

Functional Types of Lexical Bundles in Reading Texts of Malaysian University English Test: A Corpus Study

Christina Ong Sook Beng
ongsb@utar.edu.my
Universiti Tunku Abdul Rahman

Yuen Chee Keong
yuenck@ukm.edu.my
Universiti Kebangsaan Malaysia

ABSTRACT

It is widely claimed that many university students were found to be unprepared for the reading demands placed upon them. To provide students with an understanding of the features of the discourses they may encounter, this study investigates lexical bundles (LBs) used in the reading passages of Malaysian University English Test (MUET), a compulsory entrance examination for tertiary education. More specifically it aims to compare and contrast the functional types of LBs found in arts and science-based reading passages. A specialised corpus of MUET test papers made up of only the reading passages categorised into two main traditional disciplines; arts and science was built using WordSmith Tools Version 5. The lists of LBs of the identified disciplines were generated using WordSmith Tools Version 5. The generated data was then analysed qualitatively based on Hyland's (2008) Functional Taxonomy. Findings revealed that although the number of LBs in both disciplines shows significant difference, many similar LBs are found. Science-based passages employ more research-based LBs whereas arts-based texts employ more participant-oriented LBs. These findings seemed to confirm that the functions of LBs are specific to particular disciplines. Hence, explicit teaching of LBs should be considered in schools and two separate sets of MUET reading tests where the first comprises of texts adopted from the scientific and technological context while the other made up of texts from the social sciences should be designed to accommodate both arts and science stream pre-university students.

Keywords: lexical bundles (LBs); functional types; reading text; MUET; reading test

INTRODUCTION

Malaysian University English Test (henceforth MUET) is an English proficiency test taken by thousands of Malaysians annually. It is not only used as a stepping stone to tertiary education (Malaysian Examination Council, 2008) but also as a means to determine university graduates' language proficiency when they enter the workforce (Lee, 2004). All four language skills namely listening, speaking, reading and writing are tested in MUET where it measures the overall English Language proficiency of candidates in the cumulative score of the four language skills in a single Band Score – ranging from the lowest, Band 1 to the highest, Band 6 (Naginder & Rohayah, 2006).

In this study, only the reading component is analysed because “reading is clearly primary to any definition of literacy” (Venezky as cited in Lee, 2004, p. 2) which is reflected in the 40% weightage of the total marks allocated to the reading test in MUET. One of the reasons for this heavier weightage is because university students are expected to read more than to write (Lee, 2004). In addition, students need to be academically literate to succeed at universities since a major part of university education involves reading where acquisition of knowledge depends on text materials (Zuraidah, 2003). Consequently, the aim of MUET

reading syllabus is to prepare pre-university students for academic study at university where their sources of new knowledge will come from textbooks, references, journal articles and periodicals. Thus, the MUET reading test is designed to help students make the transition from ESL reading to reading for academic purposes (Zuraidah, 2003).

The Malaysian Examination Council (2008), which is responsible for conducting the test, has identified the following features of the passages used in MUET:

- i) Basic criteria for text selection:
Length (200 – 700 words), level of complexity (content and language), text type
- ii) Possible genres:
Articles from journals, newspapers and magazines, academic texts, electronic texts
- iii) Rhetorical style:
Analytical, descriptive, persuasive, argumentative, narrative

Nevertheless, MUET may not contain all the suggested genres in academic studies because the texts in MUET are sourced from a variety of reading materials and highly specialised texts are not selected (Zuraidah & David, 2005).

In general, MUET has been developed in response to particular needs and requirements which emerged with the unfolding globalization developments in the education scene “emphasising on formal linguistics knowledge and skills development” (Sarah Abedi, 2004, p. 103). Like any other admission test, MUET is a criterion-referenced test used to “ensure that individuals possess at least certain minimum knowledge or skills before they are allowed to engage in certain activities and to ensure the candidate’s ability meet specified minimum standards” (Klein, 1990, p. 6).

MUET is selected for investigation in this study for three reasons. Firstly, the reading test compared to the combined tests of the other three skills; writing, speaking and listening in MUET, carries 40% of the total marks which means it is a significantly more important skill than that of the rest. Secondly, MUET test papers have been validated; thus it can be used as certified instruments to draw conclusion about learners’ ability to read (Alderson, 1990). Thirdly, reading tests are expected to capture a range of reading processes and skills within a limited time frame, at the same time avoiding unintended biasness (Green, Unaldi & Weir, 2010, p.192).

The urgency in identifying and analysing commonly used LBs in reading tests is because many Malaysian undergraduates are still not prepared for the reading demands expected of them (Noorizah, 2006; Nambiar, 2007). A major part of university education involves reading (Zuraidah, 2003) and undergraduates have to comprehend texts beyond the printed words. They need to understand the relationship between form and function, the difference between locutionary expression and illocutionary force and the importance of language and culture (Kern, 2000). This is because one of the measures of their academic ability is their maturity in comprehending texts presented to them.

LITERATURE REVIEW

LBs are “sequences of two or more words frequently occurring in a particular setting” (Biber, Conrad & Cortes, 2004, p. 376). The most salient and defining characteristic of LB is frequency of occurrence. This feature has been highlighted by many researchers who investigated LBs in a variety of context (Dontcheva-Navratilova, 2012; Lin, 2008; Biber & Barbieri, 2007). 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”.

