

An Exploration of Lexical Bundles in Academic Lectures: Examples from Hard and Soft Sciences

Hadi Kashiha

University Putra Malaysia, Malaysia

Chan Swee Heng

University Putra Malaysia, Malaysia

Recurrent word combinations that carry out specific function have long captured the attention of many linguists. Referred to as extended collocations, lexical bundles are frequently used in spoken and written discourse, helping to shape meaning and coherence in a text or speech. Frequent use of these bundles is indicative of fluency in linguistic production. In the last two decades, lexical bundles have begun to attract considerable attention in corpus-based research; however, there is still an open question of whether they are different across disciplines in spoken discourse. This study aims to explore how four-word lexical bundles are used in 24 academic lectures of hard and soft sciences taken from the British Academic Spoken English (BASE) corpus to see the possible variations in their frequency, structure and function. The present research also attempts to discover the manifestation of the identified bundles in introduction, body and closing sections of the lectures. Findings revealed some marked variations across the two divisions; in that, lecturers in each division appeared to apply different structures and functions in the use of lexical bundles in order to convey their message, so as to be as comprehensive as possible for the learners.

Keywords: academic lectures, lexical bundles, spoken discourse

INTRODUCTION

The history of recurrent word combinations can be traced back to Jespersen (1924) and Firth (1951), who introduced the term collocation and fixed expressions. Since then, there has been an increasing interest in the research on frequent word combinations under different names such as “phraseology” (Granger & Meunier, 2008; Meunier & Granger, 2007), “formulaic language” (Schmitt & Carter, 2004; Wray 2000, 2008) and “lexical bundles” (Biber & Conrad, 1999). The focus of most of these studies was on the importance of such multi-word units in language learning as they play a central role in building fluency and confidence, mainly in spoken language (Pawley & Syder, 1983). Wray and Perkins (2000) also asserted that learning these expressions can serve as a type of short-cut to language knowledge, as they are stored in memory and retrieved later at the time of use.

One example of these recurrent word combinations is what the *Longman Grammar of Spoken and Written English* refers to as “lexical bundles.” Biber and Conrad (1999, p. 183) defined lexical bundles as “multi-word expressions which occur frequently and with accidental sequences of three or more words (e.g. *in the case of the, do you want me to*).” Basically, a lexical bundle is a combination of more than two words that co-occur more frequently than we expect by chance. Their frequent use helps to shape the meaning in a text and make difference in a register (Hyland, 2008). Therefore, a bundle such as *as can be seen* is more likely to belong to a written register, while *we are going to* is indicative of proficient language use in a spoken register like academic lecture. There are three main characteristics that distinguish lexical bundles from other types of word combinations such as idioms and collocations. The first characteristic is frequency of occurrence. Many lexical bundles have to occur more than 10 or 20 times (based on the register) in a million words to be considered for analysis, whereas an idiom such as *kick the bucket* occurs only 0.5 times in a million words (Cortes, 2004). Second, most lexical bundles are not idiomatic in nature and not perceptually salient; rather, “their meaning is transparent” (Cortes, 2004, p. 400) and can be easily understood from the

translation of individual words that make up the bundles. Finally, lexical bundles are not structurally complete. According to Biber and Conrad (1999), only 15% of the lexical bundles in conversation are complete clauses or phrases, whereas the percentage is less than 5% for academic prose.

There is a well-established belief that having a good command of such frequent word combinations could account for a mature and proficient speaker. For example, lecturers in an ESL context can build academic voice and self-confidence by using a variety of lexical bundles while giving lectures in their specific discipline. When they lecture in English, frequent use of lexical bundles would help them present the materials at a proficiency level that reflects a mature use of the language. In addition, non-native students have shown difficulty in understanding long lectures. One reason might be the fact that many of these students are not familiar with the discursive functions of the lexical bundles used in the lectures. Given such a context of language use, it is the task of English for Academic purposes (EAP) practitioners as well as discourse analysts to provide learners with strategies such as discourse function or meaning of these bundles in order to help them overcome the problem of inefficient understanding of the disciplinary lectures, as it would have repercussions on academic performance.

The Study of Lexical Bundles

In the last 20 years, several corpus-based studies have been conducted on the use of lexical bundles in a range of academic genres, but the focus of most of these studies was on academic written discourse (Adel & Erman, 2012; Cortes, 2004, 2006; Hyland, 2008; Strunkyt & Jurkūnait, 2008). Adel and Erman (2012) compared native and non-native speakers' use of lexical bundles in academic writing and found that non-native learners used a smaller proportion of lexical bundles compared to native speakers. In contrast, little attention has been given to the notion of lexical bundles in spoken discourse. Among the few studies on spoken registers, Tracy-Ventura, Cortes, and Biber (2007) compared the

frequency and use of lexical bundles in speech and writing by English and Spanish speakers and found that unlike English, lexical bundles in Spanish are more prevalent in writing than in speech. Csomay and Cortes (2010) also examined the discourse functions of lexical bundles and their relation in classroom teaching and concluded that there is a one-to-one relationship between intratextual linguistic differences and the corresponding change in discourse. Among the relevant studies, two notable works that have assisted considerably to our understanding of lexical bundles in spoken registers are the works by Biber, Conrad, and Cortes (2004) and Nesi and Basturkmen (2009).

