# A cross-disciplinary investigation of multi-word expressions in the moves of research article abstracts

Taha Omidian [a, *], Hesamoddin Shahriari [b], Anna Siyanova-Chanturia [a]

[a] School of Linguistics and Applied Language Studies, Victoria University of Wellington, New Zealand
[b] School of English Language and Literature, Ferdowsi University of Mashhad, Iran

A B S T R A C T

Conformity to the epistemological orientations of academic disciplines is often reflected in the ways in which knowledge is constructed and communicated through certain linguistic features in academic genres. This paper explores instances of such conformity in hard and soft science disciplines, as it is reflected in the use of recurrent word combinations in different rhetorical moves of research article abstracts. A corpus-driven approach was adopted to identify the most frequently occurring word combinations of different length in a corpus of 5910 abstracts from six disciplines. Following this, the identified sequences were classified according to their communicative functions in different moves of the abstract. A mixed-methods approach was then adopted through which the patterns of variation reflected in the use of the identified sequences were examined both quantitatively and qualitatively. It was found that members of different academic domains have different priorities for representing their research in academic abstracts.

© 2018 Elsevier Ltd. All rights reserved.

## 1. Introduction

Disciplinary variation emerges from conformity to the conventions and particularities that characterize the epistemo-logical orientations of an academic community (Hyland & Bondi, 2006). Such conformity to norms can manifest itself in academic texts, such as research articles (RAs), where authors adopt certain rhetorical strategies and linguistic resources to not only manage arguments in their narrative, but also express collegiality within their respective disciplines. Research has shown that these rhetorical strategies and their linguistic realizations can vary across different parts of RAs (Parkinson, 2013; Samraj, 2005; Stoller & Robinson, 2013). One section of the RA that has been an active area of research since Graetz (1985) and Swales' 1990) seminal work on genre analysis is the abstract. Evidence from several studies (e.g., Hyland & Tse, 2005; Hyland, 2004; Lorés, 2004; Pho, 2008) suggests that the abstract can clearly delineate the boundaries between disciplines and show how authors adopt different discursive strategies to promote their research within their respective academic/scientific domain. The findings from these studies reveal that, even in a compact genre such as the abstract,[1] a close investigation of the employed rhetorical strategies and their associated linguistic features could highlight the fundamental distinctions that exist between disciplines. However, despite the strong research base that the study of disciplinary variation in RAs has helped

---

* Corresponding author. Kelburn Parade, Wellington, 6012, New Zealand.
 *E-mail address:* taha.omidian@vuw.ac.nz (T. Omidian).

[1] There is a general consensus among researchers on the status of abstracts as an academic genre (e.g., Bondi & Lorés Sanz, 2014; Hyland, 2004; Swales & Feak, 2009; Ventola, 1994.).

develop, most of the literature in this area has primarily focused on the rhetorical moves of different parts of the RA, and less attention has been paid to the linguistic features frequently associated with these moves. To bridge this gap, the present study aimed to investigate disciplinary variation in terms of linguistic patterns that are frequently used to fulfill different rhetorical moves of RA abstracts. For this purpose, a bottom-up approach was adopted in which lexical bundles, as "the most frequently recurring sequences of words in a register" (Biber, 2009, p. 282), were systematically identified and subsequently analyzed in relation to the rhetorical moves and the disciplines in which they were observed.

### 1.1. Disciplinary variation

Disciplines are said to provide academic institutions with a contextualized framework that guides their research activities (Clark, 1983). In theory, the intellectual boundaries drawn by disciplinary frameworks allow members to pursue various academic goals within the same clearly-defined context. However, in practice, these boundaries are not always straightforward or easy to follow. This is partly due to the epistemological overlaps that exist between certain academic areas. As Becher (1994) argues, disciplines are distinguished by the shared features of their associated domains of enquiry and, in very many instances (e.g., Biology and Chemistry or Linguistics and Psycholinguistics), these domains tend to overlap with those of neighboring fields. Many studies have attempted to account for such similarities by proposing different classification schemes for academic areas (Berliner, 2002; Biglan, 1973a; Donald, 1990; Durrant, 2017; Hyland, 2004; Storer & Parsons, 1968). In one of the earliest of such studies, Biglan (1973a) analyzed the judgments of faculty members on the shared characteristics of 36 academic subject areas and found three 'dimensions' to be involved in shaping academics' perceptions of the disciplines under investigation. These dimensions included (a) the existence of an established body of theory, or lack thereof, in certain disciplines (b) the degree to which the area in question has applications for practical and real-world issues, and (c) the extent to which it deals with inanimate objects. According to these three dimensions, Biglan (1973a) grouped the disciplines in his study under three main academic clusters: hard-soft, applied-pure, and life-nonlife systems, respectively. Biglan's classification is important in that it indicates how the characteristics of disciplines and their perceived overlap with other academic areas may affect the scope of their members' scholarly activities (for a detailed discussion, see Biglan, 1973b).

The variations and similarities that exist between disciplines can also affect the ways in which knowledge is communicated and presented to its target audience (Becher & Trowler, 2001). Knowledge in disciplinary communities is negotiated through constructing a form of discourse that not only reflects the collective norms of the community but also adheres to the expectations of its members. Many studies to date have explored disciplinary variation in a number of academic genres by investigating the discourse strategies adopted by writers to produce meaning. Some prominent studies on this topic have included: research articles (Hyland, 2004; Swales, 2004); books (Hyland, 2004; Moore, 2002); manuscript reviews (Samraj, 2016); statements of purpose (Brown, 2004; Samraj & Monk, 2008); PhD and master's theses (Samraj, 2008; Thompson, 2000); lab reports (Braine, 1995), acknowledgments (Hyland, 2004), among others. Perhaps, the most common aspect of these distinct, yet inherently similar, studies is their focus on disciplinary variation as reflected in discursive and rhetorical strategies adopted by writers from across different disciplines. More recently, Durrant (2017) studied disciplinary variation by focusing on a different aspect of academic texts. In a corpus-based study of university student writings, Durrant adopted a bottom-up approach based on which disciplinary groupings emerged from the analysis of word combinations that are frequently used in academic essays. Durrant's initial analyses revealed four main disciplinary groupings: Science/Technology (hard), Humanities/Social Sciences (soft), Life Sciences, and Commerce groupings. The most prominent differences appeared between clusters belonging to the 'hard' versus 'soft' distinction, which also closely corresponded to the hard-soft categorization made by Biglan (1973a). Durrant further conducted a qualitative analysis of the sequences that marked the distinction between clusters belonging to these two particular categories. The results of his analyses revealed major differences between hard/soft knowledge fields in terms of the way in which the identified sequences were used by writers to fulfill different communicative functions. Taken together, Durrant's (2017) findings indicate that the analysis of the linguistics resources commonly used in discourse making can provide a reliable characterization of disciplines and their possible variations.

### 1.2. Genre analysis

According to Swales (1981, 1984, 1990), genre is a conventionally recognized template through which discourse is constructed with the aim of fulfilling the communicative purposes of social interactions. In academic contexts, genres enable writers to communicate their message in a way that is retrievable by the reader. The communicative purposes of such 'writer-reader' interactions are realized through rhetorical choices that instantiate the expectations of the target readership in a particular academic area. These rhetorical choices mainly involve decisions regarding the overall organization of the discourse, as well as the linguistic resources employed to reflect their communicative purposes (Bhatia, 1993). In his seminal work on genre analysis, Swales (1990, 2004) demonstrated that the rhetorical choices made by authors when organizing their discourse follow certain conventional patterns, referred to as 'moves'. As Swales (2004) puts it, a move is "a discoursal or rhetorical unit that performs a coherent communicative function" (p. 29). Moves often comprise a number of obligatory and/or optional 'steps', which contribute to the textual organization of the move to which they belong. Adopting this approach, a multitude of genre-based studies have since investigated the way in which discourse is formed in different sections of the RA, as one of the most common ways of disseminating knowledge in academia. The main output from this line of research has been the development of move schemes proposed for the main sections of the RA, including the abstract (e.g., Hyland, 2004;

Pho, 2008), introduction (e.g., Samraj, 2002; Swales, 1990), methods (e.g., Bruce, 2008; Lim, 2006), results and discussion (e.g., Holmes, 1997; Yang & Allison, 2003). These move classifications are important in that they indicate how the textualization of discourse is carried out within the same genre.

