Analyzing the functions of lexical bundles in undergraduate academic lectures for pedagogical use

Chen-Yu Liu, Hao-Jan Howard Chen*

National Taiwan Normal University, 162, Section 1, Heping E. Rd., Taipei City 106, Taiwan

ARTICLE INFO

ABSTRACT

Good listening comprehension contributes greatly to academic performance. One factor in developing good listening comprehension is knowing the functions of lexical bundles. While some research has been done on lexical bundles, their functions in academic lectures have not been extensively investigated. Drawing on an 8.8-million-word corpus of academic lectures across four disciplines, we investigated the functions of high-frequency lexical bundles in academic lectures and examined the similarities and differences in the four disciplines. We discovered that lexical bundles most often perform referential and stance functions in academic lectures, although their function as discourse organizers is also important. Given these and other functions of high-frequency lexical bundles, instruction targeting functions can contribute to improving core listening skills, such as *listening for details* and *listening and predicting,* and overall comprehension of lectures. Apart from important functions, we also found variations and similarities in the use of lexical bundles across disciplines, suggesting that these should be taken into account in bundle instruction to help students become more proficient in lecture comprehension.

© 2019 Elsevier Ltd. All rights reserved.

1. Introduction

The number of L2 English speaking students enrolled in universities in English-speaking countries continues to grow. To prepare for the challenge of studying in an English speaking university, L2 students often take English language courses (Flowerdew & Miller, 1997), which typically focus on academic vocabulary, grammar, and language skills (listening, speaking, reading, and writing). These courses are important as they can help students develop listening competence, which contributes greatly to academic performance (Jeon, 2007).

Among the many issues to address in helping students develop their listening comprehension, attention needs to be given to the teaching of lexical bundles, frequently occurring multi-word sequences. Research shows that lexical bundles can contribute to listening comprehension (Neely & Cortes, 2009), as they help listeners better follow the structure of a lecture (Csomay & Cortes, 2010; Neely & Cortes, 2009) and provide discourse signals to help listeners "predict the nature of upcoming ideas and information" (Nesi & Basturkmen, 2006, p. 301). An inability to recognize them affects comprehension in lectures (Goh, 2013) and leads to "additional cognitive processing demands" (Nesi & Basturkmen, 2006, p. 301). Biber and Barbieri (2007) also point out that "failure to understand their (lexical bundles') textual and interpersonal functions will obviously influence student success in dealing with both spoken/written language situations [in the university context]" (p. 284). It thus

---

follows that teaching high-frequency lexical bundles and how they function in academic lectures can have an overall benefit for student success in their degree programs.

According to Biber and Barbieri (2007), "each register employs a distinctive set of lexical bundles, associated with the typical communicative purposes of that register" (p. 265), which may also be true for academic lectures. While many studies have investigated lexical bundles and their functions in various academic registers (Biber, Conrad & Cortes, 2004; Biber & Barbieri, 2007; Simpson-Vlach & Ellis, 2010), few have primarily investigated their use in academic lectures and, of these, typically the research focus has been limited to the use of a small number of lexical bundles together with the particular functions they perform. To broaden our understanding of the use of lexical bundles and their functions in academic lectures, we have compiled and analyzed an 8.8-million-word corpus of academic lectures across four disciplines and provide a rich resource that can inform the teaching of lexical bundles in ESP/EAP lecture comprehension courses.

1.1. Lexical bundles across registers

Lexical bundles are word combinations that commonly co-occur in a register and are identified statistically, as opposed to intuitively, based on a cut-off frequency determined arbitrarily by a researcher (Cortes, 2004). Most studies have applied a normalized frequency cut-off from ten to forty times per million words. Another criterion used is the dispersion rate, which can avoid the issue of user idiosyncrasies when identifying lexical bundles that appear in a range of texts (Hyland, 2008; Salazar, 2011). One interesting approach to identifying lexical bundles was undertaken by Simpson-Vlach and Ellis (2010), who used both statistical measures and teacher insights regarding relevance to extract academic lexical bundles (or "formulas" as they termed) that they believe to be psycholinguistically salient and pedagogically useful. The authors then further categorized the lexical bundles into three lists—the Core Academic Formulas List (AFL), the Spoken AFL, and the Written AFL, which are particularly useful for EAP classroom teaching and materials development.

Compared to idioms, which have complete structures and opaque meanings (e.g. *kick the bucket*), the meaning of a lexical bundle is rather transparent and can be understood from the meaning of its individual components (e.g. *I want to show, take a look at*) (Biber, Johansson, Leech, Conrad, & Finegan, 1999). Lexical bundles have been widely researched across different registers by Biber and his colleagues (Biber & Barbieri, 2007; Biber & Conrad, 1999; Biber, Conrad, & Cortes, 2004; Biber et al., 1999), revealing important insights. For instance, findings show that lexical bundles behave differently across registers. Biber et al. (1999) found that the features of lexical bundles in academic prose differ from those found in general conversation. Biber, Conrad, and Cortes (2004) compared lexical bundles in university register (classroom teaching, textbooks, academic prose) and general conversation, and found that their use differs not only between written and spoken registers, but also across different university registers. For instance, Biber and Barbieri's (2007) research on the differences between spoken and written lexical bundles in a university context revealed that "there are fundamentally different sets of lexical bundles associated with spoken university registers in contrast to written registers" (p. 265). Given these differences, how lexical bundles are used in different registers is worthy of investigation since from an ESP perspective, each register can be considered distinct.

1.2. Discourse functions of lexical bundles

Lexical bundles in academic registers have been classified in terms of their function. For instance, Biber et al. (2004) introduced a taxonomy to categorize spoken and written lexical bundles based on their functions and the roles they serve in academic discourse. This taxonomy, which has been adopted by many studies on lexical bundles (e.g. Bal, 2010; Biber & Barbieri, 2007; Conrad & Biber, 2005; Cortes, 2004; Staples, Egbert, & McClair, 2013) consists of four main categories: stance expressions (expressing speakers'/writers' opinions or attitudes), discourse organizers (outlining or organizing texts, introducing or elaborating on topics), referential bundles (specifying focus and quantity, highlighting time or place), and special conversation functions (showing politeness, inquiring, and reporting).

To investigate the discourse functions of spoken and written lexical bundles in academic registers, Biber's taxonomy was modified by Simpson-Vlach and Ellis (2010), who adjusted some of the categories and added new subcategories. For example, they added several new subcategories, including *contrast and comparison, hedges*, and *evaluation*. The main category of *discourse organizers* substantially expands Biber et al.'s (2004) grouping, with the addition of three subcategories: *metadiscourse and textual reference*, *cause and effect*, and *discourse markers*. According to Simpson-Vlach & Ellis (2010), a taxonomy more extensive than Biber et al.'s (2004) was needed to classify the many more lexical bundles identified in their research. Furthermore, in expanding the taxonomy, the researchers clarified the functions performed by lexical bundles, making them more specific, in an effort to identify important and meaningful bundles for EAP teaching and learning.

Hyland's (2008) interest in written academic lexical bundles led him to develop a functional taxonomy different from Biber et al.'s (2004), which yielded insights into lexical bundles in published research articles. The three main categories in this taxonomy are: research-oriented (structuring activities, procedures, and topics of the text), text-oriented (associated with organization and relations between elements in the text) and participant-oriented bundles (expressing the writer's attitudes or engaging the reader). Hyland's (2008) taxonomy was further revised by Salazar (2011) to more accurately classify lexical bundles in a corpus of scientific journal papers, creating several new categories. For example, *contrastive* and *resultative* functions in Hyland's (2008) taxonomy were expanded into four functions, including *additive*, *referential*, *comparative*, and

causative. To describe the function of bundles related to stating general facts, supporting data by citing sources, or introducing writer aims, Salazar added *generalization*, *citation*, and *objectives*.

