Exploring Lexical Bundles in the Algerian Corpus of Engineering

Arab Journal of Applied Linguistics
e-ISSN 2490-4198
Vol. 3, No. 1, March 2018, 47-77

http://www.arjals.com/

Fares Rezoug[1] & Benet Vincent
Coventry University, UK

Abstract

Although lexical bundles (LBs) have attracted considerable attention in applied linguistics, their variation by discipline is an under-researched area, particularly with regard to Master's dissertations. This paper explores the frequency, structure, and function of four-word bundles used in Electrical and Electronic Engineering Master's dissertations written in English by Algerian students working in four sub-disciplines – Power Engineering, Computer Engineering, Telecommunication Engineering and Control Engineering. The LBs extracted from each sub-discipline were subjected to structural and functional analysis, applying Hyland's (2008) framework. The results were compared across the sub-disciplines and between these sub-disciplines and Hyland's (2008) findings for Electrical Engineering to indicate the main structural and functional patterns relating to the bundles retrieved. Further, a fine-grained functional analysis was undertaken which goes beyond Hyland's framework to indicate the main realisations of bundles and propose pedagogically-friendly formal-functional 'clusters' of bundles. Potential pedagogical and methodological implications are also discussed.

Keywords: Engineering Master's dissertations, English for specific purposes, lexical bundles, phraseology, corpus linguistics.

1 rezougf@uni.coventry.ac.uk

Introduction

Recurrent word combinations have attracted the attention of linguists at least since Palmer (1933) and Firth (1951), who referred to ''collocation'' and ''collocability''. However, it was not until the increased availability of computer-readable corpora in the 1980s that it became possible to investigate these combinations in detail. This type of research can be considered under the umbrella of 'phraseology', the 'tendency of words to occur in preferred sequences' (Hunston, 2002, p. 138) and typically acknowledges Sinclair's (1991) 'idiom principle', which holds that speakers and writers do not select single words at a time, but choose pre-constructed phrases to express a particular meaning.

One influential approach to investigating phraseology and thereby finding typical 'ways of saying things in a particular discourse' (Gledhill, 2000, p. 1) is lexical bundle analysis (Biber, Johansson, Leech, Conrad, & Finegan, 1999), that is, the identification and classification of the fixed-length strings of items (e.g. *the fact that the, on the other hand*) that occur most frequently in particular texts. Lexical bundles have also received attention as n-grams (e.g. Ellis, Simpson-Vlach, & Maynard, 2008), clusters (Schmitt, Grandage & Adolphs, 2004) and recurrent word combinations (Altenberg, 1998; De Cock, 1998).

Since a significant proportion of words are found to occur in recurrent bundles, lexical bundles are 'useful devices for the comprehension and construction of discourse' (Biber & Barbieri, 2007, p. 284). Biber et al. (1999, p. 995) for example, find that around 21% of words occur in such bundles in

academic prose. Such findings are used to argue that less proficient writers should gain greater awareness of the most common realisations and functions of bundles (Ädel & Erman, 2012; Nesi & Basturkmen, 2006), and also that language teachers should know which bundles are most common in target texts. However, bundles are not usually complete phrases (Biber et al., 1999; Stubbs & Barth, 2003; Biber & Barbieri, 2007), but are more profitably seen as evidence of the phraseological tendency of language; lexical bundle analysis is one way of a number of approaches to investigating conventionalised uses of language (Vincent, 2013).

Wray (2002) and Hyland (2008) point out that different academic disciplines favour different specific word combinations, whether bundles or formulas. Hyland (2008, p. 5) explains that 'gaining control of a new language or register requires a sensitivity to expert users' preferences for certain sequences of words over others that might seem equally possible'. This sensitivity is central to the creation of academic discourse and indicates the importance of research into cross-disciplinary variations (Hyland, 2008), which may inform teaching materials and approaches, especially in fields like ESP and EAP.

This study seeks to build on previous research on bundles and their discipline specificity by investigating those frequently occurring in MSc dissertations written in English by Engineering students at an institution of higher education in Algeria. We are interested in exploring how the students writing these dissertations use bundles in their writing and how they are distributed across the sub-disciplines of Engineering in the institute.

Lexical bundles in the literature

Many studies have explored the use of lexical bundles, whether across different language backgrounds (Ädel & Erman, 2012; Chen & Baker, 2010; De Cock, 2000; Römer, 2009), genres (Biber, 2006; Biber, Conrad & Cortes, 2004; Hyland, 2008; Nesi & Basturkmen, 2006; Scott & Tribble, 2006), disciplines (Cortes, 2004; Durrant, 2015) or proficiency levels (Pan, Reppen, & Biber, 2016; Staples et al., 2013). Simpson-Vlach and Ellis (2010) propose lists of bundles (their Academic Formulas List) which are most salient, and therefore potentially pedagogically useful in academic registers.

All of these studies base their analyses on the most frequently occurring lexical bundles, or 'target bundles' (Cortes, 2004, 2006). As Biber et al. (2004, p. 376), point out, such frequency data is not explanatory, but 'identifies patterns that must be explained'. Since bundles are ubiquitous in all types of text, this explanation tends to be based on classifications of bundles in terms of the structures they typically fall into and bundle functions.

One way of approaching the classification of bundles for the purposes of comparison is to consider their main structural realisations, bearing in mind that bundles, particularly three- or four-word bundles, are not usually structurally complete (Biber et al., 1999). Studies commonly follow Biber et al. (1999) in recognising that typical structural realisations of bundles vary considerably by register. This study focuses on those found in academic prose.

Biber et al. (1999) list 11 main structural realisations of lexical bundles found in academic prose and one 'other' category for less frequently occurring

types. These categories are also used by Hyland (2008), although he only finds that 7 of these occur frequently in his corpora (see Table 1).