However, as Dudley-Evans (1994) argues, decisions regarding the classification of moves in different sections of the RA, or other academic genres, are mainly made on the basis of their associated linguistic features. As Haswell (1991) points out, different genres have different preferences for the phraseology used to fulfill their communicative purposes. That is, the reason why specific genres, such as 'law reports', can be clearly distinguished from other narrative genres, such as 'fiction', is due to the linguistic features that are distinctive of their style of communication. These linguistic resources can take the form of both single and multi-word items. Lim (2006) investigated the way in which individual moves and their constituent steps are realized through lexical and syntactic choices made by authors in the Method sections of RAs. Lim found that authors use certain linguistic clues, both in the form of single and multi-word units, to indicate internal boundaries (moves) in their texts. His findings also revealed that the majority of these single word items were co-occurring with certain words, helping to convey the overall communicative aims of a specific move and/or step. Lim's (2006) findings point to a highly 'formulaic' nature of the language in academic writing.

### 1.3. Multi-word expressions and rhetorical moves

It has long been acknowledged that certain words tend to co-occur in specific configurations (Biber, Johansson, Leech, Conrad, & Finegan, 1999; Schmitt, 2004; Wray, 2002). These configurations are often referred to as *multi-word expressions* (MWEs), a terminology that encompasses a wide range of linguistic units above the single word level (as opposed to *formulaic language* which may also include single word items, Van Lancker-Sidtis & Rallon, 2004; Wray, 2002; also see Siyanova-Chanturia & Omidian, in press for a discussion). MWEs can take the form of collocations (*strong tea*), binomials (*bride and groom*), idioms (*spill the beans*), lexical bundles (*on the other hand*), and others. According to some estimates, these sequences comprise somewhere between 20 and 50 per cent of natural language production (Biber et al., 1999; Erman & Warren, 2000; Foster, 2001; Howarth, 1998; Sorhus, 1977).

Traditionally, the study of MWEs was mainly based on intuition, with researchers relying on rather impressionistic evidence for their identification (e.g., Cowie, 1994; Foster, 2001; Howarth, 1998). The focus of this traditional (also referred to as 'phraseological') approach was mainly on distinguishing compositional phrases from non-compositional ones, as well as making intuitive judgments regarding their semantic properties (e.g., literal vs figurative). Following the developments introduced by corpus-based methodologies, however, a more systematic approach emerged in which frequency and distributional properties of sequences are deemed to take precedence over intuition and perceptual judgments (e.g., Sinclair, 1991). By utilizing large text corpora and various statistical metrics, such as frequency and measures of association strengths (mutual information, t-score), proponents of this approach have developed several methodologies to study the use of MWEs in natural language. Two general approaches that have emerged from these efforts are: 'corpus-based' and 'corpus-driven' (Tognini-Bonelli, 2001). While the former approach relies on corpus evidence as a tool to validate theoretical presumptions regarding the semantic and syntactic properties of MWEs, the latter employs a more bottom-up approach in which sequences emerge from the corpus analysis with little theoretical assumptions guiding their identification (Biber, 2009). In this approach, it is the emerged linguistic patterns that form the foundation of a linguistic theory, and not vice versa.

Recently, by adopting a bottom-up (corpus-driven) approach, Cortes (2013) examined the use of lexical bundles in the moves and steps of RA introductions. She found a strong connection between the communicative functions of certain bundles and the rhetorical aims of the moves and steps in which they were found. Cortes (2013) also observed that certain bundles were used by authors to signal the onset of a move and/or step. Similarly, Le and Harrington (2015) investigated the use of MWEs in the Discussion section of quantitative RAs from the field of applied linguistics. They too found that the use of specific word combinations strongly manifested both the genre and the communicative aims of the moves in which they were used. More recently, Mizumoto, Hamatani, and Imao (2017) explored the pedagogical implications of the correspondence between lexical bundles and moves by developing a web-based tool that allows the user to search for the most frequently used bundles in different moves of applied linguistics RAs. The authors piloted the tool with eight L2 speakers of English for two to five months. Their results showed positive user feedback on the pedagogical implications of the tool and the bundle-move approach for academic writing.

These studies are important because they provide evidence for the strong connection that exists between MWEs and rhetorical moves, and argue against traditional approaches towards move analysis in which the identification of moves and their linguistic characteristics were solely based on qualitative evidence drawn from subjective observations (see Biber, Connor, & Upton, 2007, p. 36). What is largely missing in the described studies, however, is a systematic analysis of the link between rhetorical moves and their linguistic realizations across different disciplines. Specifically, more empirical evidence is needed to develop a better understanding of disciplinary writing through the lens of the strong connection that exists between function and form in academic texts (see Moreno & Swales, 2018).

## 2. Present study

The present study attempts to investigate disciplinary variation in terms of MWEs that are commonly used to realize the communicative aims of different rhetorical moves in RA abstracts. The rationale for focusing on the abstract section in the

present study is twofold. First, as Dos Santos (1996, p. 483) argues, the RA abstract is "an important site for the visibility of scientific endeavor" since it is generally the first part of the research article that readers encounter. Hence, it is very important for academic writers in various disciplines to be able to promote their study in a way that is readily recognizable by the reader. This requires the author to deploy certain rhetorical and linguistic devices that reflect the norms and expectations of the target readership in their discipline. Moreover, since authors often consider the abstract section as a brief summary of their study and its findings, the overall textual organization of the accompanying article (i.e., IMRD) is often reflected in this section. This makes the abstract an ideal site for investigating how writers in different disciplines commonly promote various parts of their study, and also which aspects of their research are often viewed as the most important promotional tool for their paper. Therefore, we believe that, due to the substantial role of abstracts in promoting and summarizing the accompanying article, the study of disciplinary variations in this particular section can reveal interesting insights into how writers in different academic areas reflect the expectations of their readers. To this end, quantitative and qualitative analyses were carried out to identify patterns of disciplinary variation in the use of recurrent word combinations in different moves of the abstract. This mixed-method approach allowed us to address the following questions:

1. How do writers from hard and soft science fields differ in terms of the distribution of the sequences they frequently use in different moves of RA abstracts?
2. How do writers from hard and soft knowledge fields differ in their use of the sequences employed to fulfill the rhetorical functions of different moves in RA abstracts?

## 3. Material and methods

### 3.1. Corpus

The study is based on a one-million-word[2] corpus consisting of 5910 RA abstracts accompanying journal articles published between 2005 and 2015. The abstracts were extracted from high-ranking, accredited journals nominated by expert informants from six different disciplines (i.e., mechanical engineering, physics, biology, sociology, marketing, and applied linguistics). The extraction of abstracts was based on a stratified random sampling through which three abstracts were randomly sampled from each issue of a journal in each year. In addition, attention was paid to extract a roughly equal number of abstracts from different journals in each discipline so as to avoid possible journal influences on writing style. Further, for comparison purposes, care was taken to reach a closely matched word count for the six selected disciplines. It is also important to note that only abstracts following a structure common to empirical studies were included in the corpus.

The disciplines were selected according to their orientation toward either the soft sciences (applied linguistics, marketing, and sociology) or the hard sciences (physics, biology, and mechanical engineering). The disciplinary orientation of each field was determined with reference to the distinction between *soft* and *hard* disciplines characterized by Biglan (1973a), Hyland (2004), and Durrant (2017). Table 1 provides a description of the corpus analyzed in this study.