Table 1 presents a summary of the functional taxonomies reviewed above, including the taxonomies of Biber et al. (2004, pp.384–388), Hyland (2008, pp. 13–14), Salazar (2011, pp. 32–33), and Simpson-Vlach and Ellis (2010, pp. 498–502).

A small number of studies have examined lexical bundles in lectures (Biber et al., 2004; Biber & Barbieri, 2007; Simpson-Vlach & Ellis, 2010); however, some issues need to be addressed. For instance, while Nesi and Basturkmen (2006) drew on a corpus containing academic lectures only, their focus was mainly on the cohesive roles of lexical bundles in lectures. Further, the size of the lecture data, such as the 1.2-million-word data of classroom teaching used by Biber et al. (2004) and Biber and Barbieri (2007), is rather limited. Additionally, the sources of data to investigate academic spoken lexical bundles lack homogeneity in that they contain non-lecture content. For instance, the academic speech corpus used in Simpson-Vlach and Ellis (2010) was comprised of the Michigan Corpus of Academic Spoken English (MICASE) plus the files of academic speech in the British National Corpus (BNC), which contain lectures, seminars, office hours, student presentations, class discussions, and lab sessions. Considering register differences in lexical bundles, a bundle list generated from a corpus containing only academic lectures together with a focus on their functions is highly desirable.

This study has two goals. The first was to identify high-frequency lexical bundles and their discourse functions from a corpus of academic lectures across four disciplines, Engineering, Humanities and Arts, Science and Math, and Social Science,

Table 1
A summary of functional taxonomies of lexical bundles (Biber et al., 2004, pp.384–388; Simpson-Vlach & Ellis, 2010, pp. 498–502; Hyland, 2008, pp. 13–14; Salazar, 2011, pp. 32–33).

Functional taxonomies of spoken and written academic bundles		
	Biber et al. (2004)	Simpson-Vlach & Ellis (2010)
Referential bundles	Specification of attributes	Specification of attributes
	- Intangible framing attributes	- Intangible framing attributes
	- Tangible framing attributes	- Tangible framing attributes
	- Quantity	- Quantity specification
	Identification/focus	Identification and focus
	Imprecision	Vagueness markers
	Time/place/text reference	Deictics and locatives
	- Place reference	Contrast and comparison
	- Time reference	
	- Text deixis	
	- Multifunctional reference	
Stance expressions	Epistemic stance	Epistemic stance
	Attitudinal/modality stance	
	- Desire	
	- Obligation/directive	Obligation and directive
	- Intention/prediction	Intention/volition, prediction
	- Ability	Expressions of ability and possibility
		Evaluation
		Hedges
Discourse organizers/Discourse organizing functions	Topic introduction	Topic introduction and focus
	Topic elaboration/clarification	Topic elaboration
		- Non-causal
		- Cause and effect
		Metadiscourse and textual reference
		Discourse markers
Special Conversational functions	Politeness	
	Simple inquiry	
	Reporting	
Functional taxonomies of written academic bundles		
	Hyland (2008)	Salazar (2011)
Research-oriented	Location	Location
	Procedure	Procedure
	Quantification	Quantification
	Description	Description
	Topic	Grouping
Text-oriented	Structuring signals	Structuring
	Framing signals	Framing
	Transition signals	Additive
	Resultative signals	Causative
		Citation
		Comparative
		Generalization
		Inferential
		Objective
Participant-oriented	Stance features	Stance
	Engagement features	Engagement
		Acknowledgment

Table 2
Number of lectures and word count in each domain and discipline in the present corpus.

Discipline/Domain	Number of Lectures	Word Count
ENGINEERING (hard-applied)		
Materials Engineering	67	404,870
Electrical Engineering	65	596,659
Mechanical Engineering	65	683,311
Computer Science	65	674,264
Subtotals	262	2,359,104
HUMANITIES AND ARTS (soft-pure)		
History	80	571,503
Literature	80	517,622
Philosophy	79	508,348
Music	70	402,589
Subtotals	309	2,000,062
SCIENCE AND MATH (hard-pure)		
Biology	70	547,356
Chemistry	72	487,061
Mathematics	70	591,711
Physics	71	531,311
Subtotals	283	2,157,439
SOCIAL SCIENCE (soft-applied)		
Economics	69	688,866
Political Science	70	409,516
Psychology	67	673,403
Law	70	602,841
Subtotals	276	2,374,626
Totals	1,130	8,891,231

leading to the compilation of a pedagogically relevant list of lexical bundles and their functions. The second was to explore to what extent the frequency of lexical bundles and their functions differ across disciplines in lectures. Some studies (Cortes, 2004; Hyland, 2008) have discovered considerable variations in the frequency of forms and functions of lexical bundles across disciplines in academic writing, and many lexical bundles have been found to be discipline-specific. Despite the great differences among the knowledge domains covered by different disciplines in lectures, we thus hypothesized that there should be a certain number of inter-disciplinary lexical bundles that are core to the register and can reflect the register nature as well.

Typically, lexical bundles are described as word sequences consisting of three or more words. In this study, however, we chose to investigate only four-word bundles because many four-word and five-word bundles "hold three-word bundles in their structures" (Cortes, 2004, p. 401). Four-word bundles are also much more frequent than five-word bundles (Hyland, 2008) and perform a wider variety of functions (Cortes, 2004). To achieve our two goals, three research questions are addressed in this study:

1) What are the high-frequency four-word lexical bundles in the academic lecture corpus? How do they compare to those in Simpson-Vlach and Ellis' (2010) Spoken Academic Formulas List?
2) What are the discourse functions of these lexical bundles? What are the functional distributions of the lexical bundles in the corpus?
3) Are there disciplinary variations in the lexical bundles in academic lectures?

2. Method

To construct our corpus of academic lectures for this investigation, we collected 1,130 academic lectures from MIT OpenCourseWare, Open Yale Courses, and university webpages on YouTube. The lectures were given in 73 different undergraduate level courses taught in the United States, with 55 being introductory courses and 18 being upper-level courses. Appendix A presents the sources of the lectures and the number of lectures collected from each source.

The online open courseware provided both lecture videos as well as transcripts. To enrich the variety of subjects in the corpus and balance the coverage of the four disciplines, we also included academic lectures provided directly by universities to YouTube. Most of the lectures included subtitles, although a few relied on subtitles generated automatically by YouTube. These were examined by the researchers to confirm their accuracy before being included in the corpus.

The transcripts of each lecture were collected and each checked manually by the researchers to ensure its suitability to be included in the corpus. Before deciding to include a transcript in the corpus, the researchers watched a short section of the lecture to check that it was recorded in the context of a regular class and was not specifically made for online use. Clues that facilitated this determination included instructors informing students that the class was to be recorded for use on an online open course and instructors walking around classrooms full of students.

Table 3

Numbers of bundle types and token frequency in each category.

Category	Bundle types	Tokens
Referential expressions	73 (43%)	24,715 (42%)
Stance expressions	66 (38%)	23,910 (40%)
Discourse organizing functions	32 (19%)	10,684 (18%)
Total	171	59,309

Since the academic lectures were drawn from a wide range of subjects, in order to properly classify them into different domains and disciplines, we consulted the disciplinary classifications in Becher (1994) and those on MIT OpenCourseWare, Open Yale Courses and other open course platforms. The collected lectures were first categorized into 16 different disciplinary domains through checking the subjects of the lectures and the schools/departments that offered such lecture series. Based on Becher's (1994) disciplinary classification that categorizes all disciplines into four disciplinary groupings—hard-pure, hard-applied, soft-pure, and soft-applied—the 16 disciplinary domains were further grouped into four main disciplines: Engineering (hard-applied), Humanities and Arts (soft-pure), Science and Math (hard-pure), and Social Science (soft-applied). Each discipline in our corpus contains four different domains. The discipline of Engineering contains the domains of Materials Engineering, Electrical Engineering, Mechanical Engineering and Computer Science. The discipline of Humanities and Arts contains History, Literature, Philosophy and Music. Biology, Chemistry, Mathematics and Physics are included in the discipline of Science and Math. Social Science includes Economics, Political Science, Psychology and Law.

