Structural Analysis of Lexical Bundles in University Lectures of Politics and Chemistry

Hadi Kashiha (Corresponding author)

Department of English, Faculty of Modern Languages and Communication, University Putra Malaysia, Malaysia

E-mail: hadi.kashiha@yahoo.com

Chan Swee Heng

Department of English, Faculty of Modern Languages and Communication, University Putra Malaysia, Malaysia

E-mail: shchan@upm.edu.my

Abstract

Referred to as extended collocations, lexical bundles are considered as a main factor in building fluency in academic discourse; helping to shape meaning and coherence in a text or speech. For decades, lexical bundles have attracted considerable amount of attention in corpus-based research in English for Academic Purposes (EAP). While, the focus of the most of the studies on lexical bundles was to explore the use of these multi-word expressions in academic written registers such as research articles, academic spoken registers such as university lectures have not received that amount of attention from the scholars. In this vein, there is still an open question of how they are structurally different across disciplines. With these concerns in mind, this study aimed to explore how lexical bundles are used structurally in a 50291 words corpus of 8 university lectures across two disciplines: chemistry and politics. To this aim, the most frequent four-word bundles in the corpus were classified according to their grammatical types to see the possible disciplinary variations in their frequency of use as well as the structure involved in their use. Results of the analysis revealed that noun phrase and prepositional phrase fragments were the most common structures in the lectures of the two disciplines, accounting for more than half of the bundles in politics. University lecturers appear to apply a variety of structures in the use of lexical bundles often peculiar to the discipline in order to convey their disciplinary messages. This would lead to the need to emphasize the instruction of the most common structures in that discipline in a way for the lectures to be as comprehensive as possible for the intended audiences.

1. Introduction

Research in the area of multi-word combinations has a long history in applied linguistics, starting with the work by Jespersen (1924) who introduced the term "collocation". Since then, many studies have attempted to single out the importance of using frequent word sequences in academic discourse, as these expressions are considered as building blocks of discourse (Biber & Barbieri, 2007) thus resulting in better fluency among speakers of a language. This type of multi-word expressions which recur frequently with a sequence of three or more words is referred to as lexical bundle (Biber & Conrad, 1999), such as *in terms of the*, found prevalently in writing, or *do you want me to* in speech. Lexical bundles, in essence, are combinations of words that normally co-occur more recurrently than we expect. This frequency of occurrence is the defining characteristic of lexical bundles and distinguishes them from other types of multi-word expressions such as idioms. For instance, the well-quoted idiom "*kick the bucket* "occurs less than 1 time in one million words, while there are lexical bundles which occur more than 20 times per million words (Cortes, 2004). Moreover, lexical bundles are not idiomatic in meaning, since unlike idioms, their meaning can be easily retrieved from the analysis of individual words which make them up.

It is widely realized that frequent use of lexical bundles can lead to better English proficiency among language learners. In other word, these bundles are considered as the main source of fluent native speech in which nonnative learners could target as useful resources in the learning of the language. Academic lecturers, especially those working in EFL/ESL contexts, can also benefit from the working knowledge of lexical bundles used in relation to disciplines as they help build up academic voice and self-confidence in developing knowledge of language use in various discourse communities so as to develop relevant materials for teaching and learning.

A considerable body of research has in fact shown that listening to long academic lectures could be a major problem for nonnative learners (Cheng, 2012; Flowerdew & Miller, 1997; Thompson, 2003; Vidal, 2003; Young, 1994). While listening to academic lectures, even proficient students who may understand the individual words uttered by the lecturer might fail to realize content information of relevant discipline-specific terms. One reason partly comes from the fact

that most of these students do not recognize the structural relations that exist between the word elements and thus this would result in only partial mastery of knowledge input. This leads back to the underlying importance of understanding the grammatical characteristics of lexical bundles used by the lecturer. The difficulties would lead to the need to explore the structures of lexical bundles in university lectures of different disciplines to provide insight towards better comprehension of academic lectures.