### 3.2. Identification of lexical bundles

Lexical bundles are the most frequently occurring type of MWEs that function as basic building blocks of discourse (Biber, Conrad, & Cortes, 2004). In order for a sequence to be considered a lexical bundle, it should meet certain frequency and dispersion criteria (Biber et al., 1999). Lexical bundles in this study were identified using WordSmith Tools 7.0 (Scott, 1996). The program was set to look for 4+-word recurring clusters. The frequency thresholds for the identification of the expressions were also set according to the length of the sequences. Since length inversely correlates with frequency (Biber et al., 1999, p. 992), the moderately high frequency cut-off point of 20 occurrences per million words (pmw) was employed for identifying 4-

**Table 1**
Composition of the corpus.

| Discipline | Texts | Words |
|---|---|---|
| Applied Linguistics (AL) | 983 | 167,016 |
| Marketing (MA) | 981 | 167,003 |
| Sociology (SO) | 988 | 167,032 |
| Physics (PH) | 983 | 167,008 |
| Biology (BI) | 989 | 167,036 |
| Mechanical Engineering (ME) | 986 | 167,012 |
| Total number of texts and words | 5910 | 1,002,107 |

---

[2] The one-million corpus threshold is now established in many studies as the basic corpus size for the identification of lexical bundles (e.g., Cortes & Csomsay, 2015; Cortes, 2013).

word lexical bundles, ten times pmw for 5-word bundles, seven times pmw for 6- and 7-word bundles, and five times pmw for longer bundles. In addition, to ensure that the identified bundles were typical of the entire corpus and not restricted to a few texts or certain writing styles, the cut-off point of 5 + texts was employed for range. That is, only those sequences that occurred in five or more texts were included in our list. Finally, for shorter sequences embedded within longer bundles, only those that met the above-mentioned inclusion criteria were included in the final list. For instance, the six-word sequence *article concludes with a discussion of* would only be considered as a lexical bundle if it was observed seven times per million words by itself and not part of the 8-word bundle *the article concludes with a discussion of the*. The final list included 198 lexical bundle types and 5759 tokens.

## 3.3. Connecting bundles to rhetorical moves

All the bundles on the final list were examined in terms of their communicative roles in different rhetorical moves of the RA abstract. For this purpose (and in line with earlier research), the identified bundles were classified by the authors according to their general functions in discourse using the functional classification schemes proposed by Biber et al. (2004) and Hyland (2008). In order to better accommodate the identified bundles, certain minor adjustments were made to these classifications. Specifically, since no examples of Hyland's (2008) *topic* bundles were found, this sub-category was excluded from the *research-oriented* category and was replaced with the *intangible framing attributes* sub-category in Biber et al.'s (2004) classification scheme. This was deemed necessary as certain bundles were found to serve the function of describing abstract and intangible characteristics of the noun that followed (e.g., *the validity of the questionnaire*). In addition, the *stance feature* sub-category in Hyland's classification was divided into the subcategories of *epistemic stance* and *attitudinal/ modality stance* bundles to better distinguish between these two functions. The final categories are presented in Table 2.

Following this, the move model for abstracts proposed by Dos Santos (1996) and Pho (2008) was adopted to connect the bundles to the rhetorical moves in which they were found. This framework, as shown in Table 3, includes five main moves (i.e., Introduction, Purpose, Methods, Results, and Conclusion) which have been found to be highly consistent across texts (Gillaerts, 2013). Following Cortes (2013), every individual token of the identified bundles was analyzed in its surrounding context and was then assigned to the moves in which it occurred. This analysis was done by two trained linguists, with each rater independently coding all the identified bundle tokens. In addition, in order to gain a better understanding of the type of discourse commonly used in the abstracts of each discipline, five published authors from the fields of engineering, physics, biology, sociology, and marketing were consulted. This consultation involved discussions about the nature of the discipline in question, and the textual framework and standards that the authors typically follow when writing abstracts in their fields.

**Table 2**
Functional classification scheme (adapted from Biber et al., 2004; Hyland, 2008).

| Research-oriented bundles |
| --- |
| *Location - Indicating Time/Place (over the course of)* |
| *Procedure (the use of the)* |
| *Quantification (in a wide range of)* |
| *Description (the size of the)* |
| *Intangible framing attributes (the impact of the)* |
| |
| Text-oriented bundles |
| *Transition signals (in addition to the)* |
| *Resultative signals (as a result of)* |
| *Structuring signals (in the present paper)* |
| *Framing signals (on the basis of the)* |
| |
| Participant-oriented bundles |
| *Epistemic stance (are more likely to)* |
| *Attitudinal/modality stance (it is possible to)* |
| *Engagement (it is well known that)* |

**Table 3**
Moves in RA abstracts (adopted from Dos Santos, 1996; Pho, 2008).

| |
| --- |
| Move 1: Situating the research (Introduction) |
| Situating the research and indicating the gap |
| Move 2: Presenting the research (Purpose) |
| Presenting the research, hypothesis raising, and indicating the purpose |
| Move 3: Describing the methodology (Methods) |
| Providing information on design and data analysis |
| Move 4: Summarizing the results (Results) |
| Presenting results and summarizing main findings |
| Move 5: Discussing the research (Conclusion) |
| Interpreting the implications of main findings and recommendations for future research |

This procedure helped to avoid any misunderstanding regarding the nature of moves in different disciplines and consequently increased the reliability of raters' judgments of the general and rhetorical functions performed by each bundle. Following this, inter-rater reliability was also calculated, which showed 99.6% agreement between the two raters in terms of their move-bundle connection identifications. The points of disagreement which involved the occurrence of 4-word bundles belonging to multiple moves in the mechanical engineering and biology disciplines were discussed with the published authors from each of these two fields and were then resolved between the two raters based on the received input.

## 4. Analysis and results

### 4.1. Quantitative analyses

As was mentioned, the identification procedure resulted in a total of 198 bundle types and 5759 bundle tokens. Table 4 provides information about the number of bundle types and tokens found in each discipline.

Following their identification, all the bundles were grouped together according to the rhetorical functions of the moves in which they were found. Table 5 illustrates the distribution of the identified bundle types in different moves of RA abstracts across the target disciplines.

The results of a chi-square test showed that there is a significant difference between soft and hard knowledge fields in terms of the distribution of bundle types occurring in different rhetorical moves of RA abstracts ($X^2 = 18.403$, $df = 4$, $p = 0.001$, Cramer's $V = 0.109$).[3] The analysis of the standardized residuals (R) in the chi-square's contingency table showed that the number of bundle types used in Move 3 (Methods) made a significant contribution to the observed difference. In fact, as shown in Table 6, only the cells under the column "Methods" have an absolute value of R greater than 1.96. These results show that the writers of abstracts in the hard cluster made significantly more use of different bundle types in Move 3 (Methods), compared to their counterparts in the soft fields.

**Table 4**
The number of bundle types and tokens used in each discipline.

| Disciplines | Types | Tokens | Type/Token |
|---|---|---|---|
| Applied Linguistics (AL) | 148 | 1045 | .14 |
| Marketing (MA) | 145 | 1163 | .12 |
| Sociology (SO) | 166 | 798 | .20 |
| Physics (PH) | 136 | 1062 | .12 |
| Biology (BI) | 111 | 595 | .18 |
| Mechanical Engineering (ME) | 142 | 1096 | .13 |

**Table 5**
Distribution of bundle types in different moves of the RA abstract across disciplines.

| | Disciplines | Introduction | Purpose | Methods | Results | Conclusion |
|---|---|---|---|---|---|---|
| Soft sciences | AL | 30 | 79 | 40 | 75 | 33 |
| | SO | 38 | 79 | 34 | 85 | 28 |
| | MA | 37 | 63 | 31 | 91 | 26 |
| **Total** | | **105** | **221** | **105** | **251** | **87** |
| Hard sciences | PH | 39 | 66 | 61 | 87 | 20 |
| | BI | 45 | 48 | 40 | 69 | 19 |
| | ME | 35 | 72 | 52 | 85 | 20 |
| **Total** | | **119** | **186** | **153** | **238** | **59** |

**Table 6**
Standardized residuals in a chi-square contingency table for bundle types in different moves.

| | | Introduction | Purpose | Methods | Results | Conclusion |
|---|---|---|---|---|---|---|
| Soft | Observed | 105 | 221 | 105 | 251 | 87 |
| | Expected | 113.02 | 205.36 | 130.18 | 246.74 | 73.67 |
| | R | −0.755 | 1.090 | **−2.207** | 0.270 | 1.552 |
| Hard | Observed | 119 | 186 | 153 | 238 | 59 |
| | Expected | 110.97 | 201.63 | 127.81 | 242.25 | 72.32 |
| | R | 0.762 | −1.100 | **2.227** | −0.273 | −1.567 |

$X^2 = 18.403$, $df = 4$, $p = 0.001$, Cramer's $V = 0.109$.