The corpus includes 262 lectures in Engineering (2,359,104 words), 309 lectures in Humanities and Arts (2,000,062 words), 283 lectures in Science & Math (2,157,439 words) and 276 lectures in Social Science (2,374,626 words). In total, the corpus contains 8,891,231 words. Table 2 below summarizes the number of lectures and the word count for each domain in each discipline.

WordSmith Tools 6 (Scott, 2011) was used to identify the four-word lexical bundles in the corpus and the extracted lexical bundles were listed according to frequency. As the set of frequency cutoffs is somewhat arbitrary (Biber et al., 1999) to focus on higher frequency bundles and to keep the number of extracted bundles manageable, this study set a relatively high threshold, with the frequency cutoff being 40 times per million words. Additionally, the criterion of text dispersion rate was set at 10% to avoid the effects of speaker idiosyncrasies (Hyland, 2008).

The extracted four-word lexical bundles were then compared with those identified by Simpson-Vlach and Ellis (2010) to explore possible differences. Since the Core AFL contains both written and spoken bundles and the Written AFL contains written bundles only, we compared the current lexical bundles with those in the Spoken AFL. Also, since the current study focused only on four-word lexical bundles, while the Spoken AFL contains 3-, 4-, and 5-word bundles, we compared the current lexical bundles with only the four-word bundles in the Spoken AFL.

To investigate the functions of the lexical bundles extracted from the current corpus, we used Simpson-Vlach and Ellis' (2010) taxonomy for bundle classification of different functions. This taxonomy was chosen because it most closely aligned with our interest in identifying the lexical bundles that would be most worthwhile to teach. Previous comparison among taxonomies showed that, although both Biber et al.'s (2004) and Simpson-Vlach and Ellis' (2010) taxonomies are potentially useful for categorizing functions of spoken academic bundles, Simpson-Vlach and Ellis' (2010) is much more extensive in terms of the variety of forms and functions compared to Biber et al.'s (2004). We also believed that a more extensive functional taxonomy would be more useful for bundle categorization and provide more insights regarding register characteristics.

In the process of classifying the bundles into different discourse functions, the bundles from the academic lectures were compared with those in the AFL. Bundles not found in the AFL were categorized according to the definitions of each functional category in Simpson-Vlach and Ellis' (2010) taxonomy. When we disagreed, we returned to the original transcripts to check how a bundle was used in context and to compare it with functional bundles identified by previous studies. Our inter-rater reliability was high (93%).

To explore disciplinary variation in the use of bundles, we identified the frequently-used bundles for each individual discipline bundles and classified them, as well.

3. Results and discussion

In response to the first research question regarding the high-frequency four-word lexical bundles in academic lectures and their similarity with those on Simpson-Vlach and Ellis'(2010) list, 171 four-word high-frequency lexical bundles in the academic lectures were extracted from the corpus with a total frequency of 59,309 tokens. These 171 lexical bundles were then compared with the Top 200 lexical bundles in Simpson-Vlach and Ellis' (2010) Spoken AFL. It was found that, while the Spoken AFL is more comprehensive—it includes three-word, four-word and five-word lexical bundles—only 70 four-word lexical bundles were included on their list. This is important because four-word lexical bundles are pedagogically more useful due to their higher frequency, more complete syntactic structure, and a wider variety of functions they can perform. Given our greater number of lexical bundles, we believe that the pedagogical value of our list is likely more useful for teaching lecture comprehension.

Our comparison showed that around one third of our lexical bundles are the same as those in the AFL. This small amount of overlap may be due to the different corpora examined, ours being comprised of only academic lectures, while theirs

Table 4
List of high-frequency lexical bundles in academic lectures categorized by their functions.

Referential expressions			
(1) Specification of attributes			
a. Intangible framing attributes			
a certain kind of	in terms of the	on the order of	the way this is
as a function of	in the case of	so in this case	the ways in which
as a kind of	in the context of	such a way that	with respect to the
in a way that	is a kind of	the direction of the	
in such a way	of the things that	the way in which	
b. Tangible framing attributes			
parts of the world	the size of the	the square root of	the value of the
the energy of the	the speed of light	the sum of the	
c. Quantity specification			
a little bit more	a lot of people	a whole bunch of	the rest of the
a little bit of	a lot of the	the mass of the	there's a lot of
(2) Identification and focus			
and one of the	for those of you	one of the things	this is just a
and so this is	how many of you	so this is a	those of you who
and this is a	is one of the	so this is the	to do with the
and this is the	one of the most	this is a very	
(3) Contrast and comparison			
do the same thing	is less than or	on the one hand	the same as the
greater than or equal	is the same as	on the other hand	to be equal to
is equal to the	less than or equal	than or equal to	
(4) Deictics and locatives			
at the beginning of	in the middle of	the beginning of the	the middle of the
at the end of	in the united states	the end of the	the right hand side
at the same time			
(5) Vagueness markers			
and so on and	and things like that	or something like that	this is kind of
this is sort of			
Stance expressions			
(1) Hedges			
a little bit about	be a little bit	is a little bit	that looks like this
(2) Epistemic stance			
does that make sense	I don't know if	it seems to me	to figure out what
(3) Obligation and directive			
and you can see	have to do is	to make sure that	you don't have to
and you have to	I want you to	what you do is	you have to be
and you want to	it has to be	you can see that	you have to do
as you can see	so you can see	you do is you	you want to do
going to have to			
(4) Expressions of ability and possibility			
should be able to	to be able to	to think about the	you can think of
(5) Intention/volition, prediction			
and I want to	I want to do	now I'm going to	we're going to be
and I'm going to	I would like to	so I want to	we're going to do
and we're going to	if you want to	so I'm going to	we're going to have
are going to be	I'm going to do	so we're going to	we're not going to
going to be a	I'm not going to	that we're going to	what I'm going to
going to be the	is going to be	that's going to be	what we're going to
going to do is	it's going to be	this is going to be	you're going to be
I am going to	not going to be	want to do is	you're going to get
I don't want to	now I want to	we are going to	you're going to have
I just want to			
Discourse organizing functions			
(1) Metadiscourse and textual reference			
and we'll talk about	going to talk about	want to talk about	we're going to talk
come back to that	say a little bit	we'll come back to	when we talk about
(2) Topic introduction and focus			
I want to show	if you look at	take a look at	what I want to
I want to talk	if you think about	to give you a	what we want to
if I have a	let me show you	want to show you	you look at the
if you have a	so if you have	what do you think	
(3) Topic elaboration			
Non-causal			
and it turns out	that is to say	turns out to be	
it turns out that	turn out to be		
(4) Discourse markers			
and by the way	by the way if	by the way this	so in other words

containing data from a wide range of sources, including seminars, lectures, student presentations and office hours. Our findings revealed that lexical bundles used in academic lectures differ substantially from those used in other spoken academic registers, which further confirms that there are register differences of lexical bundles, as also found in previous research (Biber & Barbieri, 2007; Biber et al., 1999; Biber et al., 2004).

Based on our findings, the adoption of register-specific instruction regarding lecture bundles seems reasonable, since the bundles used in lectures are very different from those used in other academic spoken registers.

3.1. The discourse functions of the lexical bundles

The second research question investigated the discourse functions of the high-frequency lexical bundles and their distributions in the corpus. The present study thus categorized the extracted 171 lexical bundles into different functional categories according to Simpson-Vlach and Ellis' (2010) functional taxonomy. Table 3 presents the number of bundle types and token frequency of each functional category found in the corpus.