The last two decades have witnessed a growing interest in the study of lexical bundles in English for Academic Purposes (Adel & Erman, 2012; Biber, Conrad & Cortes, 2004; Cortes, 2004, 2006; Hyland, 2008). Many of these studies are corpus-based, analyzing the use of lexical bundles in a variety of academic registers such as research articles, textbooks and classroom teaching. Disciplinary variations in the use of lexical bundles have also been the focus of many other studies (Cortes, 2004; Hyland, 2008), but the main purpose of these studies was to explore the bundle use in academic written discourse. Spoken discourse, on the other hand, has not received that amount of attention from the researchers. Among the few related studies on academic speech, Nesi and Basturkmen (2009) explored the cohesive role lexical bundles play in academic lectures of BASE and MICASE online corpora and found that consistency and organization of university lectures depend heavily on the frequent use of these bundles. Lecturers in academic contexts appear to make use of a range of lexical bundles to link different parts of the discourse. In their corpus, nearly all the target bundles were used to signal discourse relations with some exceptions. In some parts of his study, Khuwaileh (1999) investigated the effect of lexical bundles in lecture comprehension of Jordanian learners. Neely and Cortes (2009) examined the discourse function of five lexical bundles (*if you look at, a little bit about, a little bit of, I want you to,* and *I would like you)* which were frequently used to introduce new topics and organize the discourse in spoken academic language, and compared the lecturers' use of these bundles to those of students. They came to the conclusion that students should be taught lexical bundles in a way that would become familiar to their everyday academic lives. Through a review of a related literature, it was found that research on the disciplinary variations of lexical bundles in academic lectures is minimal. To this aim, the present study attempts to shed light on the structural characteristics of the most frequent four-word lexical bundles used in university lectures of two disciplines, namely chemistry and politics, to come up with possible similarities and differences in use. The rationale behind comparing the two disciplines results from the necessity to conduct such comparative studies on the use of different linguistic features in social and physical sciences to arrive at the better picture of the variations in the language use of hard and soft fields. Chemistry, representing hard fields, is more physical and observable, dealing with experiments and tools, while politics is a representative of soft fields and deals with human behavior. The comparison between the language use of these disciplines with such different characteristics has always been the focus of many research-based studies and would certainly provide valuable results for both students and lecturers working in ESL/EFL contexts in dealing with or presenting disciplinary materials.

2. Research objectives

The main objective of this study is to characterize the structural attributes of lexical bundles in university lectures of two disciplines, chemistry and politics. To this aim, the following research questions are presented:

1. What are the grammatical types of four-word lexical bundles in the lectures of chemistry and politics?
2. To what extent and how the university lectures in politics and chemistry differ in terms of structure or grammatical form of the lexical bundles used?

3. Corpus and method

The corpus of this study consists of 8 university lectures transcripts from the BASE online corpus across two disciplines of chemistry and politics (4 lectures from each). The two disciplines were selected on the criterion of comparing the representatives of hard and soft sciences. In addition, the 4 lectures in each discipline were the only available transcripts on transactional lectures which had the same range of words, because in order to make the data as comparable as possible, word counting is also taken into consideration. Both disciplines contain more than 25000 running words and 50291 words overall. Table 1 outlines the detailed content of the corpus.

Table1. Details of the corpus

Disciplines	No. of lectures	Word count
Chemistry	4	25106
Politics	4	25185
Total	8	50291