In one study, it was revealed that average and poor students use far less LBs (Read & Nation, 2002) in formal examinations. Read and Nation (2012) highlighted that Band 8
ISSN: 1675-8021

achievers of the International English Language Testing Systems (IELTS) use substantially more LBs in the speaking test as compared to those who obtain lower bands. This limited usage could be due to LBs not being taught formally.

Further studies on LBs in the speaking test of MUET by Lourdunathan and Menon (2005), Hafiza Aini and Hadina (2008), and Nazira and Kamaruzaman (2009) experimented on the effectiveness of repeating interaction utterances in the forms of lexical phrases and multi-word items to Malaysian students. Weak students reported that the phrases had helped to boost their confidence and maintain the flow of the discussion (Nazira & Kamaruzaman, 2009). Similarly, the teaching of interaction strategies together with useful phrases had helped learners to produce good results (Lourdunathan & Menon, 2005) while the use of the audio-lingual method in the teaching of discussion skills using repetition of short utterances prepared by teachers had helped the learners too (Hafiza Aini & Hadina, 2008).

Given the positive effect of LBs in speaking tests as indicated in the above studies, there is a need to investigate the use of LBs in reading texts. With regard to the MUET reading test, several approaches to improve students' performances have been suggested by past researchers. Firstly, the four reciprocal teaching strategies (predicting, questioning, summarising and clarifying) in the form of a dialogue between teachers and students during reading class were introduced to 68 low-proficiency level sixth form students from a Malaysian school (Tan Ooi, Tan & Norlida, 2011). Secondly, the thematic approach which aimed at broadening the students' schemata were proposed by Naginder (2004) and Naginder and Rohayah Nordin (2006). They believed, by drawing up a list of prominent themes frequently used in the MUET exam, students would be able to cope because they had acquired fundamental linguistic terms or meta-language.

Moreover, LBs in MUET reading texts are worth investigating because most past studies on MUET have been focusing on two aspects namely ways to improve test scores and approaches to prepare test-takers. This could be because MUET is a potentially high-stake test since it is a requirement for entry into a Malaysian public university (Lee, 2004).

The function of LBs in academia generally needs to be looked into after understanding the context of this study as explained earlier. Multi-word expressions in academic prose according to Biber et al. (2004), often served to bridge two phrases, using for example prepositional phrase (*in the case of*) or a noun phrase (*the base of the*). In other words, they function like scaffolding for new information (Biber & Barbieri, 2007). Hence, many researchers such as Hyland (2008), Neely and Cortes (2009), and Hyland and Tse (2009) claimed that LBs are agents that determine the familiarity of language users in a particular discourse.

With regard to functional classification of LBs, three primary discourse functions can be distinguished for LBs: (1) stance expressions, (2) discourse organisers, and (3) referential expressions (Biber & Barbieri, 2007; Biber et al., 2004). Stance bundles are often used to express a writer's evaluation of a proposition in terms of certainty or uncertainty; discourse organisers are used to structure texts where they can introduce, elaborate or make inference about a topic; and referential expressions are characterized by the function of attribute specification which are used to specify a given attribute or condition.

In addition, each discourse function in Biber and Barbieri's (2007), and Biber et al.'s (2004) LB classification consists of sub-categories which are rather specific especially stance bundle because their corpora were sampled from a large range of spoken and written activities associated with academic life. They included dialogues extracted from teaching in the classroom, office hours, study groups and on-campus service counters (Biber et al. 2004) as well as written data.

Subsequently, Hyland (2008) modified the classification by introducing sub-categories which reflected on research writing; they were research-oriented, text-oriented and

participant-oriented. Research-oriented LBs usually help writers to structure their activities and experiences of the real world; text-oriented LBs are concerned with the organization of the text and its meaning as a message or argument; participant-oriented LBs are focused on the writer or reader of the text. Dontcheva-Navratilova (2012) claimed that there was 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). Hence, they can be used interchangeably.

In view of the issue on reading difficulty among university students and benefits students can reap if they have a sound knowledge of LBs, this study aims to:

- i) investigate LBs that are commonly used in MUET
- ii) identify the functional categories of LBs in MUET
- iii) compare and contrast functional types of LBs in arts and science-based texts.

METHODOLOGY

All MUET reading test papers which amounted to 22 reading tests (since its commencement in 1999) had been purchased for this study. Except for the cloze test (which was part of the 1999 to mid-2008 MUET reading test), all the passages were used. A specialised corpus of MUET test papers made up of only the reading passages categorised into two main disciplines namely arts and science was built.

