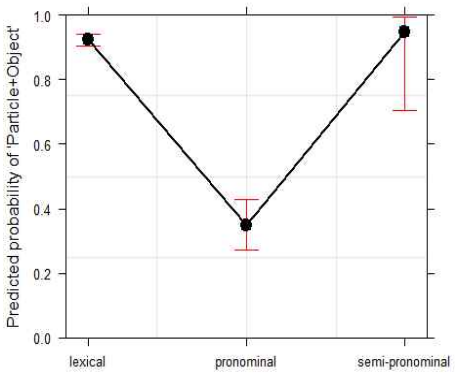

Figure 5. Effect plot for *NPType*

As you can observe, when the direct object contained a 'lexical' item, two groups of speakers used the *po* constructions more frequently. However, if the direct object was a 'pronominal', the *op* constructions were used more frequently. In addition, when the direct object included a 'semi-pronominal', the overall tendency was similar to that of 'lexical' and the *po* constructions were used more frequently.

The next factor was *PP*, which implied the directional adverbials were followed by the direct object or the particle. Figure 6 was the effect plot for this factor.

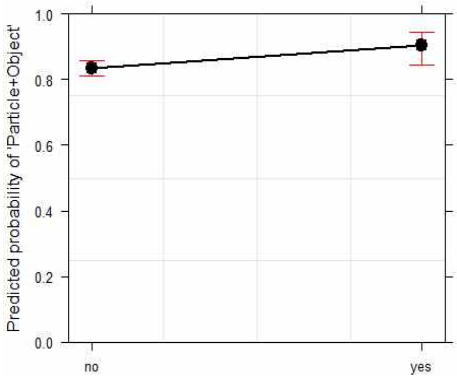

Figure 6. Effect plot for *PP*

As you can see, although the overall tendency was that the *po* constructions were used more frequently, the proportion of the *po* constructions was higher when the factor was encoded as 'yes' (which means that the given sentence contained the

directional adverbials). When the factor was encoded as 'no' (which means that the given sentence contained no directional adverbial), the proportion of the *po* constructions became slightly lower. That is, more *op* constructions were used when the factor *PP* was 'no.'

The final factor to consider was *Concreteness*, which implied whether the direct object referred to an abstract entity or a concrete entity. Figure 7 was the effect plot for this factor.

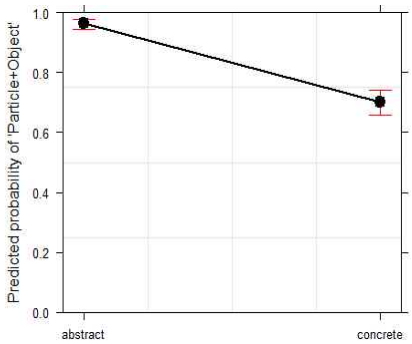

Figure 7. Effect plot for *Concereteness*

This effect plot demonstrated that the *po* constructions were employed more often when the direct object referred to an 'abstract' entity. When the factor was encoded as 'concrete' (which means that direct object referred to a 'concrete' entity), the proportion of the *po* constructions became slightly lower, while more *op* constructions were used.

4.3 Linguistic factors with different tendencies: Research question 2

As mentioned above, the above five linguistic factors were those which were applied to both groups of speakers. That is, all the other factors except *L1* were the linguistic factors that were significantly involved in the choice of alternation between the *op* constructions and the *po* constructions. Their tendencies were similar in both groups of speakers. In other words, the tendencies were observed not only in the (British) ENL speakers but also in the Korean EFL learners.

However, the following two linguistic factors showed the interactions with *L1*. That is, these linguistic factors were significantly different between two groups of

speakers. In other words, these two linguistic factors were the factors which made the non-native properties of the particle placement in the Korean EFL learners' writings.

The first linguistic factor which was examined was *L1*:*NPType*. Figure 8 shows us the association plot for *L1*:*NPType*.