---

[3] All the quantitative analyses in this study were performed using the open source programming language, R (R Core Team 2014. http://www.R-project.org/).

The token distribution of the bundles in different moves of RA abstracts also revealed a somewhat similar trend (see Table 7). A chi-square test indicated that, similar to the distribution of bundle types, there is a statistically significant difference between soft and hard clusters in terms of the distribution of bundle tokens in different moves of RA abstracts ($X^2 = 307.87$, $df = 4$, $p < 0.001$, Cramer's $V = 0.231$). In addition, the analysis of the standardized residuals in each cell showed that the number of bundle tokens used in Move 2 (Purpose), Move 3 (Methods), and Move 5 (Conclusion) significantly contributed to the rejection of the null hypothesis. As can be seen from Table 8, while abstract writers in the soft sciences made significantly more use of bundle tokens to convey the communicative aims of Move 2 (Purpose) and Move 5 (Conclusion), their counterparts in the hard cluster employed significantly more bundle tokens to explain the methodology of their study (Move 3).

Following this, the identified bundles were classified according to their general functions in discourse. As can be seen in Table 9, the differences between hard and soft clusters, in terms of the functional distribution of bundle types, were not found to be statistically significant.

However, the analysis of the functional distribution of bundle tokens in the two clusters revealed a different pattern of results. As shown in Table 10, while abstracts in the hard sciences contained more research-oriented bundle tokens, soft science abstracts were made up of more bundle tokens from the text- and participant-oriented categories.

**Table 7**
Distribution of bundle tokens in different moves of the RA abstract across disciplines.

|  | Disciplines | Introduction | Purpose | Methods | Results | Conclusion |
|---|---|---|---|---|---|---|
| Soft sciences | AL | 64 | 388 | 88 | 439 | 80 |
|  | SO | 89 | 296 | 53 | 305 | 53 |
|  | MA | 74 | 348 | 60 | 608 | 95 |
| **Total** |  | **227** | **1032** | **201** | **1352** | **228** |
| Hard sciences | PH | 57 | 228 | 265 | 455 | 32 |
|  | BI | 108 | 106 | 68 | 281 | 27 |
|  | ME | 63 | 335 | 239 | 406 | 38 |
| **Total** |  | **228** | **669** | **572** | **1142** | **97** |

**Table 8**
Standardized residuals in a chi-square contingency table for bundle tokens in different moves.

|  |  | Introduction | Purpose | Methods | Results | Conclusion |
|---|---|---|---|---|---|---|
| Soft | Observed | 227 | 1032 | 201 | 1352 | 228 |
|  | Expected | 240.64 | 899.62 | 408.82 | 1319.02 | 171.88 |
|  | R | −0.879 | **4.413** | **−10.278** | 0.907 | **4.280** |
| Hard | Observed | 228 | 669 | 572 | 1142 | 97 |
|  | Expected | 214.35 | 801.37 | 364.17 | 1174.97 | 153.11 |
|  | R | 0.931 | **−4.676** | **10.890** | −0.961 | **−4.534** |

$X^2 = 307.87$, $df = 4$, $p < 0.001$, Cramer's $V = 0.231$.

**Table 9**
Functional distribution of bundle types.

|  | Disciplines | Research-oriented | Text-oriented | Participant-oriented |
|---|---|---|---|---|
| Soft sciences | AL | 45 | 99 | 4 |
|  | SO | 47 | 112 | 7 |
|  | MA | 42 | 96 | 7 |
| **Total** |  | **134** | **307** | **18** |
| Hard sciences | PH | 46 | 85 | 5 |
|  | BI | 41 | 64 | 6 |
|  | ME | 46 | 92 | 4 |
| **Total** |  | **133** | **241** | **15** |

$X^2 = 2.46$, $df = 2$, $p = 0.2$.

**Table 10**
Functional distribution of bundle tokens.

|  | Disciplines | Research-oriented | Text-oriented | Participant-oriented |
|---|---|---|---|---|
| Soft sciences | AL | 277 | 741 | 27 |
|  | SO | 195 | 542 | 61 |
|  | MA | 226 | 884 | 48 |
| **Total** |  | **698** | **2167** | **141** |
| Hard sciences | PH | 361 | 682 | 19 |
|  | BI | 213 | 361 | 21 |
|  | ME | 392 | 687 | 17 |
| **Total** |  | **966** | **1730** | **57** |

**Table 11**
Standardized Residuals in a chi-square contingency table for functional distribution (tokens).

|      |          | Research-oriented | Text-oriented | Participant-oriented |
|------|----------|-------------------|---------------|----------------------|
| Soft | Observed | 698               | 2167          | 141                  |
|      | Expected | 868.55            | 2034.1        | 103.34               |
|      | R        | **−5.78**         | **2.94**      | **3.70**             |
| Hard | Observed | 966               | 1730          | 57                   |
|      | Expected | 795.44            | 1862.9        | 94.65                |
|      | R        | **6.04**          | **−3.07**     | **−3.86**            |

$X^2 = 116.91$, $df = 2$, $p < 0.001$, Cramer's $V = 0.14$.

The results of a chi-square test indicated that there is a significant difference, in terms of the functional distribution of bundle tokens, between the hard and soft sciences ($X^2 = 116.91$, $df = 2$, $p < 0.001$, Cramer's $V = 0.14$). The standardized residuals were again calculated to identify the cells contributing to the observed differences. As can be seen in Table 11, every cell in the contingency table contributed significantly to the observed differences, indicating that abstract writers in the hard sciences made significantly more use of research-oriented bundle tokens, whereas writers in the soft disciplines employed significantly more bundle tokens from the text- and participant-oriented categories in their abstracts. Given that such differences were not found to be statistically significant for the type distribution of these functions, it might be suggested can be argued that the frequent use of certain research-oriented bundles in hard disciplines and text- and participant-oriented bundles in soft disciplines has become conventionalized in abstracts in these fields.

Taken together, it would seem that writers of abstracts in the hard fields tend to use a large number of bundles to describe research objects, equipment and materials to specify the practical aspects of their research methods. In contrast, authors in the soft fields appear to have a tendency to use bundles that help organize a form of argument that justifies their research purpose and its findings. The following section will provide an in-depth analysis of such discourse functions in various moves of abstracts.

### 4.2. Qualitative analysis

In order to achieve a more fine-grained explanation of the discourse functions that mark the distinction between hard and soft disciplines, a qualitative analysis of the identified bundles in their surrounding context was deemed necessary. This part of our analyses focused on the usage and discourse functions of the bundles that were found to be distinctive of either hard or soft disciplines in each rhetorical move of the abstract. Distinctive bundles were defined as combinations which were found in either hard or soft disciplines and were absent in the other disciplinary group.