The majority of the lexical bundles (82%) found in the academic lectures function as *referential* and *stance expressions*, with *referential expressions* being the majority. This is not surprising given that one important goal of a course lecture is to convey content. This involves explaining key concepts, elaborating ideas, highlighting important points, making comparisons, discussing implications and applications of certain inventions or findings evaluating and giving directions. To fulfill these two functions a great variety of bundle types are used (139 of 171 bundles). These findings are consistent with Simpson-Vlach and Ellis' (2010), who also found that most bundles were *referential expressions*, such as *those of you who*. Even though slightly smaller in variety, *stance* bundles were quite frequent in the lectures.

The least and the smallest variety of lexical bundles were those functioning as *discourse organizers*, such as *if you think about*. *Discourse organizing* bundles were common at the beginning of lectures when instructors provide an agenda for the day's class, or at topic transitions when instructors want to start a new topic or to switch focus. The relatively fewer *discourse organizing* bundles identified in our corpus does not lessen their importance as signals for topic starts or changes. As indicated by Biber et al. (2004), *discourse organizing* bundles are particularly common in lecture register compared to other academic registers. By being aware of this function students can "predict the nature of upcoming ideas and information", and "make inferences about relationships" between prior and subsequent events (Nesi & Basturkmen, 2006, p. 301).

It should be noted that no lexical bundles were identified as *evaluation* and *cause and effect* bundles in topic-elaboration in the corpus. Similarly, no four-word bundles were identified as *evaluation* and *cause and effect* bundle in the Spoken AFL. It may be that instructors want to leave room for students' own interpretations without voicing their opinions too strongly. It may also be that because our corpus consists of many first year course lectures that, like introductory textbooks, present uncontested knowledge of a discipline. However, such bundles are found in the Written AFL, which indicates that the *evaluation* and *cause and effect* lexical bundles might be specific to written registers. What should be noted is that the subcategory of evaluation, which is added by Simpson-Vlach and Ellis (2010) to Biber et al.'s (2004) taxonomy to include bundles that had been categorized as *obligation/directive* bundles by Biber et al., but are "clearly evaluative, without necessarily being directive" (Simpson-Vlach & Ellis, 2010, p. 506). It should be clarified that *stance* bundles not under the subcategory of *evaluation* do not necessarily mean that they are not evaluative. Other *stance* bundles, such as *epistemic stance*, can be evaluative as well.

Table 4 provides the 171 frequent lexical bundles identified in the corpus and their different functions, with 73 categorized as *referential expressions*, 66 identified as *stance expressions*, and 32 performing *discourse organizing* functions. Lexical bundles performing the same sub-functions in the three main functional categories are further grouped and presented in alphabetical order in Table 4. The sub-functions of the lexical bundles are described in the following section and the example sentences of each functional category extracted from the corpus are presented in Appendix B.

3.2. Referential expressions

As indicated previously, most lexical bundles functioned as *referential expressions*. Following the classification of Simpson-Vlach and Ellis (2010), *Referential expressions* includes five subcategories, including *specification of attributes (intangible framing attributes, tangible framing attributes, and quantity specification), identification and focus, contrast and comparison,*

Table 5
Bundle types of each subcategory in Referential Expressions.

Subcategory	# of types	# of tokens
Specification of attributes		
Intangible framing attributes	18	4,744
Tangible framing attributes	7	1,576
Quantity specification	8	3,356
Identification and focus	15	5,907
Contrast and comparison	11	2,874
Deictics and locatives	9	4,144
Vagueness markers	5	2,114
Total	73	24,715

deictics and locatives, and vagueness markers. Table 5 summarizes the bundle types and token frequency of each subcategory of *Referential Expressions*.

In line with Simpson-Vlach & Ellis (2010), in our corpus the subcategory of *Intangible framing attributes* are the most common bundle types. These bundles "frame both concrete entities and abstract concepts or categories" (Simpson-Vlach & Ellis, 2010). They specify limiting conditions when instructors want to further elaborate on a certain concept. This result suggests that instructors make use of a wide variety of ways to specify, compare, or elaborate the differences across ideas, cases, or situations in lectures. The fact that many bundles perform this function reflects the instructional nature of the register, which focuses on delivering knowledge effectively.

The subcategory of *Identification and focus* contains the second most bundle types in r*eferential expressions*, although these bundles are actually used most frequently. These bundles draw attention to content on which the instructor intends to elaborate, or direct students' attention to key points. Examples are *one of the things, and this is the,* and *is one of the*. It was no surprise to find many bundles in this subcategory because "exemplification and identification are basic pragmatic functions in academic speech and writing" (Simpson-Vlach & Ellis, 2010, p. 504).

Only five bundles were identified as *Vagueness markers*, making it the smallest subcategory. This finding is similar to Biber et al. (2004), who list only three bundles under this subcategory (which they call *imprecision* bundles), and to Simpson-Vlach & Ellis (2010), who list only four. Two of the five bundles in our lecture corpus, "this is sort of" and "this is kind of", were not listed in previous studies, suggesting that they are especially frequent in academic lectures and, as such, pedagogically relevant.

3.3. Stance expressions

Stance expressions encompasses six subcategories: *hedges, epistemic stance, obligation and directive, expressions of ability and possibility, and intention/volition, prediction*. Table 6 summarizes the bundle types and the token frequency of each subcategory in s*tance expressions*.

The subcategory containing the most and the most frequent bundle types is *Intention/volition, prediction*, with 37 out of the 66 bundle types in this category. The function of this subcategory is to show one's intention to do something, or indicate that something is going to take place. Many lexical bundles in this subcategory also share the grammatical features of "want to" and "be going to". These bundles are especially frequently employed when instructors explicitly express their intentions or provide specific directions to students.

Modal verbs, such as 'can' and 'will,' are also found in many s*tance expressions* bundles, supporting Biber's (2006) proposition that "modals are especially common in spoken registers" (p. 103). Stance bundles with modal verbs, such as *as you can see*, are frequent, as instructors provide directions or inform students of subsequent steps, again suggesting "[the use of modals] show[s] a strong association with management/directive purposes" (Biber, 2006, p. 103).

The frequent grammatical elements common to many stance bundles, such as *want to* and modal verbs, indicate their important roles in constituting directive messages or conveying instructors' intentions in lectures. EAP instructors should thus inform students of the grammatical elements frequently included in stance bundles to help them recognize such bundles.

We also want to highlight that we classified the bundle *does that make sense* under the category of epistemic stance. This classification might be controversial considering that it is generally an expression used for comprehension check, similar to "do you understand?". However, as *does that make sense* is often used in the current corpus as a rhetorical question when lecturers are expressing their doubts or certainty regarding certain issues, and as we also referenced to the Spoken AFL when categorizing our bundles, it was finally categorized as an epistemic stance bundle following Simpson-Vlach Ellis' (2010) classification.

3.4. Discourse organizing functions

This category, which contains the fewest types of lexical bundles and is least common, includes the following four subcategories: *metadiscourse* and *textual reference, topic introduction and focus, topic elaboration* and *discourse markers*. Of these, *topic elaboration* is further divided into *non-causal* and *cause and effect*. Table 7 presents a summary of the bundle types and the token frequency of each subcategory.

Table 6
Number of bundle types in each subcategory of Stance expressions.

Subcategory	# of types	# of tokens
Hedges	4	1,175
Epistemic stance	4	866
Obligation and directive	17	4,884
Expressions of ability and possibility	4	1,410
Intention/volition, prediction	37	15,575
Total	**66**	**23,910**

Table 7
Number of bundle types in each subcategory of Discourse Organizing Functions.

Subcategory	# of types	# of tokens
Metadiscourse and textual reference	8	2,119
Topic introduction and focus	15	5,478
Topic elaboration (Non-causal)	5	2,075
Discourse markers	4	1,012
Total	**32**	**10,684**

Topic introduction and focus contains the most bundle types and is most often. Its function is similar to that of *identification and focus* in r*eferential expressions* (Simpson-Vlach & Ellis, 2010); it also performs the function of introducing topics and drawing the listeners' attention. These bundles also serve as signals, implying that the speaker is about to talk about something new or different.

