THE JOURNAL OF ASIA TEFL
Vol. 18, No. 1, Spring 2021, 142-160
http://dx.doi.org/10.18823/asiatefl.2021.18.1.9.142

The Journal of Asia TEFL

http://journal.asiatefl.org/

e-ISSN 2466-1511 © 2004 AsiaTEFL.org. All rights reserved.

A Corpus-Based Analysis of Lexical Bundles between English L1 and L2 Writers in Medical Journal Abstracts: Focusing on Structures and Functions

Eun-Soo Kim
Ewha Womans University, Korea

Eun-Joo Lee
Ewha Womans University, Korea

Despite the importance of research article abstracts and difficulties that English L2 writers encounter when writing research articles in English in the field of medicine, little research has investigated lexical bundles in medical journal abstracts. This paper explores the structural and functional characteristics of lexical bundles between two corpora consisting of medical journal abstracts written by English L1 and L2 writers, respectively. The findings, from extensive use of verb phrase-based (VP-based) and text-oriented bundles by English L2 writers, as well as that of noun phrase-based (NP-based) and research-oriented bundles by English L1 writers, show that English L2 writers lack awareness of discipline-specific bundles in medical journal abstracts. The findings of the study would be a valuable source for English L2 writers to explore the use of lexical bundles by English L1 writers in medical journal abstracts reporting clinical trials and randomized controlled trials, thereby helping them develop a better understanding of medical journal abstract conventions.

Keywords: lexical bundles, structures and functions, medical journal abstracts, clinical trials, randomized controlled trials (RCTs), a corpus-based analysis

Introduction

Lexical bundles are defined as a high frequency of multi-word sequences in a given register (Biber & Barbieri, 2007). There has been a consensus that lexical bundles help writers to familiarize themselves with the rhetorical structure, language usage, and discourse functions commonly shared among community members in specific disciplines (Cortes, 2004; Hyland, 2008). With growing attention to a genre of research articles, studies have investigated lexical bundles in the genre of research articles (e.g., Chen & Baker, 2010; Cortes, 2013; Esfandiari & Barbary, 2017; Le & Harrington, 2015; Maswana, Kanamaru, & Tajino, 2013; Omidian, Shahriari, & Siyanova-Chanturia, 2018; Pan, Reppen, & Biber, 2016; Pérez-Llantada, 2014; Qin, 2014; Saber, 2012). Given the kinds of disciplines, single disciplines such as psychology (Esfandiari & Barbary, 2017), applied linguistics (Le & Harrington, 2015; Qin, 2014), telecommunication (Pan et al., 2016), and medicine (Saber, 2012) as well as multi-disciplines (Chen & Baker, 2010; Cortes, 2013; Omidian et al., 2018; Pérez-Llantada, 2014) have been investigated. Although Saber (2012) investigated phraseological patterns in medical research articles, the study primarily focused on keywords. Therefore, additional studies are needed to examine the structural and functional characteristics of lexical bundles in the field of medicine.

Since the organization of texts and language use are extremely rhetorical in research article abstracts

© 2021 AsiaTEFL All rights reserved

(Hyland & Tse, 2005; Van Bonn & Swales, 2007), the research article abstracts would be a valuable source to investigate lexical bundles in specific disciplines. Given the part genre (Abstract), lexical bundles were investigated in doctoral dissertation abstracts in hard science disciplines (Lu & Deng, 2019) and research article abstracts in both soft science and hard science disciplines (Omidian et al., 2018). Except for the two aforementioned studies, little research has been conducted to investigate lexical bundles in research article abstracts. Particularly in the field of medicine, editors from some medical journals (e.g., *British Medical Journal*) tend to screen research articles for acceptance or rejection based on reading abstracts (Schroter & Barratt, 2004) because text structures and language use in medical research article abstracts are extremely controlled (Huckin, 2006). Despite the importance of research article abstracts in the field of medicine, there has been no research to investigate lexical bundles within medical journal abstracts.

The use of lexical bundles between English L1 and L2 student writers (Chen & Baker, 2010; Lu & Deng, 2019) was investigated as well as between English L1 and L2 expert writers (Esfandiari & Barbary, 2017; Pan et al., 2016; Pérez-Llantada, 2014; Qin, 2014). Although the need for English for research publication purposes among English L2 writers was reported in the field of medicine (Martín, Rey-Rocha, Burgess, & Moreno, 2014), few studies have explored language use between English L1 and L2 writers in the field of medicine. With growing attention to English for research publication (Flowerdew, 2008, 2015), English L2 writers have been extremely concerned about rhetorical structures and language features expected by their target journals (Cho, 2009; Martín, Rey-Rocha, Burgess, & Moreno, 2014). As observed in the studies of Nam (2017) and Shin et al. (2019), the usage patterns of lexical bundles by English L2 writers were different from those used by English L1 writers. By encouraging English L2 writers to be exposed to essential lexical bundles commonly used by English L1 writers, they could improve their proficiency in academic writing for research publication purposes in the field of medicine. Explicit awareness of fixed or semi-fixed lexical bundles in a specific discipline might empower English L2 writers to deliver information accurately and effectively, thus leading to successful communication with discourse community members. Although the Korean authors as English L2 writers in the present study were considered professional authors, they could be in the developmental stage, similar to the Korean students in the previous studies, and lack awareness of discipline-specific conventions and lexical bundles. The reader will most likely observe, from the findings of this study, distinctive usage patterns with the overuse of certain types of lexical bundles by Korean authors due to the L1 transfer.

Chen and Baker (2010) and Pan et al. (2016) found the more frequent use of phrase bundles by English L1 and clausal bundles by English L2 writers. However, the opposite findings have been observed in other studies (Esfandiari & Barbary, 2017; Lu & Deng, 2019; Pérez-Llantada, 2014). As for functions, it has been consistently found that research-oriented bundles are more frequently observed in hard disciplines, whereas text-oriented and stance bundles are more prominently found in soft disciplines (Hyland, 2008; Omidian et al., 2018). However, inconsistent results were found between English L1 and L2 writers in previous studies. The more frequent use of text-oriented and stance-oriented bundles by English L2 writers was found in Pan et al. (2016) and Lu and Deng (2019), but the opposite was found in three other studies (Chen & Baker, 2010; Pérez-Llantada, 2014; Qin, 2014). The more frequent use of text-oriented bundles and stance-oriented bundles by English L2 and L1 writers, respectively, was observed in Esfandiari and Barbary (2017). Based on previous findings, it was hypothesized that a difference between English L1 and L2 writers would exist even though they would be at the same level as expert writers. Furthermore, there would be some different characteristics of lexical bundles due to medical disciplines and medical journal abstracts.

The present study aims to explore the structural and functional characteristics of lexical bundles between English L1 and L2 writers (Koreans) by subsections of medical journal abstracts. Lexical bundles in the present study were defined as three- to nine-word sequences occurring across five different texts. Types and tokens of lexical bundles, categorized into structural and functional classification by subsections of medical journal abstracts, were examined between two corpora consisting of abstracts by native speakers of English (NSE) writers and non-native speakers of English (NNSE) writers. The

© 2021 AsiaTEFL All rights reserved

research questions are as follows:

1. What are the structural characteristics of lexical bundles in an NSE corpus and an NNSE corpus by subsections of medical journal abstracts?
2. What are the functional characteristics of lexical bundles in an NSE corpus and an NNSE corpus by subsections of medical journal abstracts?

Literature Review

Previous Research on Lexical bundles between English L1 and L2 writers

The frequency of lexical bundles and the dispersion criterion should be taken into consideration to qualify a multi-word sequence as a lexical bundle. Depending on the size of the corpus, cut-off points were somewhat arbitrary. Previous research employed the frequency threshold of 40 times (Esfandiari & Barbary, 2017; Pan et al., 2016), 20 times (Esfandiari & Barbary, 2017; Lu & Deng, 2019; Pérez-Llantada, 2014), 10 times (Esfandiari & Barbary, 2017), and seven times (Qin, 2014) per million words. In the case of the small size of the corpus amounting to less than 200,000 words, the raw frequency threshold of four times for 150,000 words was chosen by Chen and Baker (2010). As for the dispersion threshold, lexical bundles had to occur in at least five texts (Esfandiari & Barbary, 2017; Pan et al., 2016), three texts (Chen & Baker, 2010; Qin, 2014), and 10% of texts (Pérez-Llantada, 2014). Lexical bundles in the study of Lu and Deng (2019) had to occur in 20 texts for the English L1 corpus and 60 texts for the English L2 corpus. Despite the small size of the present corpus (about 100,000 words), we chose the moderately conservative cut-off point of five times (50 times per million words) and five different texts.