In their study of lexical bundles in academic registers, Biber et al. (2004) provided insight into how lexical bundles are classified structurally and functionally in academic spoken discourse. To this end, they conducted a study to find out the frequency, function and structure of the most frequent lexical bundles in two university registers: classroom teaching and textbooks and compared them to those found in their previous research on conversation and academic prose (Biber & Conrad, 1999). They found that teaching in the classroom uses twice as many bundles as conversation and about four times more than those of textbooks and academic prose.

Nesi and Basturkmen (2006) explored the cohesive role of four-word lexical bundles in 160 university lectures from two different online corpora: the British Academic Spoken English (BASE) corpus and the Michigan Corpus of Academic Spoken English (MICASE). To this aim, they examined concordances surrounding the occurrences of lexical bundles to see whether or not the bundles happened to link and relate parts of the discourse. They found that lexical bundles play a crucial role in building cohesion and signaling discourse relations in their lecture corpus. Almost all the frequent lexical bundles in their list appeared to signal discourse relations (with the exception of *is going to be* and *the result of the*). In sum, the findings of the related literature suggest that lexical bundles are frequently used in both oral and written academic discourse. Despite the importance and frequent use of lexical bundles, major questions still remain about their discipline-specific use in academic lectures, as well as their manifestation in introduction, body and closing

sections of the lectures. This study begins to answer these questions by identifying and analyzing the structure and function of the most frequent 4-word lexical bundles in academic lectures of hard and soft sciences.

CORPUS AND METHOD

The study is based on the corpus of 24 academic lecture transcripts taken from BASE corpus. BASE is an online academic corpus that includes lectures and seminars across four broad fields of studies. The data selected for the present study are transactional lectures across two broad fields of sciences: hard sciences (HS) and soft sciences (SS). Soft sciences include lectures in law, politics, and CELTE (Center for English Language Teacher Education), and lectures in hard sciences are represented by chemistry, computer, and engineering. The disciplines were selected on the basis of matching word count in each science in order to make the data as comparable as possible. The distribution of disciplines in each science and their word counts are given in Table 1 below.

TABLE 1
Academic Lectures Corpus by Discipline

	Sciences					
	Hard sciences			Soft sciences		
	Engineering	Chemistry	Computer	Law	Politics	CELTE
No. of lectures	4	4	4	4	4	4
Word count	29413	25106	21801	22047	25185	29667
Length of lecture (range)	3700-9079			4234-8243		
Total	76320			76899		

The first step was to identify the boundaries between the introduction, main body, and closing sections of the lectures. The researchers adopted a criterion

proposed by Thompson (1994) in identifying the boundary between the introduction and the main body of the lectures, with the lecturers using long pauses followed by boundary markers such as right or ok. However, a few lectures did not include any pauses or appeared to have vague boundary markers. In this case, the researchers read the transcript as much as needed to put the boundary at the point where the introduction ended and the lecturer moved into the main body of the lecture. The same procedure was used to identify the closing section, the point where the lecturer made the ground to end the lecture using some moves such as concluding the discussion, introducing the topic of the next lecture, or greeting the students before they leave the class.

To limit the scope of the research, this study focused only on 4-word lexical bundles as previous studies have shown that they are the most common and frequent strings of words in academic genre (Biber & Barbieri, 2007; Biber et al., 2004; Cortes, 2002, 2004; Hyland, 2008). Four-word lexical bundles “are far more common than 5-word strings and offer a clearer range of structures and functions than 3-word bundles” (Hyland, 2008, p. 8). In order to avoid the lecturers’ subjective influences, only those 4-word strings that occurred 10 times per hundred thousand words and in at least three different lectures were identified and analyzed as the salient lexical bundles. This study used the *Wordsmith Tools* (Scott, 2004) computer software in order to identify and make a list of the most frequent 4-word lexical bundles in each sub-corpus.

After comparing the frequency and patterns of use across the two sciences, the identified lexical bundles were then categorized both structurally according to their grammatical forms, and functionally according to their contextual meanings. This study adopted Biber et al.’s (2004) functional and structural taxonomies, as their study has been well quoted for use in analyzing oral data of this nature. According to their structural classification, lexical bundles have three main grammatical forms: 1) lexical bundles that incorporate verb phrase fragments; 2) lexical bundles that incorporate dependent clause fragments; and 3) lexical bundles that incorporate noun phrase and prepositional phrase fragments. In their functional classification, Biber et al. (2004) proposed three main discourse functions for lexical bundles: Stance bundles are defined as the

“overt expression of an author’s or speaker’s attitudes, feelings, judgments, or commitment concerning the message” (p. 386); discourse organizer bundles try to indicate the relationships between given and coming discourse; and referential bundles make direct reference to physical or abstract entities or single out some important features of an identity to be important in a way. Each of these functional categories involves several sub-categories linked to more specific discourse functions and meanings which will be explained later.

RESULTS AND DISCUSSION

Frequency of Disciplinary Bundles

The total number of different lexical bundles which met the criterion set for the corpus was 121 types, totaling 1102 individual cases. The bundle *we are going to* was by far the most frequent lexical bundle in the corpus, occurring 92 times per hundred thousand words, and was over three times as frequent as the second placed bundle, *I am going to*. This high occurrence shows the lecturers’ dependence on using engagement markers especially the first person plural (*we*) to make the students feel that they are part of the activity. Therefore, they are required to be engaged in the process of learning rather than being a mere listener. This adds to the social dimension in academic lectures which traditionally were viewed as serving monologic and transactional purposes that are often not analyzed for efforts in making contact with the audience.