The lecture transcripts were first downloaded from British Academic Spoken English (BASE) corpus and then grouped according to the disciplines. Each discipline was assigned 4 lectures to be analyzed. There were some steps that were followed in analyzing the transcripts. The first step was to set the criterion for a frequency cut-off point to come up with a list of the most salient four-word lexical bundles. There is a common consensus among the scholars that a frequency cut-off point is something arbitrary, and thus differs from one study to another, depending on factors such as the size of the corpus. For example, Biber and Barbieri (2007) argued that a four-word lexical item has to occur at least 40 times per million words in order to be called a lexical bundle. Due to the small size of the corpus, the present study took a rather conservative approach by setting the frequency point at occurring 20 times per hundred thousand words which represents at least 5 raw occurrences in the whole corpus. In addition, in order to avoid the lecturers' idiosyncratic use, a

four-word item also had to be used in at least 3 different lectures. Only those four-word strings which met the above criteria were identified as lexical bundles and thus selected for the analysis. To manage the research meaningfully, this study decided to focus on only four-word lexical bundles because previous literature have proved that four-word bundles are the most prevalent sequences of word forms in academic setting (Biber & Barbieri, 2007; Biber, Conrad, & Cortes, 2004; Cortes, 2002, 2004; Hyland, 2008). They 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). This study used a computer program AntConc to create a list of the most frequent four-word lexical bundles in each group.

Once the list of the target bundles was prepared, the identified four-word lexical bundles were then categorized structurally using the classification proposed by Biber, et al. (2004). Their structural taxonomy includes three main grammatical types for lexical bundles: 1) lexical bundles that incorporate verb phrase fragments such as *is based on the*; 2) lexical bundles that incorporate dependent clause fragments like *I want you to*; and 3) lexical bundles that incorporate noun phrase and prepositional phrase fragments such as *at the end of* (Biber et al., 2004, p.381). Each main structural type has several sub-structures which are listed in Table 2.

Table 2. Structural classifications of lexical bundles (Biber, *et al.*, 2004, p.381)

Structural categories	Sub-categories	Sample bundles
1. Lexical bundles that incorporate verb phrase fragments	1a. 1st/2nd person pronoun + VP fragment	*I'm not going to*
	1b. 3rd person pronoun + VP fragment	*and this is a*
	1c. discourse marker + VP fragment	*I mean I don't*
	1d. Verb phrase (with non-passive verb)	*have a lot of*
	1e. Verb phrase (with passive verb)	*is based on the*
	1f. yes-no question fragments	*are you going to*
	1g. WH- question fragments	*what do you think*
2. Lexical bundles that incorporate dependent clause fragments	2a. 1st/2nd person pronoun + dependent clause fragment	*I want you to*
	2b. WH-clause fragments	*when we get to*
	2c. If-clause fragments	*if we look at*
	2d. to-clause fragment	*to be able to*
	2e. That-clause fragment	*that this is a*
3. Lexical bundles that incorporate noun phrase and prepositional phrase fragments	3a. Noun phrase with of-phrase fragment	*one of the things*
	3b. Noun phrase with other post-modifier fragment	*the way in which*
	3c. Other noun phrase expressions	*a little bit more*
	3d. Prepositional phrase expressions	*at the end of*
	3e. Comparative expressions	*as well as the*

4. Result and discussion

Altogether, there were 225 individual bundles in the corpus, including 32 different bundle types. The bundles *we are going to* and *to talk about the* were the most frequently used bundles in the whole corpus, occurring 14 and 12 times in chemistry, and 18 and 15 times in politics respectively. The high occurrence of these bundles in academic lectures suggests that university lecturers rely heavily on using topic introduction markers to raise the students' awareness towards the forthcoming information. Table 3 illustrates the distribution of lexical bundles in the two disciplines. As can be seen, politics lecturers used a larger range of lexical bundles than those of chemistry, with 131 individual cases as compared with 94 respectively. Lectures in chemistry also reported a lower number of bundle types (26 different bundles); while politics lectures were found to use all the 32 different bundle types in the corpus. The higher tendency of lecturers in politics to use lexical bundles may translate into the possibility that, explanatory nature of this discipline in which many ideas needed to be connected and argued, requires the lecturers to use a variety of multi-word chunks in order to convey their disciplinary messages. The chemistry discipline, on the other hand, depends more on the reporting of or focusing on observables such as those of experiments, in which less obvious connection is involved to convey the ideas.

