



The Effectiveness of Corpus-aided Instruction to Improve Second Language College Students' Academic Writing*

Eunjeong Park**

The Ohio State University

ARTICLE INFO

Received 28 May 2019

Revised 23 July 2019

Accepted 14 August 2019

Examples in: English

Applicable Languages: English

Applicable Levels: Secondary

KEYWORD

corpus-aided instruction/

lexical bundles/

instructed second language acquisition/

코퍼스활용 교육/

어휘 번들/

학습된 제2언어습득

ABSTRACT

Park, Eunjeong. (2019). The effectiveness of corpus-aided instruction to improve second language college students' academic writing. *Modern English Education*, 20(3), 15-25.

International second language (L2) students should learn word and phrasal combinations of disciplinary conventions in academic settings as a basic principle to enhance writing skills (Cortes, 2004; Hyland, 2008a, 2008b). The main purpose of this study is to investigate the effectiveness of using lexical bundles (the combination of words) in corpus-aided instructed second language acquisition (ISLA). This study employed a quasi-experimental research design to examine the effects of using lexical bundles in the L2 writing classes. Fifty L2 undergraduate students at a Midwestern university participated in this study. The results of the quasi-experimental research design with pre- and post-tests revealed that both intentional and incidental instructional modes were effective in improving their lexico-grammatical writing gains although the intentional learning intervention has a longer effect. This study is of significance due to the pedagogical value, the usefulness of lexical bundles in language teaching and the contribution to developing the transdisciplinary framework (The Douglas Fir Group, 2016) within the field of L2 composition and corpus linguistics in ISLA.

I. INTRODUCTION

For international second language students studying in Anglophone universities where English is the medium of instruction and performance, acquiring second language (L2) academic writing skills is often a challenging task, but one that must be conquered. In addition to developing various rhetorical and grammatical skills essential in quality writing, they should learn academic English at the word and phrasal level, including frequently occurring word combinations or lexical patterns called lexical bundles (Cortes, 2004; Hyland, 2008a, 2008b). Due to the needs of international

L2 students, researchers in linguistics and education have paid attention to phraseology for effective L2 instruction. Phraseology, the study of fixed expressions and multi-word lexical units, is one of the major scholarships that Second Language Acquisition (SLA) and L2 writing researchers focus on. The exploration of phraseology is grounded on Systemic Functional Linguistics (SFL) (Halliday, 2004). SFL is an approach to linguistics that regards language as a social semiotic system. In SFL, lexicogrammar is the system of wording, representing the linguistic resources for construing meanings through words and structures, encompassing a much broader set of phenomena in phraseology (Halliday

* This study is partly based on the author's doctoral dissertation.

** Author: Eunjeong Park (The Ohio State University, Research Associate)

Eunjeong Park

Department of Teaching and Learning, The Ohio State University, 1945 High St, Columbus, OH, 43210, USA

Email: park.1752@osu.edu

& Matthiessen, 2004).

Applied linguists and researchers have investigated lexicogrammar with diverse components of a language. One of the features within lexicogrammar is a “lexical bundle” (Biber, Johansson, Leech, Conrad, & Finegan, 1999). Lexical bundles are the most frequently occurring sequences of multiple words and phrases in a written register indicating formulaicity of lexicogrammar in a language (Biber et al., 1999). Producing lexical bundles in written registers is challenging for international L2 students with different levels of writing proficiency and expertise (Cortes, 2004; Hyland, 2008a). Becoming adapted to the rhetorical styles and writing expectations of different disciplines are important in gaining membership in the academic community. Therefore, teaching lexical bundles would help L2 learners improve the use of lexical bundles, which eventually would enhance their academic writing proficiency and boost their confidence in writing. Corpus linguistics is an essential field in examining a variety of linguistic features in lexicogrammar. Corpus linguistics, which arose in the 1990s, compiles lists of various common word combinations (e.g., lexical bundles and collocations). Johns’ (1994) early work used “data-driven learning” (p. 296) to make language learning innovative in technological and methodological respects with the utilization of machine-readable text in corpus. In addition, Chun’s (2016) study demonstrated that students’ language experience is a crucial factor in language learning outcomes in data-driven learning. In the current study, therefore, corpus-aided instruction was provided with the use of predefined linguistic features from a systematic corpus analysis.

To make a connection of conceptual frameworks between applied linguistics and pedagogy, Instructed Second Language Acquisition (ISLA) resides in this study as a crucial channel of various theoretical and philosophical aspects of research. According to Loewen (2012), a concise definition of ISLA is L2 learning or acquisition as a result of teaching and instruction. ISLA is becoming more commonly used to indicate the subdomain of SLA that investigates the effects of manipulating various perspectives of the L2 learning effort, from L2 input, to language processing, to contexts of learning (Loewen, 2012). Therefore, the process of L2 acquisition is facilitated and guided by teachers or instructional materials of a target linguistic feature (e.g., lexical bundles) in L2 instruction. For the pedagogical development, the study employs notions of intentional and incidental learning in the combination of ISLA and L2 writing. Nation (2001) maintains that intentional and incidental learning as “complementary activities, each one enhancing the learning that comes from the other” (p. 232).

In light of beliefs and findings of second language acquisition and learning, overall, this study has the potential for redefining second language learning as the development of flexible meaning-making language capacities across contexts and broadening the scope of L2 writing by employing intentional and incidental learning within ISLA. Furthermore, this study adopts a “transdisciplinary

framework for SLA” (The Douglas Fir Group, 2016, p. 19) by integrating theoretical and pedagogical pieces of SFL, corpus linguistics, instructed second language acquisition, and L2 writing as foundational notions.