As for the disciplinary differences, Table 2 indicates that lectures in hard science disciplines used a greater range of lexical bundles, with 118 different types, totaling 628 individual cases. Soft science disciplines, on the other hand, had a smaller range of bundles, with only 96 types, totaling 474 tokens. This larger proportion of lexical bundles in the lectures of hard sciences seems to reflect the fact that these sciences are more empirical and thus require the lecturers to apply more visual tools or other accompaniments such as pamphlets or handouts while teaching. In this vein, lecturers in hard sciences used a

number of bundles such as *as you can see* or *going to give you* to attract the students' attention towards the tools. Soft science instructors, on the other hand, were not inclined on using such bundles as they did not seem to provide their students with much graphic tools or handouts during the lecture, which is indicative of the fact that these sciences are more textually interpretive and descriptive.

TABLE 2
Frequency of Lexical Bundles in Hard and Soft Sciences

Sciences	Bundle types	Total No. of bundles	% of total words
Hard sciences	118	628	0.41
Soft sciences	96	474	0.30

Out of 121 different bundle types identified in the whole corpus, only three bundles (*in the context of*, *one of the most*, and *I think it was*) were not used in the lectures of hard sciences, while lecturers in soft sciences did not tend to use 23 of the target bundles which seemed to be exclusively used by the hard disciplines. All the other 95 bundles were shared in the two sub-corpora. Despite some disciplinary differences, this large number of common bundle types in the two sciences emphasizes the high reliance of university lecture as a representative of academic genre on multi-word combinations in general and lexical bundles in specific in order to convey their messages. The most frequent 4-word lexical bundles in the two sciences, with the shared bundles (in bold) are listed in Table 3. As can be seen, only half of the top 10 bundles in each science occurred in the list of the other science, which is indicative of the inclination that lecturers in different disciplinary fields rely on different resources to convince their audiences and communicate their disciplinary messages.

TABLE 3

The 10 Most Frequent 4-Word Lexical Bundles in Hard and Soft Sciences

Hard sciences	Freq.	Soft sciences	Freq.
we are going to	72	we are going to	69
I am going to	28	the end of the	19
If you want to	22	I am going to	15
is going to be	16	in the context of	15
you are going to	16	as a result of	15
and you can see	15	on the basis of	12
to be able to	12	in terms of the	11
in terms of the	11	if you look at	11
I don't know if	10	at the end of	11
if you look at	10	you are going to	9

Structures of Disciplinary Bundles

From the structural point of view, lexical bundles in the two disciplinary groups offered some marked variations in relation to their grammatical types. Tables 4, 5 and 6 give the detailed percentages of the three main structures and their sub-categories in each science using Biber et al.'s (2004, p. 381) taxonomy. In general, findings showed that noun and prepositional phrase fragment were the most common structure in the soft science corpus, including 40.6% of all the bundles, compared with 35.54% in hard sciences. The high proportion of noun and prepositional phrase fragments in soft sciences might also have resulted from the expressive nature of the disciplines in this science that required the lecturers to use a greater variety of short phrases, including different nouns and prepositions, to best convey their disciplinary messages. Hard science lecturers, on the other hand, used higher percentages of dependent clause fragment, with 35.54% compared with 30.19% in soft sciences, suggesting the practical notion of hard science disciplines which entailed more demonstration or illustration. Lecturers in hard sciences seemed to use a larger

number of dependent clause structures in order to direct the students or provide more information, all containing a subject and an action verb. As for the bundles with verb phrase fragments, both groups of sciences shared an almost similar rate of use.

A more detailed look at Table 4 indicates that in the first structural type, verb phrases (*with non-passive verb*) were the most common structure in the two groups of lectures, while there was no example of a passive verb phrase in either of the sub-corpora. Since speech relies heavily on a more direct physical context to deliver the meaning or message, lecturers in the two groups chose active verbs as the preferred simple and straightforward structures to best convey their disciplinary lessons. Active verbs aid comprehension of the given lectures, as in the example below where the common active verb phrase (*going to + verb*) helps to raise the students' awareness towards the forthcoming information:

(1) here I'm just **going to give you** two lines of text if you... (*HS - Engineering*)

TABLE 4
Lexical Bundles That Incorporate Verb Phrase Fragments

Sub-categories	HS		SS	
	No.	%	No.	%
1a. 1 st /2 nd person pronoun + VP fragment	9	7.26	7	7.29
1b. 3 rd person pronoun + VP fragment	7	5.93	4	4.16
1c. Discourse marker + VP fragment	*	*	1	1.04
1d. Verb phrase (with non-passive verb)	13	11	12	12.5
1e. Verb phrase (with passive verb)	*	*	*	*
1f. yes-no question fragments	1	0.84	1	1.04
1g. WH- question fragments	3	2.54	3	3.12
Total	33	27.93	28	29.15

There was only one occurrence of the discourse marker +VP fragment (*I think it was*) in the soft science lectures but this structure was not found in hard

disciplines. It should be noted that regarding the sub-categories “1st/2nd person pronoun + VP fragment” and “3rd person pronoun + VP fragment,” the two broad disciplinary sciences indicated some similar results.

As for the second main structure, dependent clause fragments, Table 5 illustrates that lecturers in hard sciences were dominant in the use of 1st/2nd person pronoun + dependent clause/fragment; however, this structure was still the most prevalent form in the soft science corpus (16.66%). *If*-clause fragments also showed a higher percentage of use in hard sciences. No instance of *that*-clause fragments was found in the two corpora.