Table 1 The main structural types of lexical bundles in Hyland (2008)

Structure	Examples
Noun phrase + of	The end of the, the nature of the, a large number of
Other noun phrases	the fact that the, one of the most, the extent to which
Prepositional phrase + of	at the end of, as a result of, on the basis of
Other prepositional phrases	on the other hand, at the same time, with respect to the
Passive + prepositional phrase fragment	is shown in figure, is defined as the, can be found in,
Anticipatory it	it is important to, it is possible that, it was found that
Be + noun/adjective phrase	is the same as, is due to the, be the result of
Others	as shown in figure, should be noted that, is likely to be, as well as the

It can be useful to compare the structural types of bundles found in different text types and disciplines since there are associations between structural types of bundles and their functions (Hyland 2008). However, the chief area of interest in the investigation of lexical bundles is in terms of the functions they are associated with in discourse and how the analysis of these functions reveals differences in phraseology.

The functional framework for analysing bundles which is applied in this study is based on Hyland (2008). This framework is very similar to that applied in Biber et al. (2004) and Biber (2006), with three main overarching functional categories each containing a number of sub-categories (see Table 2). The 'Research-oriented' (RO) category includes bundles that 'help writers to structure their activities and experience of the real world' (Hyland, 2008, p. 13). 'Text-oriented' (TO) bundles are those 'concerned with the organization of the text and its meaning as a message or argument' (ibid.). Finally, 'Participant-oriented' (PO)

bundles include both those that refer to 'Stance' and a further subcategory of 'Engagement' bundles, which 'focus on the writer or the reader of the text' (Hyland, 2008, p. 14). This framework, like Biber's, is clearly inspired by Halliday's (1985) three-way functional analysis of language: 'Participant-oriented' relates to Halliday's 'interpersonal' meanings; 'Text-oriented' to 'textual' meanings; and 'Research-oriented' to 'ideational' meanings.

Table 2 Hyland's (2008) functional classification of lexical bundles

Research-oriented (RO)
- Location: at the beginning of, in the present study
- Procedure: the use of the, the role of the, the purpose of the, the operation of the
- Quantification: the magnitude of the, the wide range of, one of the most
- Description: the structure of the, the size of the, the surface of the
- Topic: the currency board system
Text-oriented (TO)
- Transition signals: on the other hand, in addition to the, in contrast to the
- Resultative signals: as a result of, it was found that, these results suggest that
- Structuring signals: in the present study, in the next section, as shown in figure
- Framing signals: in the case of, with respect to the, on the basis of, in the presence of
Participant-oriented (PO)
- Stance features: are likely to be, may be due to, it is possible that
-Engagement features: it should be noted that, as can be seen

Hyland's (2008) classification was chosen for this study since it is based on analysis of bundles derived from academic texts: Masters dissertations, PhD theses and research articles. The present study explores texts from the discipline of Electrical Engineering, which is also included in Hyland's corpus.

Rationale for the study

Dissertation writing is amongst the longest student-produced genres, and writing dissertations is therefore a difficult task for non-native and native speakers alike. With the growing number of students writing dissertations in English, research is increasingly seeking to identify genre and disciplinary-

related linguistic features which might then inform the teaching of genres which are valued in these disciplines (e.g. Jalali & Ghayoomi 2010). Hyland (2008) explores this variation by investigating bundles across different disciplines and genres.

However, the question of whether variation is found within disciplines is much less investigated. Moreover, since lexical bundle analysis is still a relatively new field, few studies have sought to apply Hyland's framework. In this study, we aim to explore bundle functions in Master's dissertations across four sub-disciplines of the same field of study (Electrical and Electronic Engineering). In doing so we aim to address the following research questions:

1. What are the most frequent four-word lexical bundles in each Electrical/Electronic Engineering sub-discipline? What is the extent of the overlap between these?
2. What are the main structural categories of these bundles across the sub-disciplines?
3. What are the distributions of functional categories of these four-word lexical bundles across the sub-disciplines? What are their main realisations at a sub-functional level?

The answers to these questions can help us determine the pedagogical implications.

Methodology

Data collection.

The National Institution of Electrical and Electronic Engineering (IEEE) is based at the University of Mohamed Bougara Boumerdes (UMBB) in Algeria. IEEE is, to our knowledge, the only institute in the country that offers an English Master's programme in Engineering, and the only one with English as the sole medium of instruction in a scientific discipline. IEEE has two departments (Department of Electrical Engineering and Department of Electronic Engineering) divided into four sub-disciplines: Power Engineering, Control Engineering, Computer Engineering, and Telecommunication Engineering.

After obtaining permission to collect the data from UMBB, the 72 dissertations produced at IEEE in 2014 were collected in computer-readable format. During the conversion process, 2 dissertations were corrupted and could not be processed, so they were not included in this study. All of the main textual parts of the dissertations including the Abstracts were included in the present study; Acknowledgments, Tables of Contents and References were excluded. They were compiled into a corpus of 594,599 words divided into four sub-disciplinary sub-corpora (see Table 3).

Table 3 Word counts of the corpus and sub-corpora used in the study

Departments	Electrical Engineering		Electronic Engineering		Corpus size
Sub-corpus	Power E	Control E	Computer E	Tele E	594599
Word Count	181352	130323	145147	137776	
No. of Dissertations	23	15	17	15	70

Corpus annotation.

Certain features of the dissertations cannot be read by corpus software and so were replaced by annotations. Visual data like figures and tables were replaced by the single words 'figure' or 'table', respectively. Mathematical symbols were also removed and replaced by either <formula> or by <exp> (for individual expressions). This annotation prevented the emergence of unclear or unreadable output while preserving the location of information lost in the conversion process.

Lexical bundle selection criteria.

Lexical bundles were extracted from each sub-corpus using AntConc (Anthony 2015) taking account of three main criteria: length, frequency, and dispersion/range. The 'ignore case' option available in AntConc was used to avoid separating identical bundles starting with capital letters and lower-case letters. Bundles were not counted across sentence or similar punctuation boundaries (e.g. parentheses).

Studies such as Biber et al. (1999), Biber and Barbieri (2007) and Hyland (2008) have argued that 4 words is the optimum bundle length to study since they are 'far more common than 5-wordstrings and offer a clearer range of structures and functions than 3-word bundles' (Hyland, 2008, p. 8). As illustrated in Biber et al. (1999, p. 993), 3-word bundles are commonly part of longer bundles. We therefore follow common practice in extracting 4-word bundles.