II. BACKGROUND

1. Systemic Functional Linguistics (SFL)

Systemic Functional Linguistics (SFL) is an approach to linguistics considering language as a social semiotic system; that is, the social action of meaning-making by functions and purposes in context. Halliday (2004) claims that a central theoretical principle is grounded on any act of communication involving choices. With the SFL perspective, language has evolved under the pressure of the particular functions that the language system has to serve. According to Halliday (2004), the basic functions of language are achieved via metafunctions, making sense of human experience and acting out social relationships. Halliday (2004) categorized metafunctions into three perspectives with ideational, interpersonal, and textual. The ideational function is divided into the experiential and logical. If the ideational function of grammar is performed as an action, it serves as the interpersonal metafunction as either interactive or personal. The last component is textual metafunction, a mode of meaning relating to the construction of the text. Halliday (2004) specified that the textual metafunction appears as a clearly delineated motif within the grammar because ideal and interpersonal metafunctions rely on the ability to build up sequences of discourse, organizer the discursive flow, and create cohesion and continuity. Hence, human experience is transformed into a meaning of language.

With this theoretical perspective of SFL, this study uses Halliday’s lexico-grammar (i.e., a combination of vocabulary and grammar) as a major notion. Lexicogrammar is the system of wording, representing the linguistic resources for construing meanings through words and structures. The lexicogrammar approach was adopted by the proponents of systemic functional grammar (Halliday & Matthiessen, 2004), encompassing a much broader set of phenomena in the mainstream lexicology.

2. Corpus-aided Language Learning

Corpus linguistics, the study of language in use through machine-readable texts, has offered potential research investigations of linguistic features (e.g., vocabulary, semantic domains, and grammatical structures). Through the corpus-based investigations, corpus linguists and language researchers (Biber et al., 1999; Cortes, 2004, 2006, 2013) rigorously examined the co-occurrence of seemingly similar structures and patterns, serving different functions in different contexts. Corpus data is recognized as valuable for gaining knowledge of language patterning

and perspectives on the language system (Sinclair, 2004). A great deal of corpus research has made an impact on the attention to lexical association patterns, including systematic co-occurrence of words and phrases. Corpus research with a frequency-based approach suggests new visions of existing language regularities and reveals previously unobserved language phenomena. Furthermore, corpus research represents a natural approach as regular patterns are detected in the data that are meaningful to language learners based on their adaptive behaviors.

The usefulness of corpus linguistics in teaching and learning has received growing attention and recognition (Conrad, 2005; Huang, 2011; Hunston, 2002; O’Keeffe, McCarthy, & Carter, 2007; Reppen, 2010; Sinclair, 2004). Corpus data provide descriptive insights regarding how people use different language forms at various levels of formality and how language fulfills multiple speech functions across contexts (Huang, 2011). Corpus linguistics and researchers (Boulton, 2010; Hunston, 2002; Gavioli & Aston, 2001) maintain that corpus instruction enhance second language learning and development.

3. Lexical Bundles

Language is assumed to be formulaic in nature (Simpson-Vlach & Ellis, 2010). Formulaic language is essential in academic writing, taking up 21-52.3% of written discourse (Biber et al., 1999). Lexical bundles are considered as a typical kind of formulaic language. Many researchers (Biber et al., 1999; Cortes, 2004) investigated the frequent use of lexical phrases in texts or corpora, and multi-word lexical chunks exist across various registers—conversation and academic prose. Since the initiative work of lexical bundles in the Longman Grammar, Biber et al.’s (1999) lexical bundle framework has been used in a wide range of research, such as comparing different registers of textbooks and classroom discourse (Biber et al., 2004), the use of lexical bundles in different populations, such as native versus nonnative speakers (Chen & Baker, 2010) and novices versus expert writers (Cortes, 2004; Hyland, 2008a), lexical bundles of academic disciplines (Cortes, 2004; Hyland, 2008b), and the comparison of lexical bundles in different languages, such as history articles written in English versus Spanish (Cortes, 2004, 2008).

4. Intentional and Incidental Language Learning/Acquisition

Different types of language learning can remain on a continuum between incidental and intentional since attention itself is not a dichotomous construct (Gass, 1999). Therefore, both intentional and incidental learning can occur with some degree of attention (Schmidt, 1994). Hulstijn (2003) claims that intentional or incidental learning involves some amount of attention. Unlike the argument about the language instruction on a continuum, Nation (2001) asserts that intentional and incidental learning

as “complementary activities, each one enhancing the learning that comes from the other” (p. 232). Intentional language instruction results in relatively faster learning within the assigned time than incidental instruction, leading to other learning gains (reading and writing skills) at the same time. It still seems vague and complex to differentiate one from the other. Hence, this study operationalized intervention terms based on intentional and incidental language learning, in order to explore the effectiveness of lexical bundle.

III. METHOD

The overarching question is if corpus-aided instruction with the use of lexical bundles is effective for the improvement of lexico-grammatical writing gains and academic writing in ISLA through a quasi-experimental research design. The following research questions then guide this study:

- 1) Which corpus-aided instruction (i.e., intentional or incidental) of lexical bundles is effective to improve their lexico-grammatical writing gains in a short term?
- 2) Are the lexico-grammatical writing gains retained five weeks later in a long term as a carry-over effect?

1. Participants

The target population was international L2 students studying at a U.S. Midwestern university. The students assigned to learn academic writing in ESL composition classes were sampled for this study. The target international L2 students were selected according to the availability and the research criteria as purposive sampling. This study was conducted with fifty college students taking three intact ESL composition classes. Their participation was voluntary. The homogeneity among the groups was ensured according to gender, age, and ethnicity. The population of all the three ESL composition classes was satisfied with the criteria of the homogeneity; yet, the students who did not agree on the consent were not included in the study. Fifty students participated in the control and intervention groups (i.e., 19 in the intentional group, 18 in the incidental group, and 13 in the control group). There were 22 male students (44%) and 28 female students (56%). 44 students (88%) were from China, while six (12%) of them were from Malaysia. The range of the age was 18 to 24 years old ($M = 19.38$, $SD = 1.40$).