TABLE 5
Lexical Bundles That Incorporate Dependent Clause Fragments

Sub-categories	HS		SS	
	No.	%	No.	%
2a. 1 st /2 nd person pronoun + dependent clause fragment	26	22.88	16	16.66
2b. WH-clause fragments	2	1.69	2	2.08
2c. If-clause fragments	10	8.47	7	7.29
2d. to-clause fragment	4	3.38	4	4.16
2e. That-clause fragment	*	*	*	*
Total	42	36.42	29	30.19

As mentioned earlier, the last structural category, noun phrase and prepositional phrase fragments, accounted for the most common structure in respect to the other categories in the whole corpus. Table 6 indicates that noun phrase with *of*-phrase fragment by far registered the highest percentage of occurrence in the two disciplinary sciences, with the soft science instructors using this structure slightly more (20.83%) than those of hard sciences (19.49%). This structure was mainly used to quantify (*the basis of the*), qualify (*the nature of the*) or identify the topic being discussed, as in:

(2) it's the certainly **one of the most** intractable international conflicts that we have around. (*SS - Politics*)

Regarding the other sub-structures, lectures in soft sciences also showed a higher dominance in using noun phrase with other post-modifier and prepositional phrase fragments. The latter seemed to be commonly used in the two sciences to make a logical relationship between the elements:

(3) in fact you have a rotational impact between the engine system on the one hand and the driven system **on the other hand** (*PS - Engineering*)

(4) if you have to stop caring about teaching practice at all **in order to be** a competitive researcher. (*SS - CELTE*)

The two groups of lecturers did not tend to use comparative expressions in presenting the subject materials. It would appear that these expressions are more frequently used in academic written discourse.

TABLE 6
Lexical Bundles that Incorporate Noun Phrase and Prepositional Phrase Fragments

Sub-categories	HS		SS	
	No.	%	No.	%
3a. Noun phrase with <i>of</i> -phrase fragment	23	18.61	20	20.83
3b. Noun phrase with other post-modifier	4	3.38	5	5.20
3c. Other noun phrase expressions	6	5.08	4	4.16
3d. Prepositional phrase expressions	10	8.47	10	10.14
3e. Comparative expressions	*	*	*	*
Total	43	35.54	39	40.6

Functions of Disciplinary Bundles

A framework for analyzing functional characteristics of lexical bundles in this study was also adopted from Biber et al.'s (2004) study. Biber et al.'s taxonomy has been designed on the basis of a broad corpus of written and spoken registers to include conversation, text books, service encounters and so

on. However, the focus of the present study is only on a spoken register in the form of academic lectures. Therefore, in order to fulfill the purpose of the present research and to best annotate discourse functions of some bundles in our corpus, some modifications had to be applied in their classification. The modification included a revision and addition of two sub-categories under the “referential expressions” as well as identifying a new category. The sub-category “text-deixis” in Biber et al.’s classification was changed to “topic reference” as some bundles in our corpus referred to previously discussed topics in earlier parts of the lectures or prior sessions, rather than any text or course book. Another sub-category called “preview” was also added to the referential expressions in order to classify those bundles that refer to the future plan or discussion of the lectures. Finally, a new category of “lecture-specific expressions” was devised with the aim of meeting the need of describing oral lectures as a specific genre investigated in the present study. This action appears prudent and is in line with several other studies which had introduced “research-oriented topic bundles” (Hyland, 2008), “subject-specific bundles” (Jablonakai, 2010), “subject-bound bundles” (Cortes, 2004), and “special conversational functions” (Biber et al., 2004) to give specific labels to those strings of words that are peculiar to their studies. In the same way, the main category “lecture-specific expressions” in the present study was also devised to refer to those bundles which perform specific functions according to the genre that relates to university lectures. In addition, to account for the manifestation of lexical bundles in the introduction and closing sections of lectures, there were further sub-functions of *lecture opening* and *lecture closing*. Lecture opening encompassed lexical bundles which were used in the introduction section of the lectures, while lecture closing takes care of those 4-word strings which were used as closing markers at the end of each lecture. The detailed explanation of these sub-categories will be exemplified later.

Using the devised taxonomy, the results showed some slight similarities and differences in the functional classifications of the two sciences. The first major finding was the higher concentration of referential bundles in the soft science lectures, amounting to 42.70% compared to 33.89% in hard sciences. The

expository nature of disciplines in soft sciences based on which a variety of physical abstracts and entities need to be identified or become prominent would be a major reason for this higher occurrence. On the other hand, hard sciences appeared to be based on observations and measures and less focused on identifying entities. In general, stance expressions accounted for a large proportion of bundles in the two groups. This feature seems to be shared in a variety of academic spoken registers, as Biber et al., (2004) and Biber and Barbieri (2007) also reported that in conversation, stance expressions accounted for over 60% of the bundles in their corpora. This parallel may highlight some shared characteristics between academic spoken registers. One of these characteristics commonly observed in academic lectures is the frequent use of self-reference markers to express the lecturers' attitude which include the expression of certainty/uncertainty. Results also suggested that lecturers in hard sciences registered a greater tendency to use discourse organizer bundles than those of soft sciences (27.96% as compared with 20.83%, correspondingly). Regarding the newly added category "lecture-specific expressions," findings indicated that these expressions accounted for only a small proportion of bundle types in the two sciences. Hard science lectures were found to contain almost twice as many examples of this function as those of soft sciences. Details pertaining to the distribution of each main functional category and its sub-categories will be discussed in the following sections.