The length of lexical bundles was one of the issues raised in the studies of lexical bundles. Most studies primarily focused on four-word lexical bundles because many four-word lexical bundles included three-word lexical bundles and those show a higher frequency than five-word lexical bundles. Thus, four-word lexical bundles were the most researched in previous studies (Chen & Baker, 2010; Lu & Deng, 2019; Pan et al., 2016; Pérez-Llantada, 2014). Four-word to six-word lexical bundles and only five-word lexical bundles were further examined in the studies of Esfandiari and Barbary (2017) and Qin (2014). Simpson-Vlach and Ellis (2010) included three-word lexical bundles and those were found to be noteworthy as well as four-word and five-word lexical bundles. Three-word lexical bundles such as *in terms of* are likely to be stored in the language users' long-term memory rather than four-word ones such as *in terms of the*. Lexical bundles ending with *the* not only could make the lexical bundles multifunctional but also those are not associated with "usage events in the mind of a language user" (Appel & Trofimovich, p. 27). In this study, we decided to include three-word lexical bundles and finally established the word-length at three- to nine-word lexical bundles, taking into consideration the maximum length of those found in the present corpora.

The structural classification suggested by Biber, Johansson, Leech, Conrad, and Finegan (1999) has been widely adopted with a slight modification. Other than the two studies (Esfandiari & Barbary, 2017; Pérez-Llantada, 2014), the other studies grouped structural subcategories into three categories (Chen & Baker, 2010; Lu & Deng, 2019; Pan et al., 2016; Qin, 2014): "Noun Phrase-based (NP-based)," "Prepositional Phrase-based (PP-based)," and "Verb Phrase-based (VP-based)." As for the subcategories, 10 structural subcategories were frequently observed in the previous studies: 1) noun phrase with *of* (Esfandiari & Barbary, 2017; Lu & Deng, 2019; Pan et al., 2016; Pérez-Llantada, 2014; Qin, 2014); 2) noun phrase with other postmodifiers (Esfandiari & Barbary, 2017; Lu & Deng, 2019; Pan et al., 2016; Pérez-Llantada, 2014; Qin, 2014); 3) prepositional phrase with *of* (Esfandiari & Barbary, 2017; Lu & Deng, 2019; Pan et al., 2016; Pérez-Llantada, 2014; Qin, 2014); 4) other prepositional phrases (Chen & Baker, 2010; Lu & Deng, 2019; Pan et al., 2016; Pérez-Llantada, 2014; Qin, 2014); 5) copula *be* + noun/adjective phrase (Chen & Baker, 2010; Lu & Deng, 2019; Pan et al., 2016; Qin, 2014); 6) active

© 2021 AsiaTEFL All rights reserved

verb phrase (Chen & Baker, 2010; Esfandiari & Barbary, 2017; Lu & Deng, 2019; Pan et al., 2016; Qin, 2014); 7) passive verb phrase (Chen & Baker, 2010; Esfandiari & Barbary, 2017; Lu & Deng, 2019; Pan et al., 2016; Pérez-Llantada, 2014; Qin, 2014); 8) anticipatory *it* + verb/adjective phrase (Chen & Baker, 2010; Esfandiari & Barbary, 2017; Pan et al., 2016; Pérez-Llantada, 2014; Qin, 2014); 9) *that*-clause (Chen & Baker, 2010; Esfandiari & Barbary, 2017; Lu & Deng, 2019; Pan et al., 2016; Qin, 2014), and; 10) *to*-clause (Chen & Baker, 2010; Esfandiari & Barbary, 2017; Lu & Deng, 2019; Qin, 2014). Reflecting the types of lexical bundles, other noun phrases (Lu & Deng, 2019; Qin, 2014), adverbial clause (Esfandiari & Barbary, 2017; Qin, 2014), conjunctions (Lu and Deng, 2019), and comparative expressions (Esfandiari & Barbary, 2017) were additionally provided in the previous studies.

The functional classification – referential expressions, discourse organizers, and stance expressions – proposed by Biber, Conrad, and Cortes (2004), was adopted by Chen and Baker (2010), Pérez-Llantada (2014), and Qin (2014). Pérez-Llantada (2014) and Qin (2014) replaced discourse organizers with text-organizers. Esfandiari and Barbary (2017), Lu and Deng (2019), and Pan et al. (2016) referred to the functional classification – research-oriented, text-oriented, and participant-oriented – suggested by Hyland (2008). Lu and Deng (2019) and Pan et al. (2016) adopted stance-oriented instead of participant-oriented. Regarding the functional subcategories, 14 functional subcategories were frequently noted in the previous studies: 1) location (Esfandiari & Barbary, 2017; Lu & Deng, 2019; Pan et al., 2016) or time/place (Chen & Baker, 2010; Pérez-Llantada, 2014; Qin, 2014); 2) procedure (Esfandiari & Barbary, 2017; Lu & Deng, 2019; Pan et al., 2016); 3) quantification (Chen & Baker, 2010; Esfandiari & Barbary, 2017; Lu & Deng, 2019; Pan et al., 2016; Pérez-Llantada, 2014; Qin, 2014); 4) description (Esfandiari & Barbary, 2017; Lu & Deng, 2019; Pan et al., 2016; Pérez-Llantada, 2014; Qin, 2014); 5) transition (Esfandiari & Barbary, 2017; Lu & Deng, 2019; Pan et al., 2016); 6) resultative (Esfandiari & Barbary, 2017; Lu & Deng, 2019; Pan et al., 2016); 7) structuring (Esfandiari & Barbary, 2017; Lu & Deng, 2019; Pan et al., 2016); 8) framing (Chen & Baker, 2010; Esfandiari & Barbary, 2017; Lu & Deng, 2019; Pan et al., 2016; Pérez-Llantada, 2014; Qin, 2014); 9) stance (Esfandiari & Barbary, 2017; Lu & Deng, 2019; Pan et al., 2016; Pérez-Llantada, 2014; Qin, 2014); 10) engagement (Esfandiari & Barbary, 2017; Pan et al., 2016); 11) epistemic (Chen & Baker, 2010; Pérez-Llantada, 2014; Qin, 2014); 12) identification/focus (Chen & Baker, 2010; Pérez-Llantada, 2014; Qin, 2014); 13) inference (Chen & Baker, 2010; Pérez-Llantada, 2014; Qin, 2014); 14) comparison/contrast (Lu & Deng, 2019; Pérez-Llantada, 2014; Qin, 2014). Taking into consideration the distinctive functions of lexical bundles, the researchers added topic (Chen & Baker, 2010; Esfandiari & Barbary, 2017), statistical markers (Esfandiari & Barbary, 2017), brevity/exemplification/explanation (Qin, 2014), and grouping/generalization/citation/objective (Lu & Deng, 2019) to the existing subcategories.

Previous studies have shown a significant difference between English L1 writers and L2 writers in terms of structural and functional characteristics of lexical bundles, although they were at the same level as expert writers. Inconsistent results could likely be found depending on different L1 backgrounds, the kinds of disciplines, and part-genres of research articles. There has been little research to investigate structural and functional features of lexical bundles used by Korean authors in comparison with native speakers of English authors in the field of medicine and even in research article abstracts. We investigated research article abstracts written by English L1 and L2 writers (Korean) to analyze the structural and functional characteristics of lexical bundles in the field of medicine. To avoid bias, a single academic discipline (medical fields), a single register (research article abstracts), stages of proficiency (expertise), the reputation of journals (SCI-indexed), publication dates (2008-2019), article types (clinical trials or randomized controlled trials), format (IMRD: Introduction-Method-Result- Discussion), and the number of abstracts (431 texts) were controlled in this study.

© 2021 AsiaTEFL All rights reserved

The Journal of Asia TEFL
Vol. 18, No. 1, Spring 2021, 142-160

Method

Corpora

Two corpora were constructed, consisting of medical journal abstracts by native speakers of English (NSE) writers and Korean non-native speakers of English (NNSE) writers. To determine the L1 status, NSE writers were defined as the first authors whose first and last names were considered native to English-speaking countries and those who were affiliated with an institution in the USA or the UK. The same approach was applied to the operational definition of the NNSE writers. NNSE writers were defined as those who were affiliated with an institution in Korea and whose first and last names were considered native to Korea. Table 1 shows the constituents of the NSE and NNSE corpora.

TABLE 1
Constituents of the Corpora

	NSE corpus	NNSE corpus
Number of journals	6 journals	8 journals
Article types	Clinical Trials or RCTs on humans	Clinical Trials or RCTs on humans
Publication dates	2008–2019	2008–2019
Number of abstracts	431 abstracts	431 abstracts
Mean length of abstracts	317 words	218 words
Total number of words	136,710 words	94,045 words

Abstracts were excerpted from medical research articles reporting clinical trials or randomized controlled trials (RCTs) on humans. Most clinical trials and RCTs were conducted by doctors, pharmacists, nurses, and medical professionals in clinical settings. Each corpus was composed of 431 medical journal abstracts published between 2008 and 2019. Journals were all indexed in the Science Citation Index (SCI). Abstracts in the journals followed the Introduction-Method-Results-Discussion (IMRD) format. Six journals for the NSE corpus were *The New England Journal of Medicine, The Journal of the American Medical Association, The Lancet, The British Medical Journal, Annals of Internal Medicine,* and *The American Journal of Medicine*. Eight journals for the NNSE corpus were the *Yonsei Medical Journal, Journal of Korean Medical Science, Korean Journal of Internal Medicine*, *Korean Journal of Anesthesiology, Cancer Research and Treatment, Gut liver, Clinical and Molecular Hepatology,* and *Korean Journal of Gastroenterology*.

Structural and Functional Classification

The study grouped structural subcategories under four broad categories as proposed by Pan et al. (2016) and Qin (2014): "Noun Phrase-based (NP-based)," "Prepositional Phrase-based (PP-based)," "Verb Phrase-based (VP-based)," and "Others." Due to the small number of lexical bundle types, the copula *be* + noun phrase/adjective phrase, pronoun phrase/noun phrase + *be*, and anticipatory *it* + verb phrase fragments were collapsed into *be* + (adjective/noun phrase), and adverbial clausal fragments were excluded from the study. The structural subcategories were as follows: noun phrases with *of* (NP + *of*), noun phrases with other post-modifiers (NP + OPM), other noun phrases (other NP), prepositional phrases with *of* (PP + *of*), other prepositional phrases (other PP), (verb phrase +) with active verb (active verb), (verb phrase +) with passive verb (passive verb), *be* + (adjective/noun phrase) (be + (Adj/NP)), *that*-clause, *to*-clause, and other expressions.

As for functional classification, the study categorized lexical bundles into three broad categories proposed by Hyland (2008) and Pan et al. (2016): "Research-oriented," "Text-oriented," and "Stance-oriented." Lexical bundles related to physical environments and times were categorized into place/time. The text references mentioned in Biber, Conrad, and Cortes (2004) were placed together with structuring in Hyland (2008) because of the high frequency of *the purpose/aim of* followed by *this study was.* Except

© 2021 AsiaTEFL All rights reserved

for linking phrases, tangible and intangible framings as stated by Biber et al. (2004) were replaced by framing, and this was placed together with a description and topic because of the overlapped functions of lexical bundles. The framing proposed by Hyland (2008) was related to linking phrases; thus, the framing was replaced by text framing. As proposed by Esfandiari and Barbary (2017), statistics were additionally provided in this study. Resultative in Hyland (2008) and contrast/comparison in Cortes (2004) were replaced by cause/effect and compare/contrast. The stance features in Hyland (2008) were replaced by epistemic features drawn from Biber et al. (2004). The functional subcategories were as follows: place/time, procedure, quantification, description/topic/framing, statistics, text reference/structuring, compare/contrast, cause/effect, text framing, transition, epistemic, and engagement.

Procedure

Three- to nine-word lexical bundles were extracted using the software *AntConc 3.5.7* (Anthony, 2018). The raw frequency of five times and five different texts was employed in the study. The numbers of lexical bundle types between the NSE and the NNSE corpora were as follows: 148 vs. 366 in introduction; 738 vs. 471 in methods; 925 vs. 620 in results; and 101 vs. 47 in discussion.

To minimize inflated results, the lexical bundles were filtered based on exclusion criteria. Subject-specific technical words (e.g., *coronary artery disease*) and shorter lexical bundles subsumed in longer ones showing the frequency below five after the exclusion of the same frequency were excluded, as well as lexical bundles starting with *of* (e.g., *of the trial*), ending with determiners (e.g., *patients in the*), starting and ending with *and/or* (e.g., *the safety and, or placebo for*), and including numbers (e.g., *aged # to # years*). After the exclusion, the numbers of lexical bundle types between the NSE and the NNSE corpora were as follows: 44 vs. 70 in introduction; 200 vs. 128 in methods; 220 vs. 154 in results; and 34 vs. 30 in discussion.

The categorization of lexical bundles was done by two coders. Each coder independently classified lexical bundles structurally and functionally according to each of the corpora and subsections of medical journal abstracts. Intercoder-reliability was interpreted as an "almost perfect agreement" based on average Kappa values above 0.81 (Landis & Koch, 1977). Discrepancies were discussed by focusing on the primary functions of lexical bundles to reach 100% agreement between the coders. For example, *in the management of* could be categorized as a description/topic/framing or procedure bundle. Coders focused on its primary function, classifying those as procedure bundles. Types and tokens of lexical bundles between the corpora were examined using Log-Likelihood (LL) values.

Results and Discussion

Structures of Lexical Bundles in Subsections of Abstracts

The results of the structural analysis of lexical bundles in the introduction of medical journal abstracts are presented in Table 2. Tokens of VP-based bundles were significantly higher in the NNSE corpus than the NSE corpus in the introduction of medical journal abstracts.

© 2021 AsiaTEFL All rights reserved

TABLE 2
Structural Analysis of Lexical Bundles in Introduction

Structural categories	Structural subcategories	Types (per 100,000 words)		Tokens (per 100,000 words)	
		NSE	NNSE	NSE	NNSE
NP-based		16(83)	21(116)	222(1152)	281(1552)
	NP + *of* (*the effect of*)	12(62)	13(72)	195(1012)	234(1292)
	NP + OPM (*patients with chronic*)	4(21)	2(11)	27(140)	12(66)
	Other NP (*this prospective randomized*)	0(0)	6(33)	0(0)	35(193)*
PP-based		6(31)	11(61)	116(602)	153(845)
	PP + *of* (*for the treatment of*)	1(5)	3(17)	13(68)	28(155)
	Other PP (*in patients with*)	5(26)	8(44)	103(535)	125(690)
VP-based		22(114)	38(210)	211(1095)	495(2734)*
	Active Verb (*we investigated whether*)	1(5)	5(28)	5(26)	30(166)*
	Passive Verb (*is associated with*)	3(16)	3(17)	32(166)	28(155)
	be + (Adj/NP) (*this study was*)	0(0)	2(11)	0(0)	108(597)*
	that-clause (*we hypothesized that*)	1(5)	0(0)	5(26)	0(0)
	to-clause (*this study was to*)	17(88)	28(155)	169(877)	329(1817)*
Others		0(0)	0(0)	0(0)	0(0)
	Other expressions (*None*)	0(0)	0(0)	0(0)	0(0)

*$p < 0.0001$

The result is consistent with findings that English L2 writers frequently used more clausal bundles than did English L1 writers (Chen & Baker, 2010; Lu & Deng, 2019; Pan et al., 2016). Both groups of writers used NP + *of* to provide research context (e.g., *the use of*), causative relations (e.g., *the effect of*), and research properties (e.g., *the efficacy of*) as well as other PP (*in patients with, for patients with*) and passive verbs (*is associated with*) to review what has or has yet to be investigated. NNSE writers overused the *to*-clause, *be* + (Adj/NP), active verb, and other NP compared to NSE writers.

1) NNSE Ex.1 **The aim of this study was to**(45) evaluate the efficacy of …
NSE Ex.1 **To determine whether**(40) … is more effective than …

A complete sentence with the *to*-clause with signals such as *the aim of* was frequently observed in the NNSE corpus, whereas a phrase starting with the *to*-clause (*to determine whether*) was prominently found in the NSE corpus. The numbers in parenthesis next to the underlined lexical bundles were the frequency of the lexical bundles. Compared to NSE writers, NNSE writers are more likely to repeatedly use the frame *the * of this study was to* when specifying research aims. Although the frame was prevalent in the field of medicine in the 1990s, it was rarely found in the NSE corpus. NNSE writers seem to show a lack of awareness of language use preferred by community members in the present day. The verb *determine* was reported by researchers who investigated the field of medicine (Cavalieri, 2014; Gledhill, 2000; Saber, 2012), and it was frequently found in the NSE corpus as well. However, *investigate* and *examine* were more frequent than *determine* in the NNSE corpus. The finding suggests that English L2 writers are more familiar with verbs prevalent in soft disciplines.

Due to the frequent use of *this study* and *we investigated*, lexical bundles with active verbs were more frequently observed in the NNSE corpus when presenting research aims. Personal pronouns such as *we* followed by active verbs were frequently observed in the NNSE corpus. The finding is inconsistent with that of Lu and Deng (2019), who found the frequent use of *we* followed by active verbs by English L1 writers. The finding suggests that English L2 writers in the field of medicine use the personal pronoun *we* as frequently as do English L1 writers. While NNSE writers used active verbs to indicate research aims, NSE writers sought to establish a research context before specifying research objectives. NSE writers appear to provide a research background more than do NNSE writers before specifying research aims.

© 2021 AsiaTEFL All rights reserved

2) NNSE Ex.1 **This study evaluated**(6) the efficacy and safety of …
Ex.2 **We investigated whether**(8) the addition of … can improve …
NSE Ex.1 … therapy with … **reduces the risk**(5) … but increases the risk of …

Another difference was the use of other NP and *be* + (Adj/NP) by NNSE writers to specify the research design with signals of the research objectives. The combined statements might have been caused by some characteristics of a journal (*Journal of Korean Medical Science*) published between 2008 and 2017. The specification of the research design was exclusively found in the methods of journal abstracts in the NSE corpus.

3) NNSE Ex.1 The aim of **this prospective randomized**(8), double-blind study was to compare the effects of …
Ex.2 **This study was**(102) a multicenter, randomized, open-label study to evaluate the efficacy of …

Tokens of NP-based and VP-based bundles were significantly higher in the NSE and the NNSE corpus, respectively, in terms of the methods of medical journal abstracts (see Table 3).

TABLE 3
Structural Analysis of Lexical Bundles in Methods

Structural categories	Structural subcategories	Types (per 100,000 words)		Tokens (per 100,000 words)	
		NSE	NNSE	NSE	NNSE
NP-based		96(196)	41(137)	1171(2392)	430(1439)*
	NP + *of* (*a total of*)	21(43)	12(40)	249(509)	150(502)
	NP + OPM (*patients who had*)	20(41)	9(30)	158(323)	81(271)
	Other NP (*the primary outcome*)	55(112)	20(67)	764(1561)	199(666)*
PP-based		26(53)	17(57)	267(545)	188(629)
	PP + *of* (*to one of*)	7(14)	4(13)	64(131)	36(120)
	Other PP (*in the study*)	19(39)	13(44)	203(415)	152(509)
VP-based		70(142)	67(224)	1015(2073)	851(2847)*
	Active Verb (*all patients received*)	6(12)	2(7)	93(190)	13(43)*
	Passive Verb (*were randomly assigned*)	36(73)	46(154)	468(956)	607(2031)*
	be + (Adj/NP) (*the primary outcome was*)	21(43)	8(27)	368(752)	100(335)*
	that-clause (*None*)	0(0)	0(0)	0(0)	0(0)
	to-clause (*to receive either*)	7(14)	11(37)	86(175)	131(438)*
Others		8(16)	3(10)	112(228)	28(94)*
	Other expressions (*twice daily for*)	8(16)	3(10)	112(228)	28(94)*

*$p < 0.0001$

This finding is consistent with findings that English L1 writers more frequently used phrasal bundles than did English L2 writers (Chen & Baker, 2010; Pan et al., 2016). Both groups of writers used NP + *of* to describe patients (e.g., *a total of*) and research properties (e.g., *the proportion of patients*). NP + OPM, PP + *of*, and other PP were used to describe patients in group assignments (e.g., *patients who had/were, to one of, in patients with*). NNSE writers underused other NP, active verbs, *be* + (Adj/NP), and other expressions, but overused passive verbs and *to*-clauses compared to NSE writers. The finding lends support to findings that English L2 writers more frequently used passive verb and *to*-clause bundles than English L1 writers (Lu & Deng, 2019; Pan et al., 2016). NNSE writers used more passive verbs and exclusively used *were divided into* and *were randomized into*. The exclusive use of *divided into* by English L2 writers was reported by Pan et al. (2016) and Lu and Deng (2019). It is possible that Korean and Chinese writers are familiar with *divided into* when explaining group assignments.

4) NNSE Ex.1 Patients **were divided into**(22) two groups.
Ex.2 Fifty-six patients **were randomized into**(20) two groups.

© 2021 AsiaTEFL All rights reserved

Although passive verb bundles (e.g., *were randomly assigned to, were included in*) were observed in both corpora when specifying procedures of group assignments, active verb bundles (e.g., *we randomly assigned, included all patients who*) were exclusively found in the NSE corpus. Compared to NNSE writers, NSE writers are more likely to use active verbs in the methods of medical journal abstracts. It is possible that NNSE writers are more familiar with conventions of soft disciplines where passive verbs are prevalent in the methods of research articles.

5) NSE Ex.1 **We randomly assigned**(22) 251 patients with …
Ex.2 Analysis was done by intention-to-treat analysis, which **included all patients who**(5) underwent …

Since NSE writers frequently used signals of pre-planned results (*the primary outcome/end point/endpoint was*), *be* + (Adj/NP) was more frequently observed in the NSE corpus than the NNSE corpus. Although *endpoint* was prevalent in the 1990s in the field of medicine, *end point* was observed more frequently in the NSE corpus along with *outcome*. It is possible that the NNSE writers are unaware of terminology preferred by community members in the present day.

6) NNSE Ex.1 **The primary endpoint was**(25) the incidence of …
NSE Ex.1 **The primary outcome was**(93) 6-week mortality.
Ex. 2 **The primary end point was**(43) change in …

NSE writers exclusively used other NP to indicate analysis strategies (*intention to treat*), methods of randomization (*a computer generated*), durations of follow-up (*month follow up*), and levels of care (*primary care clinics*). The finding suggests that NSE writers are more likely to report details of methods and seem to be more explicitly aware of semi-fixed bundles commonly used by community members compared to NNSE writers.

7) NSE Ex.1 Analysis was by **intention to treat**(59).
Ex.2 … were randomly assigned via **a computer generated**(15) sequence….
Ex.3 Randomized, double-blind, placebo-controlled clinical trial with 12-**month follow up**(10).
Ex.4 … randomized trial with 2-year follow up in …**primary care clinics**(9)….

Tokens of NP-based and VP-based bundles were significantly higher in the NSE and NNSE corpora, respectively, in the results of medical journal abstracts (See Table 4).

TABLE 4
Structural Analysis of Lexical Bundles in Results

Structural categories	Structural subcategories	Types (per 100,000 words)		Tokens (per 100,000 words)	
		NSE	NNSE	NSE	NNSE
NP-based		97(191)	44(132)	1439(2833)	686(2058)*
	NP + *of* (*the rate of*)	23(45)	14(42)	286(563)	179(537)
	NP + OPM (*significant difference in*)	27(53)	10(30)	267(526)	106(318)*
	Other NP (*the control group*)	47(93)	20(60)	886(1744)	401(1203)*
PP-based		28(55)	21(63)	450(886)	345(1035)
	PP + *of* (*in terms of*)	2(4)	3(9)	12(24)	26(78)
	Other PP (*between the two groups*)	26(51)	18(54)	438(862)	319(957)
VP-based		77(152)	79(237)	837(1648)	851(2553)*
	Active Verb (*did not differ*)	14(28)	10(30)	210(413)	102(306)
	Passive Verb (*was associated with*)	27(53)	17(51)	308(606)	129(387)*
	be + (Adj/NP) (*there were no*)	33(65)	48(144)	298(587)	668(2004)*
	that-clause (*None*)	0(0)	0(0)	0(0)	0(0)
	to-clause (*were assigned to receive*)	3(6)	0(0)	21(41)	0(0)*
Others		18(35)	14(42)	189(372)	161(483)
	Other expressions (*group than in*)	18(35)	14(42)	189(372)	161(483)

*$p < 0.0001$

© 2021 AsiaTEFL All rights reserved

The result is consistent with findings that English L2 writers more rely on clausal bundles than phrasal bundles (Chen & Baker, 2010; Lu & Deng, 2019; Pan et al., 2016). Both groups of writers used NP + *of* to indicate the number of patients (e.g., *a total of*) and research properties (e.g., *the rate of*) as well as other PP bundles to indicate patients in groups (e.g., *in the placebo group*), active verbs (e.g., *did not differ*), and other expressions (e.g., *group than in*) to indicate group differences and comparisons.

NSE writers used more NP + OPM and other NP to indicate patients and groups (e.g., *patients in the placebo group*), unexpected results (e.g., *adverse events in*), and statistical markers (e.g., *adjusted odds ratio*) than did NNSE writers. The finding suggests that NSE writers are more likely to report findings in greater detail than NNSE writers. A brief description of the number of patients in random assignments was more frequently observed in the NSE corpus using passive verb and *to*-clause bundles (e.g., *were randomly assigned to*) in the results of medical journal abstracts. When reporting unexpected results, passive verb (e.g., *adverse events were reported*) and active verb bundles (e.g., *adverse events occurred*) were used by NSE writers, whereas *be* + (Adj/NP) bundles (e.g., *adverse events were*) were employed by NNSE writers.

8) NNSE Ex.1 … **adverse events were**(16) diarrhea, rash, stomatitis, pruritus, and anorexia.
NSE Ex.1 Serious **adverse events were reported**(15) in 124 patients (20.2%) in the … group and 130 (21%) in the usual care group.
Ex.2 **Adverse events occurred**(18) in 45 patients (24.7%) in the … group compared with 47 (25.8%) in the … group.

Although NSE writers seemed to use more active verbs than did NNSE writers in the methods and results of medical journal abstracts, passive verbs were observed as frequently in the NSE corpus as in the NNSE corpus. There might be a specific context to determine which voice is more efficient in delivering information; thus, further research is needed in terms of the use of the active and passive voice depending on contexts. The overuse of *be* + (Adj/NP) by NNSE writers was observed. It might be due to the extensive use of *there*-patterns (e.g., *there was/were no*) when reporting results with significance.

9) NNSE Ex.1 **There were no significant differences**(41) between the two groups …
Ex.2 … showed **significant increases in**(5) …
Ex.3 The use of …. was **significantly increased in**(5) … group …
Ex.4 … **was significantly reduced**(6) in Group A …
NSE Ex.1 **There was no significant difference**(12) between groups …
Ex.2 **An increase in**(5) … was seen in the intervention group …
Ex.3 … was associated with **a reduction in**(9) …

The finding lends support to Esfandiary and Barbary's (2017) study. The finding suggests that Korean and Persian writers tend to use *there*-patterns in order to indicate whether there was a significant difference between groups. Exploring the use of NP + OPM, passive verb, and *be* + (Adj/NP) bundles, it was found that NNSE writers more frequently used plural forms (e.g., *there were*, *differences, increases*) and verb phrases (e.g., *significantly increased in, was significantly reduced*), whereas NSE writers more prominently used singular forms and noun phrases (e.g., *there was*, *an increase in, a reduction in*). As asserted by Shin, Cortes, and Yoo (2018) and Lu and Deng (2019), Korean and Chinese writers seem to have difficulty using indefinite articles and they appear to be in the developmental stage from the use of clausal bundles to phrasal bundles (Esfandiari & Barbary, 2017; Pan et al., 2016).

Table 5 presents the results of the structural analysis of the discussion section. As Table 5 shows, PP-based tokens were significantly higher in the NNSE corpus. VP-based bundles ranked the highest in the NSE corpus, whereas PP-based ones ranked the highest in the NNSE corpus. The finding of the more frequent use of VP-based bundles than NP-based ones is inconsistent with findings of previous studies (Chen & Baker, 2010; Lu & Deng, 2019; Pan et al., 2016). The more frequent use of those bundles by

© 2021 AsiaTEFL All rights reserved

NSE writers appears to reflect some characteristics of clinical trials and RCTs in medical journal abstracts when discussing the findings of the study.

TABLE 5
Structural Analysis of Lexical Bundles in Discussion

Structural categories	Structural subcategories	Types (per 100,000 words)		Tokens (per 100,000 words)	
		NSE	NNSE	NSE	NNSE
NP-based		9(51)	6(47)	102(577)	46(361)
	NP + *of* (*the use of*)	5(28)	5(39)	78(441)	36(283)
	NP + OPM (*patients with advanced*)	1(6)	1(8)	8(45)	10(79)
	Other NP (*the long term*)	3(17)	0(0)	16(90)	0(0)*
PP-based		6(34)	8(63)	107(605)	103(809)*
	PP + *of* (*in terms of*)	0(0)	2(16)	0(0)	12(94)*
	Other PP (*in patients with*)	6(34)	6(47)	107(605)	91(715)
VP-based		16(90)	16(126)	150(848)	94(739)
	Active Verb (*did not improve*)	5(28)	2(16)	47(266)	10(79)*
	Passive Verb (*was associated with*)	5(28)	1(8)	62(350)	6(47)*
	be + (Adj/NP) (*there was no*)	3(17)	8(63)	19(107)	49(385)*
	that-clause (*findings suggest that*)	1(6)	4(31)	6(34)	24(189)*
	to-clause (*studies are needed to*)	2(11)	1(8)	16(90)	5(39)
Others		3(17)	0(0)	40(226)	0(0)*
	Other expressions (*as compared with*)	3(17)	0(0)	40(226)	0(0)*

*$p < 0.0001$

NSE writers used a wider variety of active verb bundle types to provide interpretations of reported results. They seem to report negative results as frequently as positive results, whereas NNSE writers appear to report positive results more than negative results. The reporting of negative results suggests a reflection on some characteristics of clinical trials and RCTs in the international medical journal abstracts.

10) NNSE Ex.1 Our study showed that … effectively **reduced the incidence**(5) …
NSE Ex.1 First-line use of … **did not improve**(13) overall survival in patients with...
Ex.2 Among patients …, treatment compared with placebo **did not reduce**(12) the risk of …

NSE writers used *be* + (Adj/NP) to indicate limitations, whereas NNSE writers did so to mitigate their interpretation of reported results. Apart from *be* + (Adj/NP), NSE writers used other NP bundles (*the long term*) to indicate the need for further research. Compared to NNSE writers, NSE writers seem to report more frequently the constraints of the study.

11) NNSE Ex.1 … may be **safe and effective**(7) …
Ex.2 … **appears to be**(5) safe and well-tolerated, and it provides clinical benefits.
NSE Ex.1 **The study was**(6) small, open-label, …
Ex.2 Further research is needed to determine **the long-term**(6) effect of …

Regarding the use of *that*-clause bundles, NNSE writers frequently used * *suggest that* with fillers (*findings, results, date*) and anticipatory *it*-clause patterns (*it is suggested that*). The finding is consistent with those of Lu and Deng (2019) and Pan et al. (2016).

12) NNSE Ex.1 Our **findings suggest that**(7) … is an effective and tolerable option …
Ex.2 The **results suggest that**(7) … may not mainly contribute to a reduction…
Ex.3 Our **data suggest that**(5)… might be effective in terms of …
Ex.4 **It is suggested that**(5) … treatment could decrease …

© 2021 AsiaTEFL All rights reserved

It was reported that both anticipatory *it*-clause and *that*-clause patterns were more frequently used by English L2 writers (Chinese) in the study of Lu and Deng (2019), and *that*-clause patterns were more prominently observed in soft disciplines than hard disciplines (Hyland & Tse, 2005; Omidian et al., 2018). The finding suggests that Korean and Chinese writers are more familiar with conventions of soft disciplines than those of hard disciplines.

Functions of Lexical Bundles in Subsections of Abstracts

The results of the functional analysis of lexical bundles in the introduction of medical journal abstracts are presented in Table 6. As shown in Table 6, tokens of text-oriented bundles were significantly higher in the NNSE corpus than the NSE corpus in the introduction of medical journal abstracts.

