

A Corpus-based Study of the Use of English Articles by Korean EFL Learners*

Daria Soon-Young Seog
(Kyungpook National University)

Seog, Daria Soon-Young. (2018). A corpus-based study of the use of English articles by Korean EFL learners. *The Linguistic Association of Korea Journal*, 26(1), 57-73. The current paper reports on the investigation of Korean college EFL learners' use of the English articles: *the*, *a(n)*, and \emptyset (the zero or null article). A comparison analysis was conducted on the researcher compiled learner corpus, the Kyungpook National University Student English Learner Corpus-Writing (KSEL-C-W), and a native speaker (NS) corpus, the Corpus of Contemporary American English (COCA). For this study, the learner corpus, KSEL-C-W, consisted of 145 writing samples from 74 university students which were collected over three semesters. Log-likelihood calculations showed that the Korean learners in this study significantly overused *the* and significantly underused *a(n)* and \emptyset compared to the NSs of English. According to Huebner (1983a), the students in this study are in Stage 2 of the six stages in the definite article acquisition process where *the*-flooding is marked and overgeneralization occurs. Furthermore, the students in this study exhibit L2 article production behavior consistent with Stage 5, *the* > *a(n)* > \emptyset , of the five-stage developmental sequence in L2 frequency of article production proposed by Wolfe-Quintero (2000). The current study sheds some light on conducting corpora comparison analyses to supplement existing research on the L2 English article acquisition process.

Key Words: English articles, learner corpus, second language acquisition, comparison of corpora, Korean EFL learners

* The preliminary results of this study were presented on May 27th 2017 at the joint spring conference hosted by the Linguistic Association of Korea and the Society of Modern Grammar.

1. Introduction

With corpus-based studies over the last few decades leading to much better descriptions of the target language being investigated, Second Language Acquisition (SLA) and Foreign Language Teaching (FLT) researchers have been compiling learner corpora to conduct a wide range of linguistic analyses. Furthermore, corpus comparisons have provided a means for researchers to carry out quantitative and qualitative comparisons on large amounts of data between native speakers (NS) and non-native speakers (NNS) or between groups of NNS with different first languages (L1) for Contrastive Interlanguage Analysis (CIA) (Granger, 2013). Such comparisons have shed light on learner specific features as well as developmental characteristics of specific learner groups.

Having the knowledge and the tools to work with computer learner corpora (CLC) and realizing its enormous potential, second language (L2) researchers have been utilizing learner corpora to further explore and obtain insights into longstanding concerns. Among them, the English article system has been well established as one of the most difficult features for English as a Second Language (ESL) or English as a Foreign Language (EFL) learners to acquire regardless of their L1 or proficiency level (Ionin, Ko, & Wexler, 2004; Zdorenko & Paradis, 2008; among others). Especially for the L2 English learners whose L1 lacks an equivalent article system, acquiring English articles tends to be even more difficult (Ionin, Ko, & Wexler, 2004; Ionin & Montrul, 2010; Lee, 2012; Crosthwaite, 2016).

Regardless of the vast number of studies on the acquisition of the English article system by L1 and L2 English learners, there exists a lack of research comparing NS and NNS data on article use. According to Hunston (2002), by comparing the corpora of NNS with NS, a researcher can identify the learners' underuse and or overuse of a linguistic feature and examine how far and in what ways the learners deviate from the NS norms. Therefore, the current study investigates the use of English articles by Korean EFL learners by comparing a researcher compiled learner corpus with an NS corpus. Moreover, this study shows how corpora comparison studies may support existing research studies using results based on Supplied in Obligatory Contexts (SOC), Target-Like Use (TLU), and Used in Obligatory Contexts (UOC) calculations for analyses and interpretation.

