TABLE OF CONTENT

Journal of Creative Practices in Language Learning and Teaching (CPLT)
Volume 12, Number 1, 2024

It-Bundles in Applied Linguistics Literature Review Texts: A Corpus-Based Contrastive Analysis

Muna Liyana Mohamad Tarmizi*
muna4566@uitm.edu.my
Akademi Pengajian Bahasa
Universiti Teknologi MARA, Shah Alam, Malaysia

Anealka Aziz Hussin
anealka@uitm.edu.my
Akademi Pengajian Bahasa
Universiti Teknologi MARA, Shah Alam, Malaysia

Corresponding author*

Received: 2 September 2023
Accepted: 15 January 2024
Published: 25 May 2024

CITE THIS ARTICLE:

Tarmizi, M. L. M., & Aziz Hussin, A. (2024). *It*-Bundles in Applied Linguistics Literature Review Texts: A Corpus-Based Contrastive Analysis. *Journal of Creative Practices in Language Learning and Teaching*, *12*(1), 31-48. https://doi. 10.24191/cplt.v12i1.23878

ABSTRACT
Lexical bundles have been recognized as valuable tools for the expression of evaluations, opinions, and attitudes in written discourse. As an important component of fluent linguistic production, previous studies have analysed the use of lexical bundles by writers in various academic fields and registers. Despite the great attention given to lexical bundles, their application in review genres remains relatively understudied. This paper aims to explore the use of *it*-bundles as a specific structural group of lexical bundles with interpersonal functions in literature review texts written by expert writers and student writers in the field of applied linguistics. A corpus-based contrastive analysis is conducted on two specialized corpora: Expert Literature Review Corpus (ELRC) and Malaysian Literature Review Corpus (MLRC). These corpora are assessed using Hewings and Hewings' (2004) *it*-clause typology which categorizes *it*-bundles into four interpersonal roles: hedges, attitude markers, emphatics, and attribution. The results reveal that expert writers and student writers display similar preferences in the use of *it*-bundles in expressing evaluation through the high use of attitude markers. However, there are notable differences between the two groups of writers indicating distinct writing patterns between them. This study's findings can enhance

writing instruction in applied linguistics by helping students effectively utilize *it*-bundles for expressing evaluations, opinions, and attitudes in their academic writing, including literature reviews. These findings also have broader implications for academic writing and underscore the need for further investigation.

Keywords: Applied linguistics, *it*-bundles, literature review texts, corpus-based, contrastive analysis

INTRODUCTION

Review genres can be considered crucial platforms where writers argue their viewpoints, signal their commitments, and display their credibility (Hyland & Diani, 2009). Hyland and Diani (2009) assert that the review genre encompasses various types of texts that are specifically written to evaluate research, texts, and the contributions of academics. This genre involves critical assessments aimed at analyzing and appraising the quality, relevance, and impact of scholarly work and includes book reviews, book review articles, review articles, book blurbs, and literature reviews.

In the realm of academic discourse, literature reviews are a critical component of academic research, particularly in applied linguistics, where they serve as a foundational element that informs new research and contributes to the ongoing dialogue within the field. They are regarded as an essential component in scholarly articles, theses, and dissertations that summarize and evaluate relevant research to provide a comprehensive understanding of a particular topic. They are written to justify the value of research and to distinguish between what has been done and what needs to be done, which explains the necessity for a particular study to be carried out (Hart, 1998; Kwan, 2006). Literature reviews allow researchers to gain familiarity with the current knowledge in their chosen field, understand the boundaries and limitations of that field, and place their research questions into context. By connecting previous studies to the present study and evaluating previous works, writers can provide insights on gaps that need to be addressed as well as the significance and relevance of their study.

While literature reviews play an integral role in advancing knowledge, it is surprising that studies focusing specifically on the linguistic features and discourse strategies within this genre are relatively understudied. One particular area that remains underexplored is the examination of multi-word expressions that also go by various names such as ‘n-grams’, ‘fixed expressions’, formulaic language’ and ‘lexical bundles’ (Biber et al., 1999; Cortes, 2004; Chen & Baker, 2010; Wray, 2000). In this study, the term lexical bundles is used (Biber et al., 1999). Chen and Baker (2010) define lexical bundles as sequences of words that are retrieved by taking a corpus-driven approach with specified frequency and distribution criteria. They further elaborate that these recurrent sequences are fixed multi-word unit with pragmatic and discourse functions which are used and recognized by the speakers or writers of a language within a certain context.

These multi-word expressions have a significant impact on contributing to the distinction of registers and shaping text meanings, including literature reviews (Hyland, 2008). Research has

demonstrated that certain words frequently co-occur, creating strings or bundles that serve as set building blocks for constructing different types of texts (Biber et al., 2004). For instance, expressions like *as a result of*, and *it should be noted* are indicative of the academic register while *in pursuance of* and *in accordance with* are likely to distinguish legal text (Hyland, 2008).