TABLE 6
Functional Analysis of Lexical Bundles in Introduction

Functional categories	Functional subcategories	Types (per 100,000 words)		Tokens (per 100,000 words)	
		NSE	NNSE	NSE	NNSE
Research-oriented		23(119)	30(166)	330(1712)	365(2016)
	Place/time (*in the United States*)	1(5)	1(6)	6(31)	5(28)
	Procedure (*we aimed to*)	9(47)	10(55)	107(555)	87(480)
	Quantification (*None*)	0(0)	0(0)	0(0)	0(0)
	Description/topic/framing (*has not been*)	13(67)	19(105)	217(1126)	273(1508)
	Statistics (*None*)	0(0)	0(0)	0(0)	0(0)
Text-oriented		20(104)	40(221)	212(1100)	564(3115)*
	Text reference/structuring (*this study was to*)	1(5)	24(133)*	5(26)	382(2110)*
	Compare/contrast (*the relationship between*)	0(0)	1(6)	0(0)	5(28)
	Cause/effect (*is associated with*)	19(98)	15(83)	207(1074)	177(978)
	Text framing (*None*)	0(0)	0(0)	0(0)	0(0)
	Transition (*None*)	0(0)	0(0)	0(0)	0(0)
Stance-oriented		1(5)	0(0)	7(36)	0(0)
	Epistemic (*None*)	0(0)	0(0)	0(0)	0(0)
	Engagement (*the need for*)	1(5)	0(0)	7(36)	0(0)

*$p < 0.0001$

The result is consistent with the findings of Lu and Deng (2019), Pan et al. (2016), and Esfandiari and Barbary (2017). Both groups of writers employed procedure bundles (e.g., *we aimed to*) to indicate research objectives as well as cause/effect bundles to indicate causative relations (e.g., *is associated with*). Other shared lexical bundles were associated with descriptions of the research context (e.g., *has not been*) and research properties (e.g., *the efficacy of*). The subcategory that showed a significant difference between the corpora in the introduction of medical journal abstracts was the text reference/structuring.

13) NNSE Ex.1 **The aim of this study was to**(45) evaluate the effects of …
Ex.2 **The purpose of this study was to**(20) compare the efficacy of …
Ex.3 **This study aimed to**(10) evaluate the efficacy, safety, …
Ex.4 **This study was performed to**(8) evaluate the effectiveness of ...

NNSE writers exclusively used *this study* with structuring bundles such as *the aim of* and *the purpose of* when specifying research objectives. The frequent use of *this study* by English L2 writers was observed by Esfandiari and Barbary (2017); accordingly, its overuse by Korean and Persian writers might be due to L1 transfer.

While NNSE writers frequently used the phrases *the aim/purpose of* and *this study* followed by *aimed to* and *was performed to*, NSE writers prominently employed the phrases *we aimed to* and *to determine*

© 2021 AsiaTEFL All rights reserved

whether. NNSE writers appear to be inexplicitly aware of discipline-specific bundles preferred by community members in the field of medicine. NSE writers' frequent use of phrases that start with *to determine whether* could reflect some characteristics of medical journal abstracts for writers to concisely deliver research aims.

Table 7 presents the functional analysis of the lexical bundles in the methods of medical journal abstracts. As shown in Table 7, tokens of text-oriented bundles were significantly higher in the NNSE corpus. The finding is consistent with previous findings (Esfandiari & Barbary, 2017; Lu & Deng, 2019; Pan et al., 2016). Research-oriented ranked the highest in both corpora. The finding is inconsistent with those of previous studies (Esfandiari & Barbary, 2017; Lu & Deng, 2019; Pan et al., 2016). The highest frequency of research-oriented bundles in the methods of journal abstracts might be due to some characteristics of medical disciplines.

TABLE 7
Functional Analysis of Lexical Bundles in Methods

Functional categories	Functional subcategories	Types (per 100,000 words)		Tokens (per 100,000 words)	
		NSE	NNSE	NSE	NNSE
Research-oriented		193(394)	113(378)	2516(5140)	1364(4564)
	Place/time (*the end of*)	38(78)	15(50)	295(603)	132(442)
	Procedure (*were randomly assigned*)	81(165)	54(181)	1355(2768)	777(2600)
	Quantification (*years of age*)	23(47)	14(47)	291(594)	172(576)
	Description/topic/framing (*randomized controlled trial*)	51(104)	30(100)	575(1175)	283(947)
	Statistics (*None*)	0(0)	0(0)	0(0)	0(0)
Text-oriented		7(14)	15(50)	49(100)	133(445)*
	Text reference/structuring (*this study was*)	3(6)	11(37)	24(49)	109(365)*
	Compare/contrast (*were compared between*)	1(2)	2(7)	5(10)	13(44)
	Cause/effect (*the effect of*)	2(4)	0(0)	15(31)	0(0)
	Text framing (*on the basis of*)	1(2)	0(0)	5(10)	0(0)
	Transition (*as well as*)	0(0)	2(7)	0(0)	11(37)*
Stance-oriented		0(0)	0(0)	0(0)	0(0)
	Epistemic (*None*)	0(0)	0(0)	0(0)	0(0)
	Engagement (*None*)	0(0)	0(0)	0(0)	0(0)

*$p < 0.0001$

Both groups of writers used place/time bundles to describe interventions (e.g., *twice daily for*) and procedure bundles to signal pre-planned results (e.g., *the primary outcome was*), random assignments (e.g., *were randomly assigned to*), and interventions (e.g., *were randomly assigned to receive*). Similarly, they used quantification bundles to describe patients (e.g., *a total of*, *years of age*) and description/topic/framing bundles to indicate research designs (e.g., *randomized controlled trial*), matched groups (e.g., *the control group*), and types of patients (e.g., *in patients with*).

Compared to NSE writers, text reference/structuring and transition bundles were overused by NNSE writers. The result is consistent with findings that English L2 writers used more text-reference/structuring (Esfandiari & Barbary, 2017; Pan et al., 2016) and transition bundles (Lu & Deng, 2019; Pan et al., 2016) than English L1 writers. Again, it is possible that Korean and Chinese writers are familiar with conventions prevalent in soft disciplines.

To specify research designs, NNSE and NSE writers started with a sentence using *this study was* and *the study was*, respectively. Consistent with the findings of Esfandiari and Barbary (2017), English L2 writers are more likely to use *this study* rather than *the study* due to L1 transfer. The definite article is neither found in the Korean nor the Persian language, hence why both groups of writers seem to be more familiar with *this study* than *the study*.

© 2021 AsiaTEFL All rights reserved

14) NNSE Ex.1 **This study was**(18) a randomized controlled study…
NSE Ex.1 **The study was**(11) open-label with no masking of participants, clinicians, or research staff.

Table 8 shows the results of functional analysis of lexical bundles in the results of journal abstracts. As shown in Table 8, tokens of research-oriented and stance-oriented bundles were significantly higher in the NSE corpus in the results of medical journal abstracts.

TABLE 8
Functional Analysis of Lexical Bundles in Results

Functional categories	Functional subcategories	Types (per 100,000 words)		Tokens (per 100,000 words)	
		NSE	NNSE	NSE	NNSE
Research-oriented		132(259)	70(210)	1931(3801)	966(2898)*
	Place/time (*follow up period*)	10(20)	9(27)	85(167)	64(192)
	Procedure (*completed the study*)	32(63)	11(33)	355(699)	92(276)*
	Quantification (*there was no significant*)	23(45)	22(66)	311(612)	419(1257)*
	Description/topic/framing (*in the placebo group*)	32(63)	18(54)	818(1610)	302(906)*
	Statistics (*hazard ratio hr*)	35(69)	10(30)	362(713)	89(267)*
Text-oriented		85(167)	84(252)	967(1904)	1125(3375)*
	Text reference/structuring (*from the study*)	2(4)	1(3)	10(20)	6(18)
	Compare/contrast (*group than in*)	50(98)	69(207)*	565(1112)	993(2979)*
	Cause/effect (*serious adverse events*)	32(63)	11(33)	386(760)	95(285)*
	Text framing (*with respect to*)	1(2)	2(6)	6(12)	24(72)*
	Transition (*as well as*)	0(0)	1(3)	0(0)	7(21)
Stance-oriented		3(6)	0(0)	17(33)	0(0)*
	Epistemic (*more likely to*)	3(6)	0(0)	17(33)	0(0)*
	Engagement (*None*)	0(0)	0(0)	0(0)	0(0)

*$p < 0.0001$

On the other hand, tokens of text-oriented bundles were significantly higher in the NNSE corpus. Research-oriented and text-oriented ranked the highest in both the NSE and NNSE corpora. Omidian et al. (2018) found the frequent use of research-oriented bundles in hard disciplines and text-oriented bundles in soft disciplines. The finding shows that NSE writers are more likely to be aware of discipline-specific bundles and conventions prevalent in hard disciplines compared to NNSE writers.