Table 3. Distribution of lexical bundles in the two disciplines

Disciplines	Bundle types	Total No. of bundles	% of total words
Chemistry	26	94	1.88
Politics	32	131	2.62

4.1 Structures of lexical bundles in chemistry and politics

Structurally, the two disciplines showed some similarities and variations in the grammatical types of lexical bundles used by the lecturers. Tables 4, 5 and 6 illustrate the detailed percentages of the three main structural categories with their specific sub-categories in the lectures of the two disciplines using the classification proposed by Biber et al. (2004, p. 381). An initial examination of the data indicates that both politics and chemistry lectures used all the three main structural types, which shows the lecturers' flexibility in selecting different structures in constructing lexical bundles. The main purpose of this diversity of use is likely to better communicate the content of the lessons and thus help the students to comprehend the specific disciplinary messages.

Concerning the first main structural category "verb phrase fragments", Table 4 indicates that this structure was distributed almost equally in both groups of lectures. This suggests that verbal elements are the key characteristics of academic university discourse in conveying the information to the listeners. However, chemistry lecturers reported a slightly higher percentage of use than those of politics instructors (26% as compared with 23% respectively). This slight difference could be actualized by the fact that the materiality nature of chemistry required the lecturers to use more verbal phrases to guide the students towards the instruction, especially in the case of bundles that signal obligation, such as *you don't have to*, *you only need to* or direction like *it is helpful to*, *it is necessary to*.

Table 4. Lexical bundles that incorporate verb phrase fragments

Structural categories	Sub-categories	Politics	Chemistry
1. Lexical bundles that incorporate verb phrase fragments	1a. 1st/2nd person pronoun + VP fragment	*11%*	*15%*
	1b. 3rd person pronoun + VP fragment	*5%*	*3%*
	1c. Discourse marker + VP fragment	*	*
	1d. Verb phrase (with non-passive verb)	*4%*	*3%*
	1e. Verb phrase (with passive verb)	*1%*	*3%*
	1f. yes-no question fragments	*1%*	*1%*
	1g. WH- question fragments	*1%*	*1%*
Total		*23%*	*26%*

There were also some similarities and minimal differences in the sub-categories. "1st/2nd person pronoun + VP fragment" (e.g. *you are going to, I would like to*) was the most common sub-category, accounting for 15% of the bundle types in chemistry, and 11% in politics. Lecturers in the two disciplines used this structure either to initiate their discussions or to show the intention, as in:

> ***I would like to*** speak about briefly problem solving or the integrated approach. ***(Politics)***

> … but eventually ***you are going to*** saturate the surface with oxygen this is the … ***(Chemistry)***

In contrast, politics lecturers were slightly more inclined to the use of "3rd person pronoun + VP fragment" (e.g. *it is likely to, it is difficult to*) to provide more information about the topic being discussed, as in:

> that in order to build up a militarily useful nuclear arsenal ***it is necessary to*** test weapons and these tests in most countries still would have to be done physically.

Though less frequent, "verb phrase (with passive verb)" such as *is ought to be* was more dominant in the lectures of chemistry, with 3% as compared to 1% in politics. On the other hand, "Verb phrase (with non-passive verb)" such as *have a look at* was slightly more favored by politics lecturers. The two disciplines reported a similar rate of use in question fragments (yes-no question and WH-question fragments), accounting for the least proportion of bundle types. No example of "discourse marker + VP fragment" was found in the lectures of both chemistry and politics.