Stance Expressions

One of the most notable findings of the present research is the similarity in the use of stance expressions in the two science corpora, with only slight differences between some discourse functions in the disciplines. Table 7 demonstrates the percentages of stance expressions in the lectures of the two disciplinary divisions. As regards to epistemic stance expressions, the result seems to be in line with the previous study by Biber et al. (2004), who found that most epistemic stance bundles in spoken registers, such as conversation and classroom teaching, are used to show personal rather than impersonal

engagement. In the present research, there was only one impersonal epistemic stance (*the fact that the*) in each disciplinary group. All the other epistemic stance expressions were personal, using the first person pronoun (*I*) to show the lecturers' degree of uncertainty. In the following examples, the lecturers used the bundle *I don't know if* to show that they might not have adequate knowledge towards the given information and thus invited mitigation:

(5) ... into an acronym and call it a FAQ. **I don't know if** anyone's done that yet. (*SS - CELTE*)

(6) I am told it's in what is now Hanoi in Vietnam but **I don't know if** that's right or not. (*HS - Computer*)

TABLE7
Stance Bundles in Hard and Soft Science Lectures

Sub-categories	HS		SS	
	No.	%	No.	%
A. Epistemic stance	9	7.62	6	6.25
B. Attitudinal/modality stance				
B1) Desire	1	0.84	1	1.04
B2) Obligation/ directive	13	11	10	0.41
B3) Intention/ Prediction	14	11.86	13	13.54
B4) Ability	2	1.69	2	2.08
Total	39	33	32	33.33

Hard science lecturers favored epistemic stance bundles as a marker of uncertainty slightly more than their soft science colleagues. This dissimilarity in the use of epistemic bundles might have resulted from the fact that hard fields mainly focused on objects and things; therefore, the possibility of making strong claims decreased since the results and conclusions were much firmer and quicker than those of soft fields which focused on human behavior.

The bundle *I think that is* combined epistemic stance with another function of identification in hard science lectures, for example:

(7) some textbooks call it metallation, **I think that is** a confusing term because ... (*HS - Chemistry*)

The two groups of lecturers used a variety of attitudinal/modality expressions in order to express their feeling and attitude towards the occasions happening according to its related propositions: *desire, obligation, directive, intention/prediction, and ability*.

Intention/prediction seemed to be the most common proposition, amounting to 11.86% in hard and 13.54% in soft sciences. The noticeable occurrence of such bundles in the corpus could be attributed to the notion of academic lectures being teacher-centered and in a “live” situation. In this regard, students did not have time to process information, and thus needed to be informed continually of the new plan or event, as in:

(8) **we're going to be** discussing policy in seminars the lectures, **you are going to** focus very much on the law. (*SS - Law*)

(9) this value of N **is going to be** constant throughout the program. (*HS - Computer*)

The only functional difference between the two sciences regarding intention/prediction was the bundle *to look at the*. This 4-word string did not seem to function as predicting or intending in hard disciplines; however, it appeared to serve a dual function in soft disciplines, either to show the lecturers' intention or to introduce a topic, as in:

(10) we're going **to look at the** individual provisions of the Criminal Justice. (*SS - Law*)

Obligation/directive bundles also accounted for 11% occurrence in hard and 10.41% in soft sciences. Lectures in the two divisions used these bundles to either attract the students' attention towards an object or a tool according to the topic of the lectures or to make them perform some actions based on the instructions and procedures presented in the lecture. Most of the bundles found under this subcategory functioned as obligation rather than direction in the hard science corpus. This might have resulted from the physicality nature of the disciplines in this science that required the lecturers to directly or indirectly force the students to do some actions as an obligation. Three of these bundles (*so you have to, you have to do, we have to do*) were specific to hard sciences and were not found in the corpus of soft sciences. In the following example from a lecture in computer studies, the lecturer was explaining a procedure and steps in using a bar code and then used the obligation bundle *you have to do* to ask the students to perform a possible action:

(11) and what **you have to do** is to take a bar code and fold up the bit of paper ... (*HS - Computer*)

In contrast, the directing function of these bundles was more dominant in soft science lectures. Most of these bundles were personal expressions, using second person pronoun (*you*). In the following example, the lecturer addressed the students to a specific list by using the bundle *have a look at*:

(12) just **have a look at** that list and think which functions did I perform. (*SS - CELTE*)

Some directive bundles in the two sciences served another function and were used as topic introduction (will be discussed in section 3.3.2), as in:

(13) Let's **take a look at** how we solve this one. (*HS - Engineering*)

(14) **If you look at** page six in the handout I've given you... (*SS - CELTE*)

Discourse Organizers

Discourse organizer bundles played an important role in building cohesion and were used to organize the lecturers' messages as they "reflect relationships between prior and coming discourse" (Biber et al., 2004, p. 384). In general, lecturers in hard sciences tended to use discourse organizer bundles more than those of soft sciences (27.96% compared to 20.83% respectively). Table 8 shows the distribution of discourse organizer and its sub-categories across the two sciences.

