



## Lexical Frames and Errors in the Use of English Definite Article in L2 Academic Writing: A Case of English Placement Test

Haeyun Jin (Seoul National University)



This is an open-access article distributed under the terms of the Creative Commons License, which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

Received: March 24, 2023

Revised: April 12, 2023

Accepted: April 19, 2023

Haeyun Jin  
Visiting Researcher, Center for Educational Research,  
Seoul National University  
Tel: 02-880-8834  
Email: haeyunj415@gmail.com

### ABSTRACT

**Jin, Haeyun. 2023. Lexical frames and errors in the use of English definite article in L2 academic writing: A case of English placement test. *Korean Journal of English Language and Linguistics* 23, 324-341.**

The purpose of this study was to investigate the use of lexical frames in L2 academic writing across writing skill levels focusing on both the target-like frames and ungrammatical frames due to article errors. Writing samples were academic essays (N=991) composed of two levels from the corpus of the English Placement Test (EPT) compiled at a Midwestern US university. The 20 realized academic writing frames (referred to as *AW-frames*) were searched for in the lower- and higher-level sub-corpora and compared in three aspects. First, the differences in the overall frequencies of the frames were explored based on normalization and chi-square tests. Then, the variability and predictability of the retrieved frames in the two levels were compared using Gray and Biber's (2013) framework. Lastly, the cases of errors in the definite article were analyzed for both the realized and unrealized frames. The results suggested meaningful differences in the *AW-frames* identified in lower versus higher-level writing. First, higher-level learners employed *AW-frames* more frequently compared to lower-level learners. Second, higher-level learners tended to employ highly variable *AW-frames* in an unpredictable manner whereas less proficient learners incorporated the same inventory of frames in a comparatively fixed and predictable way. In terms of the misused frames, regardless of levels, L2 learners produced extensive errors in the definite article, particularly having a problem using the non-nominal head for the definite article within realized frames. Taken together, by applying the lexical-frame approach to L2 phraseology, this study helps us gain a better understanding of how L2 learners' productive use of multiword expressions differs across writing skill levels. It also provides implications for the teaching of *AW-frames* to improve L2 academic writing.

### KEYWORDS

lexical frame, L2 phraseology, L2 academic writing, definite article, error analysis

## 1. Introduction

The present study investigated the use of lexical frames in L2 academic writing focusing on analyzing both the target-like and ungrammatical use of frames due to article errors. *Lexical frame*, as one approach to identifying the recurrent combination of lexical items, refers to the *discontinuous* sequences of words, which is composed of a *frame* of words surrounding a variable slot (Gray and Biber 2013). Researchers have investigated the concept of multi-word sequences with an intervening open slot using various terms such as *collocational frameworks* (Renouf and Sinclair 1991), *phrase-frames* (Fletcher 2011, Garner 2016, Römer 2010, Stubbs 2007), and *lexical frames* (Gray and Biber 2013). Regardless of the specific labels, lexical frames can be distinguished from *lexical bundles*, the *continuous* sequences of fixed words in that it deals with non-contiguous lexical sequences that allow for item-internal variation (Biber 2009, Garner 2016, Römer 2010). For instance, the lexical bundles, *the beginning of the*, and *the case of the* can be regarded as the variants of the lexical frame, *the \* of the*. While extensive research efforts have been concentrated on L2 learners' use of continuous sequences of words, such as lexical bundles (e.g., Ådel and Erman 2012, Chen and Baker 2010, Paquot and Granger 2012), not many studies have explored lexical frames in the context of L2 academic writing. Besides, little research focused on the *accuracy* of L2 formulaic language documenting frequently attested types of errors.

In response to such a research gap, the current study focused on analyzing academic-writing frames (*AW-frames*), namely, the lexical frames that were identified as frequently occurring in academic prose by Gray and Biber (2013). As already indicated by their study, lexical frames in academic prose are typically function word frames composed only of function words such as articles and prepositions while conversation relies predominantly on verb-based frames. The present research examined these *AW-frames* focusing on the frames that include definite article *the*. More specifically, the following research questions are addressed:

1. With what frequencies do L2 learners use academic writing frames (*AW-frames*) across levels?
2. With what variability and predictability are these *AW-frames* used across levels?
3. What type of errors do the learners produce in the use of the definite article embedded in the realized and unrealized *AW-frames*<sup>1</sup>?

The first and the second research questions take the traditional Contrastive Interlanguage Analysis (CIA) approach (Gilquin and Granger 2015) to describe the variation in the use of the identified *AW-frames* (referred to as *the realized AW-frames*) between the learners placed into different levels. More specifically, the first question concerns the variation of use in terms of the overall frequencies of the realized frames. The second question delves into the variability and predictability of these frames and explores how they vary across the levels. The last question, on the other hand, centers on examining the misused *AW-frames* focusing both on the errors in the realized frames and unrealized frames due to article errors. The types of errors and more qualitative patterns in those errors were also analyzed. The ultimate goal of this investigation was to reveal the patterns of use of the lexical frames across L2 writing skill levels, as well as to document the non-target-like use of those frames focusing on the errors in the definite article.

---

<sup>1</sup> *Realized* frames refer to the frames that were successfully realized and retrieved as the list of frames. *Unrealized* frames, on the other hand, refer to the frames that were attempted but fail to be realized as lexical frames due to errors. The third research question focuses on the case of misuse in both realized and unrealized frames.

## 2. Literature Review

With growing interest in multiword expressions as an important component of fluent L2 academic writing, many researchers for the last decades have centered on the aspect of L2 phraseology in written production. These earlier studies mostly investigated the use of lexical bundles in L2 writing (Å del and Erman 2012, Chen and Baker 2010, Paquot and Granger 2012, Reppen 2009, Staples et al. 2013) while very few researchers have shown emerging interests in lexical frames or the phrase-frame approach (Garner 2016). Most of these existing studies have taken the Contrastive Interlanguage Analysis (CIA) approach (Gilquin and Granger 2015) to describe L2 phraseology by comparing the overall frequencies, structural, and functional distributions of formulaic language in L2 learner writing compared to the native norm, or between writings produced by different L2 learner population (e.g., different L1s). For instance, Garner (2016) investigated the use of lexical frames (in their terminology, *p-frames*) by German EFL learners across five proficiency levels using the samples from the *EF-Cambridge Open Language Database* (EFCAMDAT). They analyzed the most frequent frames employed by learners at different levels in terms of quantitative and qualitative aspects. The results revealed that higher-level L2 learners overall employed more function word frames when compared to lower-levels. Additionally, p-frames used by lower proficiency learners tended to be more variable, less predictable, and structurally and functionally more complex.

