

This is the accepted version of the manuscript. The final, definitive version of this paper will be published in *Language Learning* (<http://onlinelibrary.wiley.com/doi/10.1111/lang.12250/abstract>) by Wiley, All rights reserved.

## Citation

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*. doi:10.1111/lang.12250 Retrieved from <http://mizumot.com/files/LL-AWSuM.pdf>

**Title:**

**Applying the bundle-move connection approach to the development of an online writing support tool for research articles**

**Authors:**

Atsushi Mizumoto,<sup>a</sup> Sawako Hamatani,<sup>a</sup> and Yasuhiro Imao<sup>b</sup>

<sup>a</sup>Kansai University and <sup>b</sup>Osaka University

**Running head:**

USING BUNDLE-MOVE CONNECTION FOR TOOL DEVELOPMENT

**Abstract**

With advances in information and computer technology, genre-based writing pedagogy has developed greatly in recent years. In order to further this growth in technology-enhanced genre writing pedagogy, the current study developed a data-driven and theory-based practical writing support tool for research articles (RAs). This web-based, innovative tool, powered by the combination of rhetorical moves and lexical bundles, has an auto-complete feature that suggests the most frequent lexical bundles in a move within an RA section. It was developed based on the proof-of-concept of the bundle-move connection approach. Preliminary user feedback was positive

overall, and it was found that the writing support tool brought about beneficial effects that genre writing pedagogy explicitly aims to achieve. In light of these findings, the pedagogical implications of the developed tool are discussed, with particular focus on the potential role that it could play in the teaching and learning of technology-enhanced genre writing.

**Keywords:**

corpus; lexical bundles; move; genre analysis; English for academic purposes

**Acknowledgements**

The study was funded by the 2015-2016 Kansai University Outlay for Establishing Research Centers and JSPS KAKENHI Grant Numbers 17H02369 and 15K02717. We would like to thank the four anonymous reviewers and Dr. Pavel Trofimovich, editor of *Language Learning*, for their very helpful suggestions in improving the quality of this paper.

**Corresponding author:**

Atsushi Mizumoto, Faculty of Foreign Language Studies, Kansai University, 3-3-35, Yamate-cho, Suita, 564-8680 Osaka, Japan. E-mail: mizumoto@kansai-u.ac.jp

## 1. Introduction

In the ever competitive world of academic publishing (Hyland, 2015), in which “publish or perish” has been the norm across a range of disciplines, early career novice researchers, including graduate students (or sometimes undergraduate students), must take up the challenge of writing publishable research articles (RAs). If the researcher’s first language is not English, they will most likely feel at a linguistic disadvantage compared to researchers whose first language is English. It is certainly understandable that, from the earlier learning phase to becoming a full-fledged researcher, “(t)he task of learning register/genre differences is even more challenging for a non-native speaker of a language” (Biber & Conrad, 2009, p. 3).

For this reason, genre analysis has been widely used in EAP (English for Academic Purposes) research. Within genre analysis, which is a more specific form of discourse analysis (Hyland, 2013), Swales’s move framework (Swales, 1990) for research articles is the framework most employed for descriptive and pedagogical purposes. Analysis of moves, which are rhetorical and organizational structures with specific communicative functions, has been further advanced by corpus-based studies over the past 20 years (e.g., Biber, Connor, Upton, & Kanoksilapatham, 2007; Flowerdew, 2009; Hardy & Friginal, 2016; Paltridge, 2014).

Corpus-based studies also have a bottom-up influence on research and practice in EAP. That is, using EAP corpora in specific disciplines, researchers and practitioners can systematically identify the words (e.g., Coxhead, 2000; Gardner & Davies, 2014) and multi-word expressions (MWEs) (e.g.,

Ackermann & Chen, 2013; Durrant, 2009; Simpson-Vlach & Ellis, 2010) used by experts in their respective academic disciplines. Among formulaic sequences (Wray, 2002), which is an encompassing term that covers a wide range of phraseology including MWEs (although see Myles & Cordier, 2016 for a more refined construct of formulaic sequences), lexical bundles have been widely investigated in corpus-based studies (Biber, Johansson, Leech, Conrad, & Finegan, 1999). Lexical bundles are “simply sequences of word forms that commonly go together in natural discourse” (Biber et al., 1999, p. 990), such as “I don’t know what,” “can I have a,” and “do you want to” in spoken discourse, and “as a result of,” “at the same time,” and “the way in which” in written discourse. Lexical bundles are not only simple units comprising a few words, but also play an important role in reflecting the difference between disciplines or registers (e.g., Biber, Conrad, & Cortes, 2004; Durrant, 2015; Grabowski, 2015; Hyland, 2008b). That is, research on lexical bundles strongly supports the notion that investigation of lexical bundles in different academic fields reveals disciplinary variation.

Pointing out the similarities between moves and lexical bundles as the “building blocks” of discourse, Cortes (2013) recently suggested an approach to connecting lexical bundles with moves according to their functions in RA introductions. Motivated by Cortes (2013), in this study, we take this approach a step further by developing a data-driven and theory-based practical RA writing support tool. The new writing support tool realizes the concept of “lexical bundles meet moves” by providing immediate suggestions for the lexical bundles most frequently used in respective moves in

RA sections using an auto-complete feature. In developing the tool, the appropriateness of the bundle-move connection approach was first examined. A pilot study was then conducted to examine the extent to which the users could benefit from the developed tool. We thus sought to illustrate how this innovative RA writing support tool can play a supporting role in the research and pedagogical endeavor of technology-enhanced genre writing teaching and learning.

The following section reviews the relevant literature informing our tool development, focusing on (a) genre and move analysis, (b) corpus use in genre-based writing pedagogy, and (c) lexical bundles as building blocks of discourse. We then present our rationale for developing a web-based RA writing support tool.

## **2. Literature Review**

### **2.1 Genre and Move Analysis**

According to Hyland (2016), genres are “abstract, socially recognized ways of using language and represent how writers typically respond to recurring situations” (p. 120). Therefore, members of a specific genre community recognize, without much difficulty, the common structures and linguistic features required for their specific purposes, which they acquire through repeated exposure to the genre. As an alternative to process approaches to writing instruction, which regard writing as an exploratory and recursive process (Polio & Williams, 2009), genre-based pedagogy has been the topic of much research publication in L2 writing over the last few decades (Hyland, 2004; Tardy,

2009, 2012).

Of the three approaches to genre adopted in the field of applied linguistics (Hyon, 1996), namely, (a) English for specific purposes (ESP), (b) North American New Rhetoric studies (the New Rhetoric), and (c) Australian systemic functional linguistics (Systemic Functional Linguistics), the ESP approach to genre has been most vibrant due to its versatile concept of genre analysis (Bhatia, 1993; Swales, 1990). Genre analysis is the investigation of the typical and recurrent linguistic features and discourse within a given genre. Move analysis, a specific type of genre analysis (and often considered equivalent to it), is mainly employed to analyze rhetorical structures in each section of a research article (RA). Swales's move analysis (1990) is most famously exemplified through the CARS (Create a Research Space) model, which is an analytic framework for describing RA introductions. In Swales's revised CARS model for RA introductions (Swales, 2004), there are three move structures: Move 1: Establishing a territory, Move 2: Establishing a niche, and Move 3: Presenting the present work. Moves are used according to their communicative purposes, and they can be further divided into steps, which are particular constituent elements of each move to be realized (e.g., Indicating a gap, Adding to what is known, and Presenting positive justification for Move 2: Establishing a niche). The CARS model "has apