

Bundle-driven move analysis: Sentence initial lexical bundles in PhD abstracts

Abstract

Rhetorical moves are discursal units that perform communicative functions in a given genre. They have been manually identified in many previous studies taking a corpus-based approach, and these studies have provided important contributions to discourse structure theories. However, manual analysis has restricted the scale and quantity of texts under investigation and the identification of moves is likely to be influenced by researchers' prior knowledge. In contrast, a corpus-driven approach, by applying automatic computational technology to process a large number of texts, possibly results in a better representation of the moves in a genre and minimises researcher bias in move identification. This study applies one typical corpus-driven approach, a bundle-driven approach, to analyse rhetorical moves of PhD abstracts, a move-intensive genre. The study focused on 5-word sentence initial bundles. Almost all the generated bundles could be identified as move indicators. The majority of the indicated moves aligned with the moves proposed for research article abstracts in previous studies, and one new move, *Structure*, was identified. This study indicates the potential of a bundle-driven approach in exploring rhetorical moves and in examining the linguistic features of moves by means of lexical bundles. Pedagogical implications for EAP writing are also suggested.

Keywords: Moves; Lexical bundles; Corpus analysis; Abstracts; Academic writing; Postgraduates

1. Introduction

The analysis of discourse from a corpus linguistics perspective has tended to have two primary areas of focus: language structures and linguistic forms, and the functions that structures and forms play. While linguists have by and large reconciled these two, acknowledging that both are integral to the study of texts (Tannen, Hamilton, Schiffrin, & Adger, 2015), corpus linguists can still be observed focusing solely on one, or privileging one over the other. Conrad (2002) for example, describes two approaches: “studies focused on individual language features [and studies focused] on a function of language to determine how it is realised in discourse” (p. 81). However, as Gómez González, Ruiz de Mendoza Ibáñez, González García, and Downing (2014) state, “form is not independent of meaning/function or language use” (p. 1). Much is gained when these two “tendencies” in corpus analysis are seen in a complementary relationship with each other, and the researcher draws on one to help explain more fully the occurrence of the other. The present paper seeks to do this with respect to functional units (*rhetorical moves*) and their linguistic realisation.

Rhetorical moves (or *moves* as they are commonly named) are discursal units that perform different communicative functions in a given genre. *Steps* are the subdivisions of those moves. These terms were first introduced by John Swales (1981) to describe the discourse structures of research article introductions. Swales's influential work on move analysis (1981,

1990, 2004) has stimulated a large number of move studies in a wide range of academic genres, including research article abstracts (Jiang & Hyland, 2017; Omidian, Shahriari, & Siyanova-Chanturia, 2018), sections of research articles (Cotos, Huffman, & Link, 2017; Gray, Cotos, & Smith, 2020; Moreno & Swales, 2018), PhD theses (Kwan, 2006; Swales, 2004), student laboratory reports (Parkinson, 2017), university textbooks (Parodi, 2010, 2013) and Three Minute Thesis (3MT) presentations (Hu & Liu, 2018). Moves have also been analysed in a broad variety of academic disciplines, for example, Applied Linguistics (Yang & Allison, 2003; Yoon & Casal, 2020), Law (Tessuto, 2015), Engineering (Kanoksilapatham, 2005, 2007, 2015), Computer Science (Chang & Kuo, 2011), Chemistry (Stoller & Robinson, 2013) and Dentistry (Basturkmen, 2012). Manual analysis has been widely used in the move studies and researchers' linguistic and subject knowledge is crucial in identifying textual boundaries.

In terms of the linguistic realisation of moves, we can look at *lexical bundles* - multiword sequences automatically generated from a corpus solely based on their length, frequency and text distribution (Biber, Johansson, Leech, Conrad, & Finegan, 1999). They deserve specific attention in move analysis and academic writing pedagogy because many of them are comparatively frequent and tend to signpost the stages in a text. Examples include *the aim of this study*, *it should be noted that* and *in the next section*. However, they are not easy for students and writers to "notice" as most bundles "do not represent a complete structural unit" and "are not idiomatic in meaning and not perceptually salient" (Biber & Barbieri, 2007, pp. 269-270). Frequency, as "the ultimate characteristic that defines lexical bundles" (Cortes, 2013, p. 34), allows bundles to be identified automatically and then be used to explore discourse features such as moves. In the following sections, we review approaches taken to the analysis of both moves and lexical bundles with a view to justifying the bundle-driven approach adopted in this study.

1.1. Analysing rhetorical moves

Biber, Connor, and Upton (2007) propose two approaches to move analysis: the top-down approach (starting an analysis by examining communicative functions) and the bottom-up approach (beginning an analysis by examining linguistic features). However, in reality, researchers tend to engage in "various interrelated levels of analysis at the same time: identification of communicative purpose(s), schematic structure, grammatical features, lexical features, etc" (Flowerdew, 2002, p. 95). A more helpful distinction seems to be that proposed by Tognini-Bonelli (2001) who uses the terms: *corpus-based* and *corpus-driven*. The corpus-based approach starts from researchers' prior knowledge of subject matter, schematic structures and linguistic features to manually identify textual boundaries. The corpus is used to verify the pre-defined analytical framework. Such an approach has a long research tradition and many previous move analyses have taken this approach. The corpus-driven approach incorporates advanced computational technology which first automatically generates the linguistic items from texts and, on the basis of these items, functional units are identified. Move analysis using a corpus-driven approach has only fairly recently been developed. One example of such an approach is Biber, Csomay, Jones and Keck's (2007) Vocabulary-based Discourse Unit (VBDU) analysis.

Many corpus-based move studies have contributed significantly to discourse structure theories; however, they are limited by the fact that manual analysis is "highly labor-

intensive” (Biber, Connor, et al., 2007, p. 15) and thus the range and number of texts that can be analysed is relatively small. Examples are 94 abstracts in Dos Santos (1996), 60 research articles in Kanoksilapatham (2007) and 142 3MT presentations in Hu and Liu (2018). A corpus-driven approach with the use of computational technology allows researchers to examine large collections of text data such as the 3,697 PhD abstracts in this study. A large range and number of texts allow for a better representation of a genre (Biber, Csomay, et al., 2007), and the use of computers minimises the influence of a researcher’s prior knowledge on the identification of moves.

1.2. Analysing lexical bundles

Lexical bundles have been investigated in a range of different genres (e.g. research articles, theses and textbooks), and researchers have explored them by varying the type of texts in their corpus, and the length and position of their generated bundles. Lexical bundles can be generated automatically by corpus tools. Biber (2015) regards the generation approach as “an extreme corpus-driven approach”, “beginning with simple word forms” and “giving priority to frequency” (pp. 202-203). This approach has been named more specifically in Li, Franken, and Wu (2017) as a *bundle-driven approach* as the recurrent word sequences are lexical bundles. Cortes (2015) argues that the generation approach is empirical as researchers generate bundles “leaving behind their intuition or perception” (p. 205). This automatic bundle generation approach allows linguistic categories to emerge from the corpus evidence (Ädel & Erman, 2012; Granger, 2017) so that it is capable of developing categories, taxonomies and frameworks whilst still being informed by existing research and literature. The automatic approach also allows researchers to analyse large scale text data to explore the possible connections between the linguistic features and discourse functions carried by these recurrent word sequences. Granger (2018) regards the bundle-driven approach as a promising method to mine the “gold” (p.192) of metadiscursive markers. Likewise, a bundle-driven approach can be employed to explore the relationship between lexical bundles and other discourse features like rhetorical moves.