No lexical bundle was classified under *Cause and effect*, but five lexical bundles were identified as *Non-causal* (elaborating events without explicit causal relationships implied). A similar result was found in Simpson-Vlach Ellis' (2010) classification, with more bundle types being identified as *Non-causal* than *Cause and effect*.

For non-native students preparing for lectures in English, understanding the meanings of these bundles in academic lectures might not be difficult, as they contain vocabulary that should have been acquired in the early stages of their English learning (Cortes, 2004). However, problems can occur when the functions that bundles perform "sometimes are not as transparent as initially expected" (p. 421). For example, students may be familiar with the meanings of "if you think about" and "if you have a", but they may need extra help to realize the topic-introduction function these bundles can perform in academic lectures. That is, the occurrences of these bundles usually suggest that new topic is going to be introduced or important ideas are going to be highlighted. Instructors should thus help students notice the frequent use of lexical bundles in lectures and provide explicit instructions on the functions they perform (Biber & Barbieri, 2007; Cortes, 2004).

3.5. Variations of lexical bundles across disciplines

The third research question explored whether there are disciplinary variations in the use of lexical bundles in academic lectures. Of specific interest here is whether the varied use of lexical bundles across disciplines is great enough to warrant a discipline-specific teaching approach. With this in mind, the different types and numbers of high-frequency lexical bundles used in the four disciplines were compared (Table 8).

Overall, Engineering lectures had many more and a much greater variety of lexical bundles than the other three disciplines. It is interesting that the lectures in the two hard science disciplines, Engineering and Science and Math, had more bundles and bundle types than did the two soft sciences, Humanities and Arts and Social Science. The differences may be attributable to nature of the disciplines.

Language use in the hard science lectures in our corpus seems to be relatively more organized and routinized, with instructors frequently explaining formulas and theories, introducing new equipment and its applications, or explaining how an experiment should be carried out. Many lexical bundles such as *I am going to* and *this is going to* were often used to explicitly and clearly show the logical reasoning behind a formula or explain the process regarding how a result is yielded. This also reflects the "empirical" nature of the hard science lectures (Kashiha & Heng, 2013, p. 139). On the other hand, lecturers in the soft sciences did not seem to use such lexical bundles as often. This is not surprising considering the more "textually interpretive and descriptive" nature of soft science lectures (p. 140), where step-by-step and explicit explanation demonstrating how a result is produced is seldom seen.

To compare functional bundles across disciplines we identified the number of bundle types and tokens in each functional category for each discipline. The results are shown in Table 9 and in Figures 1 and 2, which display the number of functional bundle types and tokens for each discipline. Chi-square tests showed no significant differences in the variety of functional bundles across the disciplines at the .05 level, but statistically significant differences at the .01 level were found for the frequency of functional bundles across the disciplines. This indicates that, although the disciplines do not differ in terms of functional bundle types, the number of functional bundles differs significantly.

Stance bundles are the most frequent in all the disciplines, except Humanities and Arts. However, more *referential expressions* are found in Humanities and Arts. A further exploration on individual referential bundles frequently appearing in Humanities and Arts showed that many referential bundles in this discipline are *deictics and locatives*, specifying the time or the place in which historical events took place, such as *at the same time* and *at the beginning of*. Bundles performing such functions are relatively frequent in history lectures where instructors often introduce historical events after providing some general background information.

Deictics and locatives bundles and those performing the function of *identification and focus* frequently appear when instructors try to draw students' attention to a specific part of a text. Many instances are found in literature lectures where instructors highlight an author's use of a certain word or expression in the text and further elaborate on its symbolic meaning or the message implied. The high density of *deictics and locatives* and *identification and focus* expressions greatly contributes to the high frequency of *referential* bundles in Humanities and Arts.

Table 8
Summary of bundle types and tokens used in each discipline.

Discipline	Types	Tokens
Engineering	107	15,988
Humanities and Arts	46	5,992
Science and Math	89	12,867
Social Science	76	9,976
Total (Academic Lectures)	171	44,823

Table 9
Summary of the disciplinary variation by functional categorization.

Discipline	Referential expressions		Stance expressions		Discourse organizing functions	
	Type	Token	Type	Token	Type	Token
Engineering	41	6346	46	6913	20	2729
Humanities and Arts	29	3993	12	1257	5	742
Science and Math	31	4639	38	5539	20	2689
Social Science	27	3853	32	4067	17	2056
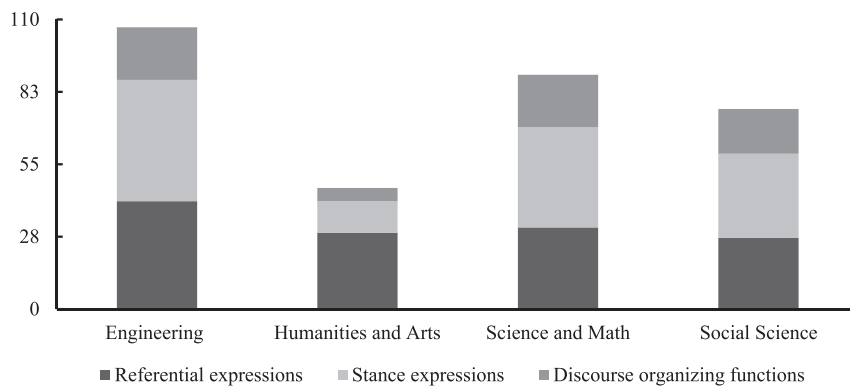
Figure 1. Functional bundle types used in each discipline.
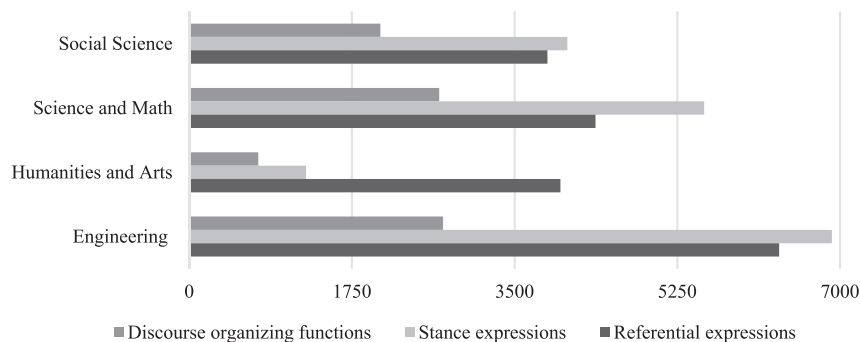
Figure 2. Frequency of functional bundles used in each discipline.

We also found a high density of *referential* bundles in Humanities and Arts performing the functions of illustrating ideas, framing issues and specifying differences among concepts. They are especially frequently used to describe intangible attributes. This finding reflects the characteristic of this discipline in which instructional content is usually heavily based on ideas or abstract concepts (Simpson-Vlach, 2006). Kashiha and Heng (2013) also pointed out the "expository nature" of soft science where "a variety of physical abstracts and entities need to be identified" (p. 146). More elaboration and description may thus be needed to facilitate students' understanding in lectures of this discipline, leading to the use of lexical bundles that perform the referential function of specifying the attributes of intangible entities.

Table 10

The top 46 most frequently-used lexical bundles in the four disciplines[a]. (Bold = bundles occurring in 4 disciplines, italic = bundles occurring in 3 disciplines).

