Pertanika J. Soc. Sci. & Hum. 25 (1): 133 – 148 (2017)

PERTANIKA JOURNALS

SOCIAL SCIENCES & HUMANITIES

Journal homepage: http://www.pertanika.upm.edu.my/

Comparing Structural and Functional Lexical Bundles in MUET Reading Test

Christina Ong Sook Beng[1]* and Yuen Chee Keong[2]

[1]Department of Languages and Linguistics, Faculty of Arts and Social Science, Universiti Tunku Abdul Rahman, 31900 UTAR, Kampar, Perak, Malaysia
[2]School of Language Studies and Linguistics, Faculty of Social Sciences and Humanities, Universiti Kebangsaan Malaysia, 43600 UKM Bangi, Selangor, Malaysia

ABSTRACT

This study aims at identifying the structural and functional types of lexical bundles (LBs) used in the reading passages of Malaysian University English Test (MUET). A specialised corpus of MUET reading passages was built. The passages were categorised into five main disciplines namely Applied Science, Pure Science, Business, Humanities and Social Science. Using WordSmith Tools version 5, the lists of frequently occurring LBs in all the five disciplines were generated. They were then sorted according to Biber, Conrad and Cortes' (2004) Structural Taxonomy and Hyland's (2008) Functional Taxonomy. Chi-square test and Fisher's exact test were adopted to determine the association between the structural and functional categories of the five disciplines. The results revealed that the number of LBs across the structural and functional categories in the five disciplines differed significantly. However, a strong association was observed between the two categories in all five disciplines where LBs incorporating noun phrases (NPs) usually performed research-oriented function whereas LBs incorporating dependent clauses (DCs) were strongly bound to text-oriented function. LBs incorporating verb phrases (VPs) on the contrary were linked to participant-oriented functions as well as other types of functions. Significant association was identified between the categories in all the disciplines. The two additional categories encompassing various structures of LBs and other functions not listed in Biber et al.'s (2004) Structural Taxonomy and Hyland's (2008) Functional Taxonomy respectively were also significantly associated. The key finding of the study was that structural categories and discourse functions are closely interrelated.

ARTICLE INFO

Article history:
Received: 2 July 2015
Accepted: 2 September 2016

Keywords: Lexical bundles (LB), MUET, reading test, reading text, structural and functional categories

E-mail addresses:
ongsb@utar.edu.my (Christina Ong Sook Beng),
yuenck@ukm.my (Yuen Chee Keong)
* Corresponding author

ISSN: 0128-7702

INTRODUCTION

The issue of text authenticity and suitability in testing has been constantly debated, which resulted in the construction of various formulae; for instance, Flesch Reading Ease (Green, Unaldi & Weir, 2010). Employing corpus linguistic tool to investigate the soundness of reading texts in an examination setting can contribute to the pool of knowledge not only in corpus linguistics study but also test design. It can also quantitatively study frequently occurring language patterns in reading texts which appear to be insufficient as highlighted by Biber et al. (2004) as cited in Green et al. (2010). The usage of corpus linguistics methodology should be able to shed light on the choice of passages for a reading test.

This present study attempts to explore shorter texts within the academic discourse. The reading passages of a national level English proficiency test, the Malaysian University English Test (henceforth MUET) has been chosen for this study. This is because test papers are certified instruments which have been validated and used to draw conclusions about readers' ability to read (Alderson, 1990). Text comprehension has always been the main concern of reading during examination. The general opinion pertaining to the definition of reading comprehension was synthesised by Masoud, Ramlee and Tengku Nor Rizan (2011, p. 98) as "the process of unlocking meaning from connected text". To achieve success in text comprehension, Ben-Anath (2005) asserted that the construction of a cognitive representation of information in the text must be coherent. Under limited time-frame, test-takers have no option but to try to make sense from the systematically structured lexis through frequently occurring patterns and to relate different parts of texts through cohesive devices. Cohesive devices function as connectors that hold between the text segments and ensure coherence in the text. This is proven by Mohamad Khatib and Mahmood Safari (2011) and Ben-Anath (2005) when they revealed the benefits readers gained from the presence of cohesive devices in reading texts.

Cohesive devices, specifically conjunction classified under grammatical devices share a few similarities with the main subject of this study, lexical bundles (LBs). Firstly, in terms of structural formation, LBs like *as well as* (adverbial phrase-AdvP) and *on the other hand* (prepositional phrase-PP) are also known as additives and adversatives conjunctions respectively. Secondly, one of the LB functional categories is specifically allocated for text organisation (e.g., *in contrast to*, *as a result of*), where it is normally used to introduce a topic, elaborate and make inference. Thirdly, Biber, Conrad and Cortes (2004) stated that past studies of cohesive devices normally focus on the usage of multi-word prefabricated expressions; likewise, studies on LBs commonly examine strings of words ranging from two to six. LBs are worth analysing because: a) multi-words assist communication by making language

more predictable to the hearer or reader (Nattinger & DeCarrico, 1992, as cited in Hyland, 2008); b) such pre-fabricated sequences help to signal the text register to readers and decrease retrieval and processing time (Wray & Perkins, 2000 as cited in Hyland, 2008) and; c) having sufficient knowledge of LBs enables learners to fully socialise in an academic setting (Cortes, 2004). Some students are incapable of integrating the meaning of separate words to arrive at the meaning of an entire sentence (Wise, 1999, as cited in Masoud et al., 2011) which eventually put them at the disadvantage because LBs are occasionally incorporated in the teaching of grammar and not explicitly taught.