The frequency criterion, on the other hand, is acknowledged to be "somewhat arbitrary" (Biber & Barbieri, 2007, p. 267; Hyland, 2008, p. 8) since

arguments can be made for a range of thresholds. Such thresholds are generally based on a normalised frequency (e.g. per million words - pmw) to allow comparisons across sub-corpora and studies. Biber et al. (1999) include bundles which occur at least 10 times pmw, but most other studies use higher cut-offs. Hyland (2008) and Cortes (2004, 2006) set a more conservative 20 instances pmw for bundle selection. Meanwhile, Biber et al. (2004), and Biber and Barbieri (2007) apply an even stricter cut-off point of 40 times pmw.

Similar variation can be found when it comes to measures of dispersion, or range, which is important to minimise the risk of one particular writer's preferences skewing the findings (Pan et al., 2016). Biber et al. (2004, p. 375) set a cut-off of occurrence in around 2% of texts, while Biber and Barbieri (2007, p. 267) set this at around 5% of their texts; Hyland (2008), only considered bundles occurring in at least 10% of his texts.

For this study, we applied quite strict frequency and range thresholds. Bundles had to occur 40 times pmw to be considered for analysis. All results were rounded to the closest figure, meaning that bundles had to occur a minimum of five times in Control Engineering and Telecommunication Engineering, and six in Computer Engineering. In terms of range, only bundles that occurred across 20% of dissertations were considered. In the case of Power Engineering, this meant a bundle had to occur in 20% of the 23 dissertations, or 4.6, which was rounded up to 5; the figure for the other sub-disciplines was occurrence in 4 of the texts.

Analysis of bundles by structure and function.

In line with the aims of this study, after lists of lexical bundles were generated and saved, each bundle was assigned a structural and then a functional classification. This often entailed an analysis of lines including the bundle from the corpus. Since two researchers were involved in this study, we were able to check each other's analyses and reach consensus.

The first, more straightforward stage was to classify each list in terms of the structural features attributable to each bundle. This classification presented relatively few problems, since it was possible in cases of doubt to compare with lists and examples in Biber et al. (1999) and Hyland (2008).

The second stage was to classify each bundle in terms of the function and sub-function it realised using the framework introduced in Table 2. The analysis in this stage was more problematic. The issue here, also noted by Ädel and Erman (2012) is that Hyland (2008) is not fully explicit regarding his framework, not providing a comprehensive list of the bundles analysed and their associated functions. This study, therefore, faced some challenges in applying Hyland's framework since not enough detail is available either in terms of descriptions of categories nor the items included in each one[2]. It was, for example, difficult to understand why *the magnitude of the* and *the size of the* are in different categories in Hyland (2008, p. 13): the former is provided as an example of 'Quantification',

[2] This is a point that is generally true for studies of bundles; with only a few exceptions where data is provided in the form of appendices (e.g. Cortes, 2013), complete lists of bundles and their functional analyses are not provided.

while the latter is in the 'Description' category. For the purposes of this study, therefore, the two sub-categories were merged.

Two further analytical issues arose which are addressed by Biber et al. (2004) and Biber (2006) but rather glossed over by Hyland (2008). The first of these regards the multi-functionality of some bundles, a good example being *at the same time*, which may refer to simultaneity or have a meaning similar to *on the other hand*. In such cases, we followed Biber (2006) in categorising such bundles according to their majority function. A second issue regards the inherent multi-functionality of certain bundles. An example of this is *can be used to*, which arguably expresses two functions at the same time; the use of *can* may be associated with the expression of 'stance' (Biber, 2006), while passive *used to* is associated with the description of procedures (Hyland, 2008). There is no totally satisfactory means of dealing with bundles of this sort. Our approach was to categorise them according to their apparent main function in context. In the case of *can be used to*, the 'procedure' meaning seemed more salient than the 'stance' meaning.

Results and discussion

The most frequently occurring bundles: comparisons across sub-disciplines.

Table 4 shows the 20 most frequently occurring bundles in each of the Engineering sub-disciplines in order of frequency, and gives an idea of the extent of the overlap across sub-disciplines. Where more than 20 are included, this indicates that all bundles in the final row had equal frequency. Bundles in **bold**

occur in the top 20 of all sub-corpora, while those in *italics* occur in three sub-disciplines.

Table 4 The 20 most frequent 4-word bundles across the four sub-disciplines

Control Engineering	Computer Engineering	Power Engineering	Telecommunication Engineering
in this chapter we	is shown in figure	as shown in figure	as shown in figure
as shown in figure	as shown in figure	is shown in figure	is shown in figure
the closed loop system	in this chapter we	as shown in fig	with respect to the
with respect to the	can be used to	of the power system	as shown in fig
as shown in the	the nios ii processor	in the case of	is defined as the
of the closed loop	one of the most	can be used to	is equal to the
in the case of	the size of the	is one of the	the total number of
can be written as	it can be seen	it is necessary to	the length of the
if and only if	the implementation of the	*the output of the*	to the number of
the position of the	the speed of the	is equal to the	can be used to
we are going to	*we are going to*	is given by the	is given by the
the length of the	the performance of the	it is possible to	is the number of
the steady state error	a wide range of	*we are going to*	the center of the
shown in the figure	is connected to the	is based on the	at the same time
can be divided into	that can be used	one of the most	the performance of the
can be used to	at the same time	the difference between the	can be written as
is shown in figure	is the number of	the effect of the	is based on the
the difference between the	speed of the motor	we can see that	the end of the
the output of the	is one of the	as well as the	the size of the
to be able to	a graphical user interface	at the end of	are shown in figure
	can be divided into	on the other hand	as a function of
	in addition to the	the voltage and current	at the end of
	in this project we		the effect of the
	nios ii based system		the upper and lower
	the output of the		
	the state of the		
	this project is to		

It is noticeable that the bundles found towards the top of all the lists involve writers referring to data contained in figures, reflecting that these Engineering students frequently present results with reference to visual data, a point also noted by Hyland (2012). We also note that the three ubiquitous bundles are all listed in the top 10 of Hyland's (2008, p. 12) list of bundles found

in Electrical Engineering. However, there is also a degree of variability across the four corpora; only three bundles are found in all four sub-corpora and only two across three of the four sub-disciplines.