2. Literature Review

2.1. The English article system

The English article system consists of the indefinite article *a(n)*, the definite article *the*, and \emptyset (the *zero* and *null*) article.¹⁾ According to Sinclair (1991), in the early 1991 version of the COBUILD corpus of 17.9 million words, *the* with a frequency rate of 25.1% was the most frequently occurring word occurrences surpassing *of* (12.6%), *and* (12.5%), and *to* (11.1%) with the indefinite article *a(n)* as the fifth most frequent item at 10.5%. As for the \emptyset article, Master (1993) reported that the \emptyset article occurred the most among the articles resulting in a frequency order of $\emptyset > the > a(n)$ (48.0% > 36.3% > 15.7%) from a learner corpus of 197,644 words. For further supporting evidence, Crosthwaite (2016) reported confirmed observations of articles as follows: *the* was 5,189, *a(n)* was 3,488, and \emptyset was 12,452. The resulting article frequency order was $\emptyset > the > a(n)$ with raw frequencies of 12,452, 5,189, and 3,488 respectively. The articles investigated were from the selection gathered by Crosthwaite from the written version of the International Corpus Network of Asian Learners of English (ICNALE; Ishikawa, 2011, 2013). Whether it is an NS corpus or a learner corpus, the English article raw frequency order was $\emptyset > the > a(n)$ (\emptyset occurred the most followed by *the* and last *a(n)*) of within the investigated corpus.

Even though articles are among the most commonly occurring words and learners encounter them very early in the L2 English learning stage, articles are one of the most difficult structural elements for ESL/EFL learners to acquire. Especially for article-less L1 learners of English, achieving mastery of the English article system may be an impossibility. Lu (2001) stated that L2 learners' English article acquisition difficulty basically stems from the fact that article choice is complicated due to stacking of multiple functions onto a single morpheme, being context-specific, and being unstressed non-salient function words with little lexical meaning. As a result, researchers are continuously investigating and examining article accuracy and usage patterns for essential information and knowledge for underlying processes and possible pedagogical implications.

1) For the purpose of this research, the \emptyset article will not be differentiated into two type: *zero* and *null*.

2.2. Article Use in Identifying Contexts

Many English article acquisition studies have used Huebner's (1983a) model, which was adopted from Bickerton's (1981) semantic wheel (shown in Figure 1), for analysis of article use in identifying noun phrase (NP) environments. According to Huebner, early morpheme studies (Brown, 1973; Dulay & Burt, 1973, 1974; Bailey, Madden, & Krashen, 1974; Larsen-Freeman, 1975) that inspected only obligatory contexts were not sufficient nor refined enough for detecting morpheme use variations in an interlanguage continuum and were incapable of providing a complete picture of the acquisition processes.

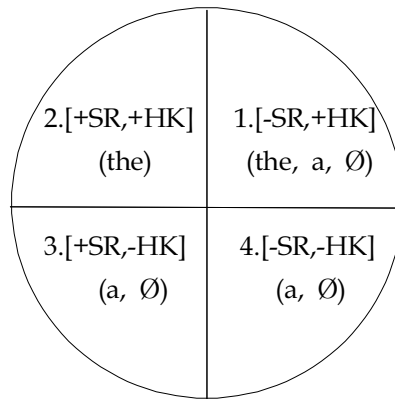


Figure 1. Bickerton's semantic wheel
(from Huebner, 1983a)

Huebner's model classifies pre-noun contexts in terms of four semantic categories, [\pm Specific Referent (\pm SR)] and [\pm Assumed Known to the Hearer (\pm HK)] as shown in Table 1 with examples from Master's (1994) test items. Huebner uses his classification to investigate and identify the interlanguage development of his subject in regard to the use of the definite article.

Table 1. Environments and examples for the semantic categories [\pm SR, \pm HK] (Lu, 2001)

Category	Article	Environment	Example (Master, 1994)	No.
1. [-SR, +HK]	<i>the, a,</i> \emptyset	Generics	The favorite food of <i>the</i> jaguar is <i>the</i>	48
			wild pig.	49
			\emptyset Wild pigs move in bands of fifteen to twenty.	50
2. [+SR, +HK]	<i>the</i>	Unique, previously mentioned, or physically present referents	What is <i>the</i> diameter of the moon?	8
			Once there were many trees here.	9
			Now, <i>the</i> trees are gone.	10
3. [+SR, -HK]	<i>a, \emptyset</i>	First-mention NPs, or NPs following existential 'has/have' or 'there is/are'	<i>The</i> air in this city is not very clean.	16
			I would like <i>a</i> cup of coffee, please.	11
			I always drink \emptyset water with my meals.	5
4. [-SR, -HK]	<i>a, \emptyset</i>	Equative NPs, or NPs in negation, question, or irrealis mode	There is <i>an</i> orange in that bowl.	1
			What is the sex of your baby? It's <i>a</i> boy!	4
			Einstein was <i>a</i> man of great intelligence.	17