To a certain extent, the structure and function of LBs resemble cohesive devices which are used to connect different parts of a text. The ability of several types of cohesive devices such as coordinating conjunctions (e.g., *and, but, because*), compound adverbs (e.g., *furthermore, nevertheless*) and continuatives (e.g., *anyway*) (Akindele, 2011) in facilitating text comprehension by reducing reading time and improving content recall is proven in several research findings quoted by Mohamad Khatib and Mahmood Safari (2011). In addition, Hyland (2008, p. 4) pointed out that, "bundles are not only central to the creation of academic discourse, but they offer an important means for differentiating written texts by discipline". Neely and Cortes (2009, p. 20) revealed that the knowledge of these chunks of language such as *that goes without saying*, *on the other hand, as it were* can "ease the problem of [listening] perception". Likewise, Hyland (2008, p. 5) claimed that "the extensive use of pre-fabricated sequences such as *it has been noted that* in academic written genres helps to signal the text register to readers and reduce processing time by using familiar patterns to link elements of new information." Therefore, the presence of familiar expressions in academic texts can empower learners to comprehend what they have read.

To date, few studies on the relationship between structural and functional categories of LBs have been investigated quantitatively because the qualitative method was usually preferred in the past. For example, Dontcheva-Navratilova (2012), Strunkyte and Jurkunaite (2008) and Biber et al. (2004) scrutinised structural and functional relationship by distinguishing the most frequently used structure(s) under each function; in other words, manual calculations were done. Thus, to strive for accuracy, the research questions of this study are as follows:

i) How frequent do the structural and functional categories of lexical bundles (LBs) occur in MUET reading texts?

ii) Is there any relationship between the structural and functional categories in MUET reading text passages among 5 different disciplines?

The last question has a set of alternate hypotheses and they are as follows:

i) Ha: There is an association between the structural and functional categories in MUET reading test passages among 5 different disciplines

   a. Ha 1: There is an association between the structural and functional categories in Applied Science.

   b. Ha 2: There is an association between the structural and functional categories in Pure Science.

   c. Ha 3: There is an association between the structural and functional categories in Business.

   d. Ha 4: There is an association between the structural and functional categories in Humanities.

   e. Ha 5: There is an association between the structural and functional categories in Social Science.

As past studies namely Dontcheva-Navratilova (2012), Strunkyte and Jurkunaite (2008) and Biber et al. (2004) had proven the relationship between structural and functional categories, the alternate hypothesis was chosen. Having done that, triangulation of data was observed. Prior to that, an overview of LBs and the two major categories was presented.

An Overview of Lexical Bundles

'Phraseology' and 'formulaic sequences/ language' are two terms used to refer to multi-word units including LBs (Chen & Baker, 2010). However, there are a few terminologies such as collocations and idioms which seem to create confusion when multi-word units are concerned.

Firstly, LBs appear to share similar characteristics with collocations as they are both made up of 2 to 6 words. According to Nesselhauf (2005, p. 25), collocations are "a type of word combination in a certain grammatical pattern and the term is used to refer to an abstract unit of language and its instantiations in texts." Menon and Mukundan (2012, p. 151) claimed that there are two types of collocations, namely: a) "lexical – two lexical elements such as in any combination with nouns, verbs, adjectives and adverbs co-occur and; b) grammatical – a lexical and a grammatical element such as preposition, article or a clause co-occur". Secondly, unlike idiomatic phrases such as *kick the bucket* and *a slap in the face* which are rarely used in natural speech or writing (Biber et al., 2004), LBs are semantically transparent and consistent, functioning as the building blocks of a coherent discourse (Hyland, 2008).

As Hyland (2008) observed, the difference between LBs and collocations, idioms or other multi-word units such as metaphors and proverbs is that LBs are identified purely on the basis of their frequency rather than their structures. Lin (2008) expressed similar view by

claiming LBs are distinguished mainly by the statistical features of high frequency and added that they are not constrained by features used to describe phraseological units. The most salient and defining characteristic of LB is frequency of occurrence as highlighted by many past researchers namely Dontcheva-Navratilova (2012), Lin (2008), Biber and Barbieri (2007) who investigated LBs in a variety of context. Biber et al. (2004, p. 376) echoed the point by stating "frequency data have additional importance for the study of LBs because they are reflection of the extent to which a sequence of words is stored and used as a prefabricated chunk". Thus, LBs in this study are not limited to collocations, idioms, phrasal verbs or any fixed prefabricated structures including cohesive devices; it takes into consideration any form of 2 to 6-words structures that occur 4 times and above in 3 or more texts.