Previous studies have also highlighted that these recurrent word combinations fulfill specific functions, such as referential, organizational, or interpersonal functions, which contribute to the discourse organization and meanings of texts (Biber et al., 2004; Hyland, 2008; Hyland, 2012). Certain multi-word expressions, such as *it is necessary to* and *are more likely to*, are utilized by writers to express their personal opinions and attitudes, thereby shaping the meanings of the texts.

Hyland (2012) emphasizes the significance of mastering multiword expressions as a key aspect of fluent linguistic production. He concurs with Chen and Baker (2010) that readers and writers who frequently use a particular genre are familiar with these multi-word expressions and he also argues that 'the absence of such clusters might reveal the lack of fluency of a novice or newcomer to that community'. In other words, writers who successfully comprehend and compose texts using lexical bundles correctly are likely to be perceived as fluent readers or writers of that language (Wright, 2019).

Despite the significant influence of lexical bundles in academic writing, the intricacy of multiword structures poses challenges for new academic writers to master (Wright, 2019). For example, Hewings and Hewings (2004) contend that many languages have no counterpart to *it*-bundles which could lead to grammatical and appropriacy errors. Furthermore, studies reveal that student writers seldom utilize lexical bundles in their writing (Cortes, 2004), and when they do, they rely on simple types of lexical bundles repeatedly (Joharry, 2021).

While much is known about lexical bundles, there remains a research gap in terms of in-depth exploration and detailed descriptions of a specific structural group called *anticipatory it* or *it*-bundles, despite their frequent occurrence in academic writing compared to other registers (Biber, 2006; Hewings & Hewings, 2004). Hewings and Hewings (2004) have demonstrated that lexical bundles starting with *it*-bundles fulfill crucial interpersonal functions, expressing opinions, providing commentary, and evaluating propositions, all of which are pertinent to academic writing. The review of the literature indicates that few studies have concentrated on the use of *it*-bundles within key academic genres (Hewings & Hewings, 2004; Jalali et al., 2009; Jalali, 2015; Jalali & Zarei, 2016). Particularly lacking are studies addressing the use of *it*-bundles in the review genre, especially those aiming to describe potential similarities and differences between expert and novice writers in using these word combinations in literature review texts.

Considering the importance of literature review texts, lexical bundles, and *it*-bundles in academic writing, this study aims to investigate the use of *it*-bundles by expert writers and student writers in applied linguistic literature review texts. Additionally, the study seeks to compare and contrast the similarities and differences found in the utilization of *it*-bundles by expert writers and student writers in terms of the interpersonal roles they play in applied linguistic literature review texts. Specifically, the objectives of the study are:

a. To analyze the use of *it*-bundles in literature reviews by student writers and expert writers.
b. To examine similarities and differences in terms of interpersonal functions of *it*-bundles used in the literature reviews by student writers and expert writers.

The study intends to answer the following research questions:

a. What are the most frequent *it*-bundles used by student writers in the literature review texts?
b. What are the most frequent *it*-bundles used by expert writers in the literature review texts?
c. What are the similarities and/or differences between student writers and expert writers in terms of interpersonal function of *it*-bundles used in the literature review texts?

LITERATURE REVIEW

Lexical bundles, commonly referred to as recurrent word combinations, play a crucial role in language use and are significant units of analysis in various linguistic disciplines. Biber and his colleagues view lexical bundles as recurrent word combinations that exhibit a higher frequency of occurrence than would be expected by chance (Biber et al., 1999). More specifically, Cortes (2004) defines lexical bundles as extended collocations, sequences of three or more words that statistically co-occur in a register. She also emphasizes that frequency of occurrences is the defining characteristic for lexical bundles that a bundle needs to occur more than 20 times in a million words to be qualified as one (Cortes, 2004) although this can vary depending on the specific study being conducted.

There are various ways of categorizing lexical bundles based on their functions and structures. Hyland (2008) for example, has categorized bundles into three groups, namely, Research-oriented bundles (description of research experiences), Text-oriented bundles (organization of the text/argument), and Participant-oriented bundles (writer/reader-focused features of the discourse). Whereas Biber and Barbieri (2007) have grouped lexical bundles based on their discursive functions such as organization of discourse, expression of stance, and reference to textual or external entities.

Interestingly, Biber et al. (2004) highlight that there is a relationship between the structural type of bundles and the role they play in discourse. For example, Anticipatory *it*-bundles such as *it is possible to, it should be noted that* are usually used to act as metadiscourse elements (Hyland, 2008) or expressions of stance (Biber et al., 1999; Jalali & Zarei, 2016). Hewings and Hewings (2004) further support that anticipatory *it*, or *it*-clauses have been found to be relatively frequent in academic writing when compared with other registers. They argue that anticipatory *it* or *it*-clauses is a feature of academic writing that functions to express opinions and to comment on and evaluate propositions by allowing writers to remain in the background. They have classified *it*-clauses into four main interpersonal functions as displayed in Table 1 below.