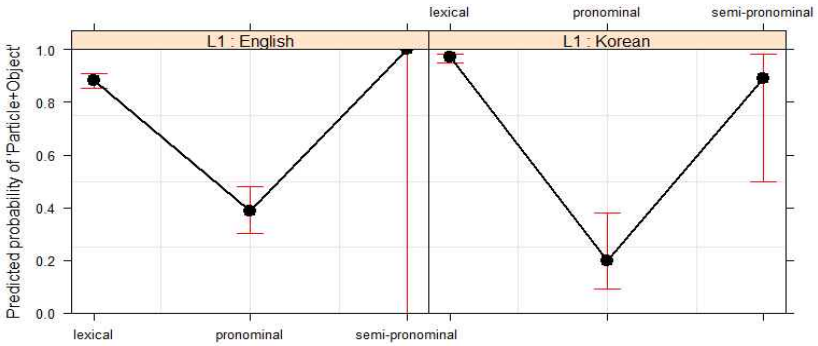

Figure 8. Effect plot for *L1:NPType*

The overall tendency was similar to that of Figure 5. However, the proportions that each construction occupies were different. When NPType is 'lexical', the Korean EFL learners adopted the *po* constructions more frequently. When NPType is 'pronominal', the (British) ENL speakers used the *op* constructions more frequently, compared with the Korea EFL learners. Finally, when NPType is 'semi-pronominal', the Korean EFL learners adopted the *po* constructions more frequently.

The next linguistic factor which was examined was *L1*:*Idiomacity*. Figure 9 shows us the association plot for *L1*:*Idiomacity*.

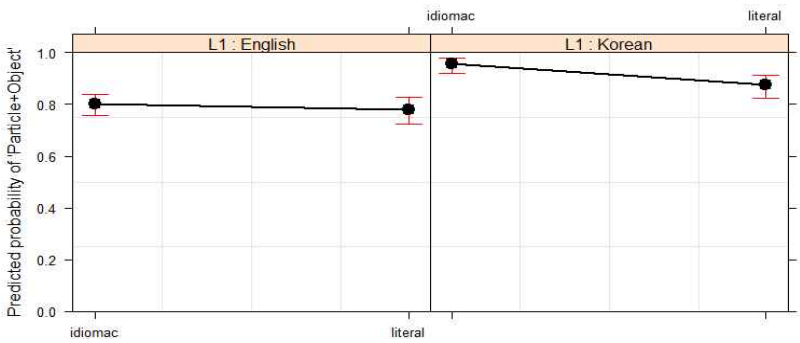

Figure 9. Effect plot for *L1:Idiomacity*

Though both groups of speakers used the *po* constructions when *Idiomacity* was 'idiomatic', the Korean EFL learners used the *po* constructions more frequently in both types of cases, compared with the tendencies of the (British) ENL speakers.

4.4 Goodness of fit

After an optimal model was constructed for the data, the goodness of fit of the model was calculated. Since a logistic regression was used in the analysis (cf. Section 4.1), *C*-statistic was used for comparison.

The *C*-statistics for the final model was 0.9378904. For the *C*-values, Harrell (2001: 248) mentioned that "*C*-values range from 0.5 to 1 and the higher the value, the better a regression model is at classifying or predicting the dependent variable; *C*-values $\geq$0.8 are generally considered good." Note that the *C*-value for our final model is 0.9378904. This suggests that our statistical model is very good for explaining the similarities and differences between the ENL speakers and the EFL learners.

5. The BP analysis: Research question 3

As Table 5 and Table 6 demonstrated, six main factors and two interactions with *L1* were statistically significant in the model (p<.05). These analysis results showed that the particle placement of Korean EFL learners were similar with those of English ENL speakers. Then, the naturally-occurring question was how much the tendency which Korean EFL learners demonstrated was similar those of English ENL speakers. To answer this question, a BP analysis was performed.

Among the linguistic factors in Table 3, the combination of *L1* and *Const* were chosen as a dependent variable and all the other factors plus *Verb* and *Particle* were used as independent variables. Figure 10 illustrates the dendrogram which resulted from the analysis (by multiscale bootstrap resampling clustering).[12]

---

12 We used Gries (2010b) in the actual analysis.

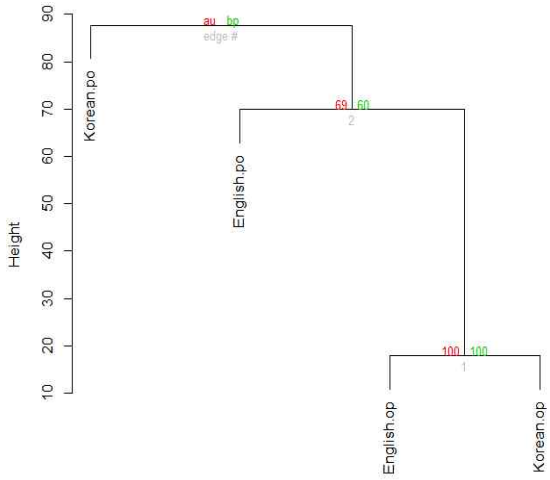

Figure 10. BP Analysis result

Here, the horizontal lines represent 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 AU (approximately unbiased) *p*-value and BP (bootstrap probability) value for each cluster.