2. Research Design and Methods

This study employed a quasi-experimental research design, more specifically a nonequivalent groups pretest-posttest, to investigate the effectiveness of corpus-aided instruction in instructed SLA. Marczyk, DeMatteo and Festinger (2005) suggest two merits of the nonequivalent groups

pretest-posttest design. First, the temporal precedence of the independent variable to the dependent variable can be made by utilizing both a pretest and a posttest. The temporal precedence enhances the researchers' confidence about the inference of changes in the outcome variable. Second, the use of a pretest allows for measuring between-group differences before exposure to the intervention (Marczyk et al., 2005). In terms of validity, the nonequivalent groups pretest-posttest design provides the potential to reduce the threat of selection bias via an examination of a group difference prior to the intervention. The study would contribute to the interdisciplinary advancement of research within the field of SLA, L2 writing, and corpus linguistics.

3. Instrumentation

1) Cloze Test for Measuring Lexico-grammatical Writing Gains

A cloze test which includes two sets of 10 questions was provided to the students. The same test questions were given three times for pre-, intermediate post-, and delayed post-tests. Cloze tests have been used to assess language competence and different skill levels of foreign/second language learners. Brown (2004) states that "the cloze procedure is perhaps deceptively simple. It is easy to develop and administer and seems to be a reliable and valid test of overall second-language proficiency" (p. 316), addressing that cloze tests can measure higher-order inter-sentential skills of both sentence-level grammar and inter-sentential level cohesion and coherence. Bachman (1985) explicates that a rational deletion procedure of a cloze test can measure textual relationships beyond clause structures and provides "the test developer a better means of controlling the specific components of language proficiency measured by the test" (p. 549). Therefore, this study employed a rational deletion cloze test, including 20 items of lexical bundle expressions (Biber et al., 1999; Salazar, 2014) (see Appendix 1). Since writing, like speaking, is a productive skill, it was assumed that the cloze procedure was suitable for the current study. Kongsuwannakul's (2017) study investigated the construct validity of a concordance-based cloze test and showed that cloze tests are appropriate for testing professional and academic English grammatical and vocabulary use. According to Kongsuwannakul (2017), the increase or gains displayed in a cloze test can lead to a better performance in general English proficiency, which can potentially lead to better academic writing skills in English. Thus, the cloze was deemed appropriate for the needs of the current study. Y. S. Ryoo (2017) also revealed that cloze test scores could serve as predictors of Korean EFL learners' written productive vocabulary with more significant differences in writing.

2) Interventions of Intentional and Incidental Lexical Bundle Learning

Each of the intentional and incidental instruction included five 20-minute lexical bundle activity sessions for five weeks

during the spring semester. The intervention classes were given the lexical bundle activities with a total of 20 lexical bundles (as indicated in Appendix 1) explicitly in intentional instruction and implicitly in incidental instruction. Lexical bundle items were selected from a learner corpus generated in a large corpus project. These bundles were focused on in the instruction so that the students gained awareness of different functional use of lexical bundles.

Although several studies (e.g., Barcroft, 2009; Nation, 2001) attempted to distinguish incidental learning from intentional learning based on focal and peripheral attention, the distinction is unclear to be applied to the educational context. Barcroft (2009) acknowledged that "different types of vocabulary learning can be viewed as points on a continuum between incidental and intentional because attention is not a dichotomous entity" (p. 85).

Therefore, this study operationalized terms of the interventions: intentional learning intervention, incidental learning intervention, and a control group. The intentional learning intervention had a clear intent with the observable evidence (e.g., explicit explanation of lexical bundles). Hanson and Padua (2014) recommend the direct and in-depth teaching of vocabulary, such as providing definitions, teaching individual words explicitly (Beck, McKeown, & Kucan, 2002). The purposes of intentional instruction included acquiring lexical bundles with explicit definitions and examples of lexical bundles and analyzing lexical bundles. The incidental learning intervention was based on Nation's (2001) three general processes of noticing, retrieval, and generation. Noticing activities were to draw attention to lexical bundles in class materials with typographical highlights, such as boldface and underlining; retrieval activities include the form-meaning mapping in the process of language reception and production; finally, generation activities concern learners' reconceptualization of the learned phraseology (Salazar, 2014). Based on the three processes of language instruction, lexical bundle activities were generated and implemented for the target students. The purposes of incidental instruction included noticing lexical bundles implicitly in the process of instruction and getting familiarized with them. Another distinction between the two instructional conditions is that the students in the intentional group knew that they would be tested with intentionality, while the ones in the incidental group did not notice any test before. A control group was designated in a regular ESL writing classroom. This control group as a reference set was compared to the two intervention groups in the quasi-experimental part (see Appendix 2).

4. Data Collection

The data collection was accomplished through the quasi-experimental design by teaching the lexical bundles to the experiment group participants in the assigned ESL composition classes. Three groups include two experiment groups (i.e., intentional and incidental interventions) and one control group (i.e., no intervention). The two experiment groups and the control group were exposed to a pre-test for their linguistic production of lexical bundles.

This nonequivalent groups pretest-posttest design lasted 12 weeks: a pre-test period (Week 1), interventions (Weeks 2-6), an immediate post-test period (Week 7), and a delayed post-test period (Week 12) as Table 1. The control group attended regular lessons in the composition class. The study also included a delayed post-test to obtain validity and credibility of the results. Lexical bundle activity materials were ready for five 20-minute micro-lessons to the experiment groups during the research semester, while the control group received regular instruction for academic writing. Pre- and post-tests of lexical bundles were given before and after the lexical bundle learning interventions. (Samples of learning objectives and activities in both intentional and incidental learning instruction were shown in Appendix 2.)