Engineering	Humanities and Arts	Science and Math	Social Science
is going to be	that is to say	**is going to be**	**is going to be**
or something like that	at the same time	or something like that	in the united states
if you have a	**at the end of**	*if you have a*	**the end of the**
if you want to	**the end of the**	**if you want to**	**if you look at**
than or equal to	the way in which	we are going to	**a little bit more**
so this is the	one of the most	so this is the	**at the end of**
a little bit of	*one of the things*	*it turns out that*	at the same time
to be able to	**a little bit of**	**a little bit of**	on the other hand
we are going to	on the one hand	*and this is the*	**a little bit about**
so this is a	**what I want to**	as a function of	*it's going to be*
if you look at	as a kind of	I am going to	going to have to
it turns out that	in the context of	**to be able to**	how many of you
a little bit more	is a kind of	*it's going to be*	so I'm going to
as a function of	**if you want to**	**if you look at**	a lot of people
and this is the	in the case of	in the case of	going to talk about
and things like that	on the other hand	*so this is a*	**if you want to**
the sum of the	to do with the	the mass of the	**to be able to**
less than or equal	in a way that	the speed of light	*are going to be*
this is going to	is one of the	you can see that	there's a lot of
going to be a	I want you to	the square root of	**a little bit of**
it's going to be	**is going to be**	going to talk about	**and I'm going to**
and I'm going to	*in terms of the*	*going to be a*	*you look at the*
so I'm going to	**to be able to**	**a little bit more**	*going to be a*
I want to do	of the things that	*in terms of the*	*so this is a*
on the order of	**if you look at**	*and I want to*	is one of the
at the end of	in the middle of	this is going to	I want to talk
going to do is	**a little bit about**	so in other words	so we're going to
in terms of the	*I want to do*	the right hand side	*one of the things*
the end of the	*and we're going to*	*going to be the*	I'm not going to
a little bit about	at the beginning of	turns out to be	that is to say
going to be the	in such a way	**and I'm going to**	what do you think
does that make sense	it seems to me	*and you can see*	*and we're going to*
the square root of	a certain kind of	**what I want to**	we'll come back to
I am going to	the beginning of the	and we're going to	in a way that
and you can see	*and you can see*	I want you to	*going to be the*
and this is a	**a little bit more**	**at the end of**	**what I want to**
are going to be	such a way that	**the end of the**	*and this is the*
so if you have	the middle of the	*are going to be*	one of the most
we're going to do	**and I'm going to**	*one of the things*	*I want to do*
what I want to	and one of the	and things like that	you don't have to
the right hand side	we're going to be	what we're going to	a lot of the
you look at the	*and I want to*	so we're going to	*it turns out that*
I'm going to do	take a look at	I just want to	we're going to talk
take a look at	*you look at the*	is the same as	want to talk about
to be equal to	the ways in which	**a little bit about**	we're going to do
and I want to	those of you who	and it turns out	*if you have a*

[a] Note: The bundles were arranged according to frequencies.

As for the high density of *stance* bundles in Engineering, Science and Math, and Social Science, we found that most perform the functions of *intention/volition, and prediction*. These bundles are usually seen when instructors try to explain complex processes by providing step-by-step guidance, such as describing how the first action should be carried out, the effects of the action, and the expected outcome derived from the process. Explicit and logical deductive structures of events or actions can be frequently seen in lectures of these three disciplines. Bundles, such as *this is going to* and *going to be a,* are frequently used in these lectures for this purpose.

Different from Engineering, and Science and Math, we found that connections between ideas and events in Humanities and Arts are sometimes less definite and less clearly defined. Relationships between concepts evolve as more ideas and views are added, and new links are established with addition of new information. In terms of knowledge construction in Humanities and Arts, instructors seem to provide more opportunities for students to connect dots and develop their own interpretations. They also tend to encourage students to consider different aspects of an issue or to help them form their own interpretations and draw their own conclusions. Since the relationship between ideas is not always defined directly and clearly and there is a great room for interpretation in this discipline, it is not surprising that not many lexical bundles function as *stance expressions*. This finding also highlights the distinct style of knowledge construction in Humanities and Arts.

Interestingly, lexical bundles that have a *discourse organizing* function are the fewest in number across all four disciplines. Despite their low frequency, the variety of discourse organizing bundles in the hard science disciplines was found to be higher

than in the soft sciences. As mentioned previously, our analysis of the current corpus showed that the instructional patterns in lectures of hard science are more routinized, such as outlining experiment procedure, describing standards for equipment operation, explaining theories, and applying formulas. Therefore, similar expressions are likely to be used by instructors in similar patterns in lectures of hard science, thus resulting in more shared bundles retrieved. For example, lecturers in the hard sciences tended to provide some empirical evidence to support their arguments or to demonstrate the ideas. Therefore, the lexical bundles *I want to show* and *let me show you* were used comparatively frequently in hard science lectures to draw students' attention to what lecturers intended to highlight.

In summary, this study revealed that lexical bundles in lectures perform very different functions across the four different disciplines. This expands on the well-established understanding that there are disciplinary variations of lexical bundles in academic writing (e.g., Cortes, 2004; Durrant, 2015; Hyland, 2008). We thus suggest that EAP/ESP teachers acknowledge disciplinary variation in instruction of bundle functions.

3.5. Cross-disciplinary bundles

While lexical bundles function differently across disciplines, it is not clear whether a discipline-specific teaching approach regarding teaching high-frequency lecture bundles should be pursued. Hyland (2008) compared the 50 most frequent bundles across four disciplines (biology, electrical engineering, applied linguistics, and business studies) in his corpus of academic writing. Only five of the bundles appeared in all the disciplines, 14 appeared in three disciplines, with 38% of the bundles shared by at least three disciplines. Because so few bundles were found to be common across disciplines, Hyland concluded that there are great disciplinary differences in terms of the use of lexical bundles in academic writing. He thus recommended that a discipline-specific teaching approach to lexical bundles should be applied. While this may be reasonable for writing, our results indicate that for EAP lecture comprehension courses such an approach may not be warranted.

To investigate if there are disciplinary differences in the use of individual lexical bundles in academic lectures, the high-frequency lexical bundles in the four disciplines were compared to identify cross-disciplinary bundles. Since only 46 high-frequency bundles were found in Humanities and Arts, this study extracted and compared the top 46 most frequently-used bundles among all the four disciplines to identify which bundles were shared by all four disciplines and which were shared by three. Table 10 presents the 46 most frequently used bundles in the four disciplines in the corpus.

As can be seen in Tables 10, 11 bundles were identified in all the disciplines with 15 appearing in three disciplines. This means that over 56% of the bundles are shared by at least three disciplines. Additionally, a higher percentage of cross-disciplinary bundles was found in academic lectures (56%), compared to Hyland's (2008) findings in academic writing, where only 38% of bundles were shared by at least three disciplines. Despite disciplinary variations in terms of the frequency of functional bundles, the lexical bundles frequently used in academic lectures are similar. It has been established that the lexical bundles used in different registers do not come from a single pool, but are "associated with the typical communicative purposes of that register" (Biber & Barbieri, 2007, p. 265). The bundles identified in our research are very different from those identified from other registers, as shown previously in the comparison of bundles in our corpus with those in the Spoken AFL, which has led us to believe there potentially is a set of lexical bundles specific to the lecture register. The identification of many inter-disciplinary bundles further supports our previous proposition and aligns with Simpson-Vlach Ellis' (2010) point that there are many academic bundles core to different disciplines. Based on the findings, although the frequencies of lexical bundles appearing in disciplines vary considerably, the items used across disciplines are similar.

Adolphs and Schmitt (2003) indicated that "spoken discourse uses a smaller variety of word types than written discourse" and "a lower percentage of lexical coverage might be required for spoken discourse than written discourse" (p. 433). Therefore, it is probable that a lower variety of lexical bundles are found in academic lectures because there are essentially fewer word types in those speech events. Swales (2005) stated in a plenary talk that "academic speech is a great leveler" which blurs disciplinary differences (as cited in Simpson-Vlach, 2006, p. 295), suggesting that more common elements could be found among disciplines in academic speech than in academic writing. This may account for the fact that many more bundles are shared across disciplines in academic lectures than in academic writing.

Since there are many cross-disciplinary bundles in academic lectures, we support Simpson-Vlach and Ellis' (2010) view that a strict discipline-specific teaching approach regarding lexical bundles might not be appropriate. Instead, instructors could focus on cross-disciplinary bundles as a starting point. See Appendix C for a list that can be used as a reference for course development.