Structural and Functional Types of Lexical Bundles

Regarded as the proponents of LBs, Biber et al. (2004) created a taxonomy encompassing both structural and functional types. According to Biber et al.'s (2004) taxonomy, LBs are divided into four major structural categories namely Noun Phrase-based, Verb Phrase-based, Prepositional Phrase-based and Dependent Clause fragments (NP-based, VP-based, PP-based and DC, henceforth). Under each category, it is broken down into several sub-categories. Thus, head words from similar parts of speech are grouped together. However, many researchers do not follow the categorisation rigidly. Lin (2008) and Bal (2010) simplified the taxonomy by grouping and analysing all phrasal structures such as NP + of phrase, PP + of phrase and so on together, followed by clausal structures consisted of adverbial clauses, verb / adjective + to clause. As for functional classification of LBs, there are three primary discourse functions namely stance expressions, discourse organisers and referential expressions (Biber et al., 2004; Biber & Barbieri, 2007). Each discourse function consists of sub-categories which are rather specific especially stance bundles; this is because their corpora were sampled from a large range of spoken and written activities associated with academic life. It appeared that dialogues extracted from classroom teaching, office hours, study groups and on-campus service counters (Biber et al., 2004) were included compared to the written register which justified the emphasis on stance bundles. Hyland (2008) therefore modified the previously-mentioned categorisation and introduced sub-categories which reflected particularly on research writing; they are research-oriented, text-oriented and participant-oriented. Dontcheva-Navratilova (2012) claimed that overlapping in the meanings and terms of both structural and functional categories used by Biber et al. (2004), Biber and Barbieri (2007) and Hyland (2008) is apparent. In other words, Biber et al. (2004) and Hyland (2008) taxonomies can be used interchangeably.

As exemplified in the three following studies, the structural and functional categories were prevalent when investigations on LBs were carried out. Examining a 3.5 million word corpus of research articles, PhD theses and Masters Dissertations in four disciplines, Hyland (2008) managed to identify 240 different 4-word bundles. The structures and functions of 4-word bundles were explored to learn the preferred form in certain disciplines (Hyland, 2008). In Strunkyte and Jurjunaite's (2008) comparative study, they revealed the commonly used structural types of LBs in humanities and natural science research articles. As for functional analysis, they found that the language of natural science research articles displayed greater precision in text structuring. Lin's (2008) research on LBs in electrical engineering introductory textbook and English for specific purposes textbook reported insufficient representations of target language in ESP textbooks. They were scrutinised based on the structural and functional categories of LBs.

In addition to the analysis of structural and functional categories, the relationship between LBs in both categories was identified. A strong relationship between structural type and discourse function of LBs was observed in Dontcheva-Navratilova (2012), Strunkyte and Jurkunaite (2008) and Biber et al. (2004) studies. All the researchers agreed that bundles incorporating NP and PP were common phrases used in the construction of referential (also known as research-oriented) bundles. They also claimed that most stance (also known as participant-oriented) bundles were made up of dependent clause fragment, specifically VP fragments. A consensus was not achieved for the structural forms of text-organising bundles but the bundles identified still fell within the structural categories proposed.

METHODOLOGY

A quantitative research design was adopted to study LBs by using different data sources. Data was firstly obtained from MUET reading test papers while the statistical data complemented the aforementioned results. The procedure for the analysis of LBs in MUET reading tests is described in the paragraphs below.

Firstly, MUET reading test papers since its commencement in 1999 which totalled to 111 reading texts were grouped into 5 disciplines as shown in Table 1. These disciplines were decided after skimming the passages which covered a variety of topics. Grouping the passages to their respective disciplines was a technique Hyland and Tse (2009), Jalali, Rasekh and Rizi (2008) and Strunkyte and Jurkunaite (2008) adopted in their studies to ensure a systematic and orderly analysis.

Table 1
Number of Texts and Word Counts according to Disciplines

Disciplines	MUET	
	Number of texts	Number of Words
Applied Science	38	4,975
Pure Science	12	2,292
Business	12	1,795
Humanities	15	2,758
Social Science	34	4,043
TOTAL	111	15,863

Secondly, using Word Smith Tools 5, the numbers of LBs in the MUET reading test corpus followed by frequency of occurrences were classified based on Biber et al. (2004) Structural Taxonomy and Hyland (2008) Functional Taxonomy. This can ensure uniformity when analysing the data.