TABLE 1
The Procedure of Interventions

Pre-test	Interventions (Intentional/incidental)	Immediate post-test	Delayed post-test
Duration	Week 1 5 days × 20 min. a session (1 session/week)	Week 7	Week 12

5. Data Analysis

The results of the pre- and post-tests were analyzed through statistical values, such as the analysis of covariance (ANCOVA) with SPSS (Version 24) to see if the corpus-aided lexical bundle instruction significantly affects the students' lexico-grammatical writing gains. ANCOVA is useful in removing variances from the dependent variable so that the results yield a stronger effect size and power (Little, Johanns, & Giordani, 2000). ANCOVA not only reduces bias, but it also improves sensitivity. Controlling for covariates tends to improve precision if the covariates are predictive of potential outcomes.

IV. RESULTS

1. The Short-term Effect of Corpus-aided Instruction

Descriptive statistics are presented for the fifty students' pre-test and two post-tests (Table 2) in response to the study's instructional interventions. 20 lexical bundles were tested for their lexico-grammatical writing gains. Each question was given one point, so that 20 points was the maximum score possible. Overall, the individual scores of the pre-test ranged from 4 to 18 ($M = 11.86, SD = 3.75$). The individual scores of the post-tests ranged from 3 to 20 ($M = 14.38, SD = 3.76$). The individual scores of the delayed post-tests ranged from 6 to 20 ($M = 14.88, SD = 3.47$). The examination of the descriptive statistics is desirable because these test scores are continuous variables. Likewise, there is value in examining the variability of individuals using the range and standard deviation (SD) results. Investigating overall scores of the continuous variables is desirable in order to monitor the significance of the effect of the interventions in the three different conditions (i.e., intentional,

incidental, and control).

TABLE 2
The Descriptive Statistics of Pre-test and Post-test

	Pre-test				Immediate post-test				Delayed post-test			
	<i>n</i>	<i>M</i>	<i>SD</i>	Min. Max.	<i>M</i>	<i>SD</i>	Min. Max.	<i>M</i>	<i>SD</i>	Min. Max.		
Intentional	19	12.21	4.26	5 18	14.47	3.44	8 20	16.21	3.36	6 20		
Incidental	18	12.83	3.10	4 17	16.17	3.10	10 20	14.89	3.55	6 20		
Control	13	10.00	3.34	5 18	11.77	3.77	3 16	12.92	2.75	7 17		
Total	50	11.86	3.75	4 18	14.38	3.76	3 20	14.88	3.47	6 20		

The analysis of covariance (ANCOVA) of the quasi-experiment (i.e., the instruction of lexical bundles implemented with the intentional and incidental learning conditions) revealed the effectiveness of the intentional and incidental instruction on the students' lexico-grammatical writing gains. Because the data met the assumption of the parametric analysis, the ANCOVA was used with the dataset. Pre-test data was employed as a covariate this time. Results of ANCOVA showed that the differences in means between groups were statistically significant ($F(2, 46) = 3.77, p = .03$) with the moderate effect size (partial $\eta^2 = .14$) ($\alpha = .05$) in Table 3. Therefore, the results presented that lexico-grammatical writing gains were significantly different as the interventions differ, controlling for the pre-test scores. Although the mean of the pre-test in the control group seemed lower than the other two groups, there was no statistical difference among the three groups. The Post-hoc test was used to check the results of multiple comparisons with the Bonferroni method which allows many comparisons for unequal sample sizes in the ANCOVA results. The effect of the immediate post-test was significantly different between only the incidental group and the control group (Marginal mean difference = 3.24, $p = .026, \alpha = .05$) in Table 4.

TABLE 3
Tests of Between-Subjects Effects in ANCOVA

Source	Type III Sum of Squares	<i>df</i>	Mean Square	F	Sig.	Partial Eta Squared
Intercept	365.15	1	365.15	38.11	.000	.45
Pretest	104.76	1	104.76	10.93	.002	.19
Intervention	72.28	2	36.14	3.77	.030	.14
Error	440.79	46	9.59			
Total	11031.0	50				

a. R Squared = .363 (Adjusted R Squared = 0.321)

b. Computed using alpha = .05

TABLE 4
Multiple Comparisons in ANCOVA

(I) Intervention type	(J) Intervention type	Mean difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
Control	Intentional	-1.79	1.15	.372	-4.65	1.05
	Incidental	-3.24*	1.18	.026	-6.17	-.30
Intentional	Control	1.79	1.15	.372	-1.05	4.65
	Incidental	-1.44	1.02	.498	-3.97	1.09
Incidental	Control	3.24*	1.18	.026	.304	6.17
	Intentional	1.44	1.02	.498	-1.09	3.97

* The mean difference is significant at the .05 level.

2. The Long-term Effect of Corpus-aided Instruction

To see the carry-over effect, ANCOVA (Analysis of Covariance) was used to compare the immediate post-test and the delayed post-test. Results of ANCOVA showed that the means between groups were not statistically significant ($F(2, 45) = 82.77, p = .010$) with the moderate effect size (partial $\eta^2 = .185$) ($\alpha = .05$) in Table 5. The results presented that lexico-grammatical writing gains are significantly different as the interventions differ.

TABLE 5
Tests of Between-Subjects Effects in ANCOVA

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Intercept	4.18	1	4.18	.52	.477	.011
Pretest	69.22	1	69.22	8.53	.005	.159
Intervention	82.77	2	41.38	5.10	.010	.185
Error	365.03	45	8.11			
Total	11.662.00	50				

a. R Squared = 0.383 (Adjusted R Squared = 0.328)

b. Computed using alpha = 0.05

The Post-hoc test was used to check the results of multiple comparisons with the Bonferroni method which allows many comparisons for unequal sample sizes. Marginal means were reported because the ANCOVA model was conducted, controlling for the immediate post-test. The effect of the delayed post-test was significantly different between the intentional group and the control group (Marginal mean difference = 2.97, $p = .033$) and between the intentional group and the incidental group (Marginal mean difference = 2.65, $p = .032$), controlling for the immediate post-test scores, as indicated in Table 6. The difference between the incidental group and the control were not statistically significant ($\alpha = .05$). From the analysis of multiple comparisons, the delayed post-test scores of lexico-grammatical writing gains in the intentional group were significantly higher than the ones in the control group; the delayed post-test scores of lexico-grammatical writing gains in the intentional group were significantly higher than the ones in the incidental group.