Table 1
Classifications of *it-bundles*

Interpersonal functions of it-clauses	Subcategories	Example realizations
hedges	1a) likelihood/possibility/certainty; importance/value/necessity etc. 1b) what a writer thinks/assumes to be/will be/was the case	it is likely, it seems improbable, it would certainly appear it could be argued, it seems reasonable to assume, it was felt
attitude markers	2a) the writer feels that something is worthy of note 2b) the writer's evaluation	it is of interest to note; it is worth pointing out; it is noteworthy it is important; it was interesting; it is surprising
emphatics	3a) the writer indicates that a conclusion/deduction should be reached; that a proposition is true 3b) the writer strongly draws the reader's attention to a point 3c) the writer expresses a strong conviction of what is possible/ important/necessary, etc.	it follows; it is evident; it is apparent it is important to stress; it should be noted; it must be recognised; it is essential to understand it is clear; it is impossible; it is safe to assume; it would be strongly desirable
attribution	4a) specific attribution (with a reference to the literature) 4b) general attribution (no referencing)	it has been proposed (+reference) it is estimated (+no reference)

Hedges
In this category, *it*-clauses or *it*-bundles play the interpersonal function of hedges that withhold writers' full commitment to the content of the extraposed subject (Hewings & Hewings, 2002). There are two subcategories which include 1a – hedges that indicate a certain degree of probability and 1b – where writers indicate the non-factual status of propositions made by marking them as a form of suggestion, assumption, and argument.

Attitude Markers
Attitude markers express writers' attitudes towards the content of the extraposed subject. There are two subcategories that distinguish different forms of expressions of writers' attitudes under this interpersonal function which are 2a – identification of information as worthy of particular attention and those which express an evaluation or assessment of how the content compares with expectations in - 2b.

Emphatics

The third interpersonal function of *it*-bundles identified here emphasises the force of writers' certainty in the content of the extraposed subject. Under this functional category, writers may indicate that a conclusion or inference should be drawn without any form of mitigation using hedges (3a). In 3b, *it*-bundles are used to forcefully draw readers' attention to some point whereas under 3c writers express strong conviction on what is possible, important, or necessary.

Attribution

Attributions are utilized to lead readers to accept the writers' judgment as being soundly based by making specific attribution (4a) with reference to literature attached to the bundles as well as using general attribution (4b) which has no such reference.

Related Studies

A number of corpus-based and comparative studies have been conducted to explore the use of lexical bundles as well as their possible similarities and/or differences in different disciplinary fields, genres, and different degrees of writing expertise. In her study that focuses on disciplinary variations in the use of multi-word expressions, Cortes (2004) found that writers in biology research articles used more lexical bundles as compared to writers from the history field. Hyland (2008) who examined the use of 4-word lexical bundles in four academic fields (electrical engineering, biology, business studies, and applied linguistics) observed disciplinary differences that some bundles used by members of a particular discipline are not found in other disciplines. He also found that writers in applied linguistics and business studies fields used lexical bundles to help frame, scaffold, and present arguments as an organized arrangement which indicates writers' awareness of the discoursal expectations of readers from their disciplines.

In the realm of student writing, multiple studies have delved into the utilization of lexical bundles among various demographics. Ang and He (2017) examined Asian college students' academic essays, employing Simpson-Vlach and Ellis' (2010) taxonomy for functional classification of lexical bundles. Their findings indicated a reluctance among these learners to employ referential expressions. Conversely, Subramaniam and Kaur (2023) scrutinized final year project reports by Malaysian Electronic Engineering students in Polytechnics, revealing a tendency for limited participant-oriented functions in passive verb bundles, as opposed to the extensive engagement and stance categories favored by L1 English learners. Joharry (2021) focused on argumentative essay writing among students across four states in Malaysia, uncovering a penchant for repetitively employing simple types of lexical bundles compared to native speaker counterparts. Notably, these studies adopt a corpus-driven approach to compare native (L1) and non-native (L2) speakers, elucidating insightful differences in lexical bundle usage among different learner groups.

A few studies that specifically focus on anticipatory it or *it*-bundles have been done to compare similarities and/or differences between writers from different degrees of writing expertise. Hewings and Hewings (2002) explored the use of *it*-clauses in expert and novice writing from the field of business studies. Their study suggests that student writers used *it*-clauses to persuade readers of the validity of claims and show more overt persuasive efforts in doing so as compared to expert writers. Jalali et al. (2009) who examined anticipatory *it*-bundles in three

corpora of research articles, doctoral dissertations, and master theses of the applied linguistics field found that student writers use fewer *it*-bundles than expert writers.

The review of the above studies indicates that there have been very few research focusing on the study of *it*-bundles within one single disciplinary area which also compares similarities and/or differences between two groups of writers from different degrees of writing expertise. This study aims to address possible intradisciplinary variations in the use of *it*-bundles between expert and student writings in the applied linguistics field.