15) NNSE Ex.1 Sixty patients were enrolled and randomized to the … group (28 of 29 patients **completed the study**(10)) or the control group (26 of 31 patients completed the study).
NSE Ex.1 Among the 509 patients randomized (mean age, 56.6 [SD, 16.8] years; 226 [44.4%] women), 480 (94.3%) **completed the trial**(18).

NSE writers used more of the procedure bundles to specify the number of patients randomized and analyzed (e.g., *completed the study*). The finding suggests that NSE writers tend to briefly mention the number of patients originally randomized and analyzed in the results of medical journal abstracts. While *study* was exclusively used by NNSE writers, *trial* was found as frequently as *study* in the NSE corpus.

Quantification and compare/contrast bundles to indicate group comparisons and group differences were more frequently observed in the NNSE corpus. While *between the two groups* and *compared to* were prevalent in the NNSE corpus, *between groups* and *compared with* were prominent in the NSE corpus. When compared to NSE writers, NNSE writers are more likely to specify the number of groups with a

© 2021 AsiaTEFL All rights reserved

definite article (e.g., *the two groups*). The more frequent use of *compared with* in the NSE corpus than the NNSE corpus could be community members' preference in the international medical journal abstracts.

16) NNSE Ex.1 However, there were no significant differences **between the two groups(**78).
Ex.2 … increased significantly, **compared to the control group(**6)…
NSE Ex.1 ... **did not differ between groups(**13) …
Ex.2 … were significantly elevated in patients who had … **as compared with(**26) those ...

NSE writers used more statistics bundles related to effect size and confidence intervals, whereas NNSE writers used more p-values. NSE writers tend to report effect size and confidence intervals, while NNSE writers tend to report *p*-values exclusively when reporting statistical significance. The reports of the effect size and confidence intervals appear to reflect some characteristics of clinical trials and RCTs in the international medical journal abstracts.

17) NNSE Ex.1 The incidence of … the … group was 8.7%, compared to 18.8% **in the control group (p(**9) = 0.137).
NSE Ex.1 39 (38%) of 103 patients with… in the … group had a 50% or greater reduction in the number of … compared with 32 (31%) of 102 patients in the … group (**adjusted odds ratio(**11) 1.28; 95% CI 0.72-2.28; p = 0.396).

Cause/effect bundles were more frequently observed in the NSE corpus than the NNSE corpus. The bundles in the NSE corpus were associated with adverse events, whereas those in the NNSE corpus were associated with signals of what was found through analyses. The finding suggests that NSE writers are more likely to report adverse events in the results of medical journal abstracts than NNSE writers.

18) NNSE Ex.1 … **analysis showed that(**7) the risk for …. was almost 60% lower among patients who received…. compared to those who received…
NSE Ex.1 Serious **adverse events occurred(**18) in five (8%) patients who received …

The more frequent use of cause/effect bundles in hard disciplines and that of contrast bundles in soft disciplines were reported in Durrant's (2017) study. The finding suggests that Korean writers in the present study tend to follow conventions of soft disciplines and seem to lack awareness of discipline-specific bundles.

Table 9 presents the results of the functional analysis of lexical bundles in the discussion of medical journal abstracts. As shown in Table 9, tokens of text-oriented bundles were significantly higher in the NSE corpus than that of NNSE corpus. Incongruent with previous findings (Lu & Deng, 2019; Pan et al., 2016), NSE writers used more text-oriented bundles than NNSE writers, especially in the discussion of medical journal abstracts. This might be due to some different characteristics involved in the discussion of medical journal abstracts. Both groups of writers used procedure (e.g., *the use of*), description/topic/framing (e.g., *in patients with*), and quantification (e.g., *there was no*) bundles to evaluate the effectiveness of treatment for a certain type of patient.

© 2021 AsiaTEFL All rights reserved

TABLE 9
Functional Analysis of Lexical Bundles in Discussion

Functional categories	Functional subcategories	Types (per 100,000 words)		Tokens (per 100,000 words)	
		NSE	NNSE	NSE	NNSE
Research-oriented		12(67)	15(118)	164(927)	154(1210)
	Place/time (*the long term*)	2(11)	1(8)	11(62)	6(47)
	Procedure (*the use of*)	2(11)	3(23)	38(215)	19(149)
	Quantification (*there was no*)	2(11)	1(8)	18(102)	6(47)
	Description/topic/framing (*in patients with*)	6(34)	10(79)	97(548)	123(967)*
	Statistics (*None*)	0(0)	0(0)	0(0)	0(0)
Text-oriented		18(101)	10(79)	202(1142)	62(487)*
	Text reference/structuring (*in this study*)	2(11)	0(0)	12(68)	0(0)
	Compare/contrast (*as compared with*)	5(28)	1(8)	72(407)	8(63)*
	Cause/effect (*was associated with*)	11(62)	7(55)	118(667)	40(314)*
	Text framing (*in terms of*)	0(0)	2(16)	0(0)	14(110)*
	Transition (*None*)	0(0)	0(0)	0(0)	0(0)
Stance-oriented		4(23)	5(39)	33(186)	27(212)
	Epistemic (*may not be*)	0(0)	2(16)	0(0)	11(86)*
	Engagement (*should not be*)	4(23)	3(23)	33(186)	16(126)

*$p < 0.0001$

NSE writers are more likely to focus on causative relations (e.g., *was associated with*) and comparisons (e.g., *as compared with*) than their counterparts. NSE writers frequently indicated an association between elements (e.g., *was not associated with*) and focused on the reduction of the possibility of bad results using *the risk of*, whereas NNSE writers focused on a decrease of events using *the incidence of*.

19) NNSE Ex.1 … it did not **reduce the incidence**(5) of …
NSE Ex.1 Among patients …, treatment compared with placebo **did not reduce**(12) **the risk of**(15) …

Text framing (e.g., *with respect to*) and epistemic bundles (e.g., *may not be*) were exclusively found in the NNSE corpus. Omidian et al. (2018) indicated that framing signals and epistemic stance were more frequently observed in soft disciplines than hard disciplines; thus, NNSE writers seem to be familiar with conventions of soft disciplines.

20) NNSE Ex.1 … is superior to it **with respect to**(7) treatment-related side effects.
Ex.2 This study suggests that lower… **may not be**(6) associated with …
NSE Ex.1 These findings suggest that … **should not be**(8) used for …

NSE writers frequently employed *should* to direct readers to engage in propositions (e.g., *should not be*) rather than epistemic bundles, although the use of engagement did not differ significantly between the corpora. Considering the thoroughness of the research design in the NSE corpus, NSE writers could be more confident in their findings than NNSE writers. This is one possible explanation for the less frequent use of hedges by NSE writers compared to NNSE writers.

Conclusion

This study explored the structural and functional characteristics of lexical bundles in the corpora. The structural analysis showed that English L2 writers are dependent on clausal bundles more than phrasal bundles. NNSE writers frequently used *to*-clause bundles for research objectives, passive verb bundles for

© 2021 AsiaTEFL All rights reserved

research methods, *be* + (Adj/NP) bundles for reporting results, and *that*-clause bundles for discussing results. On the other hand, NSE writers prominently employed NP + OPM and other NP for the methodology of randomization, level of care, signals of pre-planned results, analysis strategies, statistical markers, and adverse events for research methods and reporting results, as well as active verbs and *be* + (Adj/NP) bundles. The findings show that NSE writers are more likely to detail the methodology and results than NNSE writers and are more explicitly aware of discipline-specific conventions and lexical bundles in the field of medicine as one of the hard science disciplines. The functional analysis showed that English L2 writers rely more heavily on text-oriented than research-oriented bundles. NNSE writers extensively used text reference/structuring bundles for research aims as well as compare/contrast and quantification for reporting group differences. By contrast, NSE writers frequently used the procedure bundles for the research hypothesis, statistics bundles for statistical significance, and cause/effect bundles for adverse events. The findings indicated that NNSE writers are more familiar with conventions prevalent in soft disciplines than hard disciplines.