TABLE 1. Examples of Fields in Science and Arts-based Disciplines

Science	Arts
Agriculture	Entrepreneurship
Architecture and Design	International trade
Environmental Studies	History
Physics	Literature
Earth Sciences	Performing Arts
Chemistry	Gender Studies
Life Sciences	Geography

These disciplines were decided after skimming the passages which covered a variety of topics. To avoid discrepancy, this classification was done according to the listing by Tutor Gig Encyclopedia. Grouping the passages into their respective disciplines was a technique Hyland and Tse (2009) and Strunkyte and Jurkunaite (2008) adopted in their studies which ensured a systematic and orderly analysis. The size of the finalised corpus for investigation is as follows:

TABLE 2. Constituents of the MUET Reading Test Corpus

Corpus	Word Count	Number of Texts
MUET	15,606	111

The creation of this corpus database underwent the 3-stage process adapted from Bahiyah, et al., (2008) research design. The process is as follows:

1. *Digitization stage.* The MUET test papers in the form of books were scanned so that they are in digital form.
2. *Format conversion stage.* After scanning, a .jpeg format of the books was produced. This format was then converted to word document file and later into text files. Cleaning of the raw data was done to remove unnecessary items like graphics, instructions and multiple choice questions so that the .txt file contained only text data. The .txt files containing only reading passages were then manually adjusted again to ensure accuracy and consistency.

3. *Merging stage.* During the scanning process, for example, 1 MUET reading test paper was split into separate files where 4 passages were saved in 4 different .txt files. These different files were catalogued and merged; only then the data was ready for phase 2.

Using Scott's (2011) WordSmith Tools (WST) version 5.0, the numbers of LBs alongside their frequencies in the MUET reading test corpus were generated within the science and arts disciplines. WST is a software which allows three types of analysis namely generating wordlist, displaying the concordance of selected words and identifying keywords. However, only the first two were used in this study. The WordList tool was used as it can generate a list of all the words and word-clusters (two or more words) in a text, set out in alphabetical or frequency order (Scott, 2011). The parameters of LBs within which this study operated on are:

- i) *Cut-off frequency*

The normalised frequency threshold for large written corpora generally ranges from 20 to 40 per million words while a raw cut-off frequency for smaller corpora is from 2 to 10. The latter was adhered to because of the size of the corpus.

- ii) *Occurrence of combinations*

Combinations have to occur in at least 3-5 texts or 10% of texts to avoid idiosyncrasies from individual writers.

- iii) *Length of word combinations*

The length of word combination is usually 2-, 3-, 4-, 5-, 6-word units. However, 4-word sequences are found to be the most researched, due to the manageable size for manual categorization.

Instead of focusing on 4-word bundles as suggested by many, this study identified all bundles ranging from 2 to 6 as the corpus compiled was relatively small. Previous studies on LBs conducted by Hyland (2008), Hyland and Tse (2009) and Chen and Baker (2010), employed the size of corpora of at least one million words. As such, all 2 to 6 word bundles were taken into consideration to ensure substantial data was obtained for analysis in this study. To limit the scope, the cut-off frequency was set at 4 times with the occurrence in at least 3 texts. Thus, an "optimum" number of bundles (approximately 25 times per million words on average) as stated by Chen and Baker (2010, p. 32) was regarded as sufficient to represent the corpus being examined.

Although Biber et al. (2004) and Hyland (2008) functional taxonomies can be used interchangeably, Hyland's Taxonomy was adopted for this study instead. This is because the functional categories proposed in the taxonomy of Biber et al. (2004) dealt with conversation, classroom teaching, textbooks, and academic prose. Hyland (2008) explained that Biber et al.'s (2004) taxonomy yielded more personal, referential, and directive bundles compared to his which was more appropriate for research-focused genres. Since this study revolved around reading passages for academic purposes albeit slightly different from research articles in terms of length and organisation, Hyland's Taxonomy was considered to be more appropriate. The LBs were classified based on Hyland's (2008) Functional Taxonomy made up of three categories namely research-oriented, text-oriented and participant-oriented which are further divided into several sub-categories each.

Finally, analysis of LBs in relation to their functions was conducted. Each sub-category was examined thoroughly alongside examples extracted from the corpus to validate the analysis.

FINDINGS

A total of 15,863 words were generated first followed by the word cluster list ranked in descending order with a total of 1,359 bundles. Five word bundles did not meet the cut-off frequency, hence, they were eliminated; only 730 LBs consisting of 2, 3 and 4-word bundles were analysed. To meet the first research objective, a ranking of ten most frequently occurring LBs in the MUET reading test corpus is presented in the following table.

TABLE 3. Most frequent 2-word, 3-word and 4-word bundles in MUET reading text corpus

2-Word	Freq.	3-Word	Freq.	4-Word	Freq.
of the	388	one of the	22	in the United States	13
in the	296	the united states	20	at the University of	10
to the	138	per cent of	18	the end of the	10
it is	130	part of the	17	on the other hand	9
to be	113	be able to	16	will be able to	8
on the	102	more likely to	16	are more likely to	7
and the	99	in the united	15	as well as the	5
for the	91	it is a	15	for the first time	5
of a	84	in the world	13	a wide range of	4
in a	82	some of the	12	by the end of	4

Prior to examining the functional types of LBs, categorisation was made based on careful matching of all LBs identified with the sub-category terms used by Hyland (2008). LBs in which the structures did not correspond with meaning of any categories in the taxonomy were automatically disregarded. The numbers of occurrences for each sub-category alongside examples of functional LBs portrayed in the following table address the second research objective.