2.3. Article Acquisition Processes

One of the ground-breaking in-depth longitudinal study of L2 English acquisition is the naturalistic study conducted by Huebner (1983a) where he observed an adult Hmong speaker, Ge, with basic-level English proficiency for over one year. Huebner investigated the use of the definite article *da* by Ge and identified six stages in Ge's learning trajectory of marking *da* for the different NP environments. Through Ge's interlanguage development, Huebner's finding provided evidence of systematic variability in article use although it could not be concluded that the learning trajectory could be universal. Table 2 represents the six stages of the definite article acquisition trajectory identified by Huebner with Ge's production and corresponding NP environments.

Table 2. Six stages of definite article acquisition process identified by Huebner (1983a)

Stage	Observed behavior in NP environment
Stage 1	Ge used <i>da</i> with [+SR, +HK] NPs
Stage 2	Ge overgeneralized <i>da</i> to all NPs; <i>the</i> -flooding marked Stage 2
Stage 3	Ge screened <i>da</i> out of the [-SR, -HK] contexts that share no semantic function with the feature [+SR, +HK] but still retained the use of <i>da</i> in the other three contexts
Stage 4	Ge realized that the feature [\pm HK] served as a primary distinction for <i>da</i> marking and restricted the use of <i>da</i> with [+HK] NPs
Stage 5	Ge tested his hypothesis by using <i>da</i> again with [+SR, -HK] NPs, except the existential <i>haev(a)</i> (have a) constructions
Stage 6	Ge rejected his previous hypothesis and returned to the rule governing Stage 4 to using <i>da</i> for [+HK] NPs only

Following Huebner's longitudinal study, researchers reported supporting evidence and reached a consensus that *a* is acquired later than *the* in regard to L2 English article acquisition processes (Parrish, 1987; Master, 1987; Thomas, 1989; Chaudron & Parker, 1990). Furthermore, even though there were some discrepancies in the findings, the studies revealed that variation in article use showed some systematic patterns and thus correlating both article accuracy and frequency with L2 learner's English proficiency (Master, 1987; Thomas, 1989; Chaudron & Parker, 1990). In addition, Table 3 represents Wolfe-Quintero's (2000) five-stage developmental sequence in L2 frequency of article production based on the frequencies revealed in Kim's (2000) data from Korean learners of English as referenced by Lu (2001).

Table 3. Wolfe-Quintero's five-stage developmental sequence in L2 frequency of article production

Stage	Developmental Sequence of L2 Article Production Frequency	
Stage 1	$\emptyset > the > a$	\emptyset is the most frequent article, followed by <i>the</i> , and <i>a</i> is used only occasionally.
Stage 2	$\emptyset = the > a$	Use of <i>the</i> is frequent enough to compete with use of \emptyset .
Stage 3	$the > \emptyset > a$	<i>the</i> becomes the most frequent, and <i>a</i> is still the least frequent.
Stage 4	$the > \emptyset = a$	Use of <i>a</i> increases and appears to be as frequent as use of \emptyset .
Stage 5	$the > a > \emptyset$	<i>a</i> becomes more frequent than \emptyset , but still less frequent than <i>the</i> .

2.4. Research Questions

The current study compares the NNS learner corpus, the Kyungpook National University (KNU) Student English Learner Corpus-Writing (KSELC-W), with the NS corpus, the Corpus of Contemporary American English (COCA) and conducts Rayson and Garside's (2000) log-likelihood calculations to answer the following questions:

- (i) Which English articles, *the*, *a(n)*, and \emptyset (the *zero* article) do the Korean EFL learners in this study underuse and or overuse compared to the NS?
- (ii) How far, and in what ways, do the Korean EFL learners deviate from the NS?
- (iii) What is the developmental stage of the Korean EFL learners in this study in reference to the L2 English article acquisition process?
- (iv) Do the findings provide insights into the significance of conducting NS/NNS corpora comparison studies?