#### 4.2.1. Situating the research (move 1)

A comparison of the distinctive bundles in this move revealed key differences between abstracts in hard and soft science fields (see Appendix A for a complete list of distinctive bundles in each move). First, many of the research-oriented bundles in this move were found to be used by writers in the hard fields to specify the physical attributes of the following noun, describing different aspects of its quality, form and size (e.g., *the shape of the, the size of the, the efficiency of the, the quality of the*). These nouns were found to be research objects/contexts which were discussed in the accompanying article. Examples 1 and 2 below show how these bundles are used in context:

1) A key ingredient in such multiscale methods is the boundary condition and ***the size of the*** computational domain over which the microscale problems are solved. This problem is systematically investigated in the present paper …. (PH)
2) For such tests, a coordinate measuring machine (CMM) or other external measurement, outside of the machine tool, is required to measure ***the accuracy of the*** machined part. In this paper, a series of machining tests are …. (ME)

In addition, hard science writers also had a preference for using text-oriented bundle types in this move, compared to writers in the soft sciences. The majority of these bundles were found to be employed to situate arguments within a specific context:

3) Future detection of extensive air showers (EAS) produced by ultra-high energy cosmic particles by means of space-based fluorescence telescopes will open a new window on the universe and allow cosmic ray and neutrino astronomy at a level that is virtually impossible for ground-based detectors. ***In the context of the*** extreme universe space observatory (EUSO) project, an end-to-end simulation of EAS observation with a spatial detector has been designed. This paper investigates … (PH)
4) The most used model for predicting wear is the linear wear law proposed by Archard. A common generalization of Archard's wear law ***is based on the*** assumption that the wear rate at any point on the contact surface is proportional to the local contact pressure and the relative sliding velocity. This work focuses on … (ME)

As for the use of participant-oriented bundles in this move, abstract writers in both hard and soft fields used bundles from this category to promote and 'legitimize' their study by providing particular precedents in the same line of research:

5) Recently, some works have shown that ***it is possible to*** characterize and find the constitutive equation for brittle materials using a confined compression test … (ME)
6) Past research has shown that monetary (e.g., discounts) ***are more likely to*** be processed relative to and integrated with the original product price than are free promotions. We posit that … (MA)

We also examined the use of bundles that were found to be exclusive of Move 1, that is they appeared only in this move and were absent in other moves (see Appendix B for a complete list of exclusive bundles in each move). Almost all exclusive bundles in this move were found to convey the communicative function of indicating a gap in previous research. However, abstract writers in the hard sciences were also found to make use of a particular bundle (*it is well known that*) to express the same rhetorical aim in a somewhat different manner. As shown in the following examples, it appears that abstract writers in these disciplines use this bundle to engage the reader in recognizing the ensuing comment as an indisputable and widely-accepted view on the issue being introduced:

7) Although ***it is well known that*** the relative vibration between the tool and the work piece plays an important role in the surface generation, most of the prior work has been focused on … [ME]
8) Inferring species phylogenies is an important part of understanding molecular evolution. Even so, ***it is well known that*** an accurate phylogenetic tree reconstruction for a single gene does not always necessarily correspond to the species phylogeny … [BI]

As we can see from these examples, by using this particular bundle, the author attempts to construct a form of solidarity with the reader through appealing to shared knowledge (Hyland, 2008). However, the use of such a linguistics device was found to be completely absent in soft science abstracts. This is perhaps due to the divergent nature of the readership in soft disciplines and "fewer unequivocal bases for accepting claims" that restrain writers in these fields from making categorical statements (Hyland, 2008, p. 14).

### 4.2.2. Presenting the research (move 2)

Hard and soft science abstracts in this move were found to contain distinctive bundles from the research- and text-oriented categories, with soft science abstracts containing a few more distinctive bundle types from these two particular categories in this move. Also, the use of participant-oriented bundles in this move was found to be completely absent in both hard and soft fields.

However, once again, abstract writers in the hard disciplines were found to make use of bundles that describe the physical attributes of the following noun. Conversely, soft science writers appear to have a preference for using bundles that specify the intangible and abstract characteristics of things/processes in this move.

9) This study provides preliminary evidences for **the validity of the** Questionnaire of English Self-Efficacy scale in a sample of college students in China. [AL]

In addition, in this move, abstract writers in both domains were found to make extensive use of exclusive bundles (e.g., *this paper reports on, the aim of this study is to, objective of this paper is to*) to present the aims of their study. As was stated, such bundles were found to be exclusively linked to the rhetorical functions of a move in the abstract. Interestingly, in cases where writers were found to adopt a more direct approach to fulfilling this communicative aim, writers of abstracts in the soft fields were found to make use of the phrase "the authors" in bundles such as *the authors develop a, the authors propose a, the authors examine the, in this article the authors examine, the authors propose that.* It is worth mentioning that such bundles were completely absent in hard science abstracts.

10) The authors examine the impact of gifts on deposit balances and customer satisfaction in a longitudinal field experiment conducted at a commercial bank. They find that gifts increased … [MA]

As Bondi (2014) argues, authorial visibility in abstracts signals authors' attempt in establishing individuality as researchers and arguers. It may be that the social and discursive conventions of soft science disciplines urge writers in these fields to emphasize their own contribution to the field (Hyland & Bondi, 2006).

### 4.2.3. Describing the methodology (move 3)

Writers of abstracts in both clusters were found to make use of research-oriented bundles in this move. However, writers of abstract in the hard fields made use of various types of bundles from this category. These bundles are used to describe procedures and processes (*in the presence of a*), the shape, the quantity and the effect of the material used (*the shape of the*, *for*

*a wide range of, the influence of the*). Writers in soft disciplines, on the other hand, were found to make use of certain bundles to refer to a particular time in the process of conducting the research (*at the end of the*).

11) Using two corpora, (i) a student corpus of about 62, 000 words of postgraduate computing project reports, written **at the end of the** MSc programme … (AL).

Further, our analysis showed that only three bundles were exclusive to this move. Interestingly, all of these bundles were found to be present only in hard science abstracts (*the results are compared with, by means of the, is applied to*). The following examples indicate the rhetorical aims reflected by the use of these bundles in Move 3:

12) The effects of tube protrusion depth as well as header mass flux are investigated, and ***the results are compared with*** previous 30 channel data. [ME]
13) We calculate the partition function Z(t) going beyond the Thomas-Fermi (TF) approximation ***by means of the*** semi-classical expansion using the Winger-Kirkwood (WK) method. [PH]
14) In this study, a recently developed likelihood method for reconstructing ancestral areas (the dispersal-extinction cladogenesis [DEC] model) ***is applied to*** the angiosperm family … [BI]

### 4.2.4. Summarizing the results (move 4)

A comparison of the distinctive bundles in this move revealed that abstract writers in the soft fields made use of the bundle *are more likely to* present their results.

15) Additionally, the results of a meta-analytical study based on data from all five experiments further show that products that elicit a stronger emotional response **are more likely to** yield consistent preferences. [MA]
16) Findings suggest that older cohabiters **are more likely to** view their relationship as an alternative to marriage, whereas younger cohabiters **are more likely to** view their relationship as a prelude to it. [SO]

As can be inferred from the examples, it appears that writers in soft knowledge fields tend to avoid overgeneralizations of their results and to tone down their views when discussing their research outcomes. Hyland and Tse (2005) also observed a certain degree of tentativeness in the use of evaluative *that* clauses by writers of abstracts in soft fields when making claims. It is perhaps due to the diversity of the readership and knowledge in soft disciplines that writers in these fields tend to ensure that their interpretation of data and results is potentially falsifiable (Hyland, 2004).

Turning to the exclusive bundles in this move, we found that writers in the hard disciplines use certain bundles, such as *are in good agreement with the* and *in good agreement with*, to describe the results and main findings of their study. The analysis of the concordance lines showed that writers in hard fields use these two particular bundles to provide validity to their results. This is perhaps due to the highly competitive nature of hard science disciplines in which validity of results and newsworthiness of findings play an important role in creating a persuasive discourse:

17) At this plasma density, the average accelerating gradient is 36 GeVm-1. These results ***are in good agreement with the*** numbers obtained from particle in cell simulations describing the experiment. The linear scaling is also observed both at lower and higher … [PH]
18) We vary the radial size of the cylindrical constraint and find that our experimental results are **in good agreement with** existing predictions on the critical injection force and length of injected rod for helical buckling.

As Hyland (2004) argues, the constant progress of knowledge in hard science domains requires researchers in these fields to ensure that their research provides viable and practical solutions to current real-world issues.

Furthermore, we also found that, in this move, writers in the soft disciplines once again took a more explicit position in reporting the results of their study by employing exclusive bundles such as *the authors also find that, the authors show that the, the authors demonstrate that the*.