An incomplete sentence starting with *to*-clause (e.g., *To determine whether*) phrases in the introduction, signals of pre-planned results (e.g., *the primary outcome was*) in the methods, reports of effect size (e.g., *adjusted odds ratio*) and its precision (e.g., *95% confidence interval*), unintentional results (e.g., *serious adverse events*) in the results, and an indication of both negative and positive results in the discussion were more frequently found in the international medical journal abstracts than the Korean journal abstracts. Those might be unique characteristics of medical journal abstracts reporting clinical trials and RCTs. English L2 writers should be encouraged to acquire discipline-specific conventions and lexical bundles through an explicit instruction of language use preferred by community members in the field of medicine for research publication purposes. Further research is needed to examine one- and two-word uses as well as lexical bundles with number slots to investigate how discourse functions are delivered in various ways among community members.

This study has some pedagogical implications for teachers teaching academic English to English L2 writers who plan to write for research publication purposes, particularly in the medical discipline. Teachers should emphasize essential items to include in medical journal abstracts so that English L2 writers are more equipped with a full packet of important information to promote their own research. Secondly, teachers need to provide more tailored writing instruction to help English L2 writers deliver information in medical journal abstracts more economically and efficiently. Lastly, teachers should provide a list of medical discipline-specific bundles to help English L2 writers become more familiar with discourse conventions and lexical bundles that are common and widely accepted in the medical discipline.

Acknowledgements

The article is based on part of the first author's doctoral dissertation (E.-S. Kim 2020). We would like to thank the reviewers for their constructive and thoughtful comments on the first version of our manuscript.

The Authors

Eun-Soo Kim received her Ph.D. and M.A. in English Education at Ewha Womans University. Her main research interests include applied corpus linguistics, English for academic/specific purposes, English for medical purposes, and English for research publication purposes. She is the first author for this article.

Department of English Education
Ewha Womans University
Seoul, 120-750, Korea
Tel: +82-2-3277-2647
Email: esk1209@naver.com

© 2021 AsiaTEFL All rights reserved

Eun-Joo Lee is professor in the Department of English Education at Ewha Womans University. She received her Ph.D. in applied linguistics from UCLA and M.A. in ESL from the University of Hawaii at Manoa. Her main research interests include applied corpus linguistics, lexical processing and acquisition and second language acquisition. She is the corresponding author for this article.

Department of English Education
Ewha Womans University
Seoul, 120-750, Korea
Tel: +82-2-3277-3788
Email: eunlee@ewha.ac.kr

References

Anthony, L. (2018). *AntConc* (Version. 3.5.7) [Computer Software]. Tokyo, Japan: Waseda University.

Appel, R., & Trofimovich, P. (2017). Transitional probability predicts native and non-native use of formulaic sequences. *International Journal of Applied Linguistics, 27*, 24-43.

Biber, D., & Barbieri, F. (2007). Lexical bundles in university spoken and written registers. *English for Specific Purposes, 26*(3), 263-289.

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., Johansson, S., Leech, G., Conrad, S., & Finegan, E. (1999). *Longman grammar of spoken and written English*. Harlow: Pearson.

Cavalieri, S. (2014). Variation across disciplines: The case of applied linguistics and medicine. In M. Bonid & R. Lorés Sanz (Eds.), *Abstracts in academic discourse: Variation and change* (pp. 161-174). Berlin: Peter Lang.

Chen, Y.-H., & Baker, P. (2010). Lexical bundles in L1 and L2 academic writing. *Language Learning and Technology, 14*(2), 30-49.

Cho, D. W. (2009). Science journal paper writing in an EFL context: The case of Korea. *English for Specific Purposes*, *28*(4), 230-239.

Cortes, V. (2004). Lexical bundles in published and student disciplinary writing: Examples from history and biology. *English for Specific Purposes, 23*(4), 397-423.

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*(1), 33-43.

Durrant, P. (2017). Lexical bundles and disciplinary variation in university students' writing: Mapping the territories. *Applied Linguistics*, *38*(2), 165-193.

Esfandiari, R., & Barbary, F. (2017). A contrastive corpus-driven study of lexical bundles between English writers and Persian writers in psychology research articles. *Journal of English for Academic Purposes*, *29*, 21-42.

Flowerdew, J. (2008). Scholarly writers who use English as an additional language: What can Goffman's "Stigma" tell us? *Journal of English for Academic Purposes, 7*(1), 77-86.

Flowerdew, J. (2015). Some thoughts on English for research publication purposes (ERPP) and related issues. *Language Teaching, 48*(2), 250-262.

Gledhill, C. (2000). The discourse function of collocation in research article introductions. *English for Specific Purposes, 19*(2), 115-135.

Huckin, T. (2006). Abstracting from abstracts. In M. Hewings (Ed.), *Academic writing in context: Implications and applications* (pp. 93-103). London: Continuum.

Hyland, K. (2008). As can be seen: Lexical bundles and disciplinary variation. *English for Specific Purposes, 27*(1), 4-21.

Hyland, K., & Tse, P. (2005). Hooking the reader: A corpus study of evaluative *that* in abstracts. *English for Specific Purposes, 24*(2), 123–139.

© 2021 AsiaTEFL All rights reserved

Kim, E.-S. (2020). *A contrastive corpus-based analysis of lexical bundles between English L1 and English L2 writers in medical journal abstracts* (Unpublished doctoral dissertation). Ewha Womans University, Seoul.

Landis, J. R., Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics, 33*, 159-174.

Le, T. N. P., & Harrington, M. (2015). Phraseology used to comment on results in the Discussion section of applied linguistics quantitative research articles. *English for Specific Purposes*, *39*, 45-61.

Lu, X., & Deng, J. (2019). With the rapid development: A contrastive analysis of lexical bundles in dissertation abstracts by Chinese and English L1 doctoral students. *Journal of English for Academic Purposes, 39*, 21-36.

Martín, P., Rey-Rocha, J., Burgess, S., & Moreno, A. I. (2014). Publishing research in English-language journals: Attitudes, strategies and difficulties of multilingual scholars of medicine. *Journal of English for Academic Purposes*, *16*, 57-67.

Maswana, S., Kanamaru, T., & Tajino, A. (2013). Analyzing the journal corpus data on English expressions across disciplines. *The Journal of Asia TEFL, 10*(4), 71-96.

Nam, D. (2017). Functional distribution of lexical bundle in native and non-native students' argumentative writing. *The Journal of Asia TEFL, 14*(4), 703-716.

Omidian, T., Shahriari, H., & Siyanova-Chanturia, A. (2018). A cross-disciplinary investigation of multi-word expressions in the moves of research article abstracts. *Journal of English for Academic Purposes*, *36*, 1-14.

Pan, F., Reppen, R., & Biber, D. (2016). Comparing patterns of L1 versus English L2 academic professionals: Lexical bundles in telecommunications research journals. *Journal of English for Academic Purposes*, *21*, 60-71.

Pérez-Llantada, C. (2014). Formulaic language in L1 and L2 expert academic writing: Convergent and divergent usage. *Journal of English for Academic Purposes*, *14*, 84-94.

Qin, J. (2014). Use of formulaic bundles by non-native English graduate writers and published authors in applied linguistics. *System, 42*(1), 220-231.

Saber, A. (2012). Phraseological patterns in a large corpus of biomedical articles. In A. Boulton, S. Carter-Thomas, & E. Rowley-Jolivet (Eds.), *Corpus-informed research and learning in ESP: Issues and applications* (pp. 45-81). Amsterdam: John Benjamins.

Schroter, S., & Barratt, H. (2004). Editorial decision-making based on abstracts. *European Science Editing*, *30*(1), 4-5.

Shin, Y. K., Choi, H., Kim, D., Ko, S., Yoo, H., Yoo, H., Yoon, J., & Yoo, I. W. (2019). Syntactic complexity of recurrent multiword sequences in the writings of published authors and L1 and L2 English apprentice writers. *The Journal of Asia TEFL, 16*(2), 516-530.

Shin, Y. K., Cortes, V., & Yoo, I. W. (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.

Simpson-Vlach, R., & Ellis, N. (2010). An academic formulas list: New methods in phraseology research. *Applied Linguistics, 31*(4), 487-512.

Van Bonn, S., & Swales, J. M. (2007). English and French journal abstracts in the language sciences: Three exploratory studies. *Journal of English for Academic Purposes*, *6*(2), 93-108.

(Received October 05, 2020; Revised February 28, 2021; Accepted March 10, 2021)

© 2021 AsiaTEFL All rights reserved