TABLE 4. Number and Example of Functional LBs in MUET Reading Passages

Functions	Sub-Categories	Examples
Research	Location (N=61)	a year, about a, the end of the, the mid, the world
	Procedure (N=36)	can be used, done in, how to, made from, to find
	Quantity (N=45)	growing number of, half a, increase in, million people
	Description (N=65)	for example the, form of, known as, types of, when a
Text	Topic (N=30)	of agriculture, of energy, the disease, the government
	Transition (N=29)	and it, not only, rather than, but also, in contrast
	Resultative (N=9)	because they, caused by a, led to, seems to, so that
	Structuring (N=11)	the next, the other, there are, the first, and then
Participant	Framing (N=22)	according to the, if you, parts of, people in, since the
	Stance (N=36)	it could, might be, researchers are, the possibility
	Engagement (N=10)	we have, you can, those who, other people, we are

The five most frequently occurring functions of LBs are LBs indicating description, location, quantity, stance and procedure in different contexts can be observed from the following excerpts and concordance lines which are extracted from the corpus.

a) Description

- i. To promote accuracy in children's testimony, the questioning of children should be done by neutral parties rather than by individuals **who are** biased either towards or against believing the children's stories of abuse. (Arts)
- ii. Fractured pipes carrying water from the mains to the standpipes suck in raw sewage. **That is** why our children get sick," says Margaret Olewoch, a birth attendant who has lived in Kibera for 20 years, pointing to a leaking pipe. (Science)

LBs of this sub-category provide elaborations to the topic discussed. These LBs play important roles in guiding readers' attention to the phrases after the bundles for further understanding which can be achieved by using *wh-clause* and *that-clause* as shown in example a (i) and a (ii) above.

- iii. Apart from corals, various species of marine invertebrates **such as** sea urchins, sea cucumbers, crinoids and cushion stars, as well as marine snails and fish have sought refuge at the Aquascape reef. (Science)
- iv. Global warming plays an instrumental role in the outbreak and spread of vector-borne diseases. **For example**, the climatic changes brought on by El Nino have been closely linked to malaria. (Science)
- v. The International Trade and Industry Minister reported that export growth in 2007 emanated from both traditional and emerging markets **such as** China, Australia, United Arab Emirates and Indonesia. (Arts)
- vi. As early as kindergarten, for instance, almost all boys choose traditionally masculine occupations such as fire fighters **for example**, and most girls name traditionally female occupations such as nurse or teacher. (Arts)

Two LBs namely *such as* and *for example* as shown in example a(iii) to a(vi) are abundantly found. Unlike other bundles in this sub-category, they offer examples relevant to the context to enhance readers' comprehension.

b) Location

- i. This aggressive marketing coupled by the selling of the lollipops **over the** counter by pharmacists without requiring a prescription, has angered many individuals and organisations like the Campaign for Tobacco-Free Kids. (Science)
- ii. She had scrubbed the floor of the kitchen, washed the vessels and put them in a shining row **on the** wooden shelf, returned the short scrubbing broom to its corner and closed the kitchen window. (Arts)
- iii. In a recent study, where we interviewed young men who woke quiet villages by racing **through the** middle of the night, we found that one of their major motivations was, indeed, the desire to shock. (Arts)
- iv. The palm-size device can tell if a person or animal has contracted the H5NI form of the virus in less than 30 minutes. And it can do so even **at the** earliest stages of the disease, when a victim has yet to show any symptoms. (Science)

Among the total of 61 location-based LBs, more than half indicated location information which are either headed by a preposition as can be seen in example b(i) and (ii). Undoubtedly, most location LBs are headed by a preposition and it is a fact that common prepositions like *at*, *about* and *by* have various meanings. For example, 2-word bundle *through the* and *at the* refer to time and space as depicted in the last two instances.

c) Quantity

- | N | Concordance |
|----|---|
| 1. | about managing your emotions. One of the major qualities that make up (Arts) |
| 2. | ways to say what you need to say. Some of these phrases are mild,(Arts) |
| 3. | for years -what is different of late is the huge number of white-collar (Arts) |
| 4. | scientists could only introduce a single human gene not entire DNA (Science) |
| 5. | 100 years or so of fossil fuels that took half a billion years to form (Science) |
| 6. | future of ageing baby boomers along with a few inconsistencies. (Science) |
| 7. | crazy, came to be the most frequent acquaintances of the giants (Science) |
| 8. | "The Paradox of Choice" (2004) that too much choice is oppressive. (Arts) |

Majority of LBs in the MUET reading text corpus, similar to Lin's (2008) study specify quantity or framed concrete and abstract properties of following noun phrases. The first five concordance lines above exemplify LBs from both disciplines specifying quantity characteristics. In addition, most of the quantity specification bundles, for instance *million people, one day, thousands of* are used to describe concrete or countable nouns; whereas very few illustrate abstract nouns (See examples 6, 7 and 8 above).

d) Stance

- i. **It would** be naive, however, to assume that most advertising is deceptive. (Arts)
- ii. The increased demand is 10 also attributed to the fact that aluminium is now more expensive than tin; **it is** cheaper to use tin cans. (Arts)
- iii. In other countries, it **may be** the health of relatives, as in the following exchange between a villager and a city-bred young man... (Arts)
- iv. As a brief footnote, it **should be** noted that nation-building is a heated and even hated notion in some parts of the world. (Arts)

LBs headed by anticipatory *it* as can be seen in the first two examples and modal verbs in the last two examples are frequently used by writers to express their viewpoints. This could be linked to writers' techniques of presenting information as an opinion rather than accredited fact (Hyland 2008). As mentioned by many (Lin, 2008; Hyland, 2008; Biber & Barbieri, 2007; Biber, et al. 2004), participant-oriented bundles are not significant in written form but they are rather prominent in a wide range of spoken registers.