When it comes to the accuracy of formulaic language, little research focused on exploring frequently attested types of errors in L2 academic writing. As already noted by some researchers, it is possible that formulaic language is not as salient in L2 writing because the way L2 learners produce it is inaccurate (Shin, Cortes and Yoo 2018). In the data-driven methodologies for identifying multi-word sequences, any attempted sequence of formulaic language with an erroneous constituent would not be retrieved automatically. In particular, researchers have noted that English articles, a crucial component of English phraseology in academic prose (Gray and Biber 2013), can be notoriously challenging to use accurately for L2 learners given its extensively complex system and usage patterns (Dulay, Burt and Krashen 1982, Liu and Gleason 2002). Despite this fact, however, few scholars have verified this assumption within multiword units. There is one study by Shin et al. (2018) that aimed to fill this gap by investigating the errors in the use of English definite article in L2 academic writing using lexical bundles. Using a corpus of L2 English academic writing samples produced by Korean first-year undergraduate students (N= 2454) in an English placement test, they analyzed the frequencies of lexical bundles, their core expressions, and the types of definite article errors embedded within the lexical bundles. This study revealed L2 learners' extensive production of article errors, suggesting that infrequent use of lexical bundles may be due to the learners' attempted yet inaccurate use of bundles. However, since their research was based on the retrieved list of fixed bundles, there is the possibility that this approach may have failed to capture the errors in the common but more variable and less predictable lexical frames that are not associated with the common bundles (Gray and Biber 2013).

Thus, the current study attempted to address this gap by investigating the use of English definite article using lexical frames, which allows for an item-internal variable slot. It was expected that using the lexical frame approach would allow for locating a more complete list of errors in the use of the definite article within formulaic language as well as exploring the productivity of L2 phraseology more effectively. Additionally, given that previous research on L2 learners' article errors was mostly based on the bottom-up approach relying on the manual identification of individual instances of errors, the present approach using phraseology was expected to systematize this process and yield valuable insights to inform the teaching of L2 English grammar as well as for the improvement of L2 academic writing.

### 3. Method

#### 3.1 Corpus

The focus of the current study was to examine L2 learners' use of AW-frames in academic writing as well as article errors within those frames. Given this aim, I chose the corpus of L2 academic essays compiled in the university-wide English Placement Test (EPT) administered to incoming international students at a Midwestern US university (*Release 2.2*, January 2018). The corpus is composed of 991 academic essays totaling 363,341 words. The essays were written by undergraduate and graduate L2 English learners with various L1 backgrounds and with different writing skill levels<sup>2</sup> from Spring 2016 to January 2018. These students were required to take this test because they did not meet the exemption criteria set by the university. In terms of English proficiency, their test scores were in the range of 71-100 for TOEFL iBT, 6-7.5 for the IELTS Academic test, or 48-72 for the PTE Academic. To describe the variation of use across levels, two sub-corpora were built according to the placement levels assigned at the operational setting of the EPT. Lower-level (LL) sub-corpus is composed of the essays originally assigned Levels B, C, and D in that these essays were written by the students who did not pass the test and required additional support in academic writing. Higher-level (HL) sub-corpus is composed of the essays assigned Level Pass (exempted from taking additional ESL writing courses). By reorganizing the existing placement levels into two levels, I intended to explore how incoming university students who needed additional writing support are distinguished from those who passed the test. The description of the two sub-corpora is presented in Table 1 below.

**Table 1. Description of the Two Sub-Corpora**

	Number of Texts	Number of Words	Average Text Length
Lower-Level (LL)	657	229,999	350.07
Higher-Level (HL)	334	133,342	339.23
Total	991	363,341	

In terms of task types, the essays were written for two different types of tasks, one summary and one argumentative essay. After reading two passages expressing contrasting positions on a current issue, they wrote a one-paragraph (100-150 words) summary contrasting the positions expressed in the two reading passages (summary essay) and a four-paragraph essay (300-350 words) making an argument on the topic of the readings using information from the texts and their own experiences (argumentative essay). However, since the distinction of task types was not the main interest of the present study, the samples for analysis were not separated by task types (i.e., *text* in Table 1 thus refers to the combined response of the two essays). Besides, since the distinction between undergraduate and graduate levels was not the target of the study, the samples were not separated by students' academic status.

<sup>2</sup> In the operational setting of the EPT, students are assigned one of the three placement levels (Levels B, C/D, or Pass). Graduate and undergraduate students who earn Level B, the lowest level, are considered as requiring substantial support in writing and are placed into an ESL writing course focusing on grammar and paragraph-level writing. Undergraduate students who receive Level C are placed into a course focusing on basic academic writing skills. Graduate students who receive Level D are placed into a course focusing on research writing. Lastly, graduate and undergraduate students with Level Pass are exempted from taking any additional ESL writing courses.

### 3.2 Data Preparation and Analysis

The data preparation and analysis procedures in response to the three research questions are summarized in Table 2.

**Table 2. Research Questions, Data Preparation, and Analysis**

Research Questions	Data Preparation	Data Analysis
1. With what frequencies do L2 learners use AW-frames across levels?	Realized AW-frames <ul style="list-style-type: none"> <li>Extract the occurrences of AW-frames in each sub-corpus using the original 20 most frequent AW-frames with one variable slot (e.g., <i>in the * of</i>).</li> </ul>	<ul style="list-style-type: none"> <li>Compute the overall frequencies of the realized frames in each sub-corpus and compare quantitative patterns across levels</li> <li>Conduct chi-square tests</li> </ul>
2. With what variability and predictability are these realized AW-frames used across levels?	Realized AW-frames <ul style="list-style-type: none"> <li>Manually inspect the cases of errors in the 18 realized yet misused frame types with cataphoric <i>the</i>.</li> </ul>	<ul style="list-style-type: none"> <li>Compute the variability and predictability measures for each frame in each sub-corpus</li> <li>Classify each frame into the categories of variability and predictability</li> <li>Compare the two levels in terms of the proportion of frames by the degree of variability and predictability</li> </ul>
3. What type of errors do the learners produce in the use of the definite article embedded in the realized and unrealized AW-frames?	Unrealized AW-frames <ul style="list-style-type: none"> <li>Extract the occurrences of AW-frames in each sub-corpus using the three most frequent frames with an additional variable slot for the definite article <i>the</i> (e.g., <i>in * * of</i>).</li> <li>Manually inspect the cases of errors in the unrealized frames.</li> </ul>	<ul style="list-style-type: none"> <li>Compute the overall frequencies of the misused frames in each sub-corpus</li> <li>Classify the error types and compute the frequency of each type</li> <li>Qualitatively analyze the cases of misused frames</li> </ul>