TABLE 6
Multiple Comparisons in ANCOVA

(I) Intervention type	(J) Intervention type	MD (I-J)	Std. Error	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
Control	Intentional	-2.97*	1.12	.033	-5.74	-.19
	Incidental	-.32	1.17	1.000	-3.23	2.59
Intentional	Control	2.97*	1.12	.033	.19	5.74
	Incidental	2.65*	.991	.032	.18	5.11
Incidental	Control	.32	1.17	1.000	-2.59	3.23
	Intentional	-2.65*	.991	.032	-5.11	-.18

* The mean difference is significant at the .05 level.

The results of ANCOVA revealed several noteworthy aspects from the delayed post-test. First, the lexico-grammatical writing gains were significantly higher in the

intentional group than in the control group. Although the intentional group did not show the effect in the immediate test, it yielded a carry-over effect in the delayed post-test. Second, the results of the intentional group were also higher than the ones of the incidental group. This revealed that the intentional learning of lexical bundles would be effective in the long term. Unlike these significant effects of the intentional group, the incidental group did not present any carry-over effect, controlling for the immediate post-test scores. Overall, the intentional learning intervention has a long-term effect, compared to the incidental learning intervention and the regular class.

V. DISCUSSION

Language researchers (Byrd & Coxhead, 2010; Cortes, 2004, 2006; Coxhead & Byrd, 2007; Hyland, 2012; Simpson-Vlach & Ellis, 2010) have advocated direct instruction on lexical bundles in writing courses for L2 students. Cortes (2004) insists that “as lexical bundles are very frequent in published academic prose, it is necessary to encourage students to use these expressions” (p. 420). Hence, novice L2 writers must learn the discourse conventions (i.e., lexical bundles) of academic writing with the adequate use of lexical bundles (Biber, et al., 2013; Cortes, 2004; Pan, Reppen, & Biber, 2016).

Lexical bundles are beneficial in L2 writing instruction with the principles of frequency, range, teachability/learnability, and the usefulness in academic writing (Byrd & Coxhead, 2010; Nation, 2001). Coxhead and Byrd (2007) noted several beneficial effects of using lexical bundles in academic writing: (1) providing ready-made sets of words for crafting academic prose, (2) facilitating fluent language use, (3) recognizing L2 writers as a “member” of a discourse community, and (4) representing register-specific ways of expressing particular meanings. The results of the nonequivalent group pretest-posttest quasi-experiment in this study revealed that the incidental lexical bundle instruction had a significant effect in the immediate post-test, whereas, the intentional lexical bundle instruction had a carry-over effect in the delayed post-test.

Therefore, the quantitative results of this study corroborate that intentional language instruction of lexical bundles would be effective to international L2 college students whose writing proficiency is intermediate-low. Plenty of evidence emerged from various sources. First, the intentional lexical bundle instruction obtained a significant carry-over effect. In other words, students can retain the knowledge and skills of using lexical bundles in the long term. The findings can be aligned with prior literature about the effectiveness of intentional language learning (Hulstijn 1992; Schmitt, 2008). Schmitt (2008) affirms that intentional vocabulary learning leads to “a better chance of retention and of reaching productive levels of mastery” (p. 341). Explicit deliberate learning (i.e., intentional learning) is best for learning salient elements of word knowledge along with the rate and efficiency of

learning, while incidental learning comes from a sufficiency of time and exposure (Nation, 2011). This may be true for the findings of this study in that the interventions lasted only for five weeks. Therefore, when researchers and educators encounter the demarcation of intentional and incidental learning, it would be critical to consider how to maximize the learning conditions of different instructional interventions in research or practice.

Using lexical bundles in writing instruction is helpful in developing language learning autonomy. Johns (1994) supports learner autonomy by data-driven learning, and corpus linguists (Gavioli, 2009) also stress learner autonomy through corpus linguistics. Applying lexical bundles to different assignments is closely related to language learning autonomy in that L2 students independently explore the use of lexical bundles without the instructor's guidance and understand them in cognition. The corpus-aided lexical bundle instruction allowed self-access of the relevant materials with learner autonomy and helped the L2 students become active searchers for the application of lexical bundles to the actual use in their academic writing.

VI. CONCLUSION

This study is of significance due to the pedagogical value, the usefulness of the most frequent words in language teaching (Nation, 2001), regarding the research of L2 composition and corpus linguistics in instructed second language acquisition (ISLA). A phraseological approach associated with corpus linguistics enables us to redefine and broaden aspects of linguistic theories. Furthermore, the utility of the lexical bundle list can work as a basis for material designs and curriculum development and create a great impact on revealing the representativeness of the English language used by international college students (Byrd & Coxhead, 2010; Simpson-Vlach & Ellis, 2010). Lexical bundles have the potential to explore phraseological differences between registers and disciplines (Römer, 2009). Well-designed writing instruction can result in transfer in EFL/ESL contexts. Accordingly, intentional or incidental learning instruction maximizes learning transfer with a focus on lexical bundles.