Clearly, however, simply considering the most frequent bundles in each sub-corpus cannot provide a clear perspective on the bundles in these sub-disciplines in terms of their structural or functional features. For this, a structural and functional analysis of lexical bundles is required.

Structural features of lexical bundles: Comparisons across sub-disciplines.

As noted in Section 2.1, our investigation broadly follows the schemes set out by Biber et al. (1999) and Hyland (2008). This allows for comparison with Hyland's (2008) findings for Electrical Engineering, although some caution is needed in this respect. Firstly, his corpus is composed of Masters dissertations, PhD theses and research articles and secondly his thresholds were not as strict as the ones used in this study (see Section 3). It is, nevertheless, interesting to consider which categories differ from Hyland's (2008) findings for Engineering writing.

The differences from Hyland (2008) are most clearly seen in three main areas: passive constructions followed by prepositional phrases ('Passive + PP': *is shown in figure*); anticipatory *it* structures (*it can be seen*); and the 'Other' category (see Table 5). To some extent explanations of these findings can be sought in the 'associations' Hyland (2008) notes between structural features and main functions of bundles which will be discussed in more detail in the following section, but some initial comments can be made.

Table 5 Proportions of structural types of bundles across the corpora and in Hyland (2008)

Structure	Hyland (2008): Electrical Eng.	Power Eng.	Control Eng.	Computer Eng.	Telec. Eng.
NP + of	22.3%	26.8%	19.1%	19.1%	23.5%
Other NP	10.8%	6.2%	9.3%	10.0%	5.4%
PP + of	7.9%	9.1%	9.4%	8.4%	13.1%
Other PP	11.6%	6.5%	17.2%	16.0%	11.5%
Passive + PP	29.8%	13.6%	12.0%	16.7%	13.6%
Anticipatory it	8.4%	4.8%	1.9%	3.5%	3.0%
Others	9.2%	33.0%	31.2%	26.4%	30.0%

The far lower proportions found for 'Passive + PP' across all the sub-disciplines compared to Hyland (2008) is one interesting finding. It is, however, hard to explain without full access to Hyland's data. One possible cause of this discrepancy is the general tendency of academic prose to contain more passive structures, a tendency that has been associated with a more objective stance (Biber et al., 1999). Since Hyland's corpus contains texts written by more experienced academic writers, it makes sense for passive voice to be found more often there, although it is beyond the scope of this study to investigate this issue further.

Similar points can be made concerning the relatively low proportion of 'Anticipatory *it*' bundles. From a functional perspective these bundles tend to be (parts of) impersonal constructions which express the writer's view or address the reader; they are therefore typically categorised as Participant-oriented bundles, as in Hyland (2008). As a relatively advanced feature of academic discourse, this finding is not altogether unexpected. Anticipatory *it* is associated with more advanced academic writing; Hyland (2008) notes the comparative lack of this type of construction in his postgraduate texts.

As for the far higher proportions of 'Other' structural bundles, this difference seems largely attributable to high numbers of 'Adverbial' (*as shown in figure*) and 'be + NP/Adjective' bundles (*is equal to the*), which are apparently very useful to these students and which also feature in the functional analysis below.

Distributions of bundles by function.

This section describes the results of the comparison of bundles found in terms of their functional classification. As noted in Section 3.4, this classification proved to be a more problematic area of analysis and conclusions are consequently more tentative.

Figure 1 shows the comparison by main bundle function across the sub-disciplines investigated in this study and also the proportions reported in Hyland (2008) for Electrical Engineering texts. As we can see, the general pattern is that research-oriented (RO) bundles are the most frequent, then text-oriented (TO) bundles, with participant-oriented (PO) bundles being the least frequent.

The overall similarity in proportions across all the corpora shown in Figure 1 is interesting in view of the differences in terms of the structural types identified in the two corpora (see Section 4.2). This serves as a reminder that, while certain structures may be associated with particular functions, this is far from a one-to-one relationship. This is why it is important to take a closer look at each function in turn to examine differences at a finer level of distinction.

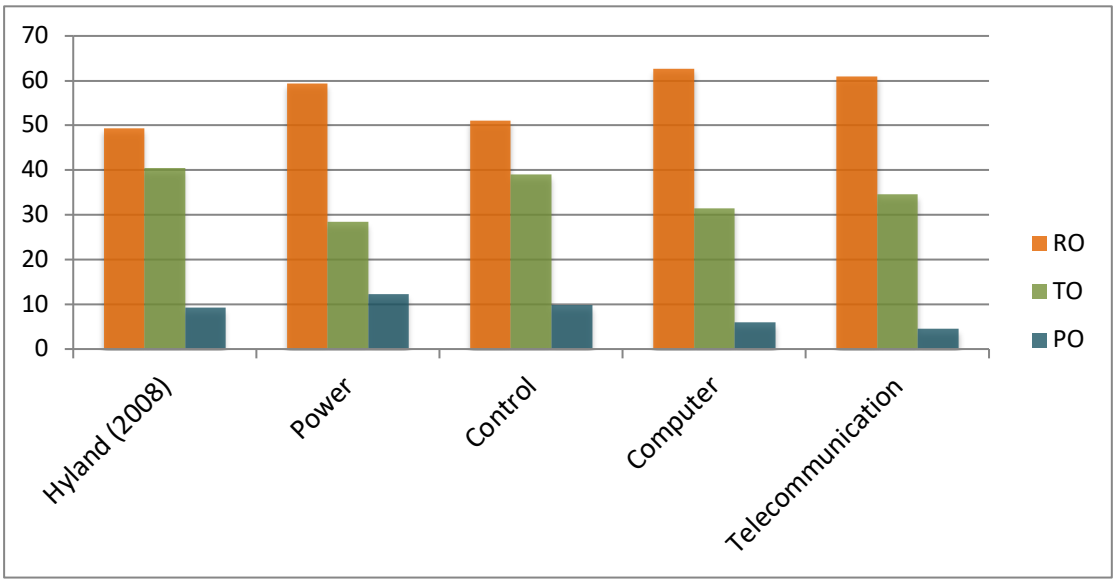

Figure 1. Distribution of functional types of bundle by percentage of total; comparison with Hyland (2008).