With regard to the capability of a bundle-driven approach, a small number of studies have used lexical bundles to explore salient discourse features such as metadiscourse shared between texts (e.g. Granger, 2017; Li et al., 2017). However, this type of research is scarce in academic writing analysis in comparison to the large number of bundle studies focusing on classifying generated bundles into existing functional categories developed by previous researchers – in particular Biber and his colleagues’ taxonomy (Biber, Conrad, & Cortes, 2003, 2004) (i.e. referential bundles, discourse organizers and stance bundles) and Hyland’s (2008a) framework (i.e. research-oriented, text-oriented and participant-oriented bundles).

1.3. Analysing the connection between lexical bundles and rhetorical moves

Lexical bundles were the starting point and original focus of this research. However, we became interested in seeing if we could not only identify lexical bundles structurally but also gain some useful insights into the functions of different bundles. Our initial analysis showed that the functions carried by the bundles indicated rhetorical moves within the PhD abstracts, which then led us to this corpus-driven move analysis approach.

Cortes (2013) provides a strong rationale for researching the connection between lexical bundles and rhetorical moves in stating that “a description of the relationship between lexical bundles and moves in a particular register could provide more evidence towards a complete

picture of the tendencies used in the organizational and lexico-grammatical patterns used to build discourse by different speech communities” (p. 36). She borrowed Swales’s (1990, 2004) frameworks for move analysis to identify the moves and steps in which her bundles occurred. Her research established a one-to-one relationship between 5+ word bundles and a single move and step but found that many 4-word bundles occurred in various moves and steps. One reason for the stronger one-to-one connection between longer bundles and moves and steps is that a greater amount of information is contained in longer bundles (Cortes, 2013). Cortes (2013) also suggests that many longer bundles (5+ words) positioned at the beginning of sentences act as triggers to start moves or steps (e.g. *the purpose of the present study, it has been suggested that*). In other words, bundles at the beginning of sentences tend to have a clearer connection with moves or steps than non-initial ones.

Cortes’s (2013) study has inspired a number of researchers (e.g. Ädel, 2014; Cotos, Huffman, & Link, 2015; Mizumoto, Hamatani, & Imao, 2017; Omidian et al., 2018; Yoon & Casal, 2020) and their investigations have extended to other linguistic forms. Ädel (2014) investigated the connection between anticipatory-*it* lexicogrammar pattern and rhetorical moves and found “relatively strong correlations between lexicogrammar and the rhetorical level” (p. 78). Cotos et al. (2015) measured the probability of the occurrence of a certain n-gram (sequences of n words) in a given move and step and successfully identified the n-grams that were better indicators of specific moves or steps. Mizumoto et al. (2017) developed a web-based writing support tool to suggest the most frequent lexical bundles in a certain move (e.g. *this study investigated the* in the move presenting the research in the abstract section). Most recently, Omidian et al. (2018) explored the disciplinary variations in the use of lexical bundles in different moves of research article abstracts. Yoon and Casal (2020) examined the association of P-frames (formulaic phrase frames) with the moves and steps in conference abstracts and observed the strong association between sentence initial P-frames and moves/steps: mono-functional frames “often occurred in a sentence-initial position and initiated the move-step” (p. 294).

Inspired by Cortes’s (2013) study, we took a bundle-driven approach to explore rhetorical moves in PhD abstracts. However, unlike Cortes’s (2013) study, we did not conduct move analysis directly on our texts. In our study, the generation of bundles came first, and the moves of the PhD abstracts emerged from the classification of the generated bundles. We wanted to see if any rhetorical moves could be aligned with the sentence initial bundles in the PhD abstracts that we had identified. In doing this, we could see the possibility of proposing linguistic features signalling specific moves. We focused on sentence initial and 5-word longer bundles with the view this would lead to identifying stronger bundle-move connections.

There are a number of reasons for choosing the PhD abstract genre. Abstracts provide us with a snapshot of “a discipline’s epistemological and social assumptions” (Hyland, 2004, p. 63), operate as “advance indicators of the content and structure of the following text” (Swales, 1990, p. 179), and feature “a promotion of oneself and one’s paper” (Hyland, 2004, p. 63). PhD abstracts, written at nearly the end of PhD journeys, summarise years of research work undertaken by emerging researchers, highlight the key points of their research, and allow PhD candidates ultimately to engage in and contribute to the dialogues in their studied areas. A linguistic reason is that PhD abstracts of around 300 words are short move-intensive texts condensed from book-length theses. Therefore, the sentence initial bundles generated from

PhD abstracts are more likely to indicate rhetorical moves. While most previous studies (e.g. Dos Santos, 1996; Hyland, 2004; Omidian et al., 2018; Pho, 2008) have chosen research article abstracts, the focus on PhD abstracts in this study provides pedagogical insights into the challenges faced by these higher degree students in an aspect of their thesis writing practices.

2. Methodology

2.1. The corpus

This study is based on the PhD abstract collections built in FLAX (<http://flax.nzdl.org>), an open-access language learning and analysis system documented in Witten, Wu, Li, and Whisler (2013). These PhD abstract collections originally come from the E-theses Online Service (ETHOS) Open Access in the British Library. The abstracts were classified into four overarching disciplines on the basis of Dewey Decimal Classification: Arts and Humanities, Social Sciences, Life Sciences and Physical Sciences. This study explored the use of sentence initial bundles specifically in the Arts and Humanities collection, which has 1,164,182 running words and 3,697 abstracts published between 2005 and 2016. The average length of the abstracts is 315 words. The Arts and Humanities collection includes disciplines such as News Media, Journalism and Publishing, Philosophy and Psychology, Religion, Language, Arts, Literature, Rhetoric and Criticism, and History. The reason for choosing this collection was that we previously investigated sentence initial bundles in the PhD theses in the discipline of General and Applied Linguistics (Li, 2016; Li et al., 2017). The focus on the Arts and Humanities collection, featuring a number of PhD abstracts of General and Applied Linguistics or the like, would allow a possible rough comparison between the findings. It should be noted that while a number of abstracts in this collection are very likely to have been written by writers with a different L1, this is not a significant factor that needs to be taken into account in our findings. This is because the focus of our study is not on highlighting the possible differences in language use by native and non-native writers but on revealing the general patterns of discourse functions and linguistic features embedded in published PhD abstracts. The bundle generation criteria, multiple occurrences across multiple texts, also help to remove individual writers' idiosyncratic usage.

2.2. Bundle identification decisions

According to Biber et al. (1999), the three key criteria for generating lexical bundles are the length of word combinations, the frequency threshold and the range of distribution. Another criterion, the position of bundles in sentences, was added in this study. Few bundle studies have distinguished lexical bundles at different positions in sentences. However, sentence initial bundles, unlike non-initial ones, overlap with sentence starters (Cortes, 2013) and thus can be seen to share similarities. Sentence starters mostly act as points of departure to orient the topic of the clause (Cortes, 2013; Flowerdew, 2013; Halliday & Matthiessen, 2004). In other words, sentence initial bundles are highly likely to start a move. In addition, sentence starters are more likely to pose challenges for writers as they need to consider cohesion and coherence of their unfolding text (Hinkel, 2004; Williams, 2003). When they start a sentence, they need to decide which information has already been mentioned, which is familiar to their

readers and what should come first; and they also need to choose a related set of concepts as the topics of their sentences (Hinkel, 2004; Williams, 2003).

As well as bringing advantages from a writer's point of view, there are advantages for the researcher in bundle identification. The focus on sentence initial bundles avoids bundle overlaps (e.g. *is one of the* and *one of the most*). Bundle overlap is fairly common as bundles are usually identified by a computer program which reads from the first word of each text in a corpus, advances one word at a time, stores every n-word sequence, and checks against its previously identified sequences. The overlaps, if not manually filtered, will inflate the frequency of bundle instances.

Five-word sentence initial bundles are the focus of this study. Many bundle studies have chosen 4-word bundles as their focus (e.g. Biber & Barbieri, 2007; Cortes, 2004; Hyland, 2008a, 2008b) as they are "much more frequent than 5-word bundles and many of them include three-word bundles in their structure" (Cortes, 2013, p. 34). However, 5-word bundles contain more information and are more likely to account for a single move rather than multiple moves (Cortes, 2013). Besides, focusing on the 5-word length excludes many 4-word set phrases, such as *at the same time* and *on the other hand*, which often occur as 4-word lexical bundles in academic writing corpora and in several moves/steps (e.g. Cortes, 2013; Hyland, 2008b; Li, 2016).

The cut-off frequency was set as 5 times per million words, less conservative than the cut-off points generally used which tend to be 10+ times per million words (see for example Ädel & Erman, 2012; Hyland, 2008a; Pan, Reppen, & Biber, 2016; Wright, 2019). A lower cut-off point is usually established for longer and less common bundles in bundle studies. For example, Biber et al. (1999) set the cut-off point at 5 times per million words for 5-word and 6-word bundles, and Cortes (2013) chose 8 times per million words for 6-word and 7-word bundles and 6 times per million words for her longer ones. No distribution threshold was set in this study because these PhD abstracts were comparatively short (i.e. 300+ words) and it was very unlikely for a bundle to reappear within one abstract. Two content-based bundles (proper names of entities and processes) were removed from the extracted bundle list: one contained the name of the methodology (e.g. *Interpretative Phenomenological Analysis (IPA) was*) and the other was the name of the service (e.g. *Child and Adolescent Mental Health*).

The findings are presented in the following two sections: establishing the framework, and applying the framework. In the first section, we explain how we initially developed a framework for our structural and functional analysis on the basis of the generated bundles. The framework partially overlaps with those suggested by previous researchers but at the same time, our framework better reflects the features of our present bundle data.

3. Findings – establishing the framework

3.1. The initial structure-based analysis

The structure-based analysis informed the first part of the framework. The structural categories and patterns that were initially used as a reference point to classify bundles were developed from the studies of Biber and his colleagues (Biber et al., 2004; Biber et al., 1999). As presented in Table 1 below, three major categories were identified from our bundle data:

NP-based, VP-based and clause-based bundles. NP-based bundles refer to any nominal phrases with post-modifier fragments, mainly *of*-phrase in this study. VP-based bundles start with infinitive verb components and *in order to explore the* was the only one to appear in our corpus. Clause-based bundles comprise clause fragments and three different patterns were identified in this study: anticipatory it + verb phrase fragment, noun phrase + verb phrase fragment and noun phrase + be.

Table 1. Structures of sentence initial bundles

Categories	Patterns	Examples
NP-based	noun phrase with post-modifier fragment	<i>The aim of this thesis</i>
VP-based	(<i>in order</i>) <i>to</i> -clause fragment	<i>In order to explore the</i>
Clause-based	anticipatory <i>it</i> + verb phrase fragment	<i>It is argued that the</i>
	noun phrase + verb phrase fragment	<i>The thesis is divided into</i>
	noun phrase + <i>be</i>	<i>This thesis is a study</i>

The pattern *noun phrase + verb phrase fragment* is not in Biber et al. (1999) but is the same pattern as that found in Cortes's (2013) longer bundles (e.g. *the objective of this study was to evaluate*). As with our sentence initial bundles, many of her longer bundles act as sentence starters. Seven out of the twelve patterns in Biber et al. (1999) were not found in our data and so were disregarded. Among these, some patterns are fairly unlikely to start a sentence (e.g. passive verb + prepositional phrase fragment, copula be + noun phrase/adjective phrase, (verb phrase +) that-clause fragment, and adverbial clause fragment); and others might not fit into the concise feature of abstract writing or are less likely to be generated as 5-word bundles (e.g. prepositional phrase with embedded of-phrase fragment, and other prepositional phrase fragment). We had no bundle that needed to be classified as other expressions, as Biber et al. (1999) had.

3.2. The initial function-based analysis

The function-based analysis informed the second part of the framework. Biber and his colleagues' taxonomy (Biber et al., 2003, 2004) (i.e. referential bundles, discourse organizer and stance bundles) and Hyland's (2008a) framework (i.e. research-oriented, text-oriented and participant-oriented bundles) have been extensively used in a wide range of bundle studies (e.g. Ädel & Erman, 2012; Biber & Barbieri, 2007; Hyland, 2008a, 2008b; Pan et al., 2016; Wright, 2019). Hyland's (2005a, 2005b) metadiscourse model was adapted in our analysis of sentence initial bundles in Chinese L2 and New Zealand L1 masters and PhD theses (Li, 2016; Li et al., 2017). None of them, however, could be applied easily to our current sentence initial bundle data as they were developed from a much wider range of registers or genres and none cater for the linguistic items appearing at different parts of sentences. It is difficult to align our specific sentence initial bundle data of PhD abstracts with these models. One example of this mismatch is that around 40% of our sentence initial bundles are NP-based bundles with the pattern noun phrase with post-modifier fragment (e.g. *The aim of this study*, *The results of this study*, and *The implications of these findings*) which would all fall into the same category as referential bundles in Biber and his colleagues' taxonomy, research-oriented bundles in Hyland's bundle framework, or endophoric markers

in Hyland’s metadiscourse model. Clearly, a more fine-grained analysis would be more productive.

With this aim, we took a corpus-driven approach to directly identify the discourse functions of the sentence initial bundles retrieved from the PhD abstract corpus. In other words, we started with specifying the functions of individual bundles with reference to their contexts and then grouped the bundles with the same functions to generate a framework. It became clear as the analysis progressed that many of the discourse functions carried by our bundles could be identified as rhetorical moves relating to the IMRD (Introduction-Methods-Results-Discussion) organisation of research articles suggested by Swales (1990). Two authors independently coded the data. The lexical items, grammatical features and the placement in the abstracts, the three indicators suggested by Swales (2004) for move analysis, were all considered as important factors during coding, together with the contexts of the generated bundles. A very high level of agreement was obtained (an inter-rater score of 95%) and the discrepancies in coding were discussed and resolved. Compared to the findings of Cortes (2013), there were only a few bundles indicating more than one move in our study. For example, the bundle *It is argued that the* indicated the moves of findings, method and implications as in (1-3).

1. *It is argued that the sector has the potential to deliver the sustainable development agenda in terms of capacity building, policy advice and knowledge management the extent to which individual organisations deliver these concepts varies.* Findings
2. *It is argued that the social meaning of reconciliation is to be understood within Paul's comprehensive vision of reconciliation: a vision grounded in the story of Christ and Paul's own reconciliation experience, substantiated by the Isaianic vision of cosmic peace, and given form and expression in a rich symbolism of reconciliation.* Method
3. *It is argued that the field may benefit from supplementing the currently favoured locally-based methods with one which seeks to link linguistic behaviour with supra-local systems of inequality.* Implications

Two possible reasons for the strong bundle-move connection in our data are the greater amount of information contained in longer bundles and the clearer move indication of sentence initial bundles. With respect to the multifunctionality of a few bundles, the most common moves were assigned to them. The above bundle *It is argued that the* was categorised as a *findings* bundle because in most cases it was used to present the findings of the research. Table 2 is a summary of the discourse functions of our sentence initial bundles.