This study also stimulates active awareness and perceptions of the prevalent usage of lexical phrases in practice. The idea of the overall research design in this study was generated from the implications of prior research (e.g., Cortes, 2006; Hyland, 2012; Simpson-Vlach & Ellis, 2010). The researchers maintained that the future research should link the constituent of applied linguistics (i.e., the analysis of lexical bundles) and the pedagogical use (i.e., the application of the use of lexical bundles) in the boundaries of composition studies and applied linguistics (Silva & Leki, 2004). Novice L2 writers must learn the discourse conventions (i.e., lexical bundles) of academic writing with the adequate use of lexical bundles (Biber et al., 2013; Cortes, 2004). Therefore, this study will contribute to the development and advancement of L2 writing

instruction and curriculum in the long term by providing effective language learning methods with the use of lexical bundles in academic writing.

Lastly, the use of corpus data can be helpful in guiding corpus-based writing pedagogy. The current study was built on the results obtained from a learner corpus of international students. This was beneficial from a pedagogical perspective, in that learner corpora involve "pedagogic mediation to contextualize the data for the students' own writing environment" (Flowerdew, 2009, p. 393) and play an important role in selecting and structuring teaching contents (Flowerdew, 2015; Granger, 2002). Swales (2019) also indicated that corpus-based research should have pedagogical value. There has been a myriad of corpus studies, but Swales argued that lists of lexical bundles based solely on frequency data may not be worth compiling if they are of little use to language teachers and learners. Thus, corpora enable L2 writing teachers to identify potential learner needs based on their use of a target language and evaluate the level of lexical and grammatical complexity of learner language.

This study may have several limitations that need to be addressed for future research. The most significant limitation of this study is external validity in light of generalizability. This study did not implement random selection in the research. The scope is limited to international college students at one Midwestern university in the United States of America. However, L2 classroom researchers should strive for "particularizability," supporting teachers through connections between the findings of research and the particularity of their lives. Secondly, as Cortes' (2006) implication in her study, the exposure to the target lexical bundles during the micro-lessons may not be sufficient for L2 students' learning transfer in the long term. This can be resolved if corpus-aided instruction is possibly adopted as an essential part of L2 writing sessions. Furthermore, cloze tests can be seen as word matching. As Kongsuwanakul (2017) said, however, gains of words and phrases in a cloze test can lead to a better performance in general English proficiency, and further academic writing skills. Lastly, the quasi-experimental design may have been weak in controlling for threats to internal validity, such as testing (Shadish et al., 2002). Testing can influence the effect on subsequent exposures to the test. In addition, this quantitative research was not sufficient in exploring why the intentional group scored high and the incidental group scored low. This inquiry should be examined through qualitative research, such as surveys or interviews with the students.

To systematize corpus-aided pedagogy for L2 students and corpus literacy for L2 teachers, several suggestions are encouraged for the future research. Authentic materials from the students' writing are recommended. The students' written products include valuable information about linguistic features and structural rhetoric that can be used in the future writing class. As this study revealed that explicit instruction would be effective to L2 learners, more research of corpus-aided instruction should be imple-

mented in developing corpus-aided pedagogy. Language teachers should approach corpus data systematically with the learners' needs. Lastly, L2 language researchers (McCarthy, 2008; O'Keeffe & Farr, 2003) have integrated corpus linguistic techniques into teacher education, setting up guidelines for teacher education in corpus literacy. Although this study touched upon corpus literacy to some extent, future research should further extend standards and guidelines of corpus use for curriculum and instruction in teacher education. As language researchers (S. Chun, 2016; S. A. Kim, 2017) addressed the significance of phraseological language instruction in English writing, the findings of this study can also be feasible and beneficial in the EFL context in Korea for Korean students' English language learning.