TABLE 8
Discourse Organizers in Hard and Soft Science Lectures

Sub-categories	HS		SS	
	No.	%	No.	%
A: Topic introduction/focus	17	14.40	8	8.33
B. Topic elaboration/clarification	16	13.55	12	12.5
Total	33	27.96	20	20.83

Regarding the first sub-category, hard disciplines were reported to use nearly twice as many topic introduction/focus bundles as soft disciplines. It might partly due to the presence of more difficult and novel content in the lectures of hard sciences which resulted in using more pauses. Consequently, the lecturers had to start a new sentence in order to avoid ambiguity in the message and promote understanding among the students. The other possible reason was the application of some instruments or tools which required the lecturers to use some topic initiating bundles in order to attract the students' attention and invite their participation towards the instrument, as in:

(15) **If you look at** that the top one has no change of direction... (*HS - Engineering*)

The second sub-category, topic elaboration/clarification, was generally used to elucidate the previous sentences used or provide more information in the lectures. In this regard, no significant difference was found between the percentages of the two groups, with almost all the bundle types occurring equally in the two disciplinary divisions. This parallel rate of use reflected that lecturers in each disciplinary science were likely to rely heavily on similar phrases and strings of words to connect their points, regardless of belonging to any disciplinary community. In other words, there are certain groups of words which are commonly used in a variety of academic registers to make some discursive relationships between the members' viewpoints. The bundle *on the other hand* was frequently used to make a logical comparison between the given and forthcoming information, as in:

(16) what they regard as justice, others **on the other hand** would argue that.... (SS - Politics)

(17) ... have a rotational impact between the engine system on the one hand and the driven system **on the other hand** (HS - Engineering)

Some bundles were used to clarify a topic or a term specific to a discipline, such as:

(18) we'll deal with beams and just to remind you **what we mean by** a beam, it's something which is long and thin ... (HS - Engineering)

All in all, the discourse organizers category accounted for the most lexical bundles with a dual function in the whole corpus, especially in the case of topic introduction/focus sub-category in hard sciences. For instance, the bundle *what I'd like to* combined topic introduction with intention stance expressions:

(19) **What I'd like to** spend the last few minutes today is just hinting at ... (HS - Engineering)

The directive bundle *if you look at* also served a topic introduction function, mostly in engineering lectures:

(20) **If you look at** the way in which cricket bats are shaped. (*HS - Engineering*)

Referential Expressions

Regarding the third functional category, one major difference was the higher concentration of referential bundles in the soft science lectures, amounting to nearly half of all the bundles. However, in hard science lectures, this category comprised almost one-third of all the bundles. The detailed information about the sub-functions with the percentages of use is given in Table 9.

TABLE 9
Referential Expressions in Hard and Soft Science Lectures

Sub-categories	HS		SS	
	No.	%	No.	%
A. Identification/focus	7	5.93	7	7.29
B. Imprecision	1	0.84	1	1.04
C. Specification of attributes				
C1) Quantity specification	7	5.93	7	7.29
C2) Tangible framing	1	0.84	1	1.04
C3) Intangible framing	9	7.62	10	10.41
D. Time/ Place/ Text reference				
D1) Place reference	5	4.23	4	4.16
D2) Time reference	2	1.69	5	5.20
D3) Topic reference	4	3.38	4	4.16
D4) Preview	2	1.69	2	2.08
D5) Multi-functional reference	2	1.69	*	*
Total	40	33.89	41	42.70

As regards to the first sub-category, identification/focus, soft science lectures

showed a slightly higher percentage of use, leading to the conclusion that these lectures comprised more expressive topics that needed to be identified in the form of giving definition or characterizing some important features, such as the bundle *one of the most* in the following example which was used to single out the importance of balance of power:

(21) He identified the balance of power as being **one of the most** crucial elements ... (*SS - politics*)

In some cases, this bundle also served a topic introduction function, as in:

(22) **One of the most** widely quoted examples that she uses is ... (*SS - CELTE*)

There was only one instance of imprecision bundles (*or something like that*) in each of the two sciences. This scarcity suggests that lecturers were almost sure of the exactness of the references they were using, or there were no supplementary references of the same type. The former can be considered as one of the major characteristics of academic genre. Soft sciences also registered a higher proportion of specification of attribute bundles, with the intangible framing attributes being the most frequent sub-function. These bundles tended to function as building some commonality for discussion in the lectures. For example, the bundle *in the context of* which was only used in soft science lectures, contributed to the explanation of contexts under which the topics were being discussed:

(23) We need to remember too much about the distributive approach that I illustrated **in the context of** the US and the Soviet Union. (*SS - politics*)

Another function in this sub-category, referred to as quantity specification, was also slightly more commonly used in soft sciences (7.29% compared with 5.93% in hard sciences). This bundle functioned as measuring the amount,

number or quantity of the attributes in the lectures:

(24) but it worries **quite a lot of** people and with some reason all the ..
(*HS - Engineering*)

(25) rather than targeting the plaintiff and making the plaintiff suffer a
little bit more under this legislation. (*SS - Law*)

The fourth sub-category, time/place/topic reference, revealed some similarities and minor disciplinary variations between the lectures. Soft science lecturers used time reference bundles about three times more than those of hard science, especially in some topics of politics and law lectures which had to do with periods of time:

(26) US had favored **for a long time** that nuclear deterrence should be ...
(*SS - Politics*)

The two disciplinary sciences demonstrated a similar scale of use concerning the revised (topic reference) and newly added (preview) sub-sections. Bundles in topic reference were used to make somehow direct or indirect reference to previously-mentioned issues in the lectures, as in:

(27) we have the catenary under self-weight of the cable, **and as I said**,
the weight of the cable actually at the ends is ... (*HS - Engineering*)

(28) **If we go back** to my early lectures when we talked about an
objective of the of the criminal law ... (*SS - Law*)