### 4.2.5. Discussing the research (move 5)

Finally, the analysis of the identified bundles in Move 5 (Conclusion) showed that certain bundles exclusive to this move were found only in one discipline (i.e., Marketing). Exclusive bundles, such as *the authors discuss the implications of these findings* and *the authors conclude with a discussion of* were only used by authors in the Marketing discipline. It is also worth noting that these two particular bundles were the longest bundles identified in the entire corpus. As illustrated in the following examples, writers of abstracts in the Marketing discipline use these bundles to briefly outline the structure of the accompanying article.

19) However, this occurred only for a product in the same category but not for a product in a different one. ***The authors discuss the implications of these findings*** for corporate brand and product brand managers and suggest directions for further research. [MA]

20) The results support explanations of chasms and technological cycles for information/entertainment products, business cycles, and technological cycles for kitchen/laundry products. ***The authors conclude with a discussion of*** the findings, contributions, and implications. [MA]

Similar rhetorical patterns were also found in other soft disciplines but with different exclusive bundles (e.g., *the article concludes with a discussion of the, implications of these findings for*). The following examples indicate the rhetorical work achieved by the use of these bundles in Applied linguistics and Sociology:

21) Negotiators used HRS strategically in combination with 'formulations' and 'accounts' to move the negotiation forward in pursuit of their goals. ***The article concludes with a discussion of the*** teaching implications of the findings and makes some concrete suggestions for teaching negotiating. [AL]

22) We also demonstrate that the effect of any of these models can be illustrated dynamically with estimates of uncertainty through a simulation procedure. The paper concludes with ***implications of these findings for*** the practical researcher. [SO]

It should further be noted that all the exclusive bundles in Move 5 were completely absent in the hard science abstracts.

## 5. Discussion

The present study set out to investigate disciplinary variation in RA abstracts as it is reflected in the use of multi-word items in different moves of this section of the RA. Our main goal was to determine the degree to which writers of abstracts in hard and soft science disciplines differ in terms of the MWEs they commonly employ to realize the communicative purposes of the abstract. To explore such differences, we adopted a mixed-methods approach through which the identified patterns of variation were examined both quantitatively and qualitatively. The complementary strength of this hybrid methodology enabled us to uncover a series of differences in focus and practice between hard and soft science fields. Below, we first discuss the patterns of variation in light of our quantitative analyses and then specify some of the major differences uncovered in our qualitative inquiry.

To explore possible differences between hard and soft science abstracts, we first compared hard and soft science disciplines in terms of the sequences they commonly use in different moves of RA abstracts. Our analyses showed that abstract writers in the hard disciplines made significantly more use of bundle types and tokens in Move 3 (methods), compared to their counterparts in the soft science fields. This finding seems to reflect the highly formulaic and conventional nature of this move in hard science abstracts. As Hyland (2015) argues, conventionality in genres is heavily influenced by the norms and expectations of the community in which they are used. In other words, it can be argued that the disciplinary demands that exist in hard science disciplines require writers in these fields to invest more rhetorical effort in describing their research methods, which in turn has resulted in the conventionalization of certain sequences in this move over time. As Gray (2015) points out, one of the important parameters that corresponds to linguistic variation between disciplines is the research paradigm within which researchers carry out their scientific activities. Gray (2015) observed that hard science RAs always contained an explicit description of data and methods which was taken as a reflection of the epistemological orientations that guide research practices in the hard sciences. As Hyland (2008) argues, writers in the hard domains have a tendency for placing emphasis on the methods, equipment and procedures used in their study to highlight the experimental basis of their research for their target readership.

In addition, we also found that abstract writers in the soft fields employed significantly more bundle tokens in Move 2 (presenting the research and its purpose) and Move 5 (discussing the research and its findings). Considering that such a difference was only observed in the use of bundle tokens (but not types) in the same moves, we can argue that abstract writers in soft science fields are less diverse, compared to their counterparts in the hard fields, in fulfilling the communicative functions of these two particular moves. This implies that writers in these fields have a tendency to make extensive use of certain bundles that are conventionally accepted for the realization of these two moves. The question to be addressed here, however, is that why such a tendency exists only in soft science abstracts. Our interpretation is that abstract writers in soft fields feel the need to ensure that the purpose and implications of their study are readily understood by its target readership. And so, they prefer to employ a set of broadly recognized sequences to realize the communicative functions of these two moves. Previous studies have shown that, due to the fuzzy boundaries of knowledge, or lack of a clear research paradigm in soft disciplines (Becher & Trowler, 2001; Gray, 2015; Hyland & Bondi, 2006; Kuhn, 2012), writers in these fields often feel the need to demarcate their studies by deploying certain discursive and meta-discursive strategies (Bondi & Cavalieri, 2012; Cavalieri, 2014; Hyland, 2004, 2008). Therefore, it can be argued that abstract writers in soft fields have a tendency to galvanize interest in their study by highlighting its purpose and implications, which entails favoring highly conventional sequences to maximize comprehensibility in these two rhetorical moves. This finding is in line with Jiang and Hyland (2017), who found that writers of abstracts in soft disciplines made extensive use of metadiscursive nouns to explain the aims of their

research. The authors maintain that, due to dealing with loosely-defined research issues and unspecified theoretical structures, writers of abstracts in soft disciplines tend to invest more rhetorical and discursive effort to acquaint their readers with the importance of their research goals and their implications.

We further analyzed the functional distribution of bundles types and tokens in hard and soft science abstracts. Our analyses showed that abstracts in hard science fields contained significantly more research-oriented bundle tokens than soft disciplines. This finding is in line with Hyland (2008) and Durrant (2017) who also found a greater concentration of research-oriented bundles in hard science texts. It is also noteworthy that the majority of these bundles in hard science abstracts were found to contribute to the description of the physical aspects of research objects/materials. Moreover, we also found that abstract writers in the soft disciplines made significantly more use of bundle tokens from text- and participant-oriented categories. As Hyland (2008) notes, this seems to reflect the more discursive nature of argument in soft disciplines in which persuasive strategies are carried out in a more interpretative, and less empiricist, fashion. The inherently argumentative nature of texts in soft science fields in different academic genres has been documented in several studies (e.g., Bondi & Lores Sanz, 2014; Durrant, 2014; Hyland, 2005; Pho, 2008; Stotesbury, 2003). Hyland (2011), for instance, demonstrated that writers in the humanities and social sciences tend to make extensive use of hedges and boosters in their research articles and also frame their arguments in a more interpretative and dialogically engaging manner, compared to their counterparts in hard fields.

Our further investigation focused on the qualitative analysis of the identified bundles in their surrounding context. Specifically, we examined those bundles that were distinctive of each cluster in a certain move, along with those that were found to be exclusively linked to one specific move. Our analyses highlighted a number of key findings. First, we found that exclusive bundles used to fulfill the communicative aims of Move 3 (describing the methodology) were only present in hard science abstracts. Interestingly, we also found that all the exclusive bundles in Move 5 (discussing the research and its findings) belonged to soft science abstracts. This finding, together with those outlined above, suggests that writers of abstracts in hard and soft fields have different priorities for promoting their studies. That is, while hard fields view their methodology as a unique 'selling-point' for their study, writers in soft fields prefer to place more emphasis on their research objectives and their possible theoretical implications.

Second, we observed key differences in the way in which the communicative aims of certain moves were fulfilled by authors in the two clusters. One interesting example is the use of hedging devices (e.g., *are more likely to*) in soft science abstracts in Move 4 (summarizing the results). Abstract writers in these fields appear to be more inclined, compared to hard science writers, to avoid overgeneralization and present their findings in a less assured way. Another key difference is the exclusive use of bundles that describe the physical attributes of research objects (e.g., *the shape of the*) in various moves by authors in the hard disciplines, which "impart[s] a greater real-world and laboratory-focused sense to writing" in these fields (Hyland, 2008, p. 14). In contrast, abstract writers in the soft disciplines showed a preference for focusing on abstract constructs by making use of bundles that specify intangible attributes of procedures and processes (e.g., *the validity of the*).