Table 2. A taxonomy of sentence initial bundles in PhD abstracts

Bundles	Moves	Examples
Background bundles	1. Providing background	<i>It has been suggested that</i>
Purpose bundles	2. Stating a goal or purpose	<i>The aim of this thesis</i>
Method bundles	3. Introducing methods	<i>The data was analysed using</i>
Findings bundles	4. Showing results or findings	<i>The thesis argues that the</i>
Implications bundles	5. Drawing implications	<i>The implications of these findings</i>
Structure bundles	6. Outlining structures	<i>The thesis is divided into</i>

It is interesting to note that the functions of a large majority of bundles (i.e. 91%), that is, number 1-5 in the table, reflected IMRD organisation and were closely related to the five moves of research article abstracts proposed in Dos Santos (1996) and Hyland (2004); the remaining 9% of the bundles (i.e. number 6 in the table) functioned as text organisers, outlining the structure of the thesis. The bundles with the first five functions were most likely to appear in a predictable part and sequence within abstracts. Structure bundles (e.g. *The thesis concludes with a*, *The second part of the* and *This is followed by an*), on the other hand, were found at various locations across the abstracts. One of the possible reasons for the need to include the structure move in the PhD abstracts could be the long length of PhD theses where writers might feel a need to inform their readers of the organisation of their theses. The second phase of the findings, which involves applying the framework, is presented in the next section.

4. Findings - applying the framework

In this section, we present the distribution of structures and functions in general, examine more closely the functions performed by different sentence initial bundles, and analyse the linguistic features of bundle-based move indicators, that is, the lexical bundles carrying the communicative functions of rhetorical moves.

4.1. Distribution of structures

There were altogether 66 5-word sentence initial bundles, 717 individual cases, generated from the 1.16 million corpus (see Appendix). Table 3 indicates the structural distribution of sentence initial 5-word bundles in the corpus. The two most common patterns were *noun phrase with post-modifier fragment* and *noun phrase + verb phrase fragment*, and both comprised around 40% of all types and instances. The patterns *anticipatory it + verb phrase fragment* and *noun phrase + be* were also popular and constituted nearly 10% of all occurrences. There was only one VP-based bundle (i.e. *In order to explore the*) in the corpus with 6 occurrences.

Table 3. Distribution of structural categories

Patterns	Categories	Types	Instances	Types%	Instances%
noun phrase with post-modifier fragment	NP-based	25	328	38%	46%
noun phrase + verb phrase fragment	Clause-based	24	255	36%	36%
anticipatory it + verb phrase fragment	Clause-based	9	73	14%	10%
noun phrase + <i>be</i>	Clause-based	7	55	11%	8%
(in order) to-clause fragment	VP-based	1	6	2%	1%
Total		66	717	100%	100%

One interesting finding is the absence of PP-based bundles (including *prepositional phrase with embedded of-phrase fragment* and *other prepositional phrase fragment*) in our bundle list, which is a prevalent category in academic prose in many previous studies (e.g. Biber et

al., 1999; Bychkovska & Lee, 2017; Hyland, 2008a; Li, 2016). For example, PP-based bundles were the most frequent form of sentence initial 4-word bundles in our PhD thesis corpora, which accounted for 42% in type and 50% in frequency (Li, 2016). One possible reason is that abstracts function as distillations of other texts and given their brevity, they contain little redundancy (Swales, 1990). They act as “hooks” to encourage readers to further explore the whole texts by highlighting only the most important information (Hyland, 2004). Therefore, details and the bundles used to express details (e.g. *in the case of*, *as a result of*, *at the beginning of*) are less likely to occur in abstracts. Another possible reason is that the 5-word length excluded many 4-word idiomatic prepositional phrases, such as *on the other hand*, *at the same time*, and *in the present study*.

The analysis of the structures of our sentence initial bundles provided us with the notable lexico-grammatical features of sentence starters in Arts and Humanities PhD abstract writing. In order to better understand the role of these bundles, we now examine the rhetorical functions of the moves in the abstracts and investigate the possible connections between them and the linguistic structures.

4.2. Distribution of functions

The analysis of the sentence initial bundles in the PhD abstracts resulted in all the generated bundles in the list being identified as move indicators. These comprised: purpose bundles (e.g. *The aim of this thesis*), implications bundles (e.g. *The implications of these findings*), findings bundles (e.g. *It was found that the*), method bundles (e.g. *Semi-structured interviews were conducted with*), structure bundles (e.g. *The thesis is divided into*), and background bundles (e.g. *A review of the literature*). It should be noted here that rhetorical moves can be realised by different linguistic units and these sentence initial bundles are only a part of a rhetorical move because as Swales (2004) suggests, a move is “a functional, not a formal, unit” (p. 229). The functional distribution is presented in Table 4.

Table 4. Distribution of functional categories

Bundles	Moves	Types	Instances	Types %	Instances %
Purpose bundles	Stating a goal or purpose	30	388	45%	54%
Implications bundles	Drawing implications	11	93	17%	13%
Findings bundles	Showing results or findings	10	93	15%	13%
Method bundles	Introducing method	6	54	9%	8%
Structure bundles	Outlining structures	6	64	9%	9%
Background bundles	Providing background	3	25	5%	3%
Total		66	717	100%	100%

The most frequent bundles were purpose bundles, accounting for about half of the total occurrences. Background bundles were the least used bundles, only accounting for 5% in type and 3% in frequency of all the bundles. One of the likely reasons for the uneven distribution of move-associated bundles is the distribution of moves across abstracts. As found by Hyland (2004), less than 5% of his article abstracts contain all the moves. The only

obligatory move identified across the studies of Dos Santos (1996), Hyland (2004) and (Pho, 2008) was the purpose move.

4.3. Linguistic features of bundle-based move indicators

We explore the linguistic features of bundle-based move indicators in this section. It is almost impossible for traditional move analysis to generalise the linguistic features of moves because of the small number of text samples used in move studies (Kanoksilapatham, 2007). The bundle-driven approach adopted in this study enabled us to investigate the possible connection between moves and the lexico-grammatical patterns (i.e. associations between words and syntactic structures) carried by sentence initial bundles, although we need to bear in mind that the sentence initial bundles only represent a fairly small proportion of possible linguistic means for the realisation of moves.

4.3.1. Purpose bundles

Purpose bundles were the most frequent bundles, with 30 different types and 388 occurrences (see Table 5). The specific goal or purpose of the PhD research was mainly realised in three patterns: *noun phrase with post-modifier fragment of*, *noun phrase + verb phrase fragment* and *noun phrase + be*. With these patterns, a narrow range of shell nouns (e.g. *aim*, *purpose*, *subject*, *focus*, *objective* and *intention*), modality/causation/effort verb phrases (e.g. *is concerned with*, *aimed to explore*, *focuses on the*, *sets out to*) and copular verb phrases (e.g. *is an attempt/investigation/examination*) was used.