To find out the relationship between both structural and functional categories, Chi-square test of independence and Fisher's exact test were adopted; hence, SPSS (version 20) was employed to perform the test. These statistical tools are able to provide precise result, unlike several previous studies namely Dontcheva-Navratilova (2012) and Strunkyte and Jurkunaite (2008) which merely grouped LBs incorporating certain structures according to their functions manually. Chi-square test of independence is deemed suitable because it is used to determine whether or not a relationship exists between two categorical variables namely structural and functional categories for this study. Values generated by this statistical tool indicated by x^2 value and p value refer to the strength of relationship and its significance, respectively. Following the rule of thumb, p value must be less than 0.01 to be considered significant. Two new categories termed other structural bundles and other functional bundles abbreviated as SO and FO respectively were included in this quantitative analysis. In other words, LBs which could not be grouped under any of the existing categories were taken into account. It must also be noted that frequency count of LBs was disregarded in this phase.

RESULTS

LBs which are known as word clusters in WST 5 were computed automatically in the main controller setting. A total of 15,863 words was generated first followed by the word cluster list ranked in descending order with a total of 1,359 bundles. The lists of LBs according to the 5 disciplines were generated after eliminating LBs which did not meet the cut-off frequency. A drastic decline was observed after categorising the LBs according to the 2 taxonomies where only 357 structural types of LBs and 354 functional types of LBs were identified as shown in Table 2.

Table 2
The number of structural and functional types of LBs

Structures	Sub-Categories
NP	NP + of (N=32)
	NP + post modifier (N=56)
	NP + be (N=12)
VP	Passive + PP (N=31)
	be +N/AdjP (N=25)
	it + V/AdjP (N=11)
PP	PP + of phrase (N=0)
	PP + NP (N=88)
Dependent Clause	VP + that clause (N=17)
	VP/Adj + to clause (N=55)
	Adverbial clause (N=30)

Functions	Sub-Categories
Research	Location (N=61)
	Procedure (N=36)
	Quantity (N=45)
	Description (N=65)
	Topic (N=30)
Text	Transition (N=29)
	Resultative (N=9)
	Structuring (N=11)
	Framing (N=22)
Participant	Stance (N=36)
	Engagement (N=10)

Figure 1 depicts the overall frequency of structural types of LBs found conforming to Biber et al. (2004) Taxonomy in the five respective disciplines. In general, PP-based LBs seemed to be dominating MUET reading texts, irrespective of disciplines. Dependent clause-based and NP-based LBs were found in all disciplines but the numbers were imbalanced where they were fewer in Business, Humanities and Pure Science texts. It can be seen that except LBs incorporating PP which recorded a high frequency count, other phrases and clause-based LBs occurred less than 30% in all the texts.

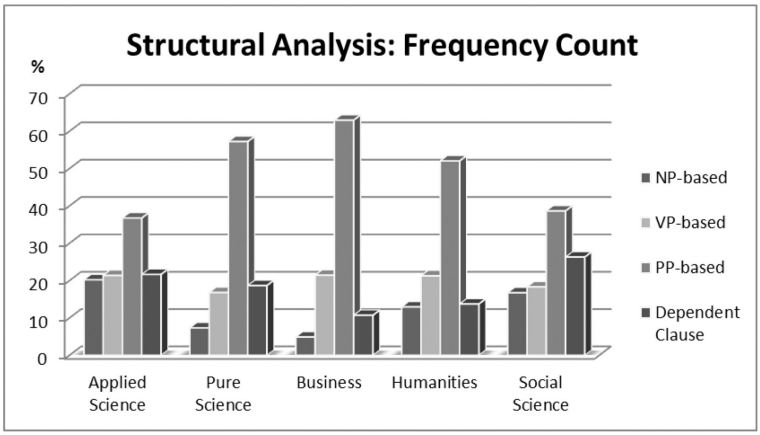

Figure 1. Frequency Count of LBs identified for Structural Analysis in Percentages

As shown in Figure 2, it is apparent that research-oriented LBs with more than half of all the bundles in the corpus were widely used regardless of disciplines. With a range of frequency from 20% to 30%, text-oriented bundles came in second. This distribution echoed Biber's claim as cited in Dontcheva-Navratilova (2012) that in academic discourse, almost 70% of the most commonly used LBs performing the function of text organisers and research expressions. The least favoured function of LBs employed in MUET reading text corpus was participant-oriented bundles because they were usually used in highly interactional discourse of conversation (Dontcheva-Navratilova, 2012).