As regards to the second main structural type, "dependent clause fragments", Table 5 shows that chemistry lectures utilized a higher proportion of use, comprising 36% of the types, while this structure only constituted 24% of the bundles in politics. The popularity of this structure in chemistry lectures could be resulted from the physical nature of this discipline based on which more consciousness-raising and pointing function are required in order to facilitate the process of comprehending the chemistry lectures. One way of raising the students' awareness is through using a variety of dependent clause structures, especially in the case of "if clause fragments" such as *if you look at, if you think about*, as in:

> ***if you think about*** it, a hundred electron volts ionization of a typical organic is about ten *(Chemistry)*

> because ***if you look at*** cyclohexane you've got the two pairs of hydrogens *(Chemistry)*

Table 5. Lexical bundles that incorporate dependent clause fragments

Structural categories	Sub-categories	Politics	Chemistry
2. Lexical bundles that incorporate dependent clause fragments	2a. 1st/2nd person pronoun + dependent clause fragment	*3%*	9%
	2b. WH-clause fragments	*9%*	6%
	2c. If-clause fragments	*6%*	11%
	2d. to-clause fragment	*6%*	10%
	2e. That-clause fragment	*	*
Total		*24%*	**36%**

The use of this sub-structure was also dominant in chemistry lectures, occurring almost double more than those of politics. In some cases, the lecturers used the first person plural "*we*" to make the learners feel intimacy and confidence in the class, as in:

> and so ***if we know the*** extinction of this, we can work out the yield of the… *(Chemistry)*

The other sub-categories of dependent clause fragments were also higher in chemistry. Table 6 indicates that lecturers in chemistry were found to use "1st/2nd person pronoun + dependent clause fragments" such as *I don't know if*, *I don't know what,* about three times more than their colleagues in politics. They were also dominant in using "to-clause fragments" like *to be able to*, *to look at the*, *to talk about the* (including 10% of the types as compared with 6% in politics). In contrast, the only sub-category which was used more in the lectures of politics was "WH-clause fragments". It constructed 9% of the bundle types in politics, while only 3% of the bundles in chemistry were constituted in this way. In comparison, neither politics nor chemistry lecturers seemed to be interested in using "that-clause fragments", since no example of this structure was found in the lectures of the two disciplines.

Results of the analysis showed that the third main structure "noun and prepositional phrase fragments" accounted for the most bundle types in the lectures of the two disciplines, including more than half of the bundles in politics and 38% of the bundles in chemistry. This high reliance of academic lecturers on these structures is indicative of the characteristics behavior of oral language in which speakers rely more on using a range of noun-preposition combinations (e.g. *at the end of, quite a long time, a little bit of*) to communicate, rather than resorting mainly to verbs. Findings seem to be in contrast with the previous studies on academic English. Earlier findings showed that academic speech primarily comprises more lexical bundles with clause fragments, while academic writing reported to use more bundles incorporating noun and prepositional phrase fragments (Biber & Conrad, 1999; Biber et al., 2004; Hyland, 2008).

In contrast to the second structural category which was particularly popular among the chemistry lecturers, the third main structure, "noun and prepositional phrase fragments" was reported to be more favored by politics lecturers. As for the disciplinary variations, lecturers in politics tended to use this structure more than those in chemistry, (53% compared with 38% in chemistry). The difference in the use shows that the language of politics is more varied than that of chemistry, with more noun and preposition expressions. This corresponds to the idea that soft science disciplines, like politics, describe human-related issues and one of the important issues regarding humans is the diversity of behavior. One way to describe or portray different characteristics of people is by using a range of noun and prepositional phrases.

A closer look at the sub-categories of this main structure (see Table 6) reveals that most of the bundles were "noun phrase with *of*-phrase fragments" such as *a little bit of, at the end of, a certain amount of, in the number of*. This structural type constituted almost a quarter of the bundles in politics. It was also considered as the most common structure in chemistry and formed 20% of the bundles.

Table 6. Lexical bundles that incorporate noun phrase and prepositional phrase fragments

Structural categories	Sub-categories	Politics	Chemistry
3. Lexical bundles that incorporate noun phrase and prepositional phrase fragments	3a. Noun phrase with *of*-phrase fragment	*25%*	*20%*
	3b. Noun phrase with other post-modifier fragment	*3%*	*3%*
	3c. Other noun phrase expressions	*8%*	*4%*
	3d. Prepositional phrase expressions	*15%*	*11%*
	3e. Comparative expressions	*2%*	*
Total		*53%*	**38%**