The data preparation was composed of two steps. First, to extract the instances of the *realized* AW-frames (i.e., successfully realized and retrieved as the list of frames) in each sub-corpus, the essay files of the two sub-corpora were uploaded separately into the *AntConc* Tools package version 3.5.2 for Windows (Anthony 2018). The list of 20 most frequent AW-frames including the article *the* from Gray and Biber's (2013) framework (see Appendix for the full list) was searched for in each sub-corpus and the output was saved as two separate Excel files. They are a set of four-word frames with a one-variable slot that occurred more than 200 times per million words in the academic writing corpus in their study. The retrieved lists were used to address the first and the second research questions on the contrastive description of the realized AW-frames across levels. Then, the cases of realized yet misused AW-frames were manually inspected to address the third research question about the types of article errors within the frames. In identifying the errors in the definite article, the current study focused only on the cataphoric use of the definite article (i.e., definite reference established by the expressions following *later* in the text) (Biber, Conrad and Leech 2002) for several reasons. First, definite article *the* in eighteen out of the 20 target AW-frames has a cataphoric function, with the noun phrase followed by a prepositional phrase (e.g., *the N of*-phrase, *the N in*-phrase). For the comparability of these frames, the error analysis focused on cataphoric use only. Second, many previous studies have noted that cataphoric use can be identified in a more straightforward manner compared to other functions of the definite article such as anaphoric (phrases with *the* referring back to a previously mentioned item) or situational use (phrases with *the* referring to an entity known from the situation) (Liu and

Gleason 2002, Chrabaszcz and Jiang 2014). For the reliability of the error coding and the scope of the current study, the identification of errors focused only on the cataphoric function.

Second, to identify the cases of *unrealized* AW-frames (i.e., attempted yet failed to be realized as lexical frames) due to article errors, the occurrences of AW-frames in each sub-corpus were first extracted using the three most frequent frames (i.e., *the \* of the*, *in the \* of*, *with the \* of*) with an additional variable slot for the definite article *the*. All possible variants of this frame were searched for in AntConc. For instance, to identify the misuse of the frame *in the \* of*, the two variants with one or two open slots (i.e., *in \* of*, *in \* \* of*) were searched for in the corpus instead of the original frame. Once these lists of frames were generated, they went through further manual refinement. Given that only the non-target-like forms are the focus of this question, target-like idiomatic use such as *in favor of* were manually inspected and excluded from further analysis. Because the identification of errors is not always straightforward, this step was conducted by referring to the semantic/pragmatic framework of article use by Bickerton (1981) as well as former empirical studies on article errors in L2 writing (Liu and Gleason 2002, Chrabaszcz and Jiang 2014, Crosthwaite 2016, Shin et al. 2018). The prepared set of data was then used to address the third research question on the quantitative and qualitative aspects of the misused frames.

The data analysis was composed of eight steps (see Table 2). First, to address the first research question, the overall frequency of the instances of the realized AW-frames was calculated for each sub-corpus and compared across the two placement levels. The frequency rates were normalized to 100,000 words to compensate for the different sizes of the two sub-corpora. Then, chi-square tests were conducted to decide the statistical significance of the frequency differences.

To address the second research question on the distribution of these frames, the variability and predictability of each frame from the retrieved list were computed using the same measures from Gray and Biber (2013): computed variability through type-token ratio (TTR) and measured predictability through the predictability score defined as “the percentage of all occurrences of a frame in which the most frequent filler occurred” (p.114). Then, all frames were classified into one of the four categories based on their TTR (Table 3) and placed into four categories according to the predictability score (Table 4). The original frameworks of variability and predictability from Gray and Biber (2013) were modified to account for the small corpus size of the current study. The variability framework with three categories was broken down into four categories to better reflect the difference between the two levels. Similarly, the scales for each of the four categories of predictability were modified given that the frames in this learner corpus are overall highly unpredictable. The distribution of the 20 frames in the two sub-corpora was compared in terms of the proportions of the frames by the degree of variability and predictability.

**Table 3. Categories for the Degree of Variability**

TTR	Categories
≤ .30	Highly Fixed
.30 - .50	Fixed
.50 - .70	Variable
≥ .70	Highly variable

**Table 4. Categories for the Degree of Predictability**

Predictability Score	Categories
≥ 40%	Highly predictable
25% - 40%	Predictable
10% - 25%	Unpredictable
≤ 10%	Highly unpredictable

Lastly, to address the third research question, the overall normalized frequencies of the realized yet misused frames were calculated for each sub-corpus. Then, sub-types of misuses were identified in an exploratory manner with reference to the previous framework identified by Shin et al. (2018) (Table 5). The cases of misused frames were first analyzed quantitatively in terms of the normalized frequencies of each error type. Then, more qualitative analysis was conducted on the misused frames with respect to the extended grammatical constructions of these frames. The same procedure was followed for the unrealized and ungrammatical frames.

**Table 5. Previous Framework for the Misuse of Frames (Shin et al. 2018)**

Types	Explanation	Example
Omission	Use of zero article where <i>the</i> is required	<i>in</i> $\emptyset$ <i>beginning of</i>
Misformation	Use of indefinite article <i>a/an</i> where <i>the</i> is required	<i>in a</i> <i>beginning of</i>
Addition	Use of definite article <i>the</i> where zero article is required	<i>the</i> <i>one of the main</i>

## 4. Results and Discussion

### 4.1 Overall Distribution of the Realized AW-frames

The current section addresses the first research question regarding the overall frequencies of the AW-frames identified in L2 learner writing (i.e., *the realized AW-frames*). Normalized frequencies per 100,000 words were calculated for the 20 target frames searched for in the LL and the HL corpus and compared across the two corpora to elucidate the differences across the levels. Table 6 summarizes the overall frequencies of the 20 realized AW-frames and the frequencies of the top 10 most frequent frames for the lower versus higher levels.

Overall, higher-level learners employed significantly more AW-frames in their writing compared to lower-level learners with a small effect size ( $\chi^2=58.75$ ,  $df=1$ ,  $p<.001$ ,  $V=.18$ ) (See Table 6). The 10 most frequent frames also show significant differences across the two levels, with significantly more frames identified in higher-level writing, although the size of the effect associated with the level difference was small ( $\chi^2=36.08$ ,  $df=1$ ,  $p<.001$ ,  $V=.16$ ).

**Table 6. Overall Frequencies of the Realized AW-Frames across Levels**

	Lower-Level (LL)	Higher-Level (HL)	$\chi^2$ value	Cramer's V
All frames	434.35	620.96	58.75***	0.18
Top 10 frames	343.48	472.47	36.08***	0.16

*Note.* All frequencies were normalized per 100,000 words and rounded to two decimal points.