Research-oriented (RO) bundles.

RO bundles clearly play an important role in all engineering dissertations, accounting for between 49% and 64% of all bundles (see Figure 1). The prominence of bundles expressing this function is unsurprising as it has already been noted in Hyland (2008). However, reporting only proportions of all bundles rather than the normalised frequencies removes a quantitative aspect of comparability. Figure 2 gives normalised figures for all bundles, providing figures also by sub-function.

The variation in the distribution of the different sub-types of RO bundles across the four sub-disciplines is most clearly seen in comparing Computer Engineering with Telecommunication Engineering. The former has a far greater proportion of 'procedure' bundles, while the latter appears to downplay

procedure but be more focused on 'description'. It is also clear that 'location'[3] is a minor sub-category in terms of frequency.

As we can see in Figure 2, the clearest distinction in terms of distribution can be seen in relation to the 'description' sub-category. Bundles in this subcategory were found principally to fall into two structural types. The first of these consists of a form of *be* followed by either a noun or adjective phrase typically used for quantitative descriptions (*is equal to the*), definition (*is a device that*) or exemplification (*is an example of*). The second, and more frequent structural type is noun phrases including *of*; nouns typically refer to quantity/amount (*a wide range of*), behaviour/performance (*the performance of the*), and size/dimensions (*the length of the*). The types of bundles found here indicate the extent to which these writers refer to quantities and calculating/measuring them and generally what they might be interested in calculating or measuring.

It seems that all of the sub-disciplines are interested in measuring performance in some way and that the bundles they use to do this converge on a limited number of forms (e.g. *the behaviour/response/speed/output of the*). The higher number of 'description' bundles found in Telecommunication can largely be attributed to those referring to numbers and/or calculations, (*is equal to the, the total number of*) or to size and dimensions (*the size/length of the*). This is a reflection of how Master's dissertations in this sub-discipline have a particular need to report calculations, quantities and dimensions.

[3] This study follows Hyland (2008) in putting under 'location' only bundles relating to location in the text, e.g. *the beginning of the* chapter

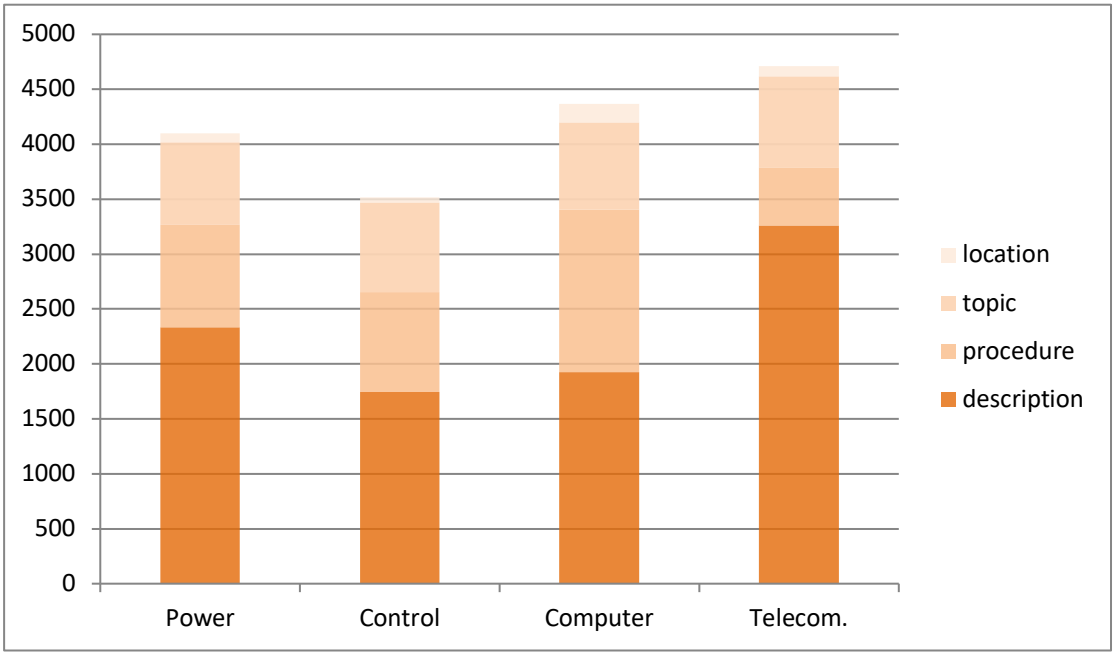

Figure 2. Normalised frequencies of bundles classified as realising the RO function for each sub-discipline, separated into sub-types.

The majority of 'procedure' bundles are based around *be used to/for/in* (eg. *can be used for/to*). The higher frequency in Computer Engineering is mainly due to bundles making specific reference to design and implementation (*the implementation of the*) and to bundles that describe more specific research processes (*is sent to the, is connected to the, control the speed of*), which are almost absent from the other sub-disciplines.

The final main point arising from Figure 2 relates to 'topic' RO bundles. This grouping of bundles relates to the specific field of the research carried out and consists of bundles with discipline-specific terminology. While the overall frequency of 'Topic' bundles remains stable across the sub-disciplines, the specific bundles show very little overlap, except in general electrical terminology (*the impedance of the*). Items occurring in more than one bundle in each sub-

discipline are shown in Table 6. These bundles indicate some of the key terms and material that writers in each sub-discipline need to master.

Table 6 Topic'-related bundles with repeated sub-discipline specific items, by sub-discipline

Power	• the current and voltage / the voltage and current / of current and voltage / of voltage and current • of the power system / the power system and / the power system is / in the power system • the transfer function of / transfer function of the
Control	• between the robot and / of the robot in / of the robot is • configuration of the robot / the robot and the • the closed loop system / of the closed loop
Computer	• the nios ii processor / on the nios ii / nios ii based system • system on a programmable / on a programmable chip / field programmable gate array
Telecom.	• of an antenna is / size of the antenna • the characteristic impedance of / the impedance of the • of the received signal / the transmitted signal and • the resonant frequency is / at the resonant frequency

Text-oriented bundles across sub-corpora.