Unexpectedly, LBs especially those headed by modals of probability are also found in science-based texts. It is understood that stance bundles are largely used to communicate uncertainty which may seem contradictory to the fact that ideas must be supported by facts in the sciences. The following concordance lines of bundles (from 1 to 5) extracted from science based-texts are proof to the claim made:

- | N | Concordance |
|----|---|
| 1. | Emotionally-induced fatigue may be compounded by sleep disturbance |
| 2. | of the punch. He feels that "there may be some advantage to the lollipop. |
| 3. | patients run a risk as some herbs may have some dangerous interactions with |
| 4. | at least once a day, transmission might be interrupted. This was easier said |
| 5. | organism, Chlamydia trachomatis, could be dispelled by antibiotics but |

e) Procedure

Because the genre of the text in this study is not research-based articles, LBs showing the ways experiments and research are conducted (Hyland 2008) cannot be found. Hence, any LB denoting a process (See examples 1 and 2) or an act is taken into consideration. The latter usually denoted by to-infinitive clause (to + Verb) is used to indicate purpose or intention of an action as shown in example 3.

1. It has since developed a process for freeing sugars in the cellulose of plant material by **the use of** acid hydrolysis. Once the sugars have been produced, Arkenol uses yeast fermentation to convert the sugars into ethanol. (Science)
2. They changed to paper cups and all napkins and office paper are now **made from** recycled paper. (Science)
3. i) This is what has happened in South Africa, where water providers, private and public, are now required by law **to provide** a basic minimum of water free of charge. (Science)
ii) This could be done by one or more of the following means: crop rotation, organic fertilization or drainage **to improve** the soil quality. (Science)

One interesting bundle worth noting is the LB *used + PP* which is abundantly found in Hyland (2008) and Lin (2008). These structures showing slight difference in its function are widely employed especially in Science related text.

4. i) The pharmaceutical industry, in particular, continues to investigate and confirm the effectiveness of many medicines and toxins **used by** indigenous peoples, and profit enormously from its commercialisation.
- ii) Malaysian rainforests support about 1200 species of plants with medicinal properties, of which 60 are commonly **used in** traditional remedies.
- iii) It is also **used to** treat skin conditions like scabies and athlete's foot.

In relation to the third research objective which is to compare and contrast functional types of LBs in science and arts-based texts, the distribution of LBs shows a similar pattern for both disciplines. Research-oriented LBs with more than 50% of occurrences come in first, followed by text-oriented bundles and the least is participant-oriented bundles with less than 20%. The claim made about the prevalence of text-oriented bundles in arts-based text by Hyland and Tse (2009) is proven in this study although the difference is only 1.1%. The only category where its occurrences in the arts-based texts greatly surpass science-based texts is participant-oriented bundles. The summary of occurrences of Hyland (2008) functional components in the two disciplines is shown below:

TABLE 5. Occurrences of Functional Categories in Science and Arts Based Texts in Percentage

		Science (%)	Arts (%)
Research-O	Location	21.2	18.7
	Procedure	7.5	6.2
	Quantity	13.1	8.3
	Description	17.8	20.5
	Topic	5.6	4.9
Text-O	Transition	9.7	9.2
	Resultative	2.6	0.7
	Structuring	3.5	6.1
	Framing	8.2	9.1
Participant-O	Stance	10.2	13.3
	Engagement	0.9	3.2
		100.0	100.0

Distinctive patterns among the disciplines are noticeable when functional sub-categories are analysed. Certain disciplines tend to adopt specific function of LBs which do not fall under the same main category. For science-based texts, LBs indicating quantity and procedure alongside resultative signals are apparent. As predicted, science-based texts adopt many LBs used to quantify concrete and abstract nouns because they rely heavily on facts and figures. Because LBs denoting ways experiments are carried out could not be found, procedure-based LBs are substituted with any LBs denoting a process or an action. Under text-oriented bundles, only resultative signals appear significant among science-based texts. The number of resultative signals in science-based texts is approximately three times more than arts-based texts.

In arts-based texts, an extraordinary trend is noted where the numbers of engagement and stance bundles grouped under participant-oriented LBs are incomparable. Stance bundles are found to be mostly used in the arts based-texts. The number of engagement bundles adopted by the arts-based texts triples the sciences. These are the only sub-categories in this study where the numbers of LBs and distribution of arts-based texts exceed the sciences. Looking at the second functional category, text-oriented bundles, structuring signals are extensively used in arts-based texts whereby they are largely used to indicate sequence in a

text. Besides that, LBs headed by non-referential *there* are grouped under this sub-category because they function to assert existence. As for research-oriented bundles, arts-based texts seem to adopt slightly more description-based bundles than science-based texts. This may be due to the abundance of LBs headed by *that clause*, *wh-clause* and two other structures namely NP + of and adverbial clause which provide a certain degree of explanations to the topics discussed.