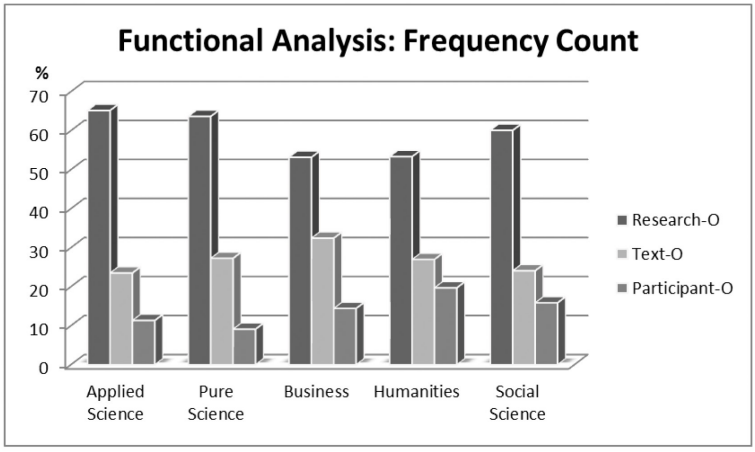

Figure 2. Frequency Count of LBs identified for Functional Analysis in Percentages

The use of an array of LBs showed that there were not only differences in disciplinary distributions (Hyland, 2008), but also within a single discipline. It can be seen that authors of the texts were attempting to convey the intended meaning through suitable language patterns. However, there were two hindrances that may affect the second phase of the analysis, context in which the LBs belonged to can make the sub-categorisation rather ambiguous and LBs possessing multiple functions could not be fitted into any of the sub-category.

Two different types of association techniques were employed due to the number of LBs identified. Apart from using Chi-square to examine the association between the structural categories and functional categories, Fisher's exact test was adopted too. Fisher's exact test is more accurate than the chi-square test when the expected numbers are small (McDonald, 2009). Hence, Fisher's exact test was used instead of chi-squared test in cases where the number of phrases identified was too few.

Table 3
Percentages of Structural Components (NP, VP, PP, DC, SO) in each of the Functional Components (RO, TO, PO, FO) – Overview of each Discipline

	NP		VP		PP		DC		SO		Result
	n	%	n	%	n	%	n	%	n	%	
Applied Science											
RO	53	30.5	13	7.5	41	23.6	46	26.4	21	12.1	82.374 [a] **
TO	9	15.5	6	10.3	12	20.7	15	25.9	16	27.6	
PO	5	20.8	14	58.3	0	0	2	8.3	3	12.5	
FO	4	5.6	23	32.4	14	19.7	10	14.1	20	28.2	
Pure Science											
RO	5	27.8	0	0	8	44.4	4	22.2	1	5.6	24.075 [b] *
TO	0	0	2	20	2	20	3	30	3	30	
PO	0	0	3	75	0	0	1	25	0	0	
FO	0	0	5	31.3	8	50	3	18.8	0	0	
Business											
RO	1	11.1	0	0	5	55.6	2	22.2	1	11.1	19.670 [b] *
TO	1	16.7	1	16.7	0	0	1	16.7	3	50	
PO	0	0	2	100	0	0	0	0	0	0	
FO	0	0	3	27.3	7	63.6	1	9.1	0	0	
Humanity											
RO	5	15.2	0	0	13	39.4	10	30.3	5	15.2	41.527 [b] **
TO	3	27.3	1	9.1	2	18.2	1	9.1	4	36.4	
PO	4	50	3	37.5	0	0	0	0	1	12.5	
FO	0	0	9	40.9	11	50	1	4.5	1	4.5	
Social Science											
RO	33	28.2	1	9	29	24.8	37	31.6	17	14.5	83.942 [a] **
TO	5	11.1	3	6.7	9	20	10	22.2	18	40	
PO	7	23.3	13	43.3	0	0	7	23.3	3	10	
FO	2	3.7	16	29.6	13	24.1	8	14.8	15	27.8	

Note.

[a]Chi-square test for independence.

[b]Fisher's exact test.

*p<.01; **p<.001.

Table 3 depicts the association between the structural categories (NP vs. VP vs. PP vs. DC vs. SO) and functional categories (RO vs. TO vs. PO vs. FO) for each of the discipline – Applied Science, Pure Science, Business, Humanity, and Social Science. SO and FO are two additional categories encompassing various structures and functions of LBs not listed in Biber et al.'s (2004) Structural Taxonomy and Hyland's

(2008) Functional Taxonomy. The associations identified in each discipline are explained thoroughly in the following paragraphs.

In Applied Science-based text, the chi-square test of independence showed a significant association between structural categories and functional categories, x^2 (447) = 82.374, p < .001. Research-oriented bundles were found significantly associated with LBs incorporating NP followed by DC, PP, SO, and the least used phrase was VP (30.5% vs. 26.4% vs. 23.6% vs. 12.1% vs. 7.5%). Next, text-oriented bundles were found significantly associated with LBs incorporating SO followed by DC, PP, NP, and the least used phrase was VP (27.6% vs. 25.9% vs. 20.7% vs. 15.5% vs. 10.3%). As for participant-oriented bundles, they were found significantly associated with LBs incorporating VP followed by NP, SO, DC, and no PP was used (58.3% vs. 20.8% vs. 12.5% vs. 8.3% vs. 0%). The additional functional categories made up of LB with other functions were found significantly associated with LBs incorporating VP, SO, PP, DC, and the least used phrase was NP (32.4% vs. 28.2% vs. 19.7% vs. 14.1% vs. 5.6%).