4. Conclusion and pedagogical implications

With a view toward providing insights into teaching lexical bundles in ESP/EAP listening comprehension courses, this study identified high-frequency lexical bundles and their functions in academic lectures and investigated disciplinary variations regarding the frequency as well as the type of functional bundles. The results showed that many more *referential* and *stance* bundles than discourse organizing bundles were used in the academic lectures we analyzed. These two types of bundles frequently occurred in general classroom activities where instructors explain ideas, give directions, make comparisons, and provide judgments. *Discourse organizing* bundles were used less often, but also played an important role in lectures to signal a new topic.

We also discovered significant variation in the frequency of functional bundles used across disciplines, suggesting that EAP/ESP teachers should take into account the disciplines of the target students when teaching lexical bundles. However, we maintain a conservative view of Hyland's (2008) proposition that because there are few cross-disciplinary bundles a strictly discipline-specific teaching approach to lexical bundles should be applied. Unlike Hyland, our comparison of high-frequency lexical bundles in the four disciplines revealed many cross-disciplinary bundles, thus supporting Simpson-Vlach Ellis' (2010) claim that there is "a common core of academic formulas that do transcend disciplinary boundaries" (p. 509). We suggest that these cross-disciplinary bundles be taught to students in courses preparing them for lectures in English first so they become familiar with these bundles' functions in academic lectures. Teachers can further introduce the primary functions that lexical bundles perform in different disciplines to students, helping them be aware of disciplinary variations. Teachers can first highlight the frequent lexical bundles on lecture transcripts from different disciplines. Students can then be asked to identify the functions of the highlighted bundles in the transcripts and compare their functions in the different disciplines. Such activities can help students see how different the functions of lexical bundles are across disciplines and, as suggested by Flowerdew (2003), be a form of data-driven learning

To support teaching efforts, we also generated a list of high-frequency lexical bundles and their functions in academic lectures. Teachers could use this list as a reference when developing materials to help students become familiar with the functions of lexical bundles in lectures. Teaching the functions of these high-frequency lexical bundles in lectures could also help to improve students' listening skills, facilitating their overall comprehension of lectures. It should be noted that explicit instruction on the functions of lexical bundles is necessary because "they (lexical bundles) are not perceptually salient, despite their high frequency" (Biber & Barbieri, 2007, p. 284). This instruction should be contextualized to deepen students' understanding of how lexical bundles perform different functions (Simpson-Vlach & Ellis, 2010). Teachers could then lead a discussion on the relationship of particular bundles to their surrounding items, helping students notice their cohesive role (Nesi & Basturkmen, 2006). Teachers could also apply a data-driven approach, asking students to examine concordance lines to explore how lexical bundles function in lectures. Moreover, teaching the functions of high-frequency lexical bundles can improve the core skills of listening comprehension, such as *listen for details* and *listen and predict* (Vandergrift & Goh, 2012). Additionally, by introducing lexical bundles that perform identification and focus functions, students can be helped to identify specific information in aural input. Teaching the intention/volition, prediction bundles can also help students anticipate what they will hear. Finally, the list of high-frequency bundles and their functions in lectures could also be used when training non-native teachers and teaching assistants to deliver lectures in English.

As with any study, this study has some limitations. This study collected academic lectures from only four disciplines, but it would have been advantageous to include lectures from other fields to make the coverage more comprehensive. We also recognize, as indicated by Biber et al. (2004), that a single lexical bundle can perform more than one function, even in one occurrence. Despite this important insight, one of our goals was to generate a pedagogically useful list for teachers to develop materials, following Simpson-Vlach and Ellis' (2010) taxonomy. In consideration of possible pedagogical applications we assigned only one function to each bundle to avoid confusion. It is therefore suggested that teachers use this list as a starting point in bundle instruction and consider whether there are other functions a particular lexical bundle might serve in the same context. Additionally, this study offers only a preliminary exploration of the disciplinary differences of bundles in terms of the frequency and bundle types used across disciplines. Future studies could usefully explore how each functional bundle type is used in individual disciplines. This study applied a rather high frequency cut-off point for bundle extraction. A lower frequency cut-off would allow a more extensive investigation into the use of lexical bundles in individual disciplines. Finally, the transcriptions analyzed in the present study might have been "cleaned up" to some extent before we retrieved them from their original sources, which means they are unable to reflect some important aspects of instructional language. For example, there are no instances of "gonna" or "wanna" in the transcripts. It is thus acknowledged that the transcripts might not necessarily precisely capture the language use. This limitation should be addressed further by thoroughly re-examining and re-editing the transcriptions to gain a more accurate picture of instructional language in undergraduate lectures.

In addition to understanding the functions of lexical bundles, non-native students' listening comprehension can be facilitated through improving the understanding of the moves of a genre (Wolff, 1989). Some scholars have researched move structures in academic lectures (Alsop & Nesi, 2014). Since lexical bundles are found to serve different functions in different positions in university classroom talk (Csomay & Cortes, 2010), one possible direction for future work would be to identify lexical bundles that frequently occur in different academic lecture moves. For example, the lexical bundles 'we'll come back to' and 'we're going to talk' may appear frequently at the beginning of lectures and 'if you think about' could appear to introduce a new idea. In addition, Neely and Cortes (2009) suggest that topic-introduction bundles can facilitate the use of top-down processing to understand lectures at a macro level. It would also be beneficial to investigate the lexical bundles of other functions in facilitating listeners' processing.

Acknowledgement

Appendix A

The sources of the lectures in the present corpus

Sources of lectures	Number of lectures
MIT	365
Yale University	491
Stanford University	148
Harvard University	22
UC Berkeley	52
UC Irvine	22
Missouri State University	30
Total	**1130**

Appendix B

Example sentences for each functional category

Referential bundles

Specification of attributes
- Intangible framing attributes
There is ***a kind of*** a vogue recurrently in the history of fiction for a kind of miraculous sense that this is just exactly the way things are.
- Tangible framing attributes
As you expect, the output is going to be ***the sum of the*** two voltages scaled appropriately.
- Quantity specification
So in the biomedical domain, where ***a lot of the*** time people are wanting to study the task of protein interaction.
Identification and focus
And I think ***one of the most*** dramatic demonstrations that a proton gradient can be a source of energy comes from this sort of thing.
Contrast and comparison
But the good citizen, Aristotle goes on to say, is not ***the same as the*** good human being, right?
Deictics and locatives
So on ***the right hand side***, this is like our other profit diagrams.
Vagueness markers
If you get degrees as high as three ***or something like that***, that tells you all of them are in one block together.

Stance expressions

Hedges
Let me talk just ***a little bit about*** over consumption and who's the number one bad guy in over consumption?
Epistemic stance
Well, I don't know. ***It seems to me*** that literary influence is not at all unlike sort of speaking or writing in the wake of a founder of discursivity, but we can let that pass.
Obligation and directive
And here ***I want you to*** realize that, even though you didn't know it, you indeed can read Italian
Expressions of ability and possibility
These are the kinds of things a geneticist does ***to be able to*** characterize mutants on a mutant hunt.
Intention/volition, prediction
You're going to see that we get some help here on the static semantics, and ***I'm going to do*** an example in a second.

Discourse organizing functions

Metadiscourse and textual reference
Then we'll see a new epistemology that stresses the laboratory bench as the source of knowledge, ***and we'll talk about*** the coming of laboratory medicine and experimental medicine as a paradigm shift.
Topic introduction and focus
And ***if you think about*** what consequences this could have, in terms of places where they have high altitude.
Topic elaboration (Non-causal)
All right, so when you do a problem, the path is going to ***turn out to be*** extremely important.
Discourse markers
And by the way, what that means is this is not a descent method.