DISCUSSION AND PEDAGOGICAL IMPLICATIONS

The frequently occurring functional categories of LBs could shed light on the testing and evaluation domain in general and of MUET in particular. Test designers must be made aware that the topic of a text selected for testing has a direct effect on the types of LBs in the text. The reading passages in this study show that many different topics are grouped under one discipline. It has been recommended that texts chosen should deal with familiar topics in learners' home culture, so that their background knowledge can compensate for linguistic difficulty (Kern, 2000) and that "subject-related texts might discriminate against individuals who happen to possess less background knowledge in a particular field" (Alderson, 2005, p. 103). In order to grasp the gist of a text, learners must at least familiarise themselves with the repertoire of LBs in that particular context. Hence, explicit teaching of LBs should be considered in schools.

As regular participants of a particular field, students may have gained control of certain LBs. This is proven when Lin (2008) and Dontcheva-Navratilova (2012) indicated that the acquisition of LBs in a discipline is a long process. It is only reasonable to test students' language ability in a specific field. To avoid defeating the purpose of a proficiency test which is to assess whether students with different language training backgrounds have reached a given level of general language ability (Alderson, Clapham & Wall, 1995), two sets of MUET reading test to accommodate both arts and science pre-university students should be designed. The first will comprise texts adopted from the scientific and technological context while the other is made up of texts from the arts and social sciences. This is because the Malaysian school education system for pre-university students is divided into the two mentioned streams; they are normally exposed to language used in textbooks, classroom discussion, teachers' notes and supplementary materials which are restricted to either one of the streams.

By designing two separate sets of MUET reading tests, the texts used may not vary greatly from one topic to another but restricted within one major discipline. For example, geography, meteorology, and architecture themed passages could be used to test the science stream students. On the contrary, passages revolving around topics such as culture, geography, communication, and language could be adopted to test the art stream students. Doing so is deemed effective because in tertiary education, students will be placed in their respective field.

The teaching and learning will revolve around that particular course chosen by the students. For instance, a psychology undergraduate is expected to learn subjects like counselling skills, child development, human personality, organisational psychology and so on. It is apparent that using zoology, chemistry, food science or other scientific-based texts to test future psychology undergraduates would be of no help to them. Looking at the data collected for this study, a preference towards scientific-based texts was shown. When more scientific texts are employed, it may appear beneficial to students who will be undertaking for instance Biotechnology but not those who will be majoring in Psychology.

In short, 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.

Frequency was one of the criteria according to Schmitt (2005) for choosing items to be taught in the classroom setting. Undoubtedly, the LBs identified in this study could be integrated in English language lessons, MUET in particular because they were identified purely based on frequency count. However, LBs were very frequent but not perceptually salient according to Dontcheva-Navrotilova (2012) and Biber and Barbieri (2007). This was proven because LBs like *on the other hand*, *more likely to*, *such as*, *thousands of*, and *involved in* were very significant; but students who have sat for MUET still have the tendency to misuse them. Dontcheva-Navrotilova (2012), and Neely and Cortes (2009) supported overt teaching of LBs because by merely exposing LBs will not result in students acquiring these bundles especially their functions (Cortes, 2004).

Explicit teaching of the functions of LBs should be considered. Students can be trained to proactively explore texts in order to be aware of their usages. LBs used may not be identical in all the fields of study. That is why Hyland (2008) highlighted the importance of learning to use frequently occurring LBs of a discipline which in his opinion could contribute to the students' sense of distinctiveness in a field. Distinguishing commonly used LBs may not create an impact on the learners because Neely and Cortes (2009, p.30) criticized that "decontextualized nature of certain corpus-based activities could create an inauthentic language learning experience." Findings of this study too do not isolate LBs; in fact disciplines in which they occurred alongside their contexts are presented. These bundles, according to Hyland and Tse (2009) help to scaffold and present arguments as they consider the discursual expectations and processing needs of a disciplinary audience. Hence, contexts in which they are used must also be indicated to address variation of discipline-specific LBs and their functions.

In addition, the texts used for such tasks should be derived from prominent themes frequently adopted in MUET reading tests. The thematic approach proposed by Naginder (2004) should also be taken into consideration. Grouping passages from similar theme and generalizing them to a specific discipline will thus ease the students when they sit for reading tests because opportunities to observe how LBs structure certain texts have been given. Such exposure could confirm that learners actually notice these LBs.

LIMITATIONS

The limitation of this study was mainly due to the size of data obtained compared to past studies such as Hyland (2008). Although the number of MUET reading test papers collected covers the span of over ten years, the amount of data collected was rather small because the passages used in MUET reading tests were relatively short.

Due to the size of the corpus (15,606 running words), instead of focusing on the researchable length of 4-word LBs (Hyland, 2008), this study took into consideration two to six-word bundles. Most of the LBs generated by WST 5 were 2-word bundles. However, 2-word bundles as mentioned by Hyland (2008) could not offer a clear range of structures compared to 4-word bundles. The categorisation of LBs became difficult due to the presence of 2-word bundles which were rather ambiguous.