REFERENCES

- Bachman, L. F. (1985). Performance on cloze tests with fixed-ratio and rational deletions. *TESOL Quarterly*, 19(3), 535-556.
- Barcroft, J. (2009). Effects of synonym generation on incidental and intentional L2 vocabulary learning during reading. *TESOL Quarterly*, 43(1), 79-103.
- Beck, I., McKeown, M., & Kucan, L. (2002). *Bringing words to life: Robust vocabulary instruction*. New York: Guilford Press.
- Biber, D., Conrad, S., & Cortes, V. (2004). If you look at . . . Lexical bundles in university teaching and textbooks. *Applied Linguistics*, 25(3), 371-405.
- Biber, D., Gray, B., & Poonpon, K. (2013). Pay attention to the phrasal structures: Going beyond T-units-A response to Weiwei Yang. *TESOL Quarterly*, 47(1), 192-201.
- Biber, D., Johansson, S., Leech, G., Conrad, S., & Finegan, E. (1999). *Longman grammar of spoken and written English*. London: Pearson.
- Boulton, A. (2010). Data-driven learning: Taking the computer out of the equation. *Language Learning*, 60(3), 534-572.
- Brown, J. D. (2004). *Language assessment: Principles and classroom practice*. London: Longman.
- Byrd, P., & Coxhead, A. (2010). On the other hand: Lexical bundles in academic writing and in the teaching of EAP. *University of Sydney Papers in TESOL*, 5, 31-64.
- Chen, Y.-H., & Baker, P. (2010). Lexical bundles in L1 and L2 academic writing. *Language Learning & Technology*, 14(2), 30-49.
- Chun, Sooin. (2016). Word knowledge expansion using different types of target words in guided data-driven learning. *Modern English Education*, 17(4), 95-118.
- Conrad, S. (2005). Corpus linguistics and L2 teaching. In E. Hinkel (Ed.), *Handbook of research in second language teaching and learning*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Cortes, V. (2004). Lexical bundles in published and student disciplinary writing: Examples from history and biology. *English for Specific Purposes*, 23, 397-423.
- Cortes, V. (2006). Teaching lexical bundles in the disciplines: An example from a writing intensive history class. *Linguistics and Education*, 17, 391-406.
- Cortes, V. (2013). The purpose of this study is to: Connecting lexical bundles and moves in research article introductions. *Journal of English for Academic Purposes*, 12, 33-43.
- Coxhead, A., & Byrd, P. (2007). Preparing writing teachers to teach the vocabulary and grammar of academic prose. *Journal of Second Language Writing*, 16(3), 129-147.
- Flowerdew, L. (2009). Applying corpus linguistics to pedagogy: A critical evaluation. *International Journal of Corpus Linguistics* 14(3), 393-417.
- Flowerdew, L. (2015). Corpus-based research and pedagogy in EAP: From lexis to genre. *Language Teaching*, 48(1), 99-116.
- Gass, S. (1999). Discussion: Incidental vocabulary learning. *Studies in Second Language Acquisition*, 21(2), 319-333.
- Gavioli, L. (2009). Corpus analysis and the achievement of learner autonomy in interaction. In L. Lombardo (Ed.), *Using corpora to learn about language and discourse* (pp. 39-71). Bern, Switzerland: Peter Lang.
- Gavioli, L., & Aston, G. (2001). Enriching reality: Language corpora in language pedagogy. *ELT Journal*, 55(3), 238-246.
- Granger, S. (2002). A bird's-eye view of learner corpus research. In S. Granger, J. Hung, & S. Petch-Tyson (Eds.), *Computer learner corpora, second language acquisition and foreign language teaching* (pp. 3-33). Philadelphia, PA: John Benjamins.
- Halliday, M. A. K. (2004). Introduction: How big is a language? On the power of language. In J. J. Webster (Ed.), *The language of science: Volume 5 in the collected works of M.A.K.* London: Continuum.
- Halliday, M. A. K., & Matthiessen, C. (2004). *An introduction to functional grammar* (3rd ed.). London: Arnold.
- Hanson, S., & Padua, J. F. M. (2014). *Teaching vocabulary explicitly*. Honolulu, HI: Pacific Resources for Education and Learning.
- Huang, L.-S. (2011). Corpus-aided language learning. *ELT Journal*, 65(4), 481-484.
- Hulstijn, J. (1992). Retention of inferred and given word meanings: Experiments in incidental vocabulary learning. In P. Arnaud & H. Béjoint (Eds.), *Vocabulary and applied linguistics* (pp. 113-125). London: Macmillan.
- Hulstijn, J. (2003). Incidental and intentional learning. In C. Doughty & M. H. Long (Eds.), *The handbook of second language acquisition* (pp. 351-381). Malden, MA: Blackwell Publishing.

- Hunston, S. (2002). *Corpora in applied linguistics*. Cambridge: Cambridge University Press.
- Hyland, K. (2008a). Academic clusters: Text patterning in published and postgraduate writing. *International Journal of Applied Linguistics*, 18(1), 41-62.
- Hyland, K. (2008b). As can be seen: Lexical bundles and disciplinary variation. *English for Specific Purposes*, 27, 4-21.
- Hyland, K. (2012). Bundles in academic discourse. *Annual Review of Applied Linguistics*, 32, 150-169.
- Johns, T. (1994). From printout to handout: Grammar and vocabulary teaching in the context of data-driven learning. In T. Odlin (Ed.), *Perspectives on pedagogical grammar* (pp. 27-45). Cambridge: Cambridge University Press.
- Kim, Sung-Ae. (2017). An analysis of Korean EFL writers' errors in the use of cohesive devices. *Modern English Education*, 18(4), 1-23.
- Kongsuwannakul, K. (2017). *Investigating the construct validity of a concordance-based cloze test: A mixed-methods study* (Unpublished doctoral dissertation). University of Leicester, England.
- Little, R. J., An, H., Johanns, J., & Giordani, B. (2000). A comparison of subset selection and analysis of covariance for the adjustment of confounders. *Psychological Methods*, 5(4), 459-476.
- Loewen, S. (2012). Instructed second language acquisition. In C. Chapelle (Ed.), *The encyclopedia of applied linguistics* (pp. 2716-2717). Hoboken, NJ: Wiley.
- Marczyk, G. R., DeMatteo, D., & Festinger, D. (2005). *Essentials of research design and methodology*. Hoboken, NJ: John Wiley & Sons.
- McCarthy, M. (2008). Accessing and interpreting corpus information in the teacher education context. *Language Teaching*, 41, 563-574.
- Nation, I. S. P. (2001). *Learning vocabulary in another language*. Cambridge: Cambridge University Press.
- O'Keefee, A., & Farr, F. (2003). Using language corpora in initial teacher education: Pedagogic issues and practical applications. *TESOL Quarterly*, 37(3), 389-418.
- O'Keefee, A., McCarthy, M., & Carter, R. (2007). *From corpus to classroom: Language use and language teaching*. Cambridge: Cambridge University Press.
- Reppen, R. (2010). *Using corpora in the language classroom*. Cambridge: Cambridge University Press.
- Ryoo, Young-sook. (2017). Predictability of the cloze test as a measure of written productive vocabulary. *Modern English Education*, 18(4), 25-45.
- Pan, F., Reppen, R., & Biber, D. (2016). Comparing patterns of L1 versus L2 English academic professionals: Lexical bundles in telecommunications research journals. *Journal of English for Academic Purposes*, 21, 60-71.
- Römer, U. (2009). English in academia: Does nativeness matter? *International Journal of English Studies*, 20(2), 89-100.
- Salazar, D. (2014). *Lexical bundles in native and non-native scientific writing*. Amsterdam, Philadelphia: John Benjamins.
- Schmidt, R. (1994). Implicit learning and the cognitive unconscious: Of artificial grammars and SLA. In N. C. Ellis (Ed.), *Implicit and explicit learning of languages* (pp. 165-210). London: Academic Press.
- Schmitt, N. (2008). Review article: Instructed second language vocabulary learning. *Language Teaching Research*, 12(3), 329-363.
- Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Belmont, CA: Wadsworth Cengage Learning.
- Silva, T., & Leki, I. (2004). Family matters: The influence of applied linguistics and composition studies on second language writing studies—past, present, and future. *The Modern Language Journal*, 88(1), 1-13.
- Simpson-Vlach, R., & Ellis, N. C. (2010). An academic formulas list: New methods in phraseology research. *Applied Linguistics*, 31(4), 487-512.
- Sinclair, J. (2004). *How to use corpora in language teaching*. Amsterdam: John Benjamins.
- Swales, J. M. (2019). The futures of EAP genre studies: A personal viewpoint. *Journal of English for Academic Purposes*, 38, 75-82.
- The Douglas Fir Group. (2016). A transdisciplinary framework for SLA in a multilingual world. *The Modern Language Journal*, 100(S1), 19-47.