3. Research Methodology

3.1. Corpora Used for the Study

The present study is based on two corpora: KSELC-W and COCA. KSELC-W, the KNU Student English Learner Corpus-Writing, is an NNS learner corpus compiled by the researcher.²⁾ KSELC-W consists of writing samples collected during three consecutive semesters, from March 2011 to June 2012. The writing samples were written during in-class timed writing sessions by KNU English education major students who were at the time juniors and seniors. From 74 students, a total of 145 handwritten writing samples were collected and transformed into an electronic form to create KSELC-W.

Although KSELC-W is a small learner corpus with 65,787 words, it is a valid learner corpora that complies with the definition of corpora which Granger (2013) suggests adopting to avoid including inappropriate data types. Granger's definition of corpora is as follows³⁾:

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

Table 4 provides detailed descriptions of the learner corpus KSELC-W. The descriptive statistics also show that the number of the distinct words in the learner corpus is 3,576. Additionally, for the 145 writing samples, the mean text length is 453.12 words per writing sample with the mean sentence length of 16.73 words per sentence.

2) The compilation of the Kyungpook National University Student English Learner Corpus-Writing (KSELC-W) is an ongoing process. The researcher is continually collecting and processing additional writing data to increase the corpus size.

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

Table 4. Descriptive statistics for the learner corpus KSELC-W

Corpus	KSELC-W
Corpus type	Learner corpus: NNS interlanguage
Type of texts	Essays
Setting	Timed in-class writing
Data collection	March 2011 - June 2012
Learner type	[-Article] L1 Korean EFL learners
Level of learners	Juniors and seniors at Dept. of English Education, Teacher's College
Number of learners	74
Number of text files	145
Number of distinct words	3,576
Mean text length in words	452.12
Mean sentence length in words	16.73
Total number of tokens	65,787

Table 5. Basic descriptive statistics of COCA

Years	Spoken	Fiction	Magazine	Newspaper	Academic	Total
1990-2015	109,391,643	104,900,827	110,110,637	105,963,844	103,421,981	533,788,932

Table 5 shows a simple description of COCA which is the NS comparison corpus utilized in this study. The Corpus of Contemporary American English (COCA) is an NS corpus created by Mark Davies, Professor of Corpus Linguistics at Brigham Young University. COCA is comprised of five subcorpora (spoken, fiction, popular magazines, newspapers, and academic texts) and the data used for this study was collections from 1990 to 2015 consisting of a total of 220,225 text files and 533,788,932 words.

3.2. Software Tools

For the purpose of analyzing and examining the text in the corpora, a text retrieval software WordSmith Tools version 7.0 is utilized in the present study for its valuable features.⁴⁾ The 'word list' function generates a complete list of words with their frequencies for analysis while the 'concord' function allows for

targeted word, phrase, and or tagged item searches. In addition, WordSmith Tools automatically computes descriptive statistics for the data being analyzed.

Knowing the value of using part-of-speech (POS) tagged learner corpora, the corpus annotation software, CLAWS (the Constituent Likelihood Automatic Word-tagging System) developed by Unit for Computer Research on the English Language (UCREL) at Lancaster University was used for POS tagging or grammatical tagging of KSELC-W. The C5 tagset which was used to POS tag the British National Corpus (BNC) and has just a little over 60 tags was used. With the conveniently POS tagged corpus, the researcher was able to search and retrieve all English articles in their original contexts using the 'concord' function and concordancing on the POS tags.

Finally, UCREL's online LL Calculator for computing log-likelihood (LL) values and Bayes Factor was used as the statistical analysis software for this study. According to Rayson and Garside (2000), the log-likelihood test can be used for corpora comparison research and is more reliable than the Pearson's chi-squared test. They also stated that the chi-squared value becomes unreliable with very low and very high frequency words as well as when comparing a relatively small corpus to a much larger one. Correspondingly, LL calculation is more appropriate for the current study investigating very high frequency words which are articles by comparing a small learner corpus, KSELC-W, with a very large NS corpus, COCA.