\*\*\*statistically significant ( $p<.001$ )

At a first glance, this result seems to contradict the trends revealed in the previous studies on L2 phraseology in

that less proficient L2 learners are generally found to show a tendency to overuse multiword expressions such as lexical bundles (Reppen 2009, Staples et al. 2013) or lexical frames (Garner 2016) compared to more proficient learners. For instance, Garner (2016) indicated a general decline in the overall frequency of the lexical frames with increased proficiency levels in German EFL writing. In fact, it is generally agreed that the heavy reliance on formulaic language is usually attested in lower-level writing (Å del and Erman 2012, Chen and Baker 2010, Reppen, 2009, Staples et al. 2013). However, the retrieved inventories of the frames in the two studies are not equivalent in nature, which needs to be considered in interpreting the overall trends. While Garner's study took an exploratory approach to identify all four-word lexical frames, the current study focused on retrieving only the function word AW-frames that were already identified as most frequent in academic prose. In fact, the majority of the frames identified in Garner's study were verb-based frames, which are much more common in conversation rather than in academic writing (Biber 2009, Gray and Bibe 2013). In terms of the function word frames, Garner also showed that learners employed slightly more function word frames at higher proficiency levels, which aligns with the current finding. Given the highly focused inventory of the target frames for the current study (i.e., the most frequent function word AW-frames including *the*), higher-level learners' more frequent uses of AW-frames revealed in this study can be understood as a manifestation of more academic nature of the language used in higher-level writing. Previous studies on the use of lexical bundles in L2 academic writing have noted a similar tendency that less proficient learners tended to have issues in using register-appropriate bundles, showing underuse of academically appropriate bundles (Adel and Erman 2012, Chen and Baker 2010, Paquot and Granger 2012). The present study demonstrated such a tendency in L2 learners' use of lexical frames, confirming that more proficient learners showed more characteristics of the phraseology of academic writing compared to their less proficient counterparts. It is important to note, however, that though statistically significant, the sizes of such differences were not large based on the small effect sizes.

#### 4.2 Variability and Predictability of the Realized AW-frames

This section addresses the second research question regarding the internal variability and predictability of the realized AW-frames. To compare the variability of the AW-frames across levels, type-token ratios (TTRs) were calculated for the 20 AW-frames identified in the first research question. They were then categorized into four categories based on the TTR. Figure 1 visualizes the distribution of the AW-frames that are highly fixed, fixed, variable, or highly variable for the two placement levels displayed by the proportion of distinct frame types.

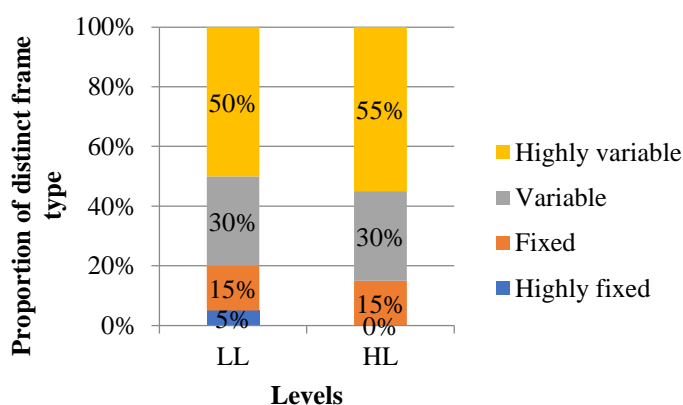


Figure 1. The Proportion of Frames by the Degree of Variability



First, in terms of the variability distribution of all 20 frames, the lower and higher-level learners display similar overall trends. In both levels, highly variable frames (e.g., *the \* and the*) are the most frequently attested type followed by variable frames (e.g., *of the \* of*). These two categories account for around 80% of all frame types in both levels indicating a tendency that both levels of students relied primarily on relatively variable frames. This overall trend is not surprising given the register characteristic of the corpus as academic writing. As already documented by Gray and Biber (2013) in their investigation of register variation of lexical frames, academic writing typically relies on lexical frames that are highly variable and predominantly composed of function words compared to conversation. Given the fact that all 20 target frames in this investigation are function word frames including the definite article *the*, the current result confirms the previous finding on the variability of AW-frames.

However, it is also noteworthy that the preponderance of variable frames is slightly more pronounced in higher-level writing. While none of the frames used by higher-level learners are categorized as *highly fixed*, lower-level students used some highly fixed frames (*that the \* of*), which accounted for 5% of all frame types. Such a level difference is further elucidated if we look at the individual AW-frames focusing on the most frequent 12 frames<sup>3</sup> sorted by the variability (i.e., TTR). Figure 2 displays the distribution of the top 12 most frequent AW-frames sorted by the degree of variability.