Appendix C

Lexical bundles shared across disciplines

Bundles shared by two disciplines	Main function/Sub-function
and things like that	Referential bundles/Vagueness markers
as a function of	Referential bundles/Specification of attributes–intangible framing attributes
at the same time	Referential bundles/Deitics and locatives
going to talk about	Discourse organizing functions/Metadiscourse and textual reference
I am going to	Stance Expressions/Intention/volition, prediction
I want you to	Stance Expressions/Obligation and directive
in a way that	Referential bundles/Specification of attributes–intangible framing attributes
in the case of	Referential bundles/Specification of attributes–tangible framing attributes
is one of the	Referential bundles/Identification and focus
on the other hand	Referential bundles/Contrast and comparison
one of the most	Referential bundles/Identification and focus
or something like that	Referential bundles/Vagueness markers
so I'm going to	Stance Expressions/Intention/volition, prediction
so this is the	Referential bundles/Identification and focus
so we're going to	Stance Expressions/Intention/volition, prediction
take a look at	Discourse organizing functions/Topic introduction and focus
that is to say	Discourse organizing functions/Topic elaboration
the right hand side	Referential bundles/Deitics and locatives
the square root of	Referential bundles/Specification of attributes–tangible framing attributes
this is going to	Stance Expressions/Intention/volition, prediction
we are going to	Stance Expressions/Intention/volition, prediction
we're going to do	Stance Expressions/Intention/volition, prediction
Bundles shared by three disciplines	**Main function/Sub-function**
and I want to	Stance Expressions/Intention/volition, prediction
and this is the	Referential bundles/Identification and focus
and we're going to	Stance Expressions/Intention/volition, prediction
and you can see	Stance Expressions/Obligation and directive
are going to be	Stance Expressions/Intention/volition, prediction
going to be a	Stance Expressions/Intention/volition, prediction
going to be the	Stance Expressions/Intention/volition, prediction
I want to do	Stance Expressions/Intention/volition, prediction
if you have a	Discourse organizing functions/Topic introduction and focus
in terms of the	Referential bundles/Specification of attributes–tangible framing attributes
it turns out that	Discourse organizing functions/Topic elaboration
it's going to be	Stance Expressions/Intention/volition, prediction
one of the things	Referential bundles/Identification and focus
so this is a	Referential bundles/Identification and focus
you look at the	Discourse organizing functions/Topic introduction and focus
Bundles shared by four disciplines	**Main function/Sub-function**
a little bit about	Stance Expressions/Hedges
a little bit more	Referential bundles/Specification of attributes - quantity specification
a little bit of	Referential bundles/Specification of attributes - quantity specification
and I'm going to	Stance Expressions/Intention/volition, prediction
at the end of	Referential bundles/Deitics and locatives
if you look at	Discourse organizing functions/Topic introduction and focus
if you want to	Stance Expressions/Intention/volition, prediction
is going to be	Stance Expressions/Intention/volition, prediction
the end of the	Referential bundles/Deitics and locatives
to be able to	Stance Expressions/Expressions of ability and possibility
what I want to	Discourse organizing functions/Topic introduction and focus

References

Adolphs, S., & Schmitt, N. (2003). Lexical coverage of spoken discourse. *Applied Linguistics, 24*(4), 425-438.

Alsop, S., & Nesi, H. (2014). The pragmatic annotation of a corpus of academic lectures. In N. Calzolari, K. Choukri, T. Declerck, H. Loftsson, B. Maegaard, J. Mariani, A. Moreno, J. Odlik, & S. Poperidis (Eds.), *Proceedings of LREC 2014, ninth international conference on language resources and evaluation* (pp. 1560-1563). Reykjavik, Iceland: European Languages Resources Association (ELRA).

Bal, B. (2010). *Analysis of four-word lexical bundles in published research articles written by Turkish scholars. (Unpublished master thesis)*. Atlanta, GA: Georgia State University.

Becher, T. (1994). The significance of disciplinary differences. *Studies in Higher Education, 19*(2), 151-161.

Biber, D. (2006). Stance in spoken and written university registers. *Journal of English for Academic Purposes, 5*(2), 97-116.

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

Biber, D., Conrad, S., & Cortes, V. (2004). If you look at the…: Lexical bundles in university teaching and textbooks. *Applied Linguistics, 25*(3), 371-405.

Biber, D., Johansson, S., Leech, G., Conrad, S., & Finegan, E. (1999). *Longman Grammar of spoken and written English*. Harlow: Longman.

Conrad, S., & Biber, D. (2005). The frequency and use of lexical bundles in conversation and academic prose. *Lexicographica, 20*, 56-71.

Cortes, V. (2004). Lexical bundles in published and student disciplinary writing: Examples from history and biology. *English for Specific Purposes, 23*(4).

Csomay, E., & Cortes, V. (2010). Lexical bundle distribution in university classroom talk. *Language and Computers, 71*(1), 153-168.

Durrant, P. (2015). Lexical bundles and disciplinary variation in university students' writing: Mapping the territories. *Applied Linguistics, 38*(2), 165-193.

Flowerdew, L. (2003). A combined corpus and systemic-functional analysis of the problem-solution pattern in a student and professional corpus of technical writing. *Tesol Quarterly, 37*(3), 489-511.

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

Goh, C. C. M. (2013). ESP and listening. In B. Paltridge, & S. Starfield (Eds.), *The handbook of English for specific purposes* (pp. 77-94). Oxford: Wiley-Blackwell.

Hyland, K. (2008). As can be seen: Lexical bundles and disciplinary variation. *English for Specific Purposes, 27*(1).

Jeon, J. (2007). *A study of listening comprehension of academic lectures within the construction-integration model (Unpublished doctoral dissertation)*. IL: The Ohio State University.

Kashiha, H., & Heng, C. S. (2013). An exploration of lexical bundles in academic lectures: Examples from hard and soft sciences. *The Journal of AsiaTEFL, 10*(4), 133-161.

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. (2006). Lexical bundles and discourse signalling in academic lectures. *International Journal of Corpus Linguistics, 11*(3), 283-304.

Salazar, D. (2011). *Lexical bundles in scientific English: A corpus-based study of native and non-native writing (Unpublished doctoral dissertation)*. University of Barcelona. Retrieved from http://www.tdx.cat/bitstream/handle/10803/52083/DJLS_DISSERTATION.pdf.

Scott, M. (2011). *WordSmith Tools version 6, liverpool*. Lexical Analysis Software.

Simpson-Vlach, R. (2006). Academic speech across disciplines: Lexical and phraseological distinctions. In K. Hyland, & M. Bondi (Eds.), *Academic discourse across disciplines* (pp. 295-316). Bern: Peter Lang.

Simpson-Vlach, R., & Ellis, N. (2010). An academic formulas list: New methods in phraseology research. *Applied Linguistics, 31*(4), 487-512.

Staples, S., Egbert, J., Biber, D., & McClair, A. (2013). Formulaic sequences and EAP writing development: Lexical bundles in the TOEFL iBT writing section. *Journal of English for Academic Purposes, 12*(3).

Swales, J. M. (2005). The flavor and structure of academic speech. In *Plenary address given at the international systemic functional conference, July 17-22*. Sydney, Australia.

Vandergrift, L., & Goh, C. (2012). *Teaching and learning second language listening: Metacognition in action*. New York: Routledge.

Wolff, D. (1989). Identification of text-type as a strategic device in L2 comprehension. In H. W. Dechert, & M. Raupach (Eds.), *Interlingual processes* (pp. 137-150). Tubingen, Germany: Gunter Nann.

Chen-Yu Liu is a PhD student of the Department of English at National Taiwan Normal University. Her research interests include corpus linguistics, second language acquisition and vocabulary acquisition.

Prof. Hao-Jan Howard Chen obtained his PhD degree from University of Pennsylvania in 1996. He is a professor of the Department of English at National Taiwan Normal University. Prof. Chen's research interests include computer-assisted language learning, corpus linguistics, second language acquisition and vocabulary acquisition.