As shown in Figure 10, *Korean.op* was combined with *English.op* first, which can be represented as {*Korean.op*, *English.op*}. This implies that the Korean *op* constructions was very close to the English *op* constructions. If the same reasoning had been applied, *Korean.po* would have been grouped with *English.po*, which can be represented as {*Korean.po*, *English.po*}. However, the result was that *English.po* was combined with {*Korean.op*, *English.op*}, which was represented as {*English.po*, {*Korean.op*, *English.op*}}. On the other hand, *Korean.po* formed another independent group by itself and it was combined with {*English.po*, {*Korean.op*, *English.op*}}. This implies that the behaviors of *Korean.po* were different from *English.po*, while *Korean.op* was closer to *English.op*.

6. Discussion

This paper adopted the concept of IL and Bates and MacWhinney's CM as two theoretical bases, and it constructed two statistical models using the corpus data. One

was for the (British) ENL speakers based on the corpus data from the ICE-GB, and the other was the Korean-English IL based on the data from the Korean component of the TOEFL11 corpus. In the constructed statistical models, linguistic factors and interactions competed for the choice of alternation. The linguistic environment was something like 'John picked ___ ___', which was the same linguistic environment both for (1a) and (1b). In this situation, twelve linguistic factor and their interactions with *L1* played either an advantageous or a disadvantageous role in the choice of alternation.[13] In order to statistically analyze the effects of each factor and interactions, two types of analyses were employed. The first one was regression analysis (GLM) and the second was the BP analysis.

In the regression analysis in Section 4, it was observed (i) that the ENL speakers used the *op* constructions more frequently than the Korean EFL learners, (ii) that six main factors significantly influenced the choice of the constructions, and (iii) that two linguistic factors had interactions with *L1*. The analysis results showed that the particle placement in the Korean EFL learners' writings were similar to that of the (British) ENL speakers.

The association plot in Figure 3 indicated that the ENL speakers used more *op* constructions than the Korean EFL learners. That is, the ENL speakers tended to put the direct object between the verb and the particle, more frequently than the Korean EFL learners. In another word, this association plot showed that the Korean EFL learners tended to avoid the separation of the particles from the co-occurring verbs than the ENL speakers. The reason for this avoidance seems to be originated from the language differences. Since Korean had no constructions which correspond to the *op* constructions, the Korean EFL learners were not familiar to this construction. Accordingly, they seemed to consider the 'verb + particle' structure as a single unit and to use the *po* constructions more frequently. However, as the level of proficiency went up, they acknowledged the fact that the *op* constructions were also possible and the frequency of the constructions increased continuously. The bar plot in Figure 2 demonstrated that this explanation was plausible, since the proportion of

---

13 For examples, the value 'abstract' in the factor *Concreteness* (Figure 7) played an advantageous role in the choice of the *po* constructions both in the ENL speakers and the Korean EFL learners. In other words, the value 'concrete' played a disadvantageous role in the choice of the *po* constructions in both groups of speakers. On the other hand, the value 'idiomatic' in the factor *L1:Idiomacity* (Figure 9) played more advantageous role in the Korean EFL learners' English than the ENL speakers' conterpart (in the choice of the *po* constructions).

the *op* constructions increased continuously, as the level of proficiency went up.

The next two factors *LengthS* and *NPType* could be related with 'end weight' and 'end focus', though they were two separate factors. According to Quirk *et al*. (1985; Chapter 18), a heavy element had to be located at the end of the sentence, which was mentioned as 'end weight.' That is, the longer a syntactic element was, the more probability the NP had that it was located at the end of the sentence. In our data, the same tendency was observed, as the effect plots in Figure 4 demonstrated. As the length of the direct object became longer, both groups of speakers employed the *po* constructions more frequently. This tendency accorded with the basic ideas of 'end weight', since a heavy NP went toward the end of the sentences.