4. Results and Discussion

4.1. Occurrence Frequencies of English Articles

The English article system consists of the indefinite article *a(n)*, the definite article *the*, and \emptyset (the *zero* and *null*) article, and they are some of the most frequently occurring words in the English language. As seen in Table 6, although \emptyset article is not listed in the word list and thus does not have a rank, it can be seen immediately from the observed occurrences that \emptyset article occurs

4) See the WordSmith Tools Manual for a complete explanation of all the different functions it is capable of conducting.

the most frequently in both the learner corpus, KSELC-W, and the NS corpus, COCA. Next, from the word list with corresponding frequency and rank, *the* is identified as the most occurring word in both corpora. Finally, *a(n)* is ranked 8th in KSELC-W and 5th in COCA. Obviously, Table 6 provides evidence in confirming that articles are among the most frequently used in the English language. Determined by the Frequency rates and the observed raw frequencies in Table 6, the article frequency order for both KSELC-W and COCA are $\emptyset > the > a(n)$. This article frequency order is the same order reported by Sinclair (1991), Master (1993), and Crosthwaite (2016).

Table 6. Raw frequencies, rank, and frequency rates of English articles in KSELC-W and COCA

Article	KSELC-W	Rank	Frequency%	COCA	Rank	Frequency%
The	4,792	1	7.28%	28,674,422	1	5.37%
A(n)	969	8	1.47%	13,379,957	5	2.51%
\emptyset	6,160		9.36%	74,133,216		14.89%
Total	11,921			116,187,595		

In order to compare the use of articles between NNS and NS, WordSmith Tools was used to extract the occurrence frequencies in the corpora. For comparative study of corpora with different sizes, the raw statistics should be normalized. Therefore, the observed original frequencies should be normalized using the following formula⁵⁾:

$$\text{Normalized Frequency } F_N = \frac{\text{Observed Frequency } F_0}{\text{Corpus Size } C} \times 100,000^*$$

* Normalizing per one hundred thousand words

Table 7 shows the observed and normalized frequencies of English article occurrences in KSELC-W and COCA.

5) Normalization formula retrieved from Department of Linguistics and Modern English Language, Lancaster University under the heading of Comparing frequencies for corpora of different sizes at https://www.lancaster.ac.uk/fss/courses/ling/corpus/blue/105_3.htm

Table 7. Occurrence frequencies of English articles in KSELC-W and COCA

Articles	Observed Frequency		Norm Frequency*	
	KSELC-W	COCA	KSELC-W	COCA
The	4,792	28,674,422	7,284	5,372
A(n)	969	13,379,957	1,473	2,507
∅	6,160	74,133,216	9,364	13,888
Total	11,921	116,187,595	18,121	21,767

* Frequency results normalized per 100,000 words and rounded to the nearest single digits

As shown above in Table 7, the normalized frequency of the per 100,000 words is 7,284 in KSELC-W compared to 5,372 in COCA; the frequency of a(n) per 100,000 words is 1,473 in KSELC-W compared to 2,507 in COCA; and the frequency of ∅ per 100,000 words is 9,364 in KSELC-W compared to 13,888 in COCA. In other words, the occurred more times in KSELC-W compared to COCA and both a(n) and ∅ occurred less times in KSELC-W compared to COCA. Moreover, the normalized frequency of article occurrences in total per 100,000 words is 18,121 in KSELC-W compared to 21,767 in COCA; in general, NNS used articles less times than NS.

4.2. Statistical Analyses

For statistical significance testing, log-likelihood values were calculated to compare the article frequencies between the two corpora as shown in Table 8.

Table 8. Log-likelihood (LL) and Bayes Factor results for English articles in KSELC-W and COCA

Articles	KSELC-W	COCA	LL*	Bayes Factor
The	4,792	28,674,422	+ 405.63**	385.53
A(n)	969	13,379,957	- 328.64**	308.54
∅	6,160	74,133,216	- 1100.11**	1080.01
Total	11,921	116,187,595	- 421.90**	401.80

* The log-likelihood value is always a positive number. The UCREL log-likelihood wizard by Rayson inserts '+' for overuse and '-' for underuse of corpus 1 (KSELC-W) relative to corpus 2 (COCA).

** LL > 15.13 is significant at p < 0.0001 level (also called the 99.99% level)

*** Bayes Factor > 10: very strong evidence against H0

In answering the first research question, the log-likelihood calculations revealed that the Korean EFL learners significantly overused *the* compared to the NS of English with LL=405.63. In contrast, the Korean EFL learners significantly underused both *a(n)* with LL=328.64 and \emptyset with LL=1389.19 compared to the NS of English.