RESEARCH METHODOLOGY

This study involved the construction of two distinct corpora tailored to the study's objectives: the Expert Literature Review Corpus (ELRC) and the Malaysian Literature Review Corpus (MLRC). The ELRC, serving as the reference corpus, comprises literature review sections extracted from journal articles sourced from prominent linguistics and applied linguistics journals such as English for Academic Purposes, Language and Communication, and Applied Linguistics. On the other hand, the MLRC serves as the target corpus and includes literature review chapters sourced from applied linguistics master theses at UiTM, UKM, and UPM. Table 2 provides a summary of key details pertaining to these corpora:

Table 2
Number of Texts and Tokens for ELRC and MLRC

Corpora	Number of Texts	Tokens
ELRC	1045	1,264,734
MLRC	90	696,494

Wordsmith Tools was used to generate *it*-bundles and identify their frequency. It has tools like KeyWords, WordList, and Concord that are used to analyze texts of different kinds and lengths. Using the WordList tool, an index list can be compiled to see word clusters or lexical bundles. The clusters or bundles can be automatically generated by imposing several restrictions which involve choosing how many words a bundle should have (in this study, 4-word bundles), how many of each bundle must be found in the corpus (minimum frequency set for this study is 5) as well as by instructing the tool to stop counting bundles at sentence breaks considering a bundle which spans across two sentences is not likely to make sense.

Functionally, Hewings and Hewings' (2002) classification of *it*-clause was used to categorize *it*-bundles into four main interpersonal functions which are hedges, attitude markers, emphatics, and attributions. Regular discussions among the researchers were conducted to ensure ongoing alignment in the classification process. Following this categorization, the results have been compared and contrasted whereby excerpts from the two corpora were selected to serve as examples as well as to examine the similarities and differences of how learners and experts use *it*-bundles in the literature review texts.

RESULTS AND DISCUSSION

The findings in this study are explained based on the Functional Typology developed by Hewings and Hewings (2002) that categorizes *it*-bundles into four main interpersonal roles; hedges, attitude markers, emphatics, and attribution. Considering the total number of words for both corpora is not the same, the frequency of occurrence for *it*-bundles found in ELRC and MLEC were normalized per 1,000 words to allow direct comparison.

An Overview of Findings: Frequency, Function, and Distribution

Based on Table 3, it can be observed that *it*-bundles identified in both corpora display the use of all four interpersonal functions of hedges, attitude markers, and emphatics with different frequencies of occurrences. The three most popular interpersonal functions of *it*-bundles in both corpora are attitude markers, hedges, and attribution whilst the least popular is emphatics. ELRC has more types of *it*-bundles with 52 different types of *it*-bundles as compared to MLRC's 42 types of *it*-bundles. Even though the frequency of occurrences of *it*-bundles in MLRC (1164) is higher than ELRC (1042) with lesser types of *it*-bundles, the total normalized frequency of *it*-bundles in ELRC is slightly higher than MLRC with 1.822 occurrences per 1000 words as compared to 1.664 occurrences per 1000 words in student writings. These results concur with the findings by Jalali et al. (2009) that published academics make noticeably heavier and more varied use of *it*-bundles than student writers.

Table 3
Functions, Types, and Frequency of Occurrences of *it*-Bundles in ELRC and MLRC

Interpersonal Functions	ELRC Type/Freq.	Freq. per 1000 words	MLRC Freq.	Freq. per 1000 words
Hedges	16/143	0.113	11/115	0.165
Attitude Markers	21/210	1.167	15/200	0.287
Emphatics	5/65	0.051	5/66	0.095
Attribution	10/103	0.08	11/201	0.285
Total	**52/1042**	**1.822**	**42/1164**	**1.664**

It-Bundles Used by Expert Writers in ELRC

Table 4 below shows the functions, types, frequency, and percentage of occurrences of *it*-bundles in ELRC. Expert writers capitalized maximally on attitude markers (20.15%) and minimally on emphatics markers (6.24%) in writing reviews of literature. Some of the most frequent *it*-bundles in the corpus include *it is necessary to* (34 times), *it is clear that* (23 times), *it should be noted* (23 times), *it is not clear* (22 times), and *it is difficult to* (22 times). This suggests that, when reviewing the literature, expert writers tend to provide their evaluations by expressing their personal attitudes towards the content of the extraposed subject. For attitude markers too, only one *it*-bundle (*it is worth noting*) falls under the subcategory 2a in which writers indicate some information is worth particular attention while most *it*-bundles found under this subcategory (2b) function as an expression of writers' evaluation using bundles with adjectives like *important, interesting, crucial, surprising*, and *interesting.*

Table 4
Functions, Types, Frequency, and Percentage of Occurrences of *it*-Bundles in ELRC