On the other hand, Quirk *et al*. (1985) mentioned that the focused element had to be located at the end of the sentences. Usually, the elements with new information such as proper nouns or indefinite NPs became a focused element, while the elements with old information such as pronouns or definite NPs could not be a focused element. Then, the general tendencies in Figure 5 and Figure 8 could be accounted for naturally. When the direct object was a 'lexical' item, the object had new information, and the object had to be located at the end of the sentences. Accordingly, the *po* constructions were more natural. When the direct object was a 'pronominal', the object corresponded to old information, and the object could not be located at the end of the sentences. Accordingly, the *op* constructions were more natural. When the direct object was a 'semi-pronominal', the general tendency was similar to that of the 'lexical' entries. It seemed that the indefinite properties of the 'semi-pronominal' items made those items closer to a 'lexical' item, rather than a 'pronominal.'

However, the same tendency could also be explained with 'end weight.' Since 'pronominals' were usually shorter than the 'lexical' items and 'semi-pronominals', when the direct object was a 'lexical' item, the *po* constructions were more natural. However, when the direct object was a 'pronominal', the *po* constructions were more natural. When the direct object was a 'semi-prnominal' item, the *po* constructions were more natural, since they were usually longer than 'pronominal.'

The other main effects (*PP* and *Concreteness*) and one interaction with *L1* (*L1*:*Idiomacity*) could be explained with the Processing Hypothesis (PH). Based on the previous studies such as Givón (1982) and Siewierska (1988), Gries (1999: 313) proposed the following hypothesis for the syntactic alternation.

(3) The processing hypothesis (PH):
By choosing one of the two constructions for an utterance *U* (along the lines predicted by the consciousness hypothesis), a speaker *S* communicates his or her idea about the amount of consciousness required by subordinating to the different processing requirements of both constructions: *S* formulates
U in such a way that he triggers mental-processing instructions in the mind of the hearer *H* and simplifies the processing of *U* (including access to the referents within *U*).

The key idea of the PH was that *S* formulated *U* such that the mental-processing by *H* could be minimum. In the particle placement, since (i) *S* strived to communicate whatever they intend to communicate with as little effort as possible and (ii) the *po* constructions was inherently easier to process, they would use the *po* constructions in situations where the processing effort associated with the utterance is already high.[14]

Now, let's see how PH can explain the tendencies of particle placement in both groups of speakers. The first linguistic factor was *PP*, which indicated whether directional adverbials followed the direct object or the particle. Even though directional adverbials may consist of adverbials only, there were some cases where the prepositional phrases presented directions. If the prepositional phrases were combined with the *po* constructions (when the value for *PP* was 'yes'), the overall order would be 'particle + NP + preposition + NP', where the first NP was a direct object and the second NP was the object of the preposition. Note that the particles were adverbs or prepositions. Then, if the word order of the sentence was 'particle + NP + preposition + NP', it would be easier for speakers or hearers to process the sentences, since similar patterns iterated in the sentences. If the prepositional phrases were combined with the *op* constructions (when the value for *PP* was 'no'), the overall order would be 'NP + particle + preposition + NP.' In this case, a different pattern appeared, and it made difficult for speakers or hearers to process the

[14] The *po* constructions are easier to process, since the verb and the particle are adjacent each other and they form a single sense unit. In the *op* constructions, the direct object goes between the verb and the particle. The verb and the particle are not adjacent, though they have to form a single semantic unit. Accordingly, the *op* constructions requires more mental process, which makes the construction not easier to process.

sentences. That's why the *po* constructions were preferred.

Now, let's go to the factor *Concreteness*. This factor had two values: 'abstract' and 'concrete.' Which one was easier to process? Of course, the latter was easier than the former. That's why the proportion of *po* constructions was higher when the direct object referred to the 'abstract' entity.

The interaction *L1*:*Idiomaciry* was similar. This factor also had two values: 'idiomatic' and 'literal.' Which one was easier to process? Of course, the latter was easier than the former. That's why the proportion of *po* constructions was higher when the phrasal verbs were used with the idiomatic usage.