The emphasis placed on 2-word bundles made it difficult to categorise certain LBs into any of the proposed sub-categories using Hyland's (2008) taxonomy. For example, *of the*, *as a*, *for the*, *in a*, were among the bundles which can be categorised in more than one functional sub-category. The most generated LBs were prepositional phrases which recorded high frequency of occurrences regardless of the disciplines. Their functions differed according to the contexts which resulted in one LB falling into more than one functional

category. This appeared to be parallel with Dontcheva-Navratilova's (2012) observation where some bundles can perform different functions in different contexts and have more than one function within a single occurrence. Hence, all the above mentioned characteristics were excluded from the analysis.

CONCLUSION

Since the objectives of the study were to determine the types and functional categories of LBs as well as their frequencies in MUET reading tests, the findings were limited to these aspects of frequently occurring LBs in the tests. Generally, science and arts-based texts preferred research-oriented LBs and participant-oriented LBs respectively. To be specific, science-based texts favoured LBs indicating location, description and quantity whereas LBs indicating description, location and stance were abundantly found in arts-based texts. Despite the claims made about the prevalence of text-oriented bundles in Social Science texts (Hyland & Tse 2009), the number of occurrences in science-based texts was slightly higher than arts-based texts. The only sub-category where its number of occurrences in the arts-based texts surpassed science-based texts was participant-oriented bundles which were relatively low. To reiterate, participant-oriented bundles were more significant in spoken form as claimed by Lin (2008), Hyland (2008), Biber and Barbieri (2007) and Biber, et al (2004). In short, the results of this study along with some other previous studies namely Biber, et al (2004), Hyland (2008) and Hyland and Tse (2009) indicated that different LBs were employed in different disciplines.

Test designers should critically review selected passages before adopting them for MUET reading test. The data (reading passages) collected for this study showed that many different topics were grouped under one discipline. In other words, the presence of new LBs in unfamiliar texts may impede students' comprehensions. Kern (2000) recommended that texts chosen should deal with familiar topics in learners' home culture, so that their background knowledge can compensate for linguistic difficulty. As for teachers, to ensure LBs do not go unnoticed, they should provide students with an understanding of the features of passages, specifically LBs that they may encounter in MUET reading test.

Interpretations of the findings of this research can lead to several suggestions for further research. Firstly, a comparative study can be adopted where a reference corpus made up of IELTS or TOEFL reading texts could be used for this purpose; MUET reading text corpus can therefore be checked against the reference corpus. The usage of reference and specialized corpora according to Gabrielatos and Sarmiento (2006, p.223) made it possible "to compare the use of language features in specific domains in relation to language used as a whole". By identifying the similarities and differences of recurrent word combinations – LBs from different English proficiency test, the commonly used language in reading tests can be determined. Secondly, to identify more commonly used LBs, larger and established corpora such as COCA (Corpus of Contemporary American English), BNC (British National Corpus) or ICE (International Corpus of English) could be used. However, only the academic sub-category should be scrutinized. It may appear too general if the whole corpus is taken into consideration as LBs used for certain purposes and genres are not made known.

REFERENCES

- Alderson, J.C. (1990). Testing Reading Comprehension Skills (part one). *Reading in a Foreign Language*. 6 (2), 425-438. Retrieved July 9, 2012 from <http://nflrc.hawaii.edu/rfl/PastIssues/rfl62anderson.pdf>
- Alderson, J.C. (2005). *Assessing Reading*. United Kingdom: Cambridge University Press.