		ELRC Type/Freq.	Freq. per 1000 words	Percentage (%)
1 Hedges	1a	7/66	0.052	6.33
	1b	9/77	0.061	7.39
	Total	**16/143**	**0.113**	**13.72**
2 Attitude Markers	2a	1/7	0.006	0.67
	2b	20/203	0.161	19.48
	Total	**21/210**	**1.167**	**20.15**
3 Emphatics	3a	3/19	0.015	1.82
	3b	1/23	0.018	2.21
	3c	1/23	0.018	2.21
	Total	**5/65**	**0.051**	**6.24**
4 Attribution	4a	10/103	0.08	9.88
	4b	0/0	0	0.00
	Total	**10/103**	**0.08**	**9.88**
Total		**52/1042**	**1.822**	**100.00**

Furthermore, 13.72% of *it*-bundles found in ELRC are used to mitigate certain propositions made by hedging them to appear as a form of suggestion, assumption, opinion, or probability. There are seven bundles under subcategory 1a that comprise expressions of likelihood and possibility for example *it is possible that, it is likely that, it may be that,* and *it is unlikely that.* Another nine *it*-bundles that fall under the subcategory of 1b are used by expert writers to express what they think or assume to be the case using bundles like *it can be assumed, it could be argued, it is expected that,* and *it can be inferred.*

When making attributions in literature review texts, expert writers use *it*-bundles to make specific attributions that have references to literature attached to them. Examples of *it*-bundles found in ELRC include *it has been found, it has been reported, it has been suggested, and it has been argued.* There are no *it*-bundles that fall under subcategory 2b – general attribution.

For interpersonal function emphatics which emphasizes the force or writers' certainty in the content of the extraposed subject, three *it*-bundles are used by expert writers to indicate a conclusion or inference should be drawn which involve bundles *it is therefore important, it is obvious that,* and *it is evident that.* There is one *it*-bundle – *it should be noted* that falls under subcategory 3b which is used to forcefully draw readers' attention to certain points. Another bundle used by expert writers to express strong conviction under subcategory 3c is *it is clear that.* A full list of 52 different types of *it*-bundles found in ELRC as well as their frequencies and normalized frequencies is summarised in Table 5.

Table 5
Types, Frequency, and Normalised Frequency per 1000 Words of *it*-Bundles in ELRC

***It*-Bundles**	**Freq.**	**Freq. per 1000 Words**
it is possible to	17	0.013
it is likely that	11	0.009
it is also possible	8	0.006
it may be that	6	0.005
it may be the	6	0.005
it is unlikely that	5	0.004
It is possible that	13	0.010
it can be assumed	8	0.006
it can/could be argued	13	0.010
it can be said	7	0.006
it is hoped that	7	0.006
it has been claimed	6	0.005
it is estimated that	6	0.005
it is expected that	6	0.005
it can be inferred	5	0.004
it is argued that	19	0.015
it is worth noting	7	0.006
it is important to	7	0.006
it is important for	9	0.007
it is important that	8	0.006
it is interesting to	7	0.006
it is crucial to	6	0.005
it is surprising that	6	0.005
it is impossible to	6	0.005
it is difficult to	22	0.017
it is not clear	22	0.017
it is essential to	17	0.013
it is unclear whether	10	0.008
it would be interesting	8	0.006
it is also important	6	0.005
It is necessary to	34	0.027
It is interesting that	6	0.005
It is of interest	5	0.004
It is not surprising	5	0.004
It is not possible	5	0.004
it is reasonable to	7	0.006
it is not uncommon	7	0.006
it is therefore important	7	0.006
it is obvious that	5	0.004
it is evident that	7	0.006
it should be noted	23	0.018
it is clear that	23	0.018
it has been found	12	0.009

it has been shown	7	0.006
it is widely acknowledged	6	0.005
it was found that	20	0.016
it has been argued	19	0.015
it has been suggested	13	0.010
it has been noted	5	0.004
it has been reported	5	0.004
it is hypothesized that	5	0.004
it can be seen	11	0.009

The following examples from this corpus can illustrate the use of some of the bundles by expert writers:

> Moreover, given the multidimensional roles of EAP teachers and the diversity of their educational backgrounds and expertise, *it is necessary to* look into TC in different educational contexts.
>
> ELRC EAP95

> *It is clear that* unless doctoral students have received explicit guidance in scientific writing, these determinants of agency may be very much out of their reach.
>
> ELRC ESP265

> *It should be noted* that most studies on ESL or EFL academic writing take ENL texts as the unquestioned language norm, thus depicting ESL or EFL texts as suffering from underuse, overuse or misuse of linguistic items.
>
> ELRC EAP4

It-Bundles Used by Student Writers in MLRC

In Table 6 below, the functions, types, frequency, and percentage of occurrences of *it*-bundles found in MLRC are displayed. It can be observed that attitude markers have the highest normalised frequency with 15 different types of bundles in this functional category. All *it*-bundles found in this interpersonal category fall under subcategory (2b) that functions as an expression of writers' evaluation. In this subcategory, student writers used bundles with adjectives like *important, crucial, essential*, *difficult, hard,* and *necessary*.