## 6. Conclusion

The intellectual boundaries drawn by disciplinary frameworks can shape the ways in which knowledge is negotiated and presented in different academic domains. The findings reported in this study support this view by underlining the ways in which knowledge is constructed within these frameworks in a given academic genre. More specifically, our study shows that members of different academic domains have different priorities for representing their research in academic abstracts. Through the methodological synergy of corpus linguistics and genre analysis, it was found that writers of abstracts in the soft sciences tend to make sure that the objective of their research and its contribution to the field is clearly expressed. On the other hand, writers of abstract in the hard sciences see it as crucial to promote their research through the lens of its methodology and apparatus.

We believe that highlighting the differences that emerge from the ways writers construct knowledge and communicate it with their readers in different academic domains has great potential to reinforce our understanding of disciplinary writing. We would also argue that studies of this kind can be of great help, especially, to graduate students and researchers at the beginning stages of their career who wish to participate in scientific activities in their field, but have not yet acquired the disciplinary knowledge required for such a feat. In addition, with the growing interest in using English for Research Publication Purposes among L2 researchers (see Flowerdew, 2015), we believe that the investigation of disciplinary conventions through the lens of their linguistic realizations in discourse can help this particular group of language users develop a strong understanding of disciplinary writing.

Future research can investigate linguistic patterns of disciplinary variation in multiple sections of the RA and conduct a cross-sectional analysis of the emerged patterns. Such a study can further inform our understanding of disciplinary variation and the construction of knowledge in different sections of the RA.

## Acknowledgements

# References

Becher, T. (1994). The significance of disciplinary differences. *Studies in Higher Education, 19*(2), 151—161. https://doi.org/10.1080/03075079412331382007.

Becher, T., & Trowler, P. (2001). *Academic tribes and territories: Intellectual enquiry and the culture of disciplines* (2nd ed.). Buckingham: Open University Press.

Berliner, D. C. (2002). Comment: Educational research: The hardest science of all. *Educational Researcher, 31*(8), 18—20. https://doi.org/10.3102/0013189X031008018.

Bhatia, V. K. (1993). *Analysing genre: Language use in professional settings*. London: Longman.

Biber, D. (2009). A corpus-driven approach to formulaic language in English. *International Journal of Corpus Linguistics, 14*(3), 275—311. https://doi.org/10.1075/ijcl.14.3.08bib.

Biber, D., Connor, U., & Upton, T. A. (2007). *Discourse on the move: Using corpus analysis to describe discourse structure*. Amsterdam/Philadelphia: John Benjamins.

Biber, D., Conrad, S., & Cortes, V. (2004). If you look at: Lexical bundles in university teaching and textbooks. *Applied Linguistics, 25*(5), 371—405.

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

Biglan, A. (1973a). The characteristics of subject matter in different academic areas. *Journal of Applied Psycholinguistics, 57*(3), 195—203.

Biglan, A. (1973b). Relationships between subject matter characteristics and the structure and output of university departments. *Journal of Applied Psychology, 57*(3), 204—213. https://doi.org/10.1037/h0034699.

Bondi, M. (2014). Changing voices: Authorial voice in abstracts. In M. Bonid, & R. Lorés Sanz (Eds.), *Abstracts in academic discourse: Variation and change* (pp. 243—269). Berlin: Peter Lang.

Bondi, M., & Cavalieri, S. (2012). The evolution of the abstract as a genre. In G. Garzone, P. Catenaccio, & C. Degano (Eds.), *Genre change in the contemporary world. Short-term diachronic perspectives* (pp. 43—55). Peter Lang: Bern.

Bondi, M., & Lorés Sanz, R. (Eds.). (2014). *Abstracts in academic discourse: Variation and change*. Berlin: Peter Lang.

Braine, G. (1995). Writing in the natural sciences and engineering. In D. Belcher, & G. Braine (Eds.), *Academic writing in a second language: Essays on research and pedagogy* (pp. 113—134). New Jersey: Ablex.

Brown, R. (2004). Self-composed: Rhetoric in Psychology personal statements. *Written Communication, 21*(3), 242—260. https://doi.org/10.1177/0741088304264338.

Bruce, I. (2008). Cognitive genre structures in methods sections of research articles: A corpus study. *Journal of English for Academic Purposes, 7*(1), 38—54. https://doi.org/10.1016/j.jeap.2007.12.001.

Cavalieri, S. (2014). Variation across disciplines. The case of applied linguistics and medicine. In M. Bonid, & R. Lorés Sanz (Eds.), *Abstracts in academic discourse: Variation and change* (pp. 161—174). Berlin: Peter Lang.

Clark, B. R. (1983). *The higher education system*. Berkeley, CA: University of California Press.

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. https://doi.org/10.1016/j.jeap.2012.11.002.

Cortes, V., & Csomay, E. (Eds.). (2015). *Corpus-based research in applied linguistics: Studies in honor of Doug Biber*. Amsterdam: John Benjamins.

Cowie, A. P. (1994). Phraseology. In R. E. Asher (Ed.), *The encyclopedia of language and linguistics* (pp. 3168—3171). Oxford: Oxford University Press.

Donald, J. G. (1990). University professors' views of knowledge and validation processes. *Journal of Educational Psychology, 82*(2), 242—249. https://doi.org/10.1037/0022-0663.82.2.242.

Dos Santos, M. B. (1996). The textual organization of research paper abstracts in applied linguistics. *Text-Interdisciplinary Journal for the Study of Discourse, 16*(4), 481—500. https://doi.org/10.1515/text.1.1996.16.4.481.

Dudley-Evans, T. (1994). Variations in the discourse patterns favoured by different disciplines and their pedagogic applications. In J. Flowerdew (Ed.), *Academic listening: Research perspectives* (pp. 46—58). Cambridge: CUP.

Durrant, P. (2014). Discipline and level specificity in university students' written vocabulary. *Applied Linguistics, 35*(3), 328—356. https://doi.org/10.1093/applin/amt016.

Durrant, P. (2017). Lexical bundles and disciplinary variation in university students' writing: Mapping the territories. *Applied Linguistics, 38*(2), 165—193. https://doi.org/10.1093/applin/amv011.

Erman, B., & Warren, B. (2000). The idiom principle and the open choice principle. *Text, 20*(1), 29—62. https://doi.org/10.1515/text.1.2000.20.1.29.

Flowerdew, J. (2015). Some thoughts on English for research publication purposes (ERPP) and related issues. *Language Teaching, 48*(2), 250—262. https://doi.org/10.1017/S0261444812000523.

Foster, P. (2001). Rules and routines: A consideration of their role in the task-based language production of native and non-native speakers. In M. Bygate, P. Skehan, & M. Swain (Eds.), *Researching pedagogic tasks: Second language learning, teaching, and testing* (pp. 75—94). Harlow, England: Longman.

Gillaerts, P. (2013). Move analysis of abstracts from a diachronic perspective: A case study. In N. L. Johannesson, G. Melchers, & B. Björkman (Eds.), *Of butterflies and birds, of dialects and genres* (pp. 49—60). Stockholm: Acta Universitatis Stockholmiensis.

Graetz, N. (1985). Teaching EFL students to extract structural Information from abstracts. In J. M. Ulign, & A. K. Pugh (Eds.), *Reading for professional purposes: Methods and materials in teaching languages* (pp. 123—135). Leuven: Acco.

Gray, B. (2015). *Linguistic variation in research articles: When discipline tells only part of the story*. Amsterdam, Netherlands: John Benjamins.

Haswell, R. (1991). *Gaining ground in college writing: Tales of development and interpretation*. Dallas: Southern Methodist University Press.

Holmes, R. (1997). Genre analysis, and the social sciences: An investigation of the structure of research article discussion sections in three disciplines. *English for Specific Purposes, 16*(4), 321—337. https://doi.org/10.1016/S0889-4906(96)00038-5.

Howarth, P. (1998). The phraseology of learners' academic writing. In A. P. Cowie (Ed.), *Phraseology: Theory, analysis, and application* (pp. 161—186). Oxford, England: Oxford University Press.

Hyland, K. (2004). *Disciplinary discourses*. Ann Arbor: University of Michigan Press.