As for the second research question, among the English articles, *the*, *a(n)*, and \emptyset (the *zero* article), the Korean EFL learners' underusage of \emptyset article deviated the most from the that of NS. Although *a(n)* was also underused by the Korean EFL learners, it deviated the least. The most interesting article was *the* since it was the only overused article compared to the NS usage data. The statistically significant overuse of *the* in the writing samples of KSEL-CW compared to the occurrences of *the* in COCA with significant underuse of *a(n)* and \emptyset indicate *the*-overgeneralization.

Consequently, the current findings are in support of previous research (Huebner, 1983a; Andersen, 1977; Master, 1987; Chaudron & Parker, 1990; Lu, 2001) on the existence of *the*-flooding. Furthermore, in answering the third research question, the Korean EFL learners are identified to be in Stage 2 of the six developmental stages defined by Huebner (1983a) in reference to the definite article acquisition process. Stage 2 is marked by *the*-flooding and *the*-overgeneralization to all NP environments.

However, in regard to the L2 English article acquisition processes, LL values calculated from the corpora comparison of KSEL-CW and COCA indicate that the Korean EFL learners' L2 article usage hierarchy is *the* > *a(n)* > \emptyset compared to the NS usage. This pattern is consistent with Stage 5, *the* > *a(n)* > \emptyset , of Wolfe-Quintero's (2000) proposal of the five-stage developmental sequence. In this light, the Korean EFL learners observed in this study are beyond the \emptyset article flooding stage and are currently going through the *the*-overgeneralization stage with the expectation of going through the *a(n)*-overgeneralization stage afterwards.

Responding to the fourth research question, the findings of the present study supported results from traditional studies dealing with obligatory contexts and targetlike use. In addition, the results evidenced in this study revealed that NS-NNS corpora comparison studies with large natural data can provide insight and supplement existing language acquisition research. Subsequently, examining

overuse and underuse patterns and relations resulting from comparisons of learner and NS corpora can confirm developmental sequences in L2 acquisition processes.

5. Limitations & Conclusion

The results of the current study revealed that NS-NNS corpora comparison studies can provide insight as well as supplement existing language acquisition research. However, the researcher needs to develop the ability to annotate the corpora accurately with details easily extractable when comparing different corpora. Dealing with a large number of occurrences, accurate annotation can tremendously expedite the data analysis process.

Another very important limitation of the present study is that the writing samples are collected from learners of only one proficiency level. However, the writing sample collection process for the KSELC-W corpus is an ongoing project, and the corpus will eventually consist of different proficiency levels. Once KSELC-W consists of subcorpora with varied proficiency levels, it is very possible that different levels and stages of the developmental sequences in L2 acquisition processes can be identified in detail. In conclusion, the current study evidenced that NS-NNS corpora comparison studies can support, contradict, supplement, and or complement existing research as well as reveal new aspects about interlanguage development.

References

- Bailey, N., Madden, C., & Krashen, S. (1974). Is there a “natural sequence” in adult second language learning? *Language Learning*, 24, 235-243.
- Bickerton, D. (1981). *Roots of language*. Ann Arbor, MI: Karoma Publishers.
- Brown, R. (1973). *A first language*. Cambridge, MA: Harvard Press.
- Chaudron, C., & Parker, K. (1990). Discourse markedness and structural markedness: The acquisition of English noun phrases. *Studies in Second Language Acquisition*, 12, 43-65.