Table 6
Functions, Types, Frequency, and Percentage of Occurrences of *it*-Bundles in MLRC

		MLRC Type/Freq.	Freq. per 1000 words	Percentage (%)
1 Hedges	1a	1/12	0.017	1.03
	1b	10/103	0.148	8.85
	Total	**11/115**	**0.165**	**9.88**

2Attitude Markers	2a	15/200	0.287	17.18
	2b	0/0	0	0.00
	Total	**15/200**	**0.287**	**17.18**
3 Emphatics	3a	3/36	0.052	3.09
	3b	1/7	0.010	0.60
	3c	1/23	0.033	1.98
	Total	**5/66**	**0.095**	**5.67**
4 Attribution	4a	6/128	0.18	11.00
	4b	5/73	0.105	6.27
	Total	**11/201**	**0.285**	**17.27**
Total		**42/1164**	**1.664**	**100**

It is also interesting to note that, when reviewing the literature, student writers prefer the use of *it*-bundles to make references by making specific attributions to the literature and by making general references. Examples of *it*-bundles found in MLRC include *it is/was found, it is/was revealed that, it was noted that,* and *it is seen as.* Five *it*-bundles under this category fall under subcategory 2b – general attribution which comprises bundles like *it was discovered that, it was suggested that, it can be seen,* and *it is said that*.

In MLRC there are ten *it*-bundles that are categorized under the subcategory of 1b whereby student writers indicate the non-factual status of a proposition by marking it as being their suggestion, contention, argument, and assumption using bundles like *it can be assumed, it could be argued, it is believed that,* and *it is hoped that.* Only one *it*-bundle falls under subcategory 1a which indicates a certain degree of likelihood and possibility in literature review texts which is *it is possible to.*

For interpersonal function emphatics which emphasizes the force or writers' certainty in the content of the extraposed subject, three *it*-bundles are used by student writers to indicate a conclusion or inference should be drawn. The bundles include *it shows that the,* and *it is evident that.* One *it*-bundle – *it should be noted* falls under subcategory 3b which is used to forcefully draw readers' attention to particular points. Another bundle used by student writers to express strong conviction is *it is clear that* which falls under subcategory 3c. A full list of 42 different types of *it*-bundles found in MLRC as well as their frequencies and normalized frequencies is presented in Table 7.

Table 7
Types, Frequency, and Normalised Frequency per 1000 Words of *it*-Bundles in MLRC

***It*-Bundles**	**Freq.**	**Freq. per 1000 Words**
it is possible to	12	0.017
it can be argued	6	0.009
it can be assumed	5	0.007
it can be inferred	7	0.010
it can/could be said	37	0.053
it is assumed that	6	0.009
it is believed that	19	0.027

it is believed to	6	0.009
it is expected that	5	0.007
it is hoped that	7	0.010
it is supposed to	5	0.007
it is also important	7	0.010
it is an important	7	0.010
it is crucial for	6	0.009
it is crucial to	8	0.011
it is difficult for	5	0.007
it is difficult to	16	0.023
it is essential for	6	0.009
it is essential to	16	0.023
it is hard to	5	0.007
it is important for	28	0.040
it is important that	7	0.010
it is important to	68	0.098
it is interesting to	5	0.007
it is necessary to	6	0.009
it is very important	10	0.014
it can be concluded	26	0.037
it is apparent that	5	0.007
it shows that the	5	0.007
it should be noted	7	0.010
it is clear that	23	0.033
it is/was found that	81	0.12
it is/was noted that	13	0.02
it is/was revealed that	22	0.03
it was concluded that	7	0.01
It is seen as	5	0.01
it was discovered that	12	0.017
it was suggested that	5	0.007
it is said that	7	0.010
it is stated that	5	0.007
it can be seen	37	0.053
it is observed that	7	0.010

The following examples from this corpus can illustrate the use of some of the bundles by student writers:

> Based on the data analysis, *it was found that* both groups showed improvement in their writing scores by the end of the term.
>
> MLRC UiTM7

> *It is important to* have the customers to developed interest towards the advertisement instead of just have casual looks on it.
>
> MLRC UKM7

> Given these points, *it can be said* the study of love that is portrayed in the anthology of Love in Penang will add to more studies of love in the Malaysian literary works.
>
> MLRC UKM18

Similarities and Differences in the Use of It-Bundles by Expert Writers and Student Writers in Literature Review Texts

There are 20 shared bundles used by both groups of writers when reviewing the literature. Nine *it*-bundles fall under the attitude markers category and seven shared *it*-bundles are categorized under interpersonal function hedges. At the same time, two shared bundles are found in both emphatics and attribution categories. A full list of 20 shared *it*-bundles found in ELRC and MLRC as well as their frequencies and normalised frequencies is presented in Table 8.