Hyland, K. (2011). Disciplines and discourses: Social interactions in the construction of knowledge. In D. Starke-Meyerring, A. Pare, N. Artemeva, M. Horne, & L. Yousoubova (Eds.), *Writing in knowledge societies* (pp. 193—214). Fort Collins, CO: The WAC Clearinghouse and Parlor Press.

Hyland, K. (2005). Stance and engagement: A model of interaction in academic discourse. *Discourse Studies, 7*(2), 173—192. https://doi.org/10.1177/1461445605050365.

Hyland, K. (2008). As can be seen: Lexical bundles and disciplinary variation. *English for Specific Purposes, 27*(1), 4—21. https://doi.org/10.1016/j.esp.2007.06.001.

Hyland, K. (2015). Genre, discipline and identity. *Journal of English for Academic Purposes, 19*, 32—43. https://doi.org/10.1016/j.jeap.2015.02.005.

Hyland, K., & Bondi, M. (Eds.). (2006). *Academic discourse across disciplines*. Bern: Peter Lang.

Hyland, K., & Tse, P. (2005). Hooking the reader: A corpus study of evaluative that in abstracts. *English for Specific Purposes, 24*(2), 123—139. https://doi.org/10.1016/j.esp.2004.02.002.

Jiang, F. K., & Hyland, K. (2017). Metadiscursive nouns: Interaction and cohesion in abstract moves. *English for Specific Purposes, 46*, 1—14. https://doi.org/10.1016/j.esp.2016.11.001.

Kuhn, T. S. (2012). *The structure of scientific revolutions* (50th Anniversary Edition). University of Chicago press.

Le, T. N. P., & Harrington, M. (2015). Phraseology used to comment on results in the discussion section of applied linguistics quantitative research articles. *English for Specific Purposes, 39*, 45—61. https://doi.org/10.1016/j.esp.2015.03.003.

Lim, J. M. H. (2006). Method sections of management research articles: A pedagogically motivated qualitative study. *English for Specific Purposes, 25*(3), 282—309. https://doi.org/10.1016/j.esp.2005.07.001.

Lorés, R. (2004). On RA abstracts: From rhetorical structure to thematic organisation. *English for Specific Purposes, 23*(3), 280—302. https://doi.org/10.1016/j.esp.2003.06.001.

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.* https://doi.org/10.1111/lang.12250.

Moore, T. (2002). Knowledge and agency: A study of 'metaphenomenal discourse' in textbooks from three disciplines. *English for Specific Purposes, 21*(4), 347–366. https://doi.org/10.1016/S0889-4906(01)00030-8.

Moreno, A. I., & Swales, J. M. (2018). Strengthening move analysis methodology towards bridging the function-form gap. *English for Specific Purposes, 50*, 40–63. https://doi.org/10.1016/j.esp.2017.11.006.

Parkinson, J. (2013). Representing own and other voices in social science research articles. *International Journal of Corpus Linguistics, 18*(2), 199–228. https://doi.org/10.1075/ijcl.18.2.02par.

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. https://doi.org/10.1177/1461445607087010.

Samraj, B. (2002). Introductions in research articles: Variations across disciplines. *English for Specific Purposes, 21*(1), 1–17. https://doi.org/10.1016/S0889-4906(00)00023-5.

Samraj, B. (2005). An exploration of a genre set: Research article abstracts and introductions in two disciplines. *English for Specific Purposes, 24*(2), 141–156. https://doi.org/10.1016/j.esp.2015.10.001.

Samraj, B. (2008). A discourse analysis of master's theses across disciplines with a focus on introductions. *Journal of English for Academic Purposes, 7*(1), 55–67. https://doi.org/10.1016/j.jeap.2008.02.005.

Samraj, B. (2016). Discourse structure and variation in manuscript reviews: Implications for genre categorization. *English for Specific Purposes, 42*, 76–88. https://doi.org/10.1016/j.esp.2015.12.003.

Samraj, B., & Monk, L. (2008). The statement of purpose in graduate program applications: Genre structure and disciplinary variation. *English for Specific Purposes, 27*(2), 193–211. https://doi.org/10.1016/j.esp.2007.07.001.

Schmitt, N. (Ed.). (2004). *Formulaic sequences: Acquisition, processing, and use.* Amsterdam: John Benjamins Publishing.

Sinclair, J. (1991). *Corpus, concordance, collocation.* Oxford: Oxford University Press.

Siyanova-Chanturia, A., & Omidian, T. (in press). Key issues in researching multi-word items. In S. Webb (Ed.), The Routledge handbook of vocabulary studies.

Scott, M. (1996). *WordSmith tools.* Oxford: Oxford University Press.

Sorhus, H. B. (1977). To hear ourselves: Implications for teaching English as a second language. *English Language Teaching Journal, 31*(3), 211–221.

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. https://doi.org/10.1016/j.esp.2012.09.001.

Storer, N., & Parsons, T. (1968). The disciplines as a differentiating force. In E. B. Montgomery (Ed.), *The foundations of access to knowledge.* Syracuse: Syracuse University Press.

Stotesbury, H. (2003). Evaluation in research article abstracts in the narrative and hard sciences. *Journal of English for Academic Purposes, 2*(4), 327–341. https://doi.org/10.1016/S1475-1585(03)00049-3.

Swales, J. (1981). *Aspects of article introductions.* Birmingham, UK: The University of Aston, Language Studies Unit.

Swales, J. (1984). Research into the structure of introductions to journal articles and its application to the teaching of academic writing. In R. Williams, & J. Kirkman (Eds.), *Common Grounds: Shared interests in ESP and communication studies* (pp. 77–86). New York NY: Pergamon Press.

Swales, J. (1990). *Genre analysis: English in academic and research settings.* Cambridge University Press.

Swales, J. (2004). *Research genres: Exploration and applications.* Cambridge: Cambridge University Press.

Swales, J. M., & Feak, C. B. (2009). *Abstracts and the writing of abstracts.* Ann Arbor: University of Michigan Press.

Thompson, P. (2000). Citation practices in PhD theses. In L. Burnard, & T. McEnery (Eds.), *Rethinking language pedagogy from a corpus perspective* (pp. 91–101). Frankfurt: Peter Lang.

Tognini-Bonelli, E. (2001). *Corpus linguistics at work.* Amsterdam: John Benjamins.

Van Lancker-Sidtis, D., & Rallon, G. (2004). Tracking the incidence of formulaic expressions in everyday speech: Methods for classification and verification. *Language & Communication, 24*(3), 207–240. https://doi.org/10.1016/j.langcom.2004.02.003.

Ventola, E. (1994). Abstracts as an object of linguistic study. In S. Cmejrková, F. Danes, & E. Havlová (Eds.), *Writing versus speaking: Language, text, discourse, communication* (pp. 333–352). Tübingen: Gunter Narr.

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

Yang, R., & Allison, D. (2003). Research articles in applied linguistics: Moving from results to conclusions. *English for Specific Purposes, 22*(4), 365–385. https://doi.org/10.1016/S0889-4906(02)00026-1.

**Taha Omidian** is a PhD student at Victoria University of Wellington. Taha's research interests include academic writing, learner corpora, and formulaic language. His recent study into the pedagogic value of multi-word expressions was published in *TESOL Journal.*

**Hesamoddin Shahriari** is an assistant professor of applied linguistics at the Department of English, Ferdowsi University of Mashhad, where he teaches undergraduate and graduate-level courses in academic writing. Hesamoddin's research interests include Corpus Linguistics, academic writing and learner English. His most recent study into the development of noun phrase complexity in academic writing was published in the *Journal of English for Academic Purposes.*

**Anna Siyanova-Chanturia** is a senior lecturer in Applied Linguistics at Victoria University of Wellington. Anna's research interests include psychological aspects of second language learning, vocabulary, multi-word expressions, and quantitative research methods. She has published in *Studies in Second Language Acquisition*, *Language Learning*, *Applied Linguistics, Journal of Experimental Psychology: Learning, Memory and Cognition,* and other venues.