Now, let's go to the BP analysis results. Figure 10 demonstrated that *Korean.op* was combined with *English.op* first, *English.po* was combined with {*Korean.op*, *English.op*}, and *Korean.po* combined with {*English.po*, {*Korean.op*, *English.op*}}. That is, the final result was {*Korean.po*, {*English.po*, {*Korean.op*, *English.op*}}}. This implies that the behaviors of *Korean.po* were different from *English.po*, while *Korean.op* closer to *English.op*. This analysis result illustrated that more complex mechanisms might be involved in the choice of alternation. Accordingly, further studies are necessary to investigated which linguistic factors were involved in the choice of alternation.

Finally, let's go to the research questions in Section 1. In the introductory section, the following questions were asked. Let's see how the analysis results in this paper answered to these questions. For the first question, Table 5 provided the answers. As observed in this table and Section 4.1, five linguistic factors (*L1*, *LengthS*, *NPType*, *PP*, and *Concreteness*) were involved with the choice of alternation. These factors were applied to both the ENL speakers and the Korean EFL learners. Their effects were illustrated with effect plots in Section 4.1. For the second question, Table 6 provided the answers. As observed in this table and Section 4.2, two linguistic factors (*NPType* and *Idiomacity*) significantly interacted with *L1*. These interactions made the Korean EFL learners' use of particle placement different from the ENL speakers' counterparts. Their interactions with *L1* were illustrated with effect plots in Section 4.2. For the third question, Figure 10 provided the answers. As observed in this figure, the behaviors of *Korean.po* were different from *English.po*, while *Korean.op* was closer to *English.op*.

7. Conclusion

In this paper, it was statistically investigated how various linguistic factors influenced the choice of particle placement both in the Korean EFL learners' English and the (British) ENL speakers' counterpart. Two corpora (the written part of the ICE-GB corpus and the Korean component of the TOEFL11 corpus) were selected in this study. After all the sentences with particles were extracted from the corpora, twelve linguistic factors were manually encoded into each sentence. Then, a GLM and a BP analysis were applied, and it was statistically analyzed which factors played a role in deciding on the choice and how they affected the choice in the two different groups of speakers. How various linguistic factors influenced the choice of alternation in particle placement was closely examined through the effect plots.

Through the analysis, the following facts were observed: (i) the Korean EFL learners used the *po* constructions more frequently than the ENL speakers, (ii) five main factors and two interactions with *L1* were statistically significant, and (iii) the behaviors of *Korean.po* were different from *English.po*, while *Korean.op* was closer to *English.op*. It was also observed that the linguistic behaviors of two groups of speakers could be explained with two principles (and-focus and end-weight) and processing hypothesis.

The analysis procedure and the analysis results in this paper 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 analysis results also showed that it was possible to statistically model and analyze the linguistic behaviors with different L1 backgrounds, along with this kind of statistical model. We hope that the developments of statistical tools like the above will ultimately put the findings within our discipline on a more solid foundation.

References

Adjemian, Christian. 1976. On the nature of Interlanguage systems. *Language Learning* 26(2): 297-320.

Atkins, Beryl. 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. Waterloo: University of Waterloo.

Bates, Elizabeth and Brian MacWhinney. 1982. Functionalist approaches to grammar. In Eric Wanner and Lila Gleitman (eds.), *Language acquisition*: *The state of the art*, 73-218. Cambridge: Cambridge University Press.

Bates, Elizabeth and Brian MacWhinney. 1989. Functionalism and the competition model. In Brian MacWhinney and Elizabeth Bates (eds.), *The cross-linguistic study of sentence processing*, 3-73. Cambridge: Cambridge University Press.

Bialystok, Elixabeth and Michael Sharwood Smith. 1985. Interlanguage is not a state of mind: An evaluation of the construct for second language acquisition. *Applied Linguistics* 6: 101-117.

Chomsky, Noam. 1957. *Syntactic structures*. Berlin: Mouton.

Chomsky, Noam. 1965. *Aspect of the theory of syntax*. Cambridge, MA: The MIT Press.

Deshors, Sandra. 2010. *Multifactorial study of the uses of may and can in French-English interlanguage*. PhD Dissertation, University of Sussex.