In Pure Science-based text, Fisher's exact test depicted a significant association between structural categories and functional categories (p = .002; 24.075). Research-oriented bundles were found significantly associated with LBs incorporating PP followed by NP, DC, SO, and no VP was used (44.4% vs. 27.8% vs. 22.2% vs. 5.6% vs. 0%). Next, text-oriented bundles were found significantly associated with LBs incorporating DC and SO followed by VP and PP, while no NP was used (30% vs. 30% vs. 20% vs. 20% vs. 0%). As for participant-oriented bundles, they were found significantly associated with LBs incorporating VP and DC while NP-based, PP-based and other structures of LBs were not used (75% vs. 25% vs. 0% vs. 0% vs. 0%). The additional functional categories made up of LB with other functions were found significantly associated with LBs incorporating PP, VP and DC while no NP and other structural form of LBs were used (50% vs. 31.3% vs. 18.8% vs. 0% vs. 0%).

In Business-based text, Fisher's exact test depicted a significant association between structural categories and functional categories (p = .007; 19.670). Research-oriented bundles were found significantly associated with LBs incorporating PP followed by DC, NP, SO, and no VP was used (55.6% vs. 22.2% vs. 11.1% vs. 11.1% vs. 0%). Next, text-oriented bundles were found significantly associated with LBs incorporating SO followed by NP, VP and DC, while no PP was used (50% vs. 16.7% vs. 16.7% vs. 16.7% vs. 0%). As for participant-oriented bundles, they were found significantly associated with LBs incorporating VP while NP-based, PP-based, DC-based and other structures of LBs were not used (100% vs. 0% vs. 0% vs. 0% vs. 0%). The additional functional categories made up of LB with other functions were found significantly associated with LBs incorporating PP, VP

and DC, while no NP and other structural form of LBs were used (63.6% vs. 27.3% vs. 9.1% vs. 0% vs. 0%).

In Humanities-based text, Fisher's exact test depicted a significant association between structural categories and functional categories (p = .000; 41.527). Research-oriented bundles were found significantly associated with LBs incorporating PP followed by DC, NP, SO, and no VP was used (39.4% vs. 30.3% vs. 15.2% vs. 15.2% vs. 0%). Next, text-oriented bundles were found significantly associated with LBs incorporating SO followed by NP, PP, VP and DC (36.4% vs. 27.3% vs. 18.2% vs. 9.1% vs. 9.1%). As for the participant-oriented bundles, they were found significantly associated with LBs incorporating NP, followed by VP and SO while PP-based and DC-based LBs were not used (50% vs. 37.5% vs. 12.5% vs. 0% vs. 0%). The additional functional categories made up of LB with other functions were found significantly associated with LBs incorporating PP, VP and DC, while no NP and other structural form of LBs were used (50% vs. 40.9% vs. 4.5% vs. 4.5% vs. 0%).

In Social Science-based text, the chi-square test of independence showed a significant association between structural categories and functional categories, x^2 (447) = 83.942, p < .001. Research-oriented bundles were found significantly associated with LBs incorporating DC followed by NP, PP, SO, and the least used phrase was VP (31.6% vs. 28.2% vs. 24.8% vs. 14.5% vs. 9%). LBs incorporating dependent clause were relatively common in performing research-oriented function. This was probably because majority of the dependent clauses consisted of WH-clause and that-clause where they were used to give new or more information. Next, text-oriented bundles were found significantly associated with LBs incorporating SO followed by DC, PP, NP, and the least used phrase was VP (40% vs. 22.2% vs. 20% vs. 11.1% vs. 6.7%). As for participant-oriented bundles, they were found significantly associated with LBs incorporating VP followed by NP, DC, SO and no PP was used (43.3% vs. 23.3% vs. 23.3% vs. 10% vs. 0%). The additional functional categories made up of LB with other functions were found significantly associated with LBs incorporating VP, SO, PP, DC, and the least used phrase was NP (29.6% vs. 27.8% vs. 24.1% vs. 14.8% vs. 3.7%).

In short, an association between the structural and functional categories in MUET reading test passages among 5 different disciplines was therefore not rejected. LBs incorporating NPs usually performed research-oriented function whereas LBs incorporating DCs were strongly bound to text-oriented function. LBs incorporating VPs on the contrary were linked to participant-oriented functions as well as other types of functions. Not only significant association was identified between the categories in all the disciplines, but also the two additional categories encompassing various structures of LBs and other functions not listed in Biber

et al.'s (2004) Structural Taxonomy and Hyland's (2008) Functional Taxonomy. It can be concluded that structural categories and discourse functions are closely interrelated.