Table 8
Types, Frequency, and Normalised Frequency per 1000 Words of Shared *it*-Bundles in ELRC and MLRC

Shared *It*-Bundles	**ELRC**		**MLRC**	
	Freq.	**Freq. per 1000 words**	**Freq.**	**Freq. per 1000 words**
it is possible to	17	0.013	12	0.017
it can be assumed	8	0.006	5	0.007
it can/could be argued	13	0.010	6	0.009
it can be said	7	37.00	37	0.053
it is hoped that	7	0.006	7	0.010
it is expected that	6	0.005	5	0.007
it can be inferred	5	0.004	7	0.010
it is important to	7	0.006	68	0.098
it is important for	9	0.007	28	0.040
it is important that	8	0.006	7	0.010
it is interesting to	7	0.006	5	0.007
it is crucial to	6	0.005	8	0.011
it is difficult to	22	0.017	16	0.023
it is essential to	17	0.013	6	0.009
it is also important	6	0.005	7	0.010
it is necessary to	34	0.027	6	0.009
it is clear that	23	0.018	23	0.033
it should be noted	23	0.018	7	0.010
it was found that	20	0.016	81	0.116
it can be seen	11	0.009	37	0.053
Total	**256**	**0.200**	**378**	**0.543**

Based on the table, it can be seen that the frequencies and normalized frequencies of the shared bundles in the student writers' corpus are higher than the expert writers' despite the smaller size of the corpus. This could indicate overuse of some *it*-bundles in MLRC *it*-bundles like *it is important to* (68 times), *it was found that* (81 times), and *it can be seen* (37 times) which occur

more than 35 times in the whole corpus. On the other hand, expert writers used not more than 20 times the listed bundles when reviewing the literature.

It is interesting to note that, even though there are some similar types of *it*-bundles that fall under the same interpersonal function category, they can be classified under different subcategories. For example, attribution markers *it was found that,* and *it can be seen* are used as general attribution in MLRC, but expert writers used these bundles as specific attribution.

> Overall, *it can be seen* that many of medical students' learning sessions or activities do require a good proficiency of English.
>
> MLRC UPM27

> From some studies on workplace communicative events, *it can be seen* that participating in meetings is one of the key spoken activities in the target situation; other common activities include giving oral presentations, telephoning, socializing, etc.
>
> ELRC ESP279

Furthermore, there are also some *it*-bundles used by expert writers to hedge and express their attitude towards certain propositions that are not found in student writers' writing. The bundles include *it is likely that, it may be that, it is worth noting, it would be interesting, it is surprising,* and *it is reasonable to.*

In MLRC, *it*-bundles for hedging are found less frequent (third highest) than in ELRC (second highest). This may suggest stronger claims are made by student writers when reviewing the literature as indicated by Hewings and Hewings (2002) in their findings, 'if statements are hedged less, then it suggests that students are making stronger claims for their validity'. This finding also concurs with Hewings and Hewings' (2004) findings where they find student writes have the tendency to make stronger claims with less hedged statements when reviewing the literature.

CONCLUSION AND IMPLICATIONS

The present study attempts to explore the most common *it*-bundles used by student writers and expert writers in literature review texts. The findings indicated that student writers could be more versatile by using a wider variety of *it*-bundles to engage interpersonally with the texts and the readers as shown by expert writers in their writings. Student writers need to also observe certain *it*-bundles to avoid overuse which could signal a limited range of expression, repetitiveness, and lack of precision when providing critical judgement in evaluating the literature. It is hopeful that the findings of the study are beneficial in constructing pedagogical instructions and guidelines for writers in their critical analysis of the literature review texts. Through clear directives on how to incorporate and leverage "*it*-bundles" within their analyses, student writers can elevate the depth and coherence of their literature reviews. This approach not only empowers students to express their ideas more effectively but also fosters a deeper understanding of how language nuances can influence the conveyance of knowledge and perspectives in academic writing. Future research

endeavors could explore the underlying factors influencing student writers' capacity to employ a broader range of "*it*-bundles" in comparison to expert writers. Additionally, conducting comparative studies across diverse disciplines may elucidate discipline-specific variations in the utilization of "*it*-bundles" within literature reviews. Such investigations could contribute significantly to our understanding of effective writing practices across academic domains.

REFERENCES

Ang, L. H., & He, M. (2017). Multi-word sequences in learner corpora: A corpus analysis of lexical bundles. *Journal of Creative Practices in Language Learning and Teaching*, *5*(1), 19-33.

Biber, D., Johansson, S., Leech, G. N., Conrad, S., & Finegan, E. (1999). *Grammar of spoken and written English*. Pearson Education Limited.

Biber, D., Conrad, S., & Cortes, V. (2004). If you look at . . .: Lexical bundles in university teaching and textbooks. *Applied Linguistics, 25*(3), 371-405. https://doi.org/10.1093/applin/25.3.371

Biber, D. (2006). *University language: A corpus-based study of spoken and written registers* (Vol. 23). John Benjamins Publishing.10.1075/scl.23Search in Google Scholar

Biber, D., & Barbieri, F. (2007). Lexical bundles in university spoken and written registers. *English for Specific Purposes*, *26*(3), 263-286.