- Crosthwaite, P. R. (2016). Definite article bridging relations in L2: A learner corpus study. *Corpus Linguistics and Linguistic Theory*. Retrieved April 24, 2017, from <https://www.degruyter.com/view/j/cllt.ahead-of-print/cllt-2015-0058/cllt-2015-0058.xml>
- Dulay, H., & Burt, M. (1973). Should we teach children syntax? *Language Learning*, 23, 235-252.
- Dulay, H., & Burt, M. (1974). Natural sequences in child second language acquisition. *Language Learning*, 24, 37-53.
- Ekiert, M. (2004). Acquisition of the English article system by speakers of Polish in ESL and EFL settings. *Teachers College, Columbia University Working Papers in TESOL & Applied Linguistics*, 4(1), 1-23.
- Gablasova, D., Brezina, V., & McEnery, T. (2017). Exploring Learner Language Through Corpora: Comparing and Interpreting Corpus Frequency Information. *Language Learning*, 67, 130-154.
- Granger, S. (2013). 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 & Philadelphia: Benjamins.
- Huebner, T. (1983a). *A longitudinal analysis of the acquisition of English*. Ann Arbor, MI: Karoma Publishers.
- Hunston, S. (2002). *Corpora in Applied Linguistics*. Cambridge: Cambridge University Press.
- Ionin, T., Ko, H., & Wexler, K. (2004). Article semantics in L2 acquisition: The role of specificity. *Language Acquisition*, 12(1), 3-69.
- Ionin, T., Ko, H., & Wexler, K. (2004). The role of semantic features in the acquisition of English articles by Russian and Korean speakers. To appear in J. M. Liceras *et al.* (Eds.), *The role of formal in second language acquisition, second language research acquisition series (Theoretical and Methodological Issues)* (pp. 000-000). Lawrence Erlbaum Associates. Retrieved April 24, 2017, from <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.518.6136&rep=rep1&type=pdf>
- Ionin, T., & Montrul, S. (2010). The role of L1 transfer in the interpretation with articles in plural definite in L2 English. *Language Learning*, 60, 877-925.
- Ishikawa, S. (2011). A new horizon in learner corpus studies: The aim of the

- ICNALE project. In G. Weir, S. Ishikawa, & K. Poonpon (Eds.), *Corpora and language technologies in teaching, learning and research* (pp. 3-11). Glasgow: University of Strathclyde Press.
- Ishikawa, S. (2013). The ICNALE and sophisticated contrastive interlanguage analysis of Asian learners of English. In S. Ishikawa (Ed.), *Learner corpus studies in Asia and the world* (pp. 91-118). Kobe: Kobe University School of Languages and Communication.
- Kim, H.-Y. (2000). *Acquisition of English nominal reference by Korean speakers*. Unpublished doctoral dissertation, University of Hawai'i at Manoa, Honolulu, HI.
- Larsen-Freeman, D. (1975). The acquisition of grammatical morphemes by adult ESL students. *TESOL Quarterly*, 9, 409-430.
- Lee, E. (2012). *Transfer at the lexical level in Korean Learners' L2 indefinite article use in English*. Unpublished doctoral dissertation, Michigan State University, Michigan.
- Lu, C. F. (2001). The acquisition of English articles by Chinese learners. *Second Language Studies*, 20(1), 43-78.
- Master, P. (1987). *A cross-linguistic interlanguage analysis of the acquisition of the English article system*. Unpublished doctoral dissertation, UCLA, Los Angeles, CA.
- Master, P. (1993). *A contrastive study of determiner usage in EST research articles*. Paper presented at the TESOL Convention, Atlanta, Georgia, 15 April.
- Master, P. (1994). The effect of systematic instruction on learning the English article system. In T. Odlin (Ed.), *Perspectives on pedagogical grammar* (pp. 229-252). New York: Cambridge University Press.
- Parrish, B. (1987). A new look at methodologies in the study of article acquisition for learners of ESL. *Language Learning*, 37(3), 361-383.
- Rayson, P., & Garside, R. (2000). Comparing corpora using frequency profiling. In *Proceedings of the Workshop on Comparing Corpora, held in conjunction with the 38th annual meeting of the Association for Computational Linguistics (ACL 2000)*, 1-6.
- Sinclair, J. M. (Ed.). (1991). *Corpus, concordance, collocation*. Oxford: Oxford University Press.
- Thomas, M. (1989). The acquisition of English articles by first- and

second-language learners. *Applied Psycholinguistics*, 10, 335-355.

Zdorenko, T., & Paradis, J. (2008). The acquisition of articles in child second language English: Fluctuation, transfer or both? *Second Language Research*, 24(2), 227-250.

Daria Soon-Young Seog

Assistant Professor

Department of English Education

Teacher's College, Kyungpook National University

80 Daehak-ro, Buk-gu, Daegu. 41566, Rep. of Korea

Office: +82-53-950-5832

E-mail: dariadks@knu.ac.kr

Received on December 31, 2017

Revised version received on March 23, 2018

Accepted on March 31, 2018