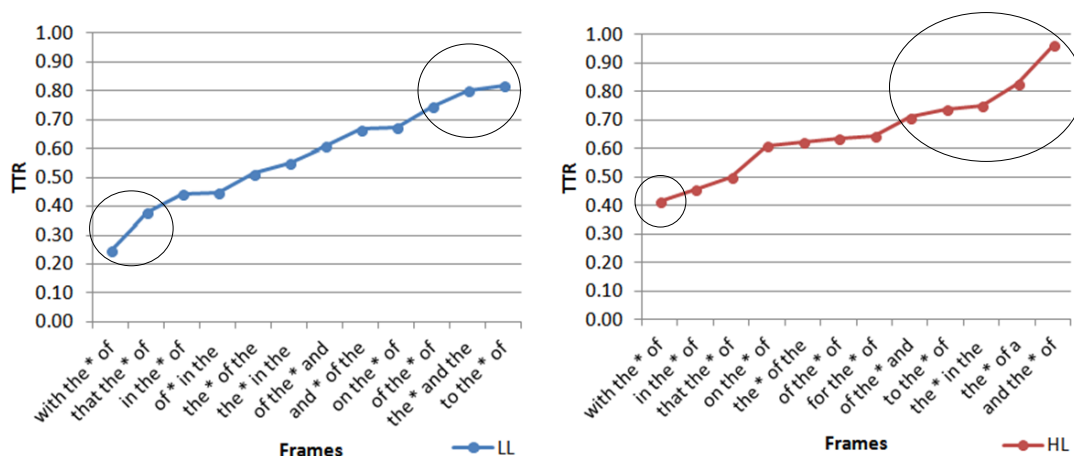


Figure 2. Distribution of Top 12 Frequent Frames by the Degree of Variability

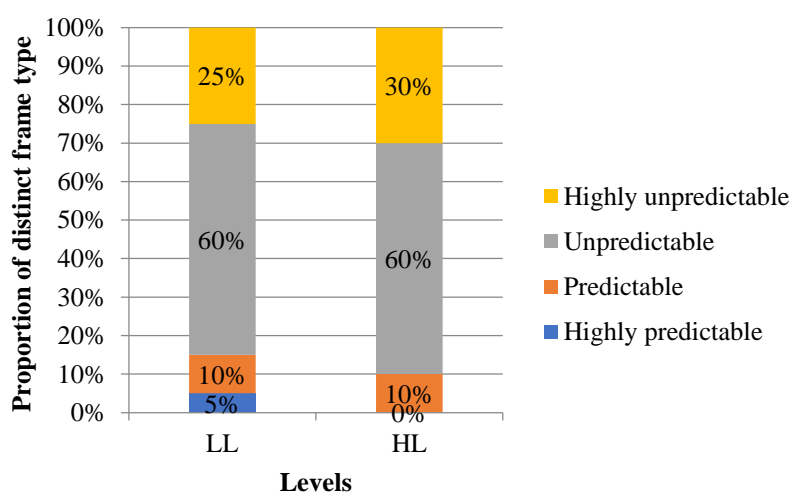
As illustrated at the right end of the red line (see the graph on the right in Figure 2), five frame types in the HL corpus are concentrated in the high TTR area (i.e., TTR over .70). We can also see that the maximum TTR value in the HL corpus (i.e., a value close to 1) is much higher than the maximum TTR of the frames in the LL corpus (graph on the left in Figure 2). This result suggests higher-level learners' tendency to rely on highly variable frames. In contrast, the LL corpus does not include as many frames with high TTR, as can be seen at the right end of the blue line (Figure 2). With regard to the relatively fixed frames with low variability, the minimum TTR value in the LL corpus is much lower than its counterpart in the HL corpus. This result indicates that the fixed frames in lower-level writing are, in fact, more fixed than those in higher-level writing (e.g., *at the \* of*), although they were

<sup>3</sup> The frame types in the lower end in terms of the rate of occurrences were excluded from this further analysis to make the analysis of variability more meaningful.

both categorized as *fixed* frames in the analysis above. Thus, this figure shows that in terms of the most frequently employed frames, the degree to which the fixedness and the variability of the frames vary across the two levels is greater, which was not as salient in the categorical analysis of the overall distributional trends shown in Figure 1. That is, unlike lower-level learners who tended to rely on a few fairly fixed frames (e.g., *with the \* of*), more proficient learners employed the frames with a wider variety of fillers occupying the variable slot.

The noted trends in the level differences are consistent with the previous findings showing that formulaic language in L2 writing is generally more variable, concomitant with the increased learner level. For instance, Garner (2016) also indicated that advanced German EFL learners tended to rely on the phrase-frames with higher variability compared to less proficient learners, suggesting that higher-level learners can employ phrase-frames more productively. The two placement levels analyzed in the current analysis did not show as striking differences as those revealed in Garner's study, presumably due to the small corpus size and more focused inventory of the target frame types. However, the performance of the L2 learners in the present study still aligns with the general trend that L2 learners employ lexical frames more productively with increased writing skills.

In terms of the predictability of the frames across levels, the predictability measures were calculated for the 20 AW-frames. They were then categorized into four categories based on the predictability score. Figure 3 summarizes the distribution of the AW-frames that are highly predictable, predictable, unpredictable, or highly unpredictable for the two placement levels displayed by the proportion of distinct frame types.



**Figure 3. The Proportion of Frames by the Degree of Predictability**

First, the predictability distribution of all 20 frames displays a similar overall trend across levels as in the variability analysis above. In both levels, unpredictable and highly unpredictable frames account for around 80% of the frames with only 10% to 15% of the frames categorized as *predictable* (see Figure 3). This result indicates that L2 learners in both levels tended to rely on unpredictable frames, although the tendency to rely on more unpredictable frames was slightly more prominent in higher-level writing.

The predictability distribution of the 12 most frequent frames points to a similar trend across levels as displayed in Figure 4. In Figure 4, we can see that the blue and the red lines (for the LL and HL corpus, respectively) generally follow similar distributional patterns within the score range of around 10% to 30% except for one frame in the LL corpus (i.e., *with the \* of*).

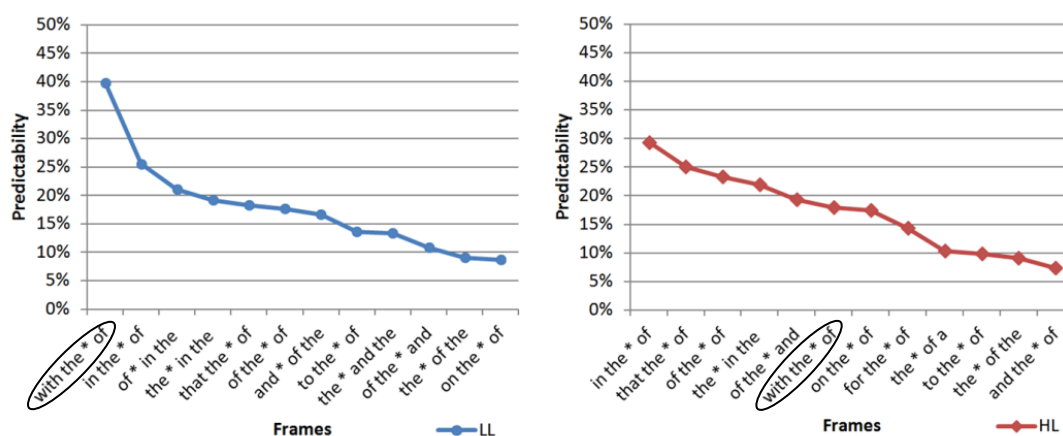


Figure 4. Distribution of Top 12 Frequent Frames by the Degree of Predictability

One interesting finding that deserves further attention concerns the frame *with the \* of*. From Figure 4, it can be seen that lower-level learners employed this frame in a fairly predictable manner as indicated by the highest predictability score among the 12 most frequent frames (i.e., predictability of around 40%). In contrast, in the HL corpus, the same frame shows a comparatively lower predictability score suggesting that higher-level learners did not use this frame in as predictable manner. At a first glance, this result would seem to contradict the variability distribution discussed in Figure 3 because this frame was found to be the most fixed frame among the top 12 frames in higher-level writing. In a logical sense, we would expect a fixed frame to be generally more predictable. In fact, in Garner's (2016) study, the distribution of the predictability of the frames across levels generally aligned with the results for the variability indicating that with increased proficiency, the variability of the frames tended to increase while the predictability tended to decrease. Interestingly, however, for several frequent frames in the higher-level writing in the present study, the variability and the predictability measures did not align.

A closer qualitative analysis of the list of fillers occupying the variable slot for the frame *with the \* of* and the extended context following the frame demonstrates that lower-level learners predominantly relied on one frequent filler, *development* (underlined in examples (1)-(2)), while higher-level learners employed this frame with several equally frequent fillers (i.e., *development*, *help*, *use*) (underlined in (3)-(4)). Another notable difference relates to the noun phrase that follows the frame (italicized in (1)-(4)). Interestingly, in the LL corpus, the phrase *with the development of* is typically followed by a fixed set of noun phrases (e.g., either *artificial intelligence* or *technology*), suggesting that these phrases were presumably used as a memorized chunk or a borrowed phrase from the prompt rather than as a genuine production of the frame *with the \* of*. On the other hand, in the HL corpus, the noun phrase following the frame comes from a much wider range of inventories (e.g., *various applications* in (3), *internet* in (4)), mostly occurring only once. This finding suggests that higher-level learners tended to use this frame more productively by varying the fillers as well as the following noun phrase based on their communicative needs in the discourse context (Römer 2009).