As indicated in Figure 1, Text-Oriented (TO) bundles make up between 28% and 39% of bundles found across all the sub-corpora and are therefore an important resource for Masters dissertation writers. Figure 3 shows the normalised frequencies of these bundles and their distributions by sub-discipline in terms of the sub-categories proposed by Hyland (2008) shown in Table 2.

The overall frequencies shown in Figure 3 indicate a degree of variation, with Control and Telecommunication Engineering showing higher overall usage of TO bundles, and Computer and Power Engineering having considerably lower frequencies. This suggests that dissertation writers in the latter two sub-

disciplines do not work as hard to guide the reader through their texts or do not rely on as much conventionalised phraseology to do so.

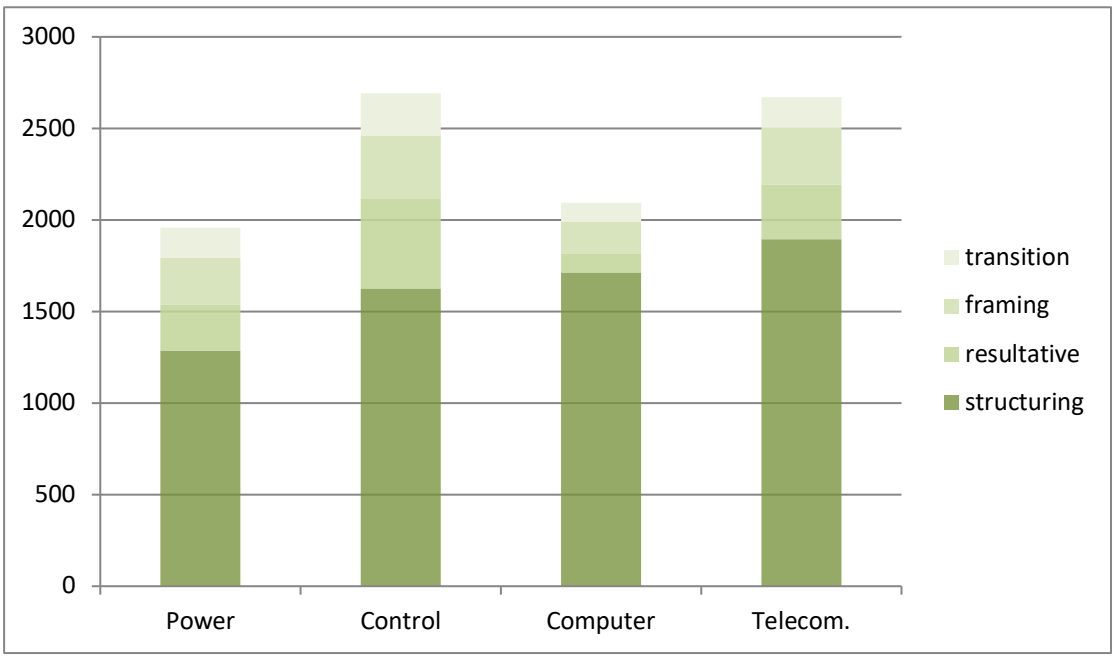

Figure 3. Normalised frequencies (pmw) of TO bundles by sub-discipline and by TO sub-category.

Bundles functioning as 'transition signals', that is, those making 'additive or contrastive links between elements' (Hyland, 2008, p. 14) such as *on the other hand* are relatively infrequent across all the sub-disciplines and do not seem to have a great significance. 'Framing signals' are those that 'situate arguments by specifying limiting conditions' (Hyland, 2008, p.14). This is also an infrequent group with a low number of bundle types. These findings support Hyland's (2008) observation that these signals are more commonly found in disciplines with wider, less focused readerships such as Applied Linguistics.

'Resultative' bundles are also relatively infrequent, but there is a wider degree of variation. Following Hyland (2008), these bundles are of two main

types: those that express a cause-effect relationship such as *is due to the* and those that report results, for example *it has been found*. Neither of these types considered separately are widely found in any of the sub-disciplines.

The most significant sub-category of TO bundles in terms of frequency across all the sub-disciplines is 'structuring signals' which are of two main types. The first comprises bundles which refer the reader to figures or tables. Figure 4 presents the main variations on this bundle type, indicating how the choices at each step in the phrase are relatively limited. These 'shown' bundles are prevalent across all the sub-disciplines – in particular in Power and Telecommunication Engineering.

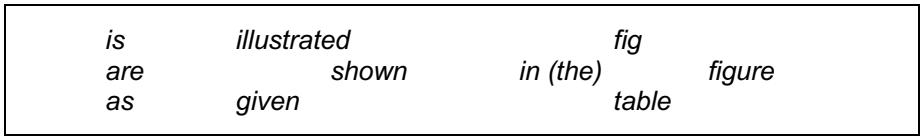

Figure 4. Schematic representation of structuring signals based around shown.

The second type of structuring signal bundles are those forming part of expressions used to draw the reader's attention to either the whole work or part of it to summarise what is said there (e.g. *in the next chapter*) or to point out the aims or objectives of the study (*the aim of this*). In contrast with the first type, the distribution of this second type of bundles is rather skewed, being more commonly found in Control and Computer Engineering dissertations than in Telecommunication Engineering (over twice the frequency) and Power Engineering (over four times as frequent). This is surprising since there is nothing to suggest that Power or Telecommunication Engineering should avoid such signals.

Participant-oriented bundles across sub-corpora.

Participant-oriented (PO) bundles are not frequently found across the sub-corpora, the highest proportion being in Power Engineering, with 12% (see Figure 1). Hyland (2008) also finds a relative lack of PO bundles in Engineering. PO bundles can be divided into two main types. The first of these is bundles expressing *stance*, which writers use to express judgements of likelihood and possibility (*the fact that the*), affective judgements[4] and their level of commitment to a proposition (*is considered to be*). The second type is 'Engagement' bundles, by means of which 'writers intervene to actively address readers as participants in the unfolding discourse' (Hyland, 2008, p.18), for example *it is important to*.