- Alderson, J.C., Clapham, C. & Wall, D. (1995). *Language Test Construction and Evaluation*. United Kingdom: Cambridge University Press.
- Bahiyah Abdul Hamid., Mohd. Subakir Mohd Yasin, Kesumawati Abu Bakar, Yuen, C.K. & Azhar Jalaluddin. (2008). Linguistic Sexism and Gender Role Stereotyping in Malaysian English Language Textbooks. *GEMA Online® Journal of Language Studies*. 8(2), 45-78.
- Biber, D. & Barbieri, F. (2007). Lexical Bundles in University Spoken and Written Registers. *English for Specific Purposes*. 26(2007), 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.
- Chen, Y-H & Baker, P. (2010). Lexical Bundles in L1 and L2 Academic Writing. *Language Learning & Technology*. 14(2), 30-49. Retrieved February 4, 2012 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(2004), 397-423.
- Dontcheva-Navratilova, O. (2012). Lexical Bundles in Academic Texts by Non-Native Speakers. *Brno Studies in English*. 38 (2). Retrieved June 1, 2012 from http://www.phil.muni.cz/plonedata/wkaa/BSE/BSE_2012/BSE_2012-38_2_-XX_Dontcheva-Navratilova.pdf
- Gabrielatos, C. & S, Sarmiento. (2006). Central Modals in an Aviation Corpus: Frequency and Distribution. *Porto Alegre*. 8(2), 215-240.
- 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.
- Hafiza Aini Hassan & Hadina Habil. (2008). Comparison of genre analysis approach and audio-lingual method in the teaching of discussion skills in the context of MUET speaking examination. Retrieved July 23, 2012 from [http://eprints.utm.my/7881/1/EDUPRES_\(F1\)_6.pdf](http://eprints.utm.my/7881/1/EDUPRES_(F1)_6.pdf)
- Hyland, K. (2008). As Can Be Seen: Lexical Bundles and Disciplinary Variation. *English for Specific Purposes*. 27(2008), 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.
- Kern, R. (2000). *Literacy and Language Teaching*. Hong Kong: Oxford University Press.
- Klein, T.W. (1990). Characteristics which differentiate criterion-referenced from norm-referenced tests. Retrieved March 8, 2012 from http://www.eric.ed.gov/ERICWebPortal/search/detailmini.jsp?_nfpb=true&_ERICExtSearch_SearchValue_0=ED324327&ERICExtSearch_SearchType_0=no&accno=ED3243279
- Lee, K.S. (2004). Exploring Connection between the Testing Of Reading And Literacy: The case of the MUET. *GEMA Online® Journal of Language Studies*. 4(1), 1-12.
- Lin, C. (2008). An Investigation of Lexical Bundles in Electrical Engineering Introductory Textbooks and ESP Textbooks. Unpublished MA thesis, Carleton University, U.K.
- Lourdunathan, J. & Menon, S. (2005). Developing Speaking Skills through Interaction Strategy Training. *The English Teacher*. 34, 1-18. Retrieved July 23, 2012 from <http://www.melta.org.my/ET/2005/DEVELOPING%20SPEAKING%20SKILLS%20THROUGH%20INTERACTION.pdf>
- Malaysian Examination Council. (2008). *Malaysian University English Test (MUET)*. Malaysia: Majlis Peperiksaan Malaysia.
- Naginder Kaur. (2004). Begin at the Beginning: A Thematic Approach to MUET. In Sarah Abedi Abdullah & Malachi Edwin Vethamani, (Eds.). *Preparing Students for the*

- Malaysian University English Test* (pp. 23-34). Malaysia: Sasbadi MELTA ELT Series.
- Naginder, K. & Rohayah, N. (2006). A Case for Reconstruction of the Pedagogy of the Malaysian University English Test (MUET) through thematic units instruction. *JIRSEA*. 4(1), 5-16. Retrieved December 6, 2011 from <http://www.seairweb.info/journal/JIRSEA%20v4%20n1%202006%20Papers.pdf>
- Nambiar, R. M. K. (2007). Enhancing Academic Literacy among Tertiary Learners: A Malaysian Experience. *3L: Language Linguistics Literature®*, *Southeast Asian Journal of English Language Studies*. 13, 77-94.
- Nazira Osman & Kamaruzaman Jusoff. (2009). Routinizing Lexical Phrases on Spoken Discourse. *International Education Studies*. 2(2), 188-191. Retrieved March 12, 2012 from <http://ccsenet.org/journal/index.php/ies/article/viewFile/1726/1660>
- Neely, E & Cortes, C. (2009). A Little Bit About: Analyzing and Teaching Lexical Bundles in Academic Lectures. *Language Value*. 1(1), 17-38.
- Noorizah Mohd Noor. (2006). Reading Academic Text: Awareness and Experiences among University ESL Learners. *GEMA Online® Journal of Language Studies*. 6(2), 65-78.
- Read, J & Nation, P. (2002). An investigation of the lexical dimension of the IELTS speaking test. *IELTS Research Reports Volume 6*. Retrieved March 8, 2012 from http://www.ielts.org/pdf/Vol6_Report7.pdf
- Sarah Abedi Abdullah. (2004). MUET & IELTS Preparation: Can One Size Fit Both? In Sarah Abedi Abdullah & Malachi Edwin Vethamani, (Eds.). *Preparing Students for the Malaysian University English Test* (pp. 100-127). Malaysia: Sasbadi MELTA ELT Series.
- Schmitt, N. (2005). *Vocabulary in Language Teaching*. United Kingdom: Cambridge University Press.
- Scott, M. (2011). *WordSmithTools*. Version 5.0. Oxford: Oxford University Press.
- Strunkyte, G. & Jurkūnaite, E. (2008). *Written Academic Discourse: Lexical bundles in Humanities and Natural Sciences* (Unpublished bachelor's thesis). Vilnius University, Vilnius.
- Tan Ooi, L.C., Tan, K.E. & Norlida Ahmad. (2011). Effects of reciprocal Teaching Strategies on Reading Comprehension. *The Reading Matrix*. 11(2), 140-149.
- Zuraidah Mohd Don. (2003). Malaysian University English Test: Issues and Concerns. *Studies in Foreign Language Education*. 18, 17-32.
- Zuraidah Mohd Don & Maya Khemlani David. (2005). The testing of Literacy Skills in an ESL Environment. In Chan Swee Heng & Malachi Edwin Vethamani, (Eds.). *ELT Concerns in Assessment* (pp. 111-126). Malaysia: Sasbadi MELTA ELT Series.

ABOUT THE AUTHORS

Christina Ong Sook Beng is a lecturer at the Department of Languages and Linguistics, Universiti Tunku Abdul Rahman.

Yuen Chee Keong (Ph.D) is an Associate Professor at the School of Language Studies and Linguistics, Universiti Kebangsaan Malaysia.