Table 5. Purpose bundles

Patterns	Sentence initial bundles	Instances
<i>noun phrase with post-modifier fragment of</i>	The aim of this thesis	49
	The aim of this study	40
	The purpose of this thesis	22
	The aim of this research	20
	The purpose of this study	18
	The main aim of this	16
	The primary aim of this	12
	The purpose of this research	11
	The aim of the thesis	9
	The aim of the study	9
	The overall aim of this	8
	The aim of the current	7
	The subject of this thesis	7
	The aim of the research	7
	The focus of this thesis	7
	The objective of this study	6
The intention of this thesis	6	
<i>noun phrase + verb phrase fragment</i>	This thesis is concerned with	20
	This study aimed to explore	19
	This thesis focuses on the	17
	The present study aimed to	9
	This thesis sets out to	9

	This study aims to investigate	8
	This thesis is based on	7
<i>noun phrase + be</i>	This thesis is a study	12
	This is a study of	8
	This thesis is an attempt	7
	This thesis is an investigation	6
	This thesis is an examination	6
<i>(in order) to-clause fragment</i>	In order to explore the	6

Both present and past tenses were used to indicate purpose but simple present tense was more common. Past tense was only used in two bundles with the subject *study* (i.e. *This study aimed to explore*, *The present study aimed to*). This finding supports Pho's (2008) finding of her 30-abstract small corpus: subjects such as *This article* and *This paper* in her study and *This thesis* in our study were followed by present tense verbs to refer to "the immediate physical object in front of the reader" (Pho, 2008, p. 241). Subjects like *This study* and *The present study* were accompanied by both past and present tense verbs to report the past research experience or to highlight the value of the research. Swales (1990) highlights the popularity of present tense in abstracts, particularly in the commentary parts (e.g. purpose statements), and discusses the different functions of present and past tense: present tense indicates the contemporary nature and generalisability of the research; whereas past tense reflects the researchers' awareness of the one-off nature of their work and expresses some caution to make generalisations and claims that are overstated.

4.3.2. Implications bundles

Implications bundles, as the second most frequent bundle category, slightly exceeded the number of findings bundles. As shown in Table 6 below, the bundles covered two main structural categories: NP-based bundles (*noun phrase with post-modifier fragment*) and clause-based bundles (*noun phrase + be*, *anticipatory it + verb phrase fragment* and *noun phrase + verb phrase fragment*).

Table 6. Implications bundles

Patterns	Sentence initial bundles	Instances
<i>noun phrase with post-modifier fragment</i>	The findings of this study	14
	The findings of the study	12
	The implications of these findings	8
	The findings of this thesis	8
	The results of this study	8
	Implications for clinical practice and	6
<i>noun phrase + be</i>	This thesis is the first	10
<i>anticipatory it + verb phrase fragment</i>	It is hoped that this	8
<i>noun phrase + verb phrase fragment</i>	This thesis proposes a new	7
	This research contributes to the	6
	Further research is needed to	6

In the first category, the statements of implications began with the shell nouns *findings* or *results*, summarising the preceding texts on research findings; the second part of these sentences elaborated the implications of the findings, as in (4) below.

4. ***The findings of this study add to the literature on EAL academic reading by lending empirical support to a componential approach to the teaching and testing of reading.***

The implications function of clause-based bundles was realised in a variety of ways with the indicators embedded in the bundles. For example, a range of evaluative lexical items *first*, *new* and *contribute to* were selected to highlight the significance of the current research. The verb *hope* was used to tentatively predict the impact of the research in the real-world context, as in (5). The signal of *future research* was followed by the recommendations offered to other researchers. The tense used in the implications bundles was present tense, consistent with Pho's (2008) finding.

5. ***It is hoped that this study can help understanding of the complicated decision-making process that parents go through and highlight the importance of empathetic and sensitive treatment by staff.***

4.3.3. Findings bundles

Findings bundles were another frequently appearing category in the corpus because findings is a key move incorporated in most abstracts (Hyland, 2004; Pho, 2008). In Table 7 below, one can see the frequent use of reporting verbs (e.g. *discuss*, *argue*, *conclude*, *show* and *suggest*) to denote different degrees of certainty in the two clause-based patterns: *anticipatory it + verb phrase fragment* and *noun phrase + verb phrase fragment*. Both patterns allow writers to depersonalise their arguments with either dummy *it* or noun phrases (e.g. *the findings*, *the thesis*, or *the study*) used as the subject. Except for the two bundles with the verb *found* (*It was found that the*, *It was also found that*), the tense in all the other bundles was present tense to generalise the results of the PhD studies and to present important findings.

Table 7. Findings bundles

Patterns	Sentence initial bundles	Instances
<i>anticipatory it + verb phrase fragment</i>	It is argued that the	11
	It was found that the	9
	It was also found that	8
	It is concluded that the	7
	It is argued that this	6
	It is suggested that the	6
<i>noun phrase + verb phrase fragment</i>	The findings are discussed in (relation to)	17
	The thesis argues that the	13
	This thesis argues that the	10
	The study shows that the	6

4.3.4. Method bundles

The six method bundles all fell into the category of *noun phrase + verb phrase fragment* bundles (see Table 8 below).

Table 8. Method bundles

Patterns	Sentence initial bundles	Instances
<i>noun phrase + verb phrase fragment</i>	Semi-structured interviews were conducted with	24
	Semi-structured interviews were carried out	6
	Interviews were transcribed and analysed	6
	The research was conducted in	6
	The data was analysed using	6
	Semi-structured interviews were used to	6

The use of verbs in these bundles indicated the physical and mental research activities (e.g. *conduct*, *transcribe* and *analyse*). The dominant use of past tense documented the research activities completed in relation to the question of what was done during the research. The consistent use of passive voice removed the agents (i.e. the PhD researchers) and placed the object of the activities (e.g. *Semi-structured interviews*, *The research* and *The data*) at the beginning of the sentences. The impersonality of these bundles signaled the PhD researcher's attempt to stress the objectivity of their research procedures. The use of past tense and passive voice in our sentence initial bundles aligns with the findings of Dos Santos (1996) and Pho's (2008) research article abstract analysis.

4.3.5. Structure bundles

As presented in Table 9, the pattern *noun phrase + verb phrase fragment* dominated structure bundles with the use of verb phrases *is divided into*, *concludes with* and *is followed by* to indicate the organisation of the PhD theses. Present tense was also used in structure bundles to refer to the physical object: the thesis or part of the thesis.

Table 9. Structure bundles

Patterns	Sentence initial bundles	Instances
<i>noun phrase + verb phrase fragment</i>	The thesis is divided into	17
	The thesis concludes with a	12
	This is followed by an	10
	This is followed by a	8
<i>noun phrase with post-modifier fragment of noun phrase + be</i>	The second part of the	11
	Section A is a review	6

4.3.6. Background bundles

Only three background bundles were identified in the corpus. They are *It has been suggested/argued that* and *A review of the literature* (see Table 10).

Table 10. Background bundles

Patterns	Sentence initial bundles	Instances
<i>anticipatory it + verb phrase fragment</i>	It has been suggested that	11
	It has been argued that	7

Both anticipatory-*it* bundles used present perfect tense (i.e. *has been suggested/argued*) to highlight the arguments proposed in the literature that started in the past and continue to have an influence in the present. Gledhill (2000) considers the role of this construction as the orientation of research aims - "past results framed in terms of a new (present tense) research direction" (p. 122). As can be seen in example (6), the first sentence with the bundle *It has been argued that* concluded the thesis writer's discussion of the previous literature and set the scene for the second sentence, which presented the aim of the current research.

6. *It has been argued that both mainstream and feminist theory and research that considers questions of ageing and old(er) age, contains biases due to being predicated on middle-aged and problem-focused perspectives. The aim of this investigation was to use a material-discursive epistemology to consider how women over seventy years old, and living in an inner London borough, themselves experience and understand their positions as older women, in order to redress these biases.*

The bundle *A review of the literature* synthesised the research in the literature by summarising the shared feature(s) of the past investigations so that a knowledge gap could be identified in the field of study to justify the innovation of the present study.