Figure 5 shows the distribution of PO bundles across the four sub-disciplines. Power Engineering has more than double the frequency of Computer Engineering and Telecommunication Engineering. While Power and Computer Engineering show even distributions of the two sub-types of PO bundles, Control and Telecommunication Engineering students have a stronger tendency to use stance bundles. This contrasts with Hyland's (2008) finding of a preponderance of engagement bundles in hard sciences and in particular in student genres, although Hyland points out this may be due to the first language of the writers (all his student texts were collected in Hong Kong).

[4] No clear examples of this type were found in the corpora; this is indeed to be expected since academic prose tends to avoid explicit reference to affect (e.g. *it is good that*).

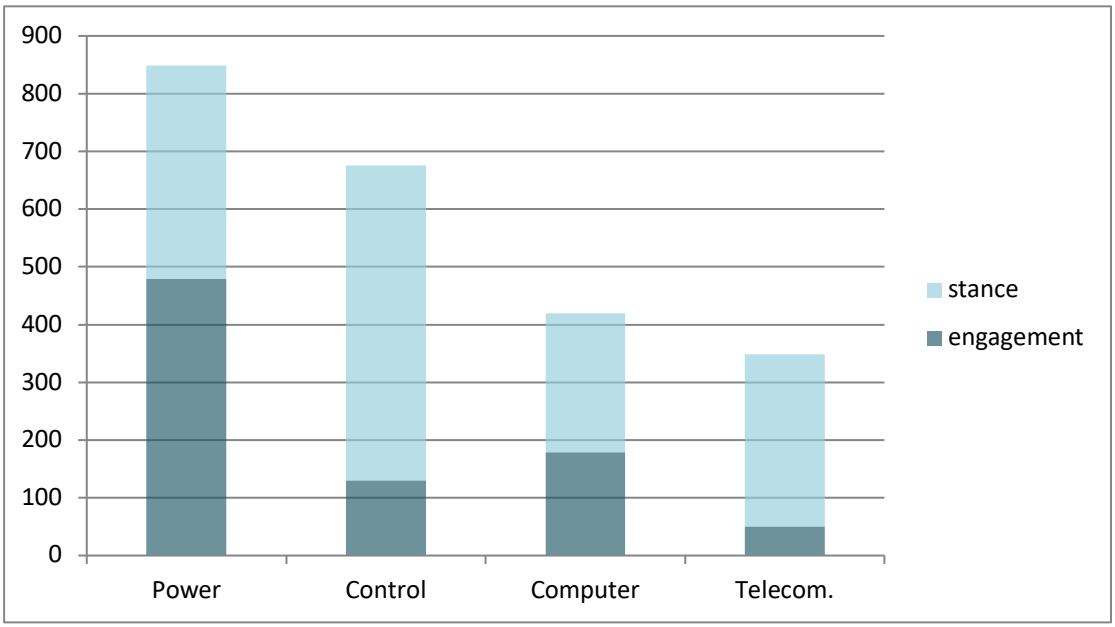

Figure 5. Distribution of PO bundles across the four sub-disciplines (normalised frequencies pmw).

Engagement bundles are favoured most by writers of Power Engineering dissertations (see Figure 5). These bundles are used either to draw the reader's attention to writer interpretations of data from a figure or table (*it can be seen*) or to emphasise the importance of a particular step or of understanding a point (*it is important to*) (Hyland, 2008). The first of these uses is not found in Telecommunication Engineering.

Bundles involving the expression of stance are found across all the sub-disciplines in higher frequencies than expected, particularly in Control and Telecommunication Engineering. These bundles mainly involve the expression of epistemic modality, that is, the degree of certainty that the writer attributes to a proposition (Lyons, 1977); included in this category is reference to ability (*to be able to*) as a 'special case of possibility' (Quirk et al., 1985, p.221). Whether the bundles concerned express high certainty (*the fact that the*) or are more tentative

(*it is possible to*), the variety of bundle types is small, following a general pattern amongst these dissertation writers to use a small repertoire of bundles to express a specific meaning.

Implications of the study.

The foregoing discussion indicates that a clearer picture of both differences and commonalities across the sub-disciplines investigated in this study emerges at the level of the functional sub-categories. It is perhaps in the area of RO bundles that the differences involved seem most meaningful, for example, the relatively high numbers of RO bundles in Telecommunication Engineering that relate to quantification and description. In terms of commonalities, we can note the preponderance of TO bundles functioning as structuring signals, an important communicative function across all the sub-corpora investigated here.

The finer-grained approach taken in this study has also identified 'clusters' of bundles, such as those based around *shown* and its synonyms (Figure 4). Indeed, the association of form and function – inspired by work in phraseology (e.g. Sinclair, 1996; Stubbs, 2002; Hoey, 2005) – seems a useful step if we take the position that bundles are worthy of pedagogical attention. This, we feel, represents a development on Simpson-Vlach and Ellis's (2010, pp.498-502) means of presentation of their Academic Formulas List. That is, an implication of this study is that it is important not only to present the main bundle types expressing a specific meaning, but also to draw attention to the formal similarities between them. Other examples of 'bundle clusters' we have seen include bundles based around the verb *use* (*can be used to/for/in*, *it/which is used to*)

which are useful in describing procedures, and the cluster based around the semantic set of numbers/calculations (*the number/sum/ratio/value of the*) relating to quantification.

We should also bear in mind Simpson-Vlach and Ellis's (2010, p. 502) recommendation that findings of this sort are best viewed as 'a resource for developing teaching materials based on further contextual research [...] rather than a resource for teaching itself'. That is, pedagogical treatments may be more effective if they draw on phraseological research such as Sinclair (1991, 1996), Stubbs (2002) and Hoey (2005) and investigate specific co-texts of bundles and bundle clusters, such as which words typically precede *is/are shown in fig/figure* or the most common sentence position of *as shown in fig/figure*. Indeed, as noted in the Introduction, while bundles are a useful approach to phraseology, their limitations in terms of fixedness suggest we should also be looking beyond bundles to more variable phraseological phenomena (Vincent, 2013).