- (1) **With the development of *artificial intelligence***, robots can do more work that human cannot do. [LL]
- (2) **With the development of *technology***, life becomes more leisure. [LL]
- (3) **With the help of *various applications***, learning how to use an ipad is very quick and easy. [HL]
- (4) **With the use of *internet***, my village is now able to learn new things and use social media. [HL]

The noted qualitative finding aligns with the previously found developmental sequence in L2 phraseology (Ellis 2002) that lower-level L2 learners tend to rely heavily on memorization and highly fixed expressions, gradually moving toward more native-like self-constructed language with increased proficiency. Ellis noted that there exists a clear developmental process in which “learners move from fixed formulaic phrases to limited frame patterns to fully productive patterns” (Garner 2016, p. 33). While the variability and predictability measures of the frames fail to capture notable differences between the two levels quantitatively (i.e., both levels employed AW-frames in a generally variable and unpredictable manner), qualitative analysis demonstrates that the two levels were clearly differentiated in terms of the ability to use lexical frames as a productive scheme for constructing language. Considering that productive use of lexical frames in various contexts entails the ability to understand and deconstruct the underlying structure of the frame (Myles, Hooper and Mitchell 1998), L2 learners from the current study demonstrate differences in terms of this cognitively demanding and complex ability.

### 4.3 Erroneous Uses of the Definite Article within AW-frames

The current section addresses the third research question regarding the errors in the use of the definite article embedded within the AW-frames. The first section deals with the errors found in the realized frame types. The next section discusses the unrealized AW-frames due to article errors focusing on the three most frequent frame types as target frames.

#### 4.3.1 Realized yet ungrammatical frames

Table 7 presents the normalized frequencies of the overall instances of the realized yet misused frames as well as the frequencies of those misused frames categorized into three types of errors in the use of the definite article.

**Table 7. Frequencies of the Realized Yet Misused AW-Frames across Error Types and Levels**

Error Types	Lower-Level (LL)	Higher-Level (HL)	$\chi^2$ value	Cramer's V
Omission	17.4	7.50	6.01*	0.19
Addition	1.74	7.00	6.36*	0.19
Non-nominal head	37.40	13.50	16.85***	0.32
Total	56.54	28.00	14.93***	0.30

*Note.* All frequencies were normalized per 100,000 words and rounded to two decimal points.

\*\*\*statistically significant ( $p < .001$ ), \* ( $p < .05$ )

In Table 7, it is first noticeable that lower-level learners overall produced significantly more article errors within the realized frames, compared to higher-level learners with a moderate effect size ( $\chi^2=14.93$ ,  $df=1$ ,  $p<.001$ ,  $V=.30$ ). This result suggests that particularly for the less proficient L2 learners, the fact that the AW-frames were realized in writing does not necessarily mean that they were in target-like forms. This finding yields meaningful insight into the existing literature of L2 phraseology given that previous research mostly investigated underuse or overuse patterns in multiword expressions assuming that successfully retrieved expressions are grammatical. For instance, when Garner (2016) compared the frequencies of phrase-frames across proficiency levels, the accuracy of these frames was not taken into consideration. The current finding suggests that L2 learners may, in fact, produce

extensive errors within lexical frames. Since they were successfully retrieved as a frame, such non-target-like forms may go undetected.

The retrieved list of misused frames was categorized into three sub-types of errors; *omission*, *addition*, and *non-nominal head* errors, as exemplified in (5)-(7), respectively. First, the realized frames missing the obligatory definite article are categorized as omission errors. The phrase *and majority of the* in (5) is a realization of the frame *and \* of the*. Since the learner failed to provide the obligatory definite article before *majority*, it is marked as a case of an omission error. The second type, addition error is the opposite case of omission error in which definite article *the* is supplied within the frame where zero article is grammatical. The phrase *is the one of* in (6) is a variant of the frame *is the \* of*. However, since placing definite article is ungrammatical in this context, it is categorized as an addition error. Lastly, non-nominal head error refers to the case when the non-nominal element such as a verb or an adjective follows the definite article. This is an indirect type of error in that the presence or absence of the definite article itself is not problematic. In (7), the phrase *the efficient of the*, which is a realization of the frame *the \* of the* is marked as this error type since the adjective *efficient* should be replaced by the noun *efficiency*.

- (5) Elizabeth Woyke did a survey **and 0 majority of the** student find it comforting to use them regularly. [LL] [*and \* of the*, Omission]  
 (6) Text two **is the one of** samples. [LL] [*is the \* of*, Addition]  
 (7) So, the iPad increase **the efficient of the** study and made the study get more interesting. [LL] [*the \* of the*, Non-nominal head]

In terms of the distribution of the three error types across levels (Table 7), lower-level learners produced significantly more omission ( $\chi^2=6.01$ ,  $df=1$ ,  $p<.05$ ,  $V=.19$ ) and non-nominal head errors ( $\chi^2=16.85$ ,  $df=1$ ,  $p<.001$ ,  $V=.32$ ) than higher-level learners. What is particularly noticeable in this table is that lower-level learners produced an extensive number of frames with a non-nominal head when compared to their higher-level counterparts, associated with a moderate effect size. This result points to a meaningful finding that placing a nominal element after the definite article in the cataphoric *the*-phrase (*the*-phrase followed by an *of*-phrase modifier, e.g., *the \* of the*), which is assumed to be quite basic grammatical knowledge, can be challenging for less proficient L2 learners. It is also noteworthy that this is the most prominent type of error in both levels accounting for over 50 percent of the errors in the realized frames. This result indicates that L2 learners tend to have the problem of using a non-nominal head for the definite article in this particular structure regardless of levels. Interestingly, this error type has not been documented, to the best of my knowledge, in previous studies that investigated the errors produced by L2 learners in the formulaic language (e.g., lexical bundles) (Huang, 2015; Shin et al. 2018). For instance, Shin et al. (2018) investigated article errors embedded in lexical bundles focusing only on the attempted yet unrealized bundles, but failed to capture this error type. Huang (2015) explored more general errors embedded in the use of lexical bundles by L2 learners, but have not identified this error type. Therefore, the current finding provides valuable insight as to the additional aspects to focus on in the teaching of AW-frames for L2 learners, namely the appropriate use of nominal head after definite article, in addition to the types of errors already identified in previous studies.