DISCUSSION

Research-oriented bundles were composed of mostly NP, followed by DC, PP, other structures, while the least used phrase was VP. The high percentage of NP in research-oriented function may be due to the abundantly used noun phrases to describe location and quantity (for example, *the future, the end of, the United States* and *the region,* while *a single, thousands of, the rise* are examples of LBs referring to quantity). The richness of NP in the scientific field has been proven by many including Menon and Mukundan (2012) who found that noun/noun and noun/adjective syntactic combinations were predominant in a science textbook corpus. Moreover, Hyland (2008) stated that noun phrases + of structures were prominent in research-oriented functions. A year later, Hyland and Tse (2009) reported a heavier use of research-oriented bundles in Science and Engineering texts. LBs incorporating DC and PP were relatively common in performing research-oriented function. This was probably because majority of the dependent clauses consisted of WH-clause and that-clause where they were used to give new or more information. With regard to the association between PP and research-oriented function, similar result was reported in the studies by Strunkyte and Jurkunaite (2008) and Dontcheva-Navratilova (2012).

Text-oriented bundles were realised in all four structural types and other structures. The most common bundles were those incorporating other structures and DC; bundles incorporating PP came in third, followed by NP and VP. As most dependent clauses were made up of transition signals such as *because they, as well as*, and *but it*, their roles in ensuring cohesion and coherence of the reading passages were apparent. Interestingly, most LBs performing text-oriented functions were found headed by conjunctions such as *and, or, but* that could not be grouped under any structural categories in Biber et al.'s (2004) Taxonomy. Thus, LBs headed by conjunction-*and* which were classified under other structural category may be the reason of the high percentage generated.

It is worth noting that Lin (2008), Hyland (2008), Biber and Barbieri (2007), Biber, et al. (2004) stressed that participant-oriented bundles are not significant in written form but they are rather prominent in a wide range of spoken registers. Participant oriented bundles were the only functional category that did not incorporate any PP. They were made up of mainly VP and NP while the least structural types incorporated were DC and other structures. The strong association between VP and participant functions could be attributed to the extensive use of anticipatory *it* in stance bundles. Moreover, LBs incorporating VP headed by modal verbs, for instance, *may*

have, might be and *should be* were also found in stance bundles. The probable reason for NP being equally prominent in participant-oriented functions was due to almost all LBs indicating engagement were NP-based.

Other functional bundles, similar to research and text-oriented bundles were realised in all five structural types. They consisted of mainly VP, other structural types, PP, DC and a few NPs. A similar trend was observed in all five disciplines; LBs incorporating VPs were usually linked to describing functions other than research, text and participant. The high percentage of LBs incorporating VP, PP and DC in performing functions other than the listed ones may indicate that the existing functional categories proposed by Hyland (2008) Functional Taxonomy were inadequate. Moreover, the strong association between other structural types and other functional categories signify that more explorations need to be done.

CONCLUSION

Focusing on the structures and functions of LBs, this study managed to shed light on commonly used LBs in MUET reading texts. The language pattern in MUET reading text corpus was relatively rich in sequences of words as depicted by the range of structures and functions of LBs identified. The results of this study along with some other previous studies, namely, Jalali et al. (2008), Hyland (2008), Strunkyte and Jurkunaite (2008), indicated that different LBs were employed in different disciplines and genres. At the same time the categories were found to be significantly associated in all the five disciplines despite the limitation of small corpus built.

Data of this study (reading passage) showed that many different topics were grouped under one discipline, which means the presence of new LBs in unfamiliar texts may impede students' comprehensions. Hock as cited in Alderson (2005) found that comprehension of a discipline-related text was relevant to knowledge of the subject area and level of proficiency. It was reinforced by Alderson (2005, p. 103) who asserted that, "subject-related texts might discriminate against individuals who happen to possess less background knowledge in a particular field". In order to grasp the gist of a text, learners must at least familiarize themselves with the repertoire of that particular context. For instance, the occurrence of NP in science-based texts in this study is considerably higher compared to other forms of LBs. Menon and Mukundan (2012) reported a similar result in their collocational study. Hence, students majoring in science and technology must be equipped with knowledge of NP structure. On the contrary, arts-based texts depicted a higher usage of participant-oriented bundles in this study; thus, these students have to be constantly exposed to LBs used for expressing personal opinion and engaging readers.

The field in which a text is adopted influences text comprehension because every discipline is believed to possess a

list of regularly used LBs. Although results from this study showed many identical LBs were used in five different disciplines, the association between their functions and structural types differed greatly across the disciplines. For instance, research-oriented bundles in Applied Science texts were significantly associated with NP compared to dependent clause in Social Science-based texts and PP in the remaining disciplines. These results echoed Hyland and Tse (2009), Hyland (2008), Jalali et al. (2008) and Cortes (2004) studies where LBs occurred and behaved differently in diverse disciplinary environments. Unlike previous studies, the association between the two mentioned categories was identified quantitatively by using Chi-square and Fisher's test; this signified higher accuracy compared to results obtained from manual calculations.