The second common sub-category of noun and prepositional phrase fragments was "prepositional phrase expressions" such as *in a different way, in the first place, for a short time*. Politics, as well yielded a bigger proportion of this structure (15%) compared to chemistry (11%). This greater dominance was also the same for other sub-categories. For example, the structure "other noun phrase expressions" like *the first thing to, a few more examples*, was used in politics double more than chemistry (8% compared with 4% respectively). There was no example of "comparative expressions" like *as well as a* in the chemistry lectures, while this structure comprises 2% of the bundles in politics. Finally, "noun phrase with other post-modifier fragment" such as *a little bit about* was distributed similarly in the two disciplines, comprising 3% of the bundle types, like in:

> what i'd like to do now is to say ***a little bit about*** water water is probably the most important system ***(Chemistry)***

> it's ***the way in which*** you achieve your objectives power ***(Politics)***

5. Conclusion

The main purpose of the present study was to identify the structural characteristics of the most frequent four-word lexical bundles in university lectures of two disciplines, politics and chemistry. In order to pinpoint any similarities and differences, the frequency as well as the structural analysis was conducted. The analysis of the frequency indicated that, overall, the four-word lexical bundles were more frequently used in the lectures of politics, which suggests that lecturers in politics relied more heavily on the use of multi-word expressions in presenting their disciplinary materials in English.

Structurally, the two disciplines were found to use the three main structural categories in a distinctive way. Noun and prepositional phrase fragments were the most frequent structural type used in the lectures of the two disciplines, with *of*-phrase fragment being the most prevalent sub-category. However, this structure was more favored by the politics lecturers, which is indicative of the fact that the language of the lectures in politics was more varied than those of chemistry. In contrast, dependent clause fragments were more popular among chemistry lecturers. For example, they appeared to use a range of *to*-clause structures such *to look at the* or *to talk about the* in order to attract the students' attention towards the coming information. The first structural category, verb phrase fragments, was, however, reported to be used almost at a similar rate in the two disciplines. This shows the dependence of academic lectures on the use of a variety of verb phrases to convey their messages in their specific disciplines.

Multi-word expressions are commonly used in university lectures. The findings of this study would contribute to the need to learn the structural characteristics of multi-word sequences in general and lexical bundles in specific in order to ease the problem of lecture comprehension, as lexical bundles are building blocks of the academic discourse, contributing to coherence in speech. Therefore, explicit teaching of their structures would be a useful approach in the language classroom to facilitate the acquisition process of these formulaic expressions in academic setting.

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 (2) 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.

Cheng, S. W. (2012). That's it for today: Academic lecture closings and the impact of class size. *English for Specific Purposes,* 31, 234–248.

Cortes, V. (2002). Lexical bundles in published and student academic writing in history and biology. Unpublished Doctoral Dissertation, Northern Arizona University.

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.

Flowerdew, J., & Miller, L. (1997). The teaching of academic listening comprehension and the question of authenticity. *English for Specific Purposes*, 16(1), 27–46.

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

Jespersen, O. (1924). *The philosophy of grammar*. London: Allen & Unwin.

Khuwaileh, A.A. (1999). The role of chunks, phrases and body language in understanding co-ordinated academic lectures. *English for Applied Studies,* Jordan University of Science and Technology. 249-260.

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

Nesi, H. , & Basturkmen, H. (2009). Lexical bundles and discourse signalling in academic lectures. *Lexical Cohesion and Corpus Linguistics*, *17*, 23.

Thompson, S. E. (2003). Text-structuring metadiscourse, intonation and the signalling of organization in academic lectures. *Journal of English for Academic Purposes*, *2*(1), 5–20.

Vidal, K. (2003). Academic listening: A source of vocabulary acquisition?. *Applied Linguistics*, *24*(1), 56-89.

Young, L. (1994). University lectures – macro-structure and micro-features. In J. Flowerdew (Ed.), *Academic listening* (pp. 159–176). Cambridge: Cambridge University Press.