#### 4.3.2 Unrealized frames due to article errors

This last section discusses unrealized frames due to the errors in the definite article. As already mentioned, the

article errors discussed in this section focus only on the errors in the cataphoric use of *the*<sup>4</sup>. Table 8 summarizes the normalized frequencies of the attempted yet unrealized AW-frames in the LL and HL corpus for the three most frequent target frames (i.e., *with the \* of*, *the \* of the*, *in the \* of*).

**Table 8. Frequencies of the Unrealized AW-Frames across Levels**

Target Frame	Lower-Level (LL)	Higher-Level (HL)	$\chi^2$ value	Cramer's V
<i>with the * of</i>	9.57	7.50	0.41	0.05
<i>the * of the</i>	26.96	14.25	6.12*	0.19
<i>in the * of</i>	8.26	15.75	4.30*	0.17
Total	44.78	37.50	1.07	0.08

Note. All frequencies were normalized per 100,000 words and rounded to two decimal points.

\*statistically significant ( $p < .05$ )

As can be seen in Table 8, the overall frequencies of the unrealized frames are not significantly different across levels ( $\chi^2=1.07$ ,  $df=1$ ,  $p > .05$ ,  $V= .08$ ) although lower-level learners produced slightly more errors compared to higher-level learners. This result should be interpreted with caution since only a subset of the target frames was analyzed for the scope of the present study. In terms of the individual frames, the frequency of the unrealized frame for *the \* of the* is significantly higher in the LL corpus compared to the HL corpus with a small effect size ( $\chi^2=6.12$ ,  $df=1$ ,  $p < .05$ ,  $V= .19$ ), while the opposite pattern was observed for the frame *in the \* of* ( $\chi^2=4.30$ ,  $df=1$ ,  $p < .05$ ,  $V= .17$ ).

One question of interest, however, is whether these frames were unrealized due to similar types of errors. Further analysis of the sub-types of errors in the unrealized frames reveals different types of errors from what was discovered in the realized frames above. Table 9 displays the distribution of the three types of errors that emerged from the unrealized frames by the normalized frequency of each error type across levels.

**Table 9. Frequencies of the Unrealized AW-Frames across Levels**

Error Types	Lower-Level (LL)	Higher-Level (HL)	$\chi^2$ value	Cramer's V
Omission	37.39	33.75	0.31	0.05
Misformation	5.22	3.00	0.94	0.08
Non-nominal head	2.17	0.75	1.03	0.08
Total	44.78	37.50	1.06	0.08

Note. All frequencies were normalized per 100,000 words and rounded to two decimal points.

First, as Table 9 shows, learners produced three types of errors in the use of the definite article, *omission*, *misformation*, and *non-nominal head* errors. As the results of the chi-square tests and Cramer's V effect sizes indicate, no statistically significant differences were found in the frequencies of each error type across levels.

<sup>4</sup> Thus, in the case of the frame *the \* of the*, I focused on finding the errors related to the first definite article, but not the second definite article.

Examples (8)-(10) illustrate each error type.

- (8) On the other hand,  $\emptyset$  author of the second text sees many disadvantages concerned about iPads. [HL] [*the* \* *of the*, Omission]
- (9) As we all know, most people are always **in a state of** depressing as they have to do the work they dislike to earn more money. [LL] [*in the* \* *of*, Misformation]
- (10) That would leave humans **with  $\emptyset$  lose of** drive and identity which would be total chaos. [HL] [*with the* \* *of*, Non-nominal head]

One notable difference from the errors identified in the realized frames is that learners produced misformation errors using indefinite articles *a* or *an* in the place of the definite article, as illustrated in example (9). The phrase *in a state of* is presumably attempted as a variant of the frame *in the* \* *of* but failed to be realized due to the ungrammatical use of an indefinite article. Another noticeable aspect in the frequencies of each error type is that the omission of the obligatory *the* is the most salient type of error at both levels as indicated by the highest frequencies among the three types (Table 9). This result suggests that both lower- and higher-level learners' attempts to use AW-frames failed most likely due to the use of zero articles when the definite article is required.

This result echoes previous findings that zero articles tend to be overused in the place of the definite article by L2 English learners, in particular, by those learners whose L1 does not have the article system (Chrabaszcz and Jiang 2014). However, the current literature is still inconclusive about the effect of article-less L1 on the correct use of English definite article. In terms of article errors within multi-word expressions, Shin et al. (2018) found a similar tendency that almost all cases of unrealized lexical bundles were due to the omission of the obligatory definite article. Thus, the current finding generally confirms previous findings of article errors and gives additional insight that correct placement of obligatory definite article within lexical frames can be challenging to achieve for L2 learners. However, this statement is still inconclusive since the current study focused only on the cataphoric function of the definite article using only a subset of the target frames.

## 5. Conclusion

The present research investigated the use of recurrent, discontinuous frames in L2 academic writing focusing on analyzing both the target-like and ungrammatical use of academic writing frames. By applying the lexical frame approach to L2 phraseology, which allows for an item-internal variation (Biber 2009), this study helps us gain a better understanding of how L2 learners' productive use of multiword expressions differs across writing skill levels.

The results of the first research question showed that higher-level learners employed AW-frames more frequently compared to lower-level learners. This finding is in line with the findings from Granger (2016), which showed that German L2 English learners employed more function word frames at higher proficiency levels when compared to lower-levels. This result echoes similar tendencies noted in the previous studies on the use of lexical bundles in L2 academic writing: less proficient L2 learners tended to have problems in using register-appropriate bundles, showing underuse of academically appropriate bundles (Å del and Erman 2012, Chen and Baker 2010, Paquot and Granger 2012). Overall, the present finding can be understood as a manifestation of the more academic nature of the phraseology of higher-level writing compared to their less proficient counterpart.

In terms of the variability and predictability of the frames, the results of the second research question demonstrated that higher-level learners tended to employ highly variable AW-frames in an unpredictable manner whereas less proficient learners incorporated the same inventory of frames in a comparatively fixed and predictable

way. The noted level differences are consistent with Garner's (2016) finding although the two levels in the current study showed less striking differences presumably due to the small corpus size. Garner also found that concomitant with the increased L2 proficiency, learners tended to rely on more variable frames in a less predictable manner. Overall, findings from the current study serve to confirm the developmental sequence of L2 phraseology (Ellis 2002), highlighting that L2 learners move from highly fixed expressions to more self-constructed language with increased proficiency. By applying the lexical frame approach, the present study attested that the command of skillfully using lexical frames as a productive scheme for constructing language can be a signaling factor of a competent L2 English writer.