In contrast, lecturers utilized preview bundles to make reference to the topics which would be explained later in the following session or the next lectures. These bundles were mostly found in the body section of the lectures, using the first person plural *we* to engage the students:

(29) **We'll get on to** that, it's a very good question the question is ... but **we'll get on to** why the problem with this ... (*HS - Computer*)

(30) **We'll come back to** many of these ideas later on in the course particularly ... (*SS - Law*)

No lexical bundle was found to function as multi-functional reference in soft science lectures, but there were some examples of this function in hard sciences. These bundles were used to refer to a specific period of time, particular place or topic depending on the context, such as the bundle *at the end of* which had three different functions:

(31) spending just a couple of minutes **at the end of** this session talking about it. (*HS - Engineering*)

(32) ...the output that you should be producing has no trailing spaces **at the end of** a line and the last line should be ... (*HS - Computer*)

(33) We can calculate this nevertheless **at the end of** all that the important message is ...(*HS - Engineering*)

Lecture-Specific Expressions

TABLE 10
Lecture-Specific Expressions in Hard and Soft Science Lectures

Sub-categories	HS		SS	
	No.	%	No.	%
A. Lecture opening	3	2.54	1	1.04
B. Lecture closing	3	2.54	2	2.08
Total	6	5.08	3	3.12

The last category which was devised to classify those lexical bundles

specific to academic lectures included two major discourse functions: lecture opening and lecture closing. The two sub-functions were devised in order to answer the question on how lexical bundles are manifested in different sections of the lectures. In total, lecturers in hard sciences appeared to have a higher interest in using these bundles.

As is shown in Table 10, there were three examples of lecture opening expressions in hard sciences, while the bundle *we were looking at* was the only expression of this type in soft sciences. These expressions were normally used in the introduction section to initiate the lecture by providing overviews to previously related topics, as in:

(34) Last week **we were looking at** how long it takes for cases to
now let's just ... (*SS - Law*)

The bundle *we said last week* combined lecture opening with topic-reference function in hard science corpus:

(35) as **we said last week** we started with a simple possible system...(*HS - Engineering*)

There were only a few examples of lecture closing expressions in the corpus of the two sciences, comprising only 2.54% of all the bundles in hard and 2.08% in soft sciences. As the name suggests, these bundles were used to announce the end of the lecture:

(36) okay right so **that's the end of** today, on Monday we will be talking about (*SS - Law*)

The bundle *thank you very much* was found to serve a dual function in soft science lectures. Apart from signaling the end of the lecture by showing politeness (example 37), in a few cases, it was used as a starting point and functioned as lecture opening (example 38).

(37) so much for today, **thank you very much** for your attention I do apologize for ...(*SS- Politics*)

(38) **Thank you very much** for all of you and CELTE members for inviting as a visiting fellow... (*SS - CELTE*)

Distributional Pattern of Cross-Sectional Bundles in Academic Lectures

The findings of the present study showed that the distribution of lexical bundles is also different according to their discourse structure. The majority of the identified bundles were located in the main body of the lectures, comprising 83% of the total number of bundles in the corpus. Such high occurrence may reflect the major focus of academic lectures which would be on the main content and subject materials. In contrast, introduction and closing sections of the lectures by far had fewer bundles. As for the disciplinary variation, the results showed no significant difference between the two groups of lectures regarding the distribution of lexical bundles in introduction, body and closing sections.

Each section of the lecture accounted for some structural and functional attributes which more or less were present in other sections. Concerning the largest section of the lectures, findings indicated that many examples of stance expressions, discourse organizers, and referential expressions were frequently used in the body section where the focus was mainly on the description of topics. Such various descriptions required a range of structures to be constructed. In our corpus, the body section of the lectures encompassed all the structural types introduced in the taxonomy. In some cases, there was a one-to-one relationship between the form and function of the bundles in that a particular structure was used to serve certain function. For instance, noun and prepositional phrase fragments were the most common structures in the production of referential expressions (*in the case of*). Verb phrases and dependent clause fragments, on the other hand, were mostly used to construct stance expressions (*I think that is*) and discourse organizers (*I would like to*).

As for the introduction section of the lectures, all the bundles were verb phrases, either with a non-passive verb (*thank you very much*) or preceded by a connector (*and there is a*). They only served lecture specific functions. Bundles in the closing section also served the same function but were mostly dependent clauses, except for the bundle *at the end of* which was a noun phrase and functioned as time-reference.

CONCLUSION

The main objective of the present study was to investigate the degree to which multi-word expressions can contribute to a better linguistic description of lecture presentation by applying a corpus driven approach to identify and analyze the most frequent 4-word lexical bundles in academic lectures of hard and soft sciences. The findings revealed that lexical bundles were frequently used in academic lectures and accounted for a large proportion of lecturers' speech. In addition, lecturers in each field of study relied on a variety of forms and functions to communicate their arguments, convince the students and get their disciplinary messages across. This indicates that disciplinary variation could impinge on the choices of the lexical bundles. On the whole, hard disciplines offered a larger number of lexical bundle use compared to that of soft disciplines. There were also some notable similarities and differences in the structural types of the bundles used. Noun and prepositional phrase fragments were the most commonly used structures in soft science corpus, while these structures along with dependent clause fragments were the most prevalent in hard sciences. As regards to the functional perspectives of the lexical bundles used in the lectures, findings reported the higher concentration of referential expressions in the soft science lectures, whereas lecturers in hard sciences showed a greater tendency towards the use of discourse organizers. The newly added category, lecture specific expressions, accounted for only a partial proportion of bundle types in the two sciences and was found to be more common in hard disciplines.