Having to sit for a reading proficiency test where the texts adopted deal with the area of their studies may reduce students' level of anxiety. Knowing the structures and functions of frequently occurring LBs in five different disciplines as highlighted in this study could assist the test-takers in comprehending passages dealing with prominent topics. Furthermore, preparation of pre-university students for tertiary education could be done effectively with more focus being placed on a specific area. As such, the main objective of MUET, which is to measure the English language proficiency of pre-university students for entry into tertiary education, could be achieved more reliably.

REFERENCES

Akindele, J. (2011). Cohesive devices in selected ESL academic papers. *African Nebula,* (3), 99-112. Retrieved from nobleworld.biz/images/Akindele_AN3.pdf

Alderson, J. C. (1990). Testing reading comprehension skills (part one). *Reading in a Foreign Language, 6*(2), 425-438. Retrieved from http://nflrc.hawaii.edu/rfl/PastIssues/rfl62anderson.pdf

Alderson, J. C. (2005). *Assessing reading*. United Kingdom: Cambridge University Press.

Bal, B. (2010). *Analysis of four-word lexical bundles in published research articles written by Turkish scholars*. (Unpublished Master Thesis). Georgia State University. Retrieved from http://digitalarchive.gsu.edu/cgi/viewcontent.cgi?article=1001&context=alesl_theses

Ben-Anath, D. (2005). The role of connectives in text comprehension. *Teachers College, Columbia University Working Papers in TESOL and Applied Linguistics, 5*(2), 1-27. Retrieved from http://journals.tc-library.org/index.php/tesol/article/download/84/87

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

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., Conrad, S., Reppen, R., Byrd, P., Helt, M., Clark, V., ... Urzua, A. (2004). *Representing Language Use in the University: Analysis of the TOEFL 2000 Spoken and Written Academic Language Corpus*. (ETS TOEFL Monograph Series, MS-25). Princeton, NJ: Educational Testing Service.

Chen, Y. H., & Baker, P. (2010). Lexical bundles in L1 and L2 academic writing. *Language Learning and Technology, 14*(2), 30–49. Retrieved from http://llt.msu.edu/vol14num2/chenbaker.pdf

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

Dontcheva-Navratilova, O. (2012). Lexical bundles in academic texts by non-native speakers. *Brno Studies in English, 38*(2), 37-58. Retrieved from http://www.phil.muni.cz/plonedata/wkaa/BSE/BSE_2012/BSE_2012-38_2_-XX_Dontcheva-Navratilova.pdf

Green, A., Unaldi, A., & Weir, C. (2010). Empiricism versus connoisseurship: Establishing the appropriacy of texts in tests of academic reading. *Language Testing, 27*(2), 191–211.

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

Hyland, K., & Tse, P. (2009). Academic lexis and disciplinary practice: Corpus evidence for specificity. *International Journal of English Studies (IJES), 9*(2), 111-129.

Jalali, H., Rasekh, A. E., & Rizi, M. T. (2008). Lexical bundles and intradisciplinary variation: The case of Applied Linguistics. *Iranian Journal of Language Studies (IJLS), 2*(4), 447-484. Retrieved from http://core.kmi.open.ac.uk/download/pdf/1053500

Khatib, M., & Safari, M. (2011). Comprehension of discourse markers and reading comprehension. *English Language Teaching, 4*(3), 243-250. Retrieved from http://www.ccsenet.org/journal/index.php/elt/article/view/11899/8357

Lin, C. (2008). *An investigation of lexical bundles in electrical engineering introductory textbooks and ESP textbooks*. (Master Thesis). Carleton University.

McDonald, J. H. (2009). *Handbook of biological statistics* (2nd Ed.). Baltimore: Sparky House Publishing.

Menon, S., & Mukundan, J. (2012). Collocations of high frequency noun keywords in prescribed science textbooks. *International Education Studies, 5*(6), 149-160. http://dx.doi.org/10.5539/ies.v5n6p149

Nattinger, J., & DeCarrico, J. (1992). *Lexical phrases and language teaching*. Oxford: OUP.

Neely, E., & Cortes, C. (2009). A little bit about: Analyzing and teaching lexical bundles in academic lectures. *Language Value, 1*(1), 17-38.

Nesselhauf, N. (2005). *Collocations in a learner corpus*. Philadelphia: John Benjamins.

Strunkyte, G., & Jurkūnaite, E. (2008). *Written academic discourse: Lexical bundles in humanities and natural sciences*. (Bachelor Thesis). Vilnius University.

Wise, B. W. (1999). The promise and limits of phonological training for children with specific reading disabilities. *SIG 1 Perspectives on Language Learning and Education, 6*(1), 22-25.

Wray, A., & Perkins, M. R. (2000). The functions of formulaic language: An integrated model. *Language and Communication, 20*(1), 1-28.

Zoghi, M., Mustapha, R., & Mohamad, T. N. R. B. T. (2011). Getting to know L2 poor comprehenders. *English Language Teaching, 4*(1), 98-104.