In terms of the misused frames, regardless of levels, L2 learners produced extensive errors in the use of the definite article within both realized and unrealized frames. This finding yields meaningful insight into the literature of L2 phraseology, which mostly investigated multiword expressions assuming that successfully retrieved expressions are grammatical. It was shown that particularly for the less proficient learners the fact that the frames were successfully realized does not necessarily mean that they were in the target-like forms. More specifically, L2 learners showed the problem of using a non-nominal head for the definite article within realized frames, which is a meaningful finding that has not been documented in the literature. In terms of the unrealized frame due to an article error, both levels of learners predominantly produced omission errors, which parallels previously noted findings on article errors within formulaic language that suggested that L2 learners whose L1 does not have the article system tended to use zero articles in the place of definite article *the* (Chrabaszcz and Jiang 2014, Roberston 2000, Shin et al. 2018). For instance, Shin et al. (2018) revealed that almost all cases of unrealized lexical bundles produced by Korean English learners were due to the omission of definite article. The present study adds to the literature the finding that correct placement of obligatory definite article within lexical frames can be challenging to achieve for L2 learners.

Taken together, the present study offers novel insights into what constitutes article errors in both realized frames and attempted yet unrealized frames for L2 learners. Unlike previous studies on article errors that mostly focused on analyzing the placement of the article itself, the current investigation found that the use of a non-nominal head within frames can be problematic, particularly for lower-level learners, by taking the lexical frame approach. This finding also provides valuable implications for the aspects to focus on in the teaching of AW-frames to L2 learners, namely the appropriate use of a nominal head after the definite article, in addition to the types of errors already identified in the previous literature.

Several suggestions for further research are derived from the limitations of the present study. First, due to the small sample size, the categorical analysis of the variability and predictability of the frames failed to capture meaningful quantitative differences between the two levels. The original frameworks of variability and predictability from Gray and Biber (2013) were modified to account for the small corpus size. Therefore, a replication study using a larger corpus is recommended to address this issue. In addition, due to practical constraints, the error analysis of the unrealized frames focused only on the cataphoric use of the definite article, based on only a subset of the target frames. The current findings herein could be supplemented by future research that adds the exploration of the errors in other functions of the definite article such as anaphoric or situational use.



## References

- Ädel, A. and B. Erman. 2012. Recurrent word combinations in academic writing by native and non-native speakers of English: A lexical bundles approach. *English for Specific Purposes* 31, 81–92.
- Anthony, L. 2018. *AntConc* (Version 3.5.2) [Computer software]. Tokyo, Japan: Waseda University Available at <http://www.laurenceanthony.net/>
- Biber, D. 2009. A corpus-driven approach to formulaic language in English: Multi-word patterns in speech and writing. *International Journal of Corpus Linguistics* 14(3), 275–311.
- Biber, D., S. Conrad, and G. Leech. 2002. *Longman student grammar of spoken and written English*. London: Longman.
- Bickerton, D. 1981. *Roots of Language*. Ann Arbor, MI: Karoma Press.
- Chen, Y. and P. Baker. 2010. Lexical bundles in L1 and L2 academic writing. *Language Learning and Technology* 14(2), 30–49.
- Chrabaszcz, A. and N. Jiang. 2014. The role of the native language in the use of English nongeneric definite article by L2 learners: A cross-linguistic comparison. *Second Language Research* 30(3), 351–379.
- Crosthwaite, P. 2016. L2 English article use by L1 speakers of article-less languages. *International Journal of Learner Corpus Research* 2(1), 68-100.
- Dulay, H., M. Burt, and S. Krashen. 1982. *Language two*. Oxford: Oxford University Press.
- Ellis, N. C. 2002. Frequency effects in language processing: A review with implications for theories of implicit and explicit language acquisition. *Studies in Second Language Acquisition* 24, 143–188.
- Fletcher, W. 2011. *Phrases in English*. Available at <http://phrasesinenglish.org>.
- Garner, J. R. 2016. A phrase-frame approach to investigating phraseology in learner writing across proficiency levels. *International Journal of Learner Corpus Research* 2(1), 31-67.
- Gray, B. and D. Biber. 2013. Lexical frames in academic prose and conversation. *International Journal of Corpus Linguistics* 18(1), 109-136.
- Gilquin, G. and S. Granger. 2015. Learner Language. In D. Biber and R. Reppen, eds., *The Cambridge Handbook of English Corpus Linguistics*, 418-435. Cambridge: Cambridge University Press.
- Huang, K. 2015. More does not mean better: Frequency and accuracy analysis of lexical bundles in Chinese EFL learners' essay writing. *System* 53, 13–23.
- Liu, D., and J. Gleason. 2002. Acquisition of the article the by nonnative speakers of English. *Studies in Second Language Acquisition* 24(1), 1–26.
- Myles, F., J. Hooper and R. Mitchell. 1998. Rote or rule? Exploring the role of formulaic language in classroom foreign language learning. *Language Learning* 48(3), 323–363.
- Paquot, M. and S. Granger. 2012. Formulaic language in learner corpora. *Annual Review of Applied Linguistics* 32, 130-149.
- Renouf, A. and J. McH. Sinclair. 1991. “Collocational frameworks in English”. In K. Aijmer and B. Altenberg, eds., 128-143. *English Corpus Linguistics*. London: Longman.
- Reppen, R. 2009. Exploring L1 and L2 writing development through collocations: a corpus-based look. In A. Barfield, and H. Gyllstad, eds., *Researching Collocations in Another Language*, 49-59. London: Palgrave Macmillan.
- Römer, U. 2009. The inseparability of lexis and grammar: Corpus linguistic perspectives, *Annual Review of Applied Linguistics* 7, 140–162.
- Römer, U. 2010. Establishing the phraseological profile of a text type: The construction of meaning in academic

- book reviews. *English Text Construction* 3(1), 95–119.
- Shin, Y. K., V. Cortes and I. W. Yoo. 2018. Using lexical bundles as a tool to analyze definite article use in L2 academic writing: An exploratory study. *Journal of Second Language Writing* 39, 29-41.
- Staples, S., J. Egbert, D. Biber. and A. McClair. (2013). Formulaic sequences and EAP writing development: lexical bundles in the TOEFL iBT writing section. *Journal of English for Academic Purposes* 12(3), 214-225.
- Stubbs, M. 2007. An example of frequent English phraseology: Distributions, structures and functions. In R. Facchinetti, ed., *Corpus Linguistics 25 Years On*, 89-105. Amsterdam: Rodopi.

Examples in: English

Applicable Languages: English

Applicable Level: Tertiary

### Appendix

#### List of 20 Most Frequent AW-frames from Gray and Biber (2013)

*the \* of the*  
*in the \* of*  
*of the \* of*  
*the \* of a*  
*to the \* of*  
*and the \* of*  
*on the \* of*  
*of the \* and*  
*for the \* of*  
*at the \* of*  
*with the \* of*  
*a \* of the*  
*by the \* of*  
*that the \* of*  
*the \* and the*  
*and \* of the*  
*is the \* of*  
*the \* in the*  
*the \* of this*  
*of \* in the*