The dissimilarity in the form and function of lexical bundles in academic lectures of different disciplinary sciences emphasizes pedagogical implications of these formulaic expressions for EAP practitioners. It is essential to raise the students' awareness towards the use of these multi-word sequences as they are frequently used in university setting and functionally positioned in the discourse. EAP course designers should also be aware of disciplinary similarities and differences in order to relegate them properly in their syllabus and teaching materials using the language specific approach. English as a Second Language (ESL) students often need an intensive orientation in learning the language to complement their disciplinary studies, and this awareness of lexical bundles could go a long way in helping them in their academic studies. In a sense, the recognition of such formulaic use of language when it becomes deep seated could then become easily retrievable from memory. Hearing them uttered leads to automatic association between form and meaning and achieving this purpose would be one of the ultimate aims in learning. An overt display of lexical bundles sourced from real life examples could form an engaging platform for discussion and learning. It would also provide specific oral training in the manner of listening to how oral deliveries (with focus on lexical bundles) are made and how students themselves could also speak in such academic settings.

THE AUTHORS

Hadi Kashiha is a Ph.D candidate in applied linguistics at University Putra Malaysia. His main research interests cover formulaic language, academic writing, genre and discourse studies.

Email: hadi.kashiha@yahoo.com

Chan Swee Heng teaches as a professor in the Department of English at Universiti Putra Malaysia. Her main research interests are in the areas of assessment, writing, text and discourse studies.

Email: shchan@upm.edu.my

REFERENCES

- Adel, A., & Erman, B. (2012). Recurrent word combinations in academic writing by native and non-native speakers of English: A lexical bundles approach. *English for Specific Purposes, 31*, 81–92.
- 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. (1999). Lexical bundles in conversation and academic prose. In H. Hasselgard & S. Oksefjell (Eds.), *Out of corpora* (pp. 181–190). Amsterdam: Rodopi.
- Biber, D., Conrad, S., & Cortes, V. (2004). If you look at...: Lexical bundles in university teaching and textbooks. *Applied Linguistics, 25*, 371–405.
- Cortes, V. (2002). *Lexical bundles in published and student academic writing in history and biology*. (Unpublished doctoral dissertation), Northern Arizona University, Arizona.
- Cortes, V. (2004). Lexical bundles in published and student disciplinary writing: examples from history and biology. *English for Specific Purposes, 23*, 397–423.
- Cortes, V. (2006). Teaching lexical bundles in the disciplines: An example from a writing intensive history class. *Linguistics and Education, 17*(4), 391–406.
- Csomay, E., & Cortes, V. (2010). Lexical bundle distribution in university classroom talk. In Gries, S. T., Wulff, S. & Davies, M. (Eds.), *Corpus linguistic applications: Current studies, new directions* (pp. 153-168). New York: Rodopi.
- Firth, J. R. (1951). *Modes of meaning. Essays and studies*, Oxford: Oxford University Press.
- Granger, S., & Meunier, F. (2008). Phraseology in language learning and teaching. Where to from here? In Meunier, F. & Granger, S. (Eds.), *Phraseology in foreign language learning and teaching* (pp. 247-252). John Benjamins.
- Hyland, K. (2008). As can be seen: Lexical bundles and disciplinary variation. *English for Specific Purposes, 27*, 4-21.
- Jablonkai, R. (2010). English in the context of European integration: A corpus-driven analysis of lexical bundles in English EU documents. *English for Specific Purposes, 29*, 253–267.
- Jespersen, O. (1924). *The philosophy of grammar*. London: Allen & Unwin.

- Meunier, F., & Granger, S. (2007). *Phraseology in foreign language learning and teaching*. Amsterdam: John Benjamins Publishing.
- Nesi, H., & Basturkmen, H. (2006). Lexical bundles and discourse signaling in academic lectures. *International Journal of Corpus Linguistics*, 11, 147–168.
- Pawley, A., & Syder, F. (1983). Two puzzles for linguistic theory: Nativelike selection and nativelike fluency. In J. Richard & R. Schmidt (Eds.), *Language and communication* (pp. 191–226). London: Longman.
- Schmitt, N., & Carter, R. (2004). Formulaic sequences in action: An introduction. In N. Schmitt (Ed.), *Formulaic sequences: Acquisition, processing, and use* (pp. 1–22). Philadelphia: John Benjamins.
- Scott, M. (2004). *Wordsmith Tools* version 4.0. Oxford: Oxford University Press.
- Strunkyt, G., & Jurkūnait, E. (2008). *Written academic discourse: Lexical bundles in humanities and natural sciences*. (Unpublished master's thesis). Vilnius University, Lithuania.
- Thompson, S. (1994). Frameworks and contexts: A genre-based approach to analysing lecture introductions. *English for Specific Purposes*, 13(2), 171–186.
- Tracy-Ventura, N., Cortes, V., & Biber, D. (2007). Lexical bundles in Spanish speech and writing. In P. Giovanni (Ed.), *Working with Spanish corpora*. (pp. 354- 375). London: Continuum.
- Wray, A. (2000). Formulaic sequences in second language teaching: Principle and practice. *Applied Linguistics*, 21(4), 463- 489.
- Wray, A. (2008). *Formulaic language: Pushing the boundaries*. Oxford: Oxford University Press.
- Wray, A., & Perkins, M. (2000). The functions of formulaic language. *Language and Communication*, 20, 1–28.