A further implication of this study relates to the application of functional frameworks such as those in Biber (2004, 2006) and Hyland (2008). The difficulty for studies which seek to apply these frameworks regards how the classifications proposed map on to specific forms not exemplified in the studies. As Ädel & Erman (2012) note, it is important to be clear about the issues faced in categorisation since otherwise comparability across studies is either difficult or impossible. Another option is to provide access to the full list of bundles and

their categorisations in a study by adding an appendix. Bearing in mind this issue, we have decided to make our own bundle analysis available[5].

6. Conclusion

This paper has explored the structural and functional patterns of four-word lexical bundles in successful Algerian Engineering Masters dissertations. The findings for the bundles found in these Master's dissertations are generally consistent with earlier results (Hyland, 2008) in that certain features of more 'expert' academic writing are found less frequently. While such analysis 'offer[s] an important means of differentiating written texts by discipline' (Hyland 2008, p.4), it is more open to debate whether it provides the same sort of differentiation at a sub-disciplinary level based on our findings. Nevertheless, it is clear that the findings of bundle analysis can inform pedagogical interventions. One approach that seems particularly promising in this respect is the grouping of bundles with similar functions into 'bundle clusters'. This, we believe, offers a way of extending the use of an approach originally intended for analysis of different discourses.

[5] The full analysis is downloadable from this site:
http://acebundles.benetvincent.coventry.domains

References

Ädel, 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.

Altenberg, B. (1998). On the phraseology of spoken English: the evidence of recurrent word-combinations. In A. P. Cowie (Eds.), *Phraseology: theory analysis and applications* (pp. 101-122). Oxford: Oxford University Press.

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

Biber, D. (2006). University language: A corpus-based study of spoken and written registers. Amsterdam: Benjamin.

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

Biber, D., Conrad, S., Reppen, R., Byrd, P., Helt, M., Clark, V., Cortes, V., Csomay, E., & Urzua, A. (2004). *Representing language use in the university: Analysis of the TOEFL 2000 Spoken and Written Academic Language Corpus. TOEFL Monograph Series MS 25.* Princeton, New Jersey: Educational Testing Service.

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

Chen, Y.-H., & Baker, P. (2010). Lexical bundles in L1 and L2 academic writing. *Language Learning and Technology, 14*(2), 30-49.

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

Cortes, V. (2013). The purpose of this study is to: Connecting lexical bundles and moves in research article introductions. *Journal of English for Academic Purposes, 12*(1), 33-43.

De Cock, S. (1998). A recurrent word combination approach to the study of formulae in the speech of native and non-native speakers of English. *International Journal of Corpus Linguistics*, *3*(1), 59–80.

De Cock, S. (2000). Repetitive phrasal chunkiness and advanced EFL speech and writing. In C. Mair & M. Hundt (Eds.), *Corpus linguistics and linguistic theory* (pp. 51–68). Amsterdam: Rodopi.

Durrant, P. (2015). Lexical Bundles and Disciplinary Variation in University Students' Writing: Mapping the Territories. *Applied Linguistics, 38*(2), 165-193.

Ellis, N. C., Simpson-Vlach, R. & Maynard, C. (2008). Formulaic language in native and second language speakers: Psycholinguistics, corpus linguistics, and TESOL. *TESOL Quarterly*, *42*(3), 375–396.

Firth, J. (1951). Modes of meaning. In Firth, J. (Ed.), *Papers in linguistics: 1934-1951* (pp. 190–215). Oxford University Press.

Gledhill, C. (2000). The discourse function of collocation in research article introductions. *English for Specific Purposes*, *19*(2), 115-135.

Halliday, M. A. K. (1985). *An introduction to functional grammar*. London: Arnold.

Hunston, S. (2002). *Corpora in applied linguistics*. Cambridge: Cambridge University Press.

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

Hyland, K. (2012). Bundles in academic discourse. *Annual Review of Applied Linguistics*, 32, 150-169.

Jalali, H. & Ghayoomi, S. (2010). A comparative qualitative study of lexical bundles in three academic genres of applied linguistics. *Modern Journal of Applied Linguistics, 2*(4): 323-333.

Lyons, J. (1977). *Semantics*. Cambridge: Cambridge University Press.

Nesi, H., & Basturkmen, H. (2006). Lexical bundles and discourse signalling in academic lectures. *International Journal of Corpus Linguistics, 11*(3), 283–304.

Palmer, H. (1933). *Second Interim report on English collocations*. Tokyo: Kaitakusha.

Pan, F., Reppen, R., & Biber, D. (2016). Comparing Patterns of L1 Versus L2 English Academic Professionals: Lexical Bundles in Telecommunications Research Journals. *Journal of English for Academic Purposes*, 21, 60-71.

Quirk, R., Greenbaum, S., Leech, G. & Svartvik, J. (1985). *A comprehensive grammar of the English language.* Harlow: Longman.

Römer, U. (2009). The inseparability of lexis and grammar corpus linguistic perspectives. *Annual Review of Cognitive Linguistics*, *7*(1), 140-162.

Schmitt, N., Grandage, S., & Adolphs, S. (2004). Are corpus-derived recurrent clusters psycholinguistically valid?. In N, Schmitt (Eds.), *Formulaic sequences* (pp. 127–152). Amsterdam: John Benjamins Publishing.

Scott, M. & Tribble, C. (2006). *Textual patterns.* Amsterdam: John Benjamins Publishing.

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

Sinclair, J. (1996). The search for units of meaning. *Textus*, *9*, 75-106.

Staples, S., Egbert, J., Biber, D., & McClair, A. (2013). Formulaic sequences in EAP writing development: lexical bundles in TOEFL iBT writing section. *Journal of English for Academic Purposes*, *12*(3), 214-225.

Stubbs, M. (2002). *Words and phrases.* Oxford: Blackwell. University Press.

Stubbs, M. & Barth, I. (2003). Using recurrent phrases as text-type discriminators: a quantitative method and some findings. *Functions of Language, 10*(1), 61-104.

Vincent, B. (2013). Investigating academic phraseology through combinations of very frequent words: A methodological exploration. *Journal of English for Academic Purposes, 12*(1), 44-56.

Wray, A. (2002). *Formulaic language and the lexicon*. Cambridge: Cambridge.