5. Discussion

In many studies, rhetorical moves have been analysed manually taking a corpus-based approach on a small number of texts, while in others, linguistic items such as lexical bundles have been extracted automatically across a large scale of texts taking a corpus-driven approach. Few studies have combined these two approaches to discourse analysis and "[a]s a result, we know little at present about the general patterns of discourse organization across a large representative sample of texts from a genre" (Biber, Connor, et al., 2007, p. 11). In this study, we took a bundle-driven approach to identify the rhetorical moves associated with the 5-word sentence initial bundles generated from our Arts and Humanities PhD abstract corpus consisting of 3,000+ texts.

Two main outcomes can be highlighted. First, the bundle-driven approach has proven to be a promising approach to investigate the connection between bundles and moves. A strong bundle-move connection was able to be established because of the selection of a move-intensive genre (i.e. PhD abstracts), the generation of longer bundles (i.e. 5-word bundles), and the focus on sentence initial ones.

The choice of a move-intensive genre is a significant decision to make. Biber, Connor, et al. (2007) argue that "more specialized corpora are more appropriate for the study of discourse structure" (p. 18). In other words, the more specific the texts in a corpus are, the more likely the corpus analysis is to inform its discourse structure. For example, the sentence initial bundles generated from the PhD thesis corpora in our previous study (Li, 2016) are considerably different from the bundles identified in the current study, although both studies focus on a single PhD genre in one discipline. Most bundles in the PhD thesis corpora were

metadiscourse indicators (e.g. *On the other hand*, *It is possible that* and *In the case of*), acting as discourse frames to relate to new information or as interactional devices to mark the involvement of the writer and the reader. They aligned in the most part with Hyland's (2005a, 2005b) metadiscourse models, whereas the bundles in this study were better related to the moves of abstracts. We were more likely to generate move-related bundles from PhD abstracts than from PhD theses, because of the shorter length of the individual texts and the more intensive and consistent moves of abstracts. Similarly, Cortes's (2013) bundle-move connection study is based on her corpus of research article introductions, not whole research articles.

Besides the selection of texts, as discussed in the introduction, the length and position of lexical bundles are also important factors affecting the results of the analysis. As Cortes (2013) comments, shorter bundles are likely to occur in several moves or steps. The 5-word bundles in our study were more likely to be associated with a single move as longer bundles contain more specific information and 5-word strings largely exclude many 4-word set phrases (e.g. *on the other hand*), many of which are general transition markers. Sentence initial position is another feature of our bundles and sentence initial bundles are more likely to indicate the start of moves.

The second major outcome of this study is the generation of linguistic features of these bundle-based move indicators. The combination of lexico-grammatical patterns, verb tenses and aspects of voices identified using the bundle-driven approach in this study signals the stages of the PhD abstracts. The move *purpose* is mainly realised in the patterns noun phrase with post-modifier fragment of (e.g. *The aim of this thesis*), noun phrase + verb phrase fragment (e.g. *This thesis is concerned with*) and noun phrase + be (e.g. *This thesis is a study*), with the selection from a small range of nouns (e.g. *aim, purpose, thesis, study* and *research*). Mention of research method is frequently associated with noun phrase + verb phrase fragment bundles (e.g. *Interviews were transcribed and analysed*) to emphasise physical and mental research activities. *Findings* are presented using the two clause-based patterns anticipatory it + verb phrase fragment (e.g. *It is argued that the*) and noun phrase + verb phrase fragment (e.g. *The thesis argues that the*). Both patterns allow writers to distance themselves from their arguments and present the findings in a more neutral and impersonal way. In addition, the grammatical patterns of some moves are fairly similar. The move *implications* echoes *purpose* with most bundles having the same patterns: noun phrase with post-modifier fragment and noun phrase + verb phrase fragment to discuss the scope or contribution of the PhD research, though the shell nouns perform different linking functions, linking back in *implications* and forward in *purpose*. The pattern of the move *structure* is similar to *method*: noun phrase + verb phrase fragment. Both moves highlight activities, whether they are activities performed on the thesis (e.g. *The thesis is divided into*) or physical activities in the real world (e.g. *Semi-structured interviews were conducted with*). The similarity between the different functions also explains why, compared to Biber et al. (1999), we have identified a far smaller range of grammatical patterns in this study.

Besides lexico-grammatical patterns, verb tenses and aspects of voice were also found as signals of rhetorical moves in our PhD abstracts. The tenses used in the bundles were found generally to be consistent for each move: present tense for *purpose, findings, implications* and *structure* to assume generalisability or to refer to the immediate physical object, *the thesis*, past tense for *method* to report completed research activities, and present perfect tense

for *background* to indicate the influence of previous research. Passive voice is a key feature of *method* moves; active voice, on the other hand, is frequently found across the other moves. However, a few anticipatory-*it* bundles with passive voice (e.g. *It has been suggested that, It is argued that the*) appear in the *background*, *findings* and *implications* moves.

The identified rhetorical moves together with their bundle-based indicators not only contribute to our knowledge of the PhD abstract genre but also have important pedagogical implications for EAP writing. As Flowerdew (2013) suggests, “Knowing how to perform a genre . . . , involves knowing both its schematic structure, or staging, on the one hand, and the specific form-function correlations of each stage, on the other” (p.148). English for Academic Purposes instructors and thesis supervisors can refer to the six identified moves and their bundle-based indicators as an instructional framework to raise the rhetorical awareness of PhD students and to scaffold their abstract writing with a range of ready-made move-related bundles functioning as “islands of reliability” (Dechert, 1984). For instance, the *purpose* move can be introduced together with two frames:

- 1) *The aim/purpose/focus/objective/intention of this thesis/study/research or*
- 2) *This thesis is a(n) study/attempt/investigation/examination.*

Given the changing tense use across moves, this is another aspect that students could be supported to notice and apply in their own work. Patterns at the micro-level of the PhD abstract can be generalised to the much larger thesis text to some extent.

One caveat in using the framework of moves and their linguistic realisation is that the bundles represent a very small proportion of the text data in the corpus, and only 5-word combinations occurring 5+ times at the beginning of sentences have been considered. Thus, other ways of signalling moves may have gone unnoticed, for example, individual words, and shorter or longer lexical chunks used with lower frequencies or in the non-initial position of sentences.

In sum, while the bundle-driven approach has generated a useful and usable framework that can potentially support thesis writers, there are other ways to extend and supplement it. The bundle-driven approach adopted in our study is only “one way of a number of approaches to investigating conventionalised uses of language” (Rezoug & Vincent, 2018, p. 49).

6. Conclusion

This study combined the two perspectives of discourse analysis, on linguistic items and on discourse structures, taking a bundle-driven approach to explore rhetorical moves of PhD abstracts. The combination allowed us to automatically analyse discourse structures and to access a large volume of texts. Our study suggests that exploring discourse structures by means of a bundle-driven approach can be productive both in terms of providing insights into methods of analysis and a pedagogical framework or tool. With regard to future studies, the bundle-driven move analysis approach needs to be further developed. We did not address variations across disciplines in this study. As a number of researchers have shown (Cortes, 2004; Durrant, 2015; Hyland, 2008b), many lexical bundles are to an extent discipline-specific. Thus, it would be fruitful to explore this further with respect to the specific bundles in PhD abstracts. Disciplinary patterns and preferences for the realisation of moves would

have important pedagogical implications. The exploratory bundle-driven approach could possibly complement traditional corpus-based move analysis to some extent and yield more fruitful findings in exploring the use of prefabricated language. For example, move analysis could start with a large number of texts using the bundle-driven approach to obtain a general picture of the potential moves in the investigated genre and then a small sample of texts could be carefully analysed to triangulate the finding and to identify other move-related linguistic features. Future exploration of lexical bundles is likely to deepen our understanding of conventionalisation of phraseology used to realise rhetorical moves in different genres, and thus provide frameworks or tools that can support writers.