APPENDIX 1

Rational Deletion Cloze Test in the Quasi-experimental Design

Direction: There are two separate texts with 10 blanks in the total of 20 questions. Fill in the blanks with appropriate options provided. You have 20 minutes to complete them.

<p>International students have been steadily increasing in the U. S. colleges and universities. According to <i>Open Doors 2015</i>, which (1) _____ comprehensive information resource on international students studying at higher education institutions in the U.S., international students (2) _____ study in the United States by 10 % growth over the prior year in 2014-2015. From this statistical fact, (3) _____ assume that international students are considered as crucial for the economic and social impact on the U.S. (4) _____ contribution was more than 30.8 billion dollars to the economic vigor of U.S. higher education institutions was also described in <i>Open Doors</i>. (5) _____ that they have contributed to scientific research and diverse perspectives into the U.S. classrooms, (6) _____ benefits for global careers and business relationships. (7) _____ international students may not be successfully adapted to the target culture and language. (8) _____ that people should know is how previous studies described this issue. Prior research has shown that (9) _____ international students encounter challenges in their adaptation process while studying in a foreign country. Due to the increasing population of international students and their contribution to higher education in the U.S., (10) _____ explore how international students at the universities in English-speaking countries are adjusted to the target language and culture in their academic lives.</p>			
(1) it is possible to	(2) a large number of	(3) as well as the	(4) on the other hand
(5) it is important to	(6) it is no doubt	(7) one of the things	(8) is known as the
(9) the fact that the	(10) were more likely to		
<p>To date, little research has been conducted for international students' academic adaptation through the mixed-methods research paradigm. (11) _____ mixed-methods design has rarely been used in the field of second language writing. (12) _____ will examine international graduate students' academic adaptation scales via survey research and explore how they are academically adapted in the U. S. college setting via qualitative research; (13) _____ interviews, observations, and field notes. (14) _____ this study pursues investigating how quantitative and qualitative results can explain international students' academic adaptation through a corpus-based analysis. This study will present several research revenues. (15) _____ researchers should better understand the nature of the students' language use. Thus, it is necessary to delve into how they receive the input and yield the output via writing (16) _____ their process of academic writing. Second, in this sense, (17) _____ use of multiword phrases can manifest the degree of academic adaptation through writing. In brief, the study investigates (18) _____ international students' academic adaptation scales and the perception of language proficiency via survey and qualitative research. In (19) _____ textual analysis, the corpus data will be examined as one of the academic literacy adaptation indicators. From this research, (20) _____ that the quantitative and qualitative results show international students' academic adaptation in different angles.</p>			
(1) the relationship between the	(2) at the same time	(3) we wish to reveal	
(4) in this study, we	(5) when it comes to	(6) the case of a	(7) in other words, a
(8) we assume that the	(9) first of all, the	(10) for example, this includes	

APPENDIX 2

Learning Objectives and Activities in the Intentional and Incidental Instruction

1. Learning Objectives and Activities in the Intentional Instruction

Session	Objectives	Activities
1	Define lexical bundles	Explicit instruction about the definition and examples of lexical bundles
2	Recognize lexical bundles in the lexical bundle list	Explicit instruction about patterns and functions of lexical bundles
3	Recognize the function of lexical bundles	Activity focusing on functional lexical bundles
4	Recognize the structure of lexical bundles	Activity of matching part of speech in lexical bundles
5	Analyze lexical bundles in the concordancing program	Activity of analyzing functional lexical bundles

2. One activity example in the intentional instruction is as follows:

1. Present the definition and examples of lexical bundles

(1) Lexical bundles: The most frequent sequences of words in a register (Biber, Conrad, & Cortes, 2004), and combinations of three or more words that frequently occur in a language or a given register (Biber et al., 1999)

(2) The examples of lexical bundles

<i>a large number of</i> <i>a lot of time</i> <i>as far as the</i> <i>at the same time</i> <i>at the end of</i> <i>are more likely to</i>	<i>on the other hand</i> <i>is one of the</i> <i>in terms of the</i> <i>first of all the</i> <i>last but not least</i> <i>when it comes to</i>
--	---

2. Write down 5 short sentences with lexical bundles.

3. Learning Objectives and Activities in the Incidental Instruction

Session	Objectives	Activities
1	Notice multiword phrases	Activity of noticing similarity and difference of multiword phrases
2	Distinguish multiword phrases in the lexical bundle list	Activity of distinguishing and responding to lexical bundle-embedded questions
3	Retrieve the multiword phrases from previous sessions	Activity of composing a short comparison paragraph with multiword phrases
4	Retrieve functional multiword phrases	Activity of composing a short essay with functional phrases
5	Generate student's own writing with multiword phrases	Activity of composing free writing with any meaningful multiword phrases

4. One activity example in the incidental instruction is as follows:

1. Come up with an example in your own writing that you are currently working on, where it would be possible to use two of the following multiword expressions.

The examples of lexical bundles

- *are more likely to*
- *it is true that*
- *last but not least*
- *on the other hand*
- *is one of the*
- *a strong relationship with*