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

Cortes, V. (2004). Lexical bundles in published and student disciplinary writing: Examples from history and biology. *English for Specific Purposes, 23*(4), 397-423. https://doi.org/10.1016/j.esp.2003.12.001

Hart, C. (1998). *Doing a literature review: Releasing the social science research imagination*. Sage Publications.

Hyland, K. (2008). Academic clusters: Text patterning in published and postgraduate writing. *International Journal of Applied Linguistics, 18*(1), 41–62. https://doi.org/10.1111/j.1473-4192.2008.00178.x

Hyland, K., & Diani, G. (2009). Introduction: Academic evaluation and review genres. In K. Hyland & G. Diani (Eds.), *Academic Evaluation*. Palgrave Macmillan. https://doi.org/10.1057/9780230244290_1

Hyland, K. (2012). Bundles in academic discourse. *Annual Review of Applied Linguistics, 32*, 150-169. https://doi.org/10.1017/S0267190512000037

Hewings, M., & Hewings, A. (2002). "It is interesting to note that...": A comparative study of anticipatory 'it' in student and published writing. *English for Specific Purposes, 21*(4), 367-383. https://doi.org/10.1016/S0889-4906(01)00016-3

Hewings, M., & Hewings, A. (2004). Impersonalizing stance: A study of anticipatory 'it' in student and published academic writing. In C. Coffin, A. Hewings & K. O'Halloran (Eds.), *Applying English Grammar: Functional and Corpus Approaches* (pp. 101–116). Hodder Arnold.

Jalali, H., Rasekh, A. E., & Rizi, M. T. (2009). Anticipatory 'it' lexical bundles: A comparative study of student and published writing in applied linguistics. *Iranian Journal of Language Studies, 3*(2), 177-194.

Jalali, H. (2015). A corpus-driven study of it lexical bundles in applied linguistics postgraduate genres. *Journal of Applied Linguistics and Language Research*, *2*(1), 36-49.

Jalali, H., & Zarei, G. (2016). Published vs. postgraduate writing in applied linguistics: The case of lexical bundles. *Applied Research on English Language*, *5*(1), 1-16.

Joharry, S. A. (2021). Repetitive bundles in Malaysian learner writing. *Australian Review of Applied Linguistics, 44*(1), 82-108. https://doi.org/10.1075/aral.19028.joh

Kwan, B. S. (2006). The schematic structure of literature reviews in doctoral theses of applied linguistics. *English for Specific Purposes, 25*(1), 30-55. https://doi.org/10.1016/j.esp.2005.06.001

Simpson-Vlach, R., & Ellis, N. C. (2010). An academic formulas list: New methods in phraseology research. *Applied Linguistics, 31*(4), 487-512. https://doi.org/10.1093/applin/amp058

Subramaniam, R., & Kaur, S. (2023). Specialised learner corpus research: A review for future directions of the global and Malaysian contexts. *TEFLIN Journal: A Publication on the Teaching & Learning of English*, *34*(1), 176-193. http://dx.doi.org/10.15639/teflinjournal.v34i1/176-193

Wray, A. (2000). Formulaic sequences in second language teaching: Principle and practice. *Applied Linguistics, 21*(4), 463–489. https://doi.org/10.1093/applin/21.4.463

Wright, H. R. (2019). Lexical bundles in stand-alone literature reviews: Sections, frequencies, and functions. *English for Specific Purposes*, *54*, 1-14. https://doi.org/10.1016/j.esp.2018.09.001

Conflict of Interest

The author declared no potential conflicts of interest concerning the research, authorship, and publication of this article.

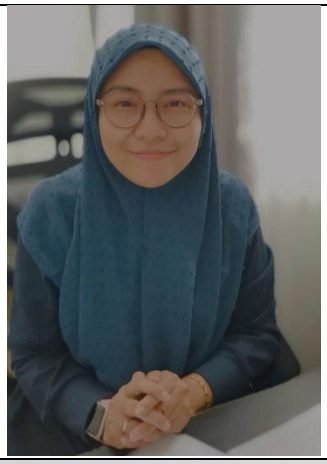
Acknowledgement

The authors would like to thank UiTM, UPM, and UKM Office of Postgraduate Studies for their support and assistance in the data collection process.

Authors' Contributions

Muna Liyana carried out the research plan and conducted data collection and analysis. Prof. Dr. Anealka Aziz Hussin provided feedback, reviewed, and assisted in completing the manuscript.

About the Authors

	**Muna Liyana Mohamad Tarmizi** is a lecturer in Akademi Pengajian Bahasa (APB) UiTM. Her areas of interest are applied linguistics, corpus linguistics, and pragmatics.
	**Anealka Aziz (PhD)** is an accomplished academician and researcher in the field of Education, focusing on e-learning, corpus-based studies, readability studies, and ELT materials development. She has been affiliated with the Akademi Pengajian Bahasa at Universiti Teknologi MARA (UiTM) in Shah Alam, Selangor since 1997.