Appendix Supplementary data

Data related to this study can be found at

https://osf.io/2fm8k/?view_only=f7e407f825074043a90990754466a0ce

Acknowledgements

We would like to thank Professor Nigel Harwood and the anonymous reviewers for their valuable time and invaluable comments. We would also like to thank Professor Ian H. Witten from the School of Computing and Mathematical Sciences for his generous support, Dr Alannah Fitzgerald for her help with corpus building, and to our Research Programmer, Anupama Krishnan.

References

- Ädel, A. (2014). Selecting quantitative data for qualitative analysis: A case study connecting a lexicogrammatical pattern to rhetorical moves. *Journal of English for Academic Purposes*, 16, 68-80.
doi:<https://doi.org/10.1016/j.jeap.2014.09.001>
- 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.
doi:10.1016/j.esp.2011.08.004
- Basturkmen, H. (2012). A genre-based investigation of discussion sections of research articles in Dentistry and disciplinary variation. *Journal of English for Academic Purposes*, 11(2), 134-144.
doi:<https://doi.org/10.1016/j.jeap.2011.10.004>
- Biber, D. (2015). Corpus-based and corpus-driven analysis of language variation and use. In B. Heine & H. Narrog (Eds.), *The Oxford handbook of linguistic analysis* (pp. 193-224). Oxford, England: Oxford University Press.

- Biber, D., & Barbieri, F. (2007). Lexical bundles in university spoken and written registers. *English for Specific Purposes*, 26(3), 263-286. doi:10.1016/j.esp.2006.08.003
- Biber, D., Connor, U., & Upton, T. A. (2007). *Discourse on the move: Using corpus analysis to describe discourse structure*. Amsterdam, the Netherlands: John Benjamins Publishing Company.
- Biber, D., Conrad, S., & Cortes, V. (2003). Lexical bundles in speech and writing: An initial taxonomy. In G. N. Leech, T. McEnery, A. Wilson, & P. Rayson (Eds.), *Corpus linguistics by the lune*. New York, NY: Peter Lang.
- Biber, D., Conrad, S., & Cortes, V. (2004). If you look at ...: Lexical bundles in university teaching and textbooks. *Applied Linguistics*, 25(3), 371-405. doi:10.1093/applin/25.3.371
- Biber, D., Csomay, E., Jones, J. K., & Keck, C. (2007). Introduction to the identification and analysis of vocabulary-based discourse units. In D. Biber, U. Connor, & T. A. Upton (Eds.), *Discourse on the move: Using corpus analysis to describe discourse structure* (pp. 155-173). Amsterdam, the Netherlands: John Benjamins Publishing Company.
- Biber, D., Johansson, S., Leech, G., Conrad, S., & Finegan, E. (1999). *Longman grammar of spoken and written English*. London, England: Longman.
- Bychkovska, T., & Lee, J. J. (2017). *At the same time*: Lexical bundles in L1 and L2 university student argumentative writing. *Journal of English for Academic Purposes*, 30(Supplement C), 38-52. doi:10.1016/j.jeap.2017.10.008
- Chang, C.-F., & Kuo, C.-H. (2011). A corpus-based approach to online materials development for writing research articles. *English for Specific Purposes*, 30(3), 222-234. doi:<https://doi.org/10.1016/j.esp.2011.04.001>
- Conrad, S. (2002). Corpus linguistic approaches for discourse analysis. *Annual Review of Applied Linguistics*, 22, 75-95. doi:10.1017/S0267190502000041
- Cortes, V. (2004). Lexical bundles in published and student disciplinary writing: Examples from history and biology. *English for Specific Purposes*, 23(4), 397-423. doi:10.1016/j.esp.2003.12.001
- 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. doi:10.1016/j.jeap.2012.11.002
- Cortes, V. (2015). Situating lexical bundles in the formulaic language spectrum. In V. Cortes & E. Csomay (Eds.), *Corpus-based research in applied linguistics: Studies in honor of Doug Biber* (Vol. 66, pp. 197-216). Amsterdam, the Netherlands: John Benjamins Publishing Company.
- Cotos, E., Huffman, S., & Link, S. (2015). Furthering and applying move/step constructs: Technology-driven marshalling of Swalesian genre theory for

- EAP pedagogy. *Journal of English for Academic Purposes*, 19, 52-72.
doi:<https://doi.org/10.1016/j.jeap.2015.05.004>
- Cotos, E., Huffman, S., & Link, S. (2017). A move/step model for methods sections: Demonstrating rigour and credibility. *English for Specific Purposes*, 46, 90-106. doi:<https://doi.org/10.1016/j.esp.2017.01.001>
- Dechert, H. W. (1984). Second language production: Six hypotheses. In H. W. Dechert, D. Mohle, & M. Raupach (Eds.), *Second language productions* (pp. 211-230). Tübingen, Germany: Gunter Narr Verlag.
- Dos Santos, M. B. (1996). The textual organization of research paper abstracts in applied linguistics. *TEXT*, 16(4), 481-499.
doi:10.1515/text.1.1996.16.4.481
- Durrant, P. (2015). Lexical bundles and disciplinary variation in university students' writing: Mapping the territories. *Applied Linguistics*, 1-30.
doi:10.1093/applin/amv011
- Flowerdew, J. (2002). Genre in the classroom: A linguistic approach. In A. M. Johns (Ed.), *Genre in the classroom: Multiple perspectives* (pp. 91-104). Mahwah, NJ: Lawrence Erlbaum Associates.
- Flowerdew, J. (2013). *Discourse in English language education*. London, England: Routledge.
- Gledhill, C. (2000). The discourse function of collocation in research article introduction. *English for Specific Purposes*, 19(2), 115-135.
doi:10.1016/S0889-4906(98)00015-5
- Gómez González, M. Á., Ruiz de Mendoza Ibáñez, F. J., González García, F., & Downing, A. (2014). *The functional perspective on language and discourse: Applications and implications*: Amsterdam, the Netherlands: John Benjamins Publishing Company.
- Granger, S. (2017). Academic phraseology: A key ingredient in successful L2 academic literacy. In R. V. Fjeld, K. Hagen, B. Henriksen, S. Johansson, S. Olsen, & J. Prentice (Eds.), *Academic language in a Nordic setting: Linguistic and educational perspectives* (Vol. 9, pp. 9-27). Oslo, Norway: University of Oslo.
- Granger, S. (2018). Tracking the third code: A cross-linguistic corpus-driven approach to metadiscursive markers. In A. Čermáková & M. Mahlberg (Eds.), *The Corpus Linguistics Discourse: In honour of Wolfgang Teubert* (pp. 185-204). Amsterdam, the Netherlands: Benjamins.
- Gray, B., Cotos, E., & Smith, J. (2020). Combining rhetorical move analysis with multi-dimensional analysis: Research writing across disciplines. In U. Römer, V. Cortes, & E. Friginal (Eds.), *Advances in corpus-based research on academic writing: Effects of discipline, register, and writer expertise* (pp. 137-168). Amsterdam, the Netherlands: John Benjamins Publishing Company.
- Halliday, M. A. K., & Matthiessen, C. M. I. M. (2004). *An introduction to functional grammar* (3rd ed.). London, England: Arnold.