Deshors, Sandra. 2014. Constructing meaning in L2 discourse: The case of modal verbs and sequential dependencies. In Dylan Glynn and Mette Sjölin (eds.), *Subjectivity and epistenicity: stance strategies in discourse and narration*, 329-348. Lund: Lund University Press.

Deshors, Sandra and Stefan Gries. 2014. A case for the multifactorial assessment of learner language: The uses of may and can in French-English interlanguage. In Dylan Glynn and Justyna Robinson (eds.), *Corpus methods for semantics: quantitative studies in polysemy and synonymy*, 179-204. Amsterdam: John Bejamins.

Givón, Talmy. 1992. The grammar of referential coherence as mental processing instructions. *Linguistics* 30: 5-55.

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

Gries, Stefan. 1999. Particle movement: A cognitive and functional approach, *Cognitive Linguistics* 10(2): 105-145.

Gries, Stefan. 2001. A multifactorial analysis of syntactic variation: Particle movement revisited. *Journal of Quantitative Linguistics* 8(1): 33-50.

Gries, Stefan. 2010a. Behavioral profiles: A fine-gained and quantitative approach in corpus-based lexical semantics. *The Mental Lexicon* 5(3): 323-346.

Gries, Stefan. 2010b. Behavioral profiles 1.01: A program for R 2.7.1 and higher.

Gries, Stefan and Naoki Otani. 2010. Behavioral profiles: A corpus-based perspective on synonymy and antonymy. *ICAME Journal* 34: 121-150.

Hanks, Patrick. 2000. Do word meanings exist? *Computers and the Humanities* 34(1/2): 205-215.

Harrell, Frank. 2001. *Regression modeling strategies with applications to linear models, logis-*

tic regression, and survival analysis. Berlin: Springer.

Jarvis, Scott. 2000. Methodological rigor in the study of transfer: Identifying L1 influence in the Interlanguage lexicon. *Language Learning* 50(2): 245-309.

Langacker, Ronald. 2000. *Grammar and conceptualization*. Berlin: Mouton.

Lee, Gyu-Hyeong, Ha-Eung Kim, and Yong-hun Lee. 2015. A Multifactorial Analysis of English Particle Movement in Korean EFL Learners' Writings. *Proceedings of the 29th Pacic Asia Conference on Language, Information, and Computation* (*PACLIC-29*), 116-124. Shanghai, China.

Lee, Yong-hun. 2007. *Corpus analysis using NLPTools and their applications*: *Applications to linguistic research, English education, and textbook evaluation*. Seoul: Cambridge University Press.

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

Nelson, Gerald, Sean Wallis, and Bas Aarts. 2002. *Exploring natural language*: *Working with the British component of the international corpus of English*. Amsterdam: John Benjamins.

Odlin, Terence. 1989. *Language transfer*. Cambridge: Cambridge University Press.

Odlin, Terence. 2003. Cross-linguistic Influence. In Catherine Doughty and Michael Long (eds.), *The Handbook of Second Language Acquisition*, 436-386. New York: Blackwell.

Quirk, Randolph, Sidney Greenbaum, Geoffrey Sidney, and Jan Svartvik. 1985. *A comprehensive grammar of the English language*. Harlow: Longman.

R Core Team. 2016. *R*: *A language and environment for statistical computing*. Vienna: R Foundation for Statistical Computing.

Selinker, Larry. 1969. Language transfer. *General Linguistics* 9(2): 67-92.

Siewierska, Anna. 1988. *Word order rules*. New York: Croom Helm.

Yoon, Tae-Jin and Yong-hun Lee. 2016. A multifactorial analysis of can and may in Korean EFL learners' writings. *English Language and Linguistics* 22(1): 117-136.

Yong-hun Lee
Department of English Language and Literature
Chungnam National University
99 Daehak-ro, Yuseng-gu, Daejeon 34134, Korea
E-mail: yleeuiuc@hanmail.net

Ha-Eung Kim
Department of English Language and Literature
Hannam University
70 Hannam-ro, Daedeok-gu, Daejeon 34430, Korea
E-mail: tankkh@hanmail.net

Gyu-Hyeong Lee
Department of English Language and Literature
Hannam University
70 Hannam-ro, Daedeok-gu, Daejeon 34430, Korea
E-mail: gyuhyung73@nevar.com

Received: 2016. 07. 13.
Revised: 2016. 09. 20.
Accepted: 2016. 09. 20.