- Hinkel, E. (2004). *Teaching academic ESL writing: Practical techniques in vocabulary and grammar*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Hu, G., & Liu, Y. (2018). Three minute thesis presentations as an academic genre: A cross-disciplinary study of genre moves. *Journal of English for Academic Purposes*, 35, 16-30.
doi:<https://doi.org/10.1016/j.jeap.2018.06.004>
- Hyland, K. (2004). *Disciplinary discourses: Social interactions in academic writing*. Ann Arbor, MI: University of Michigan Press.
- Hyland, K. (2005a). *Metadiscourse: Exploring interaction in writing*. London, England: Continuum.
- Hyland, K. (2005b). Stance and engagement: A model of interaction in academic discourse. *Discourse Studies*, 7(2), 173-192.
- Hyland, K. (2008a). Academic clusters: Text patterning in published and postgraduate writing. *International Journal of Applied Linguistics*, 18(1), 41-62. doi:10.1111/j.1473-4192.2008.00178.x
- Hyland, K. (2008b). As can be seen: Lexical bundles and disciplinary variation. *English for Specific Purposes*, 27(1), 4-21. doi:10.1016/j.esp.2007.06.001
- Jiang, F., & Hyland, K. (2017). Metadiscursive nouns: Interaction and cohesion in abstract moves. *English for Specific Purposes*, 46, 1-14.
doi:10.1016/j.esp.2016.11.001
- Kanoksilapatham, B. (2005). Rhetorical structure of biochemistry research articles. *English for Specific Purposes*, 24(3), 269-292.
doi:<https://doi.org/10.1016/j.esp.2004.08.003>
- Kanoksilapatham, B. (2007). Rhetorical moves in biochemistry research articles. In D. Biber, U. Connor, & T. A. Upton (Eds.), *Discourse on the move: Using corpus analysis to describe discourse structure* (pp. 73-119). Amsterdam, the Netherlands: John Benjamins Publishing Company.
- Kanoksilapatham, B. (2015). Distinguishing textual features characterizing structural variation in research articles across three engineering sub-discipline corpora. *English for Specific Purposes*, 37, 74-86.
doi:<https://doi.org/10.1016/j.esp.2014.06.008>
- Kwan, B. S. C. (2006). The schematic structure of literature reviews in doctoral theses of applied linguistics. *English for Specific Purposes*, 25(1), 30-55.
doi:<https://doi.org/10.1016/j.esp.2005.06.001>
- Li, L. (2016). *Sentence initial bundles in L2 thesis writing: A comparative study of Chinese L2 and New Zealand L1 postgraduates' writing*. (Doctor of Philosophy (PhD) Doctoral), University of Waikato, Hamilton, New Zealand. Retrieved from <http://hdl.handle.net/10289/10862>
- Li, L., Franken, M., & Wu, S. (2017). Bundle -driven metadiscourse analysis: Sentence initial bundles in Chinese and New Zealand postgraduates' thesis writing. In C. Hatipoglu, E. Akbas, & Y. Bayyurt (Eds.), *Metadiscourse in written genres: Uncovering textual and interactional aspects of texts* (pp. 251-283). Frankfurt am Main, Germany: Peter Lang.

- Mizumoto, A., Hamatani, S., & Imao, Y. (2017). Applying the bundle–move connection approach to the development of an online writing support tool for research articles. *Language Learning*, 67(4), 885-921. doi:10.1111/lang.12250
- Moreno, A. I., & Swales, J. M. (2018). Strengthening move analysis methodology towards bridging the function-form gap. *English for Specific Purposes*, 50, 40-63. doi:<https://doi.org/10.1016/j.esp.2017.11.006>
- Omidian, T., Shahriari, H., & Siyanova-Chanturia, A. (2018). A cross-disciplinary investigation of multi-word expressions in the moves of research article abstracts. *Journal of English for Academic Purposes*, 36, 1-14. doi:<https://doi.org/10.1016/j.jeap.2018.08.002>
- 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. doi:10.1016/j.jeap.2015.11.003
- Parkinson, J. (2017). The student laboratory report genre: A genre analysis. *English for Specific Purposes*, 45, 1-13. doi:<https://doi.org/10.1016/j.esp.2016.08.001>
- Parodi, G. (2010). The rhetorical organization of the textbook genre across disciplines: A 'colony-in-loops'? *Discourse Studies*, 12(2), 195-222. doi:10.1177/1461445609356500
- Parodi, G. (2013). Genre organization in specialized discourse: Disciplinary variation across university textbooks. *Discourse Studies*, 16(1), 65-87. doi:10.1177/1461445613496355
- Pho, P. D. (2008). Research article abstracts in applied linguistics and educational technology: A study of linguistic realizations of rhetorical structure and authorial stance. *Discourse Studies*, 10(2), 231-250. doi:10.1177/1461445607087010
- Rezoug, F., & Vincent, B. (2018). Exploring lexical bundles in the Algerian corpus of engineering. *Arab Journal of Applied Linguistics*, 4(4), 47-77.
- Stoller, F. L., & Robinson, M. S. (2013). Chemistry journal articles: An interdisciplinary approach to move analysis with pedagogical aims. *English for Specific Purposes*, 32(1), 45-57. doi:<https://doi.org/10.1016/j.esp.2012.09.001>
- Swales, J. M. (1981). *Aspects of article introductions*. Birmingham, England: The University of Aston.
- Swales, J. M. (1990). *Genre analysis: English in academic and research settings*. Cambridge, England: Cambridge University Press.
- Swales, J. M. (2004). *Research genres: Exploration and applications*. Cambridge, England: Cambridge University Press.

- Tannen, D., Hamilton, H. E., Schiffrin, D., & Adger, C. T. (2015). *The handbook of discourse analysis* (2nd ed.): Chichester, England: Wiley Blackwell.
- Tessuto, G. (2015). Generic structure and rhetorical moves in English-language empirical law research articles: Sites of interdisciplinary and interdiscursive cross-over. *English for Specific Purposes*, 37, 13-26. doi:<https://doi.org/10.1016/j.esp.2014.06.002>
- Tognini-Bonelli, E. (2001). *Corpus linguistics at work*. Amsterdam, the Netherlands: John Benjamins Publishing Company.
- Williams, J. M. (2003). *Style: Ten lessons in clarity and grace* (7th ed.). New York, NY: Addison-Wesley.
- Witten, I. H., Wu, S., Li, L., & Whisler, J. L. (2013). *The book of FLAX: A new approach to computer-assisted language learning* (2nd ed.). Hamilton, New Zealand: University of Waikato.
- Wright, H. R. (2019). Lexical bundles in stand-alone literature reviews: Sections, frequencies, and functions. *English for Specific Purposes*, 54, 1-14. doi:10.1016/j.esp.2018.09.001
- Yang, R., & Allison, D. (2003). Research articles in applied linguistics: Moving from results to conclusions. *English for Specific Purposes*, 22(4), 365-385. doi:[https://doi.org/10.1016/S0889-4906\(02\)00026-1](https://doi.org/10.1016/S0889-4906(02)00026-1)
- Yoon, J., & Casal, J. E. (2020). P-frames and rhetorical moves in applied linguistics conference abstracts. In U. Römer, V. Cortes, & E. Friginal (Eds.), *Advances in corpus-based research on academic writing: Effects of discipline, register, and writer expertise* (pp. 281-305). Amsterdam, the Netherlands: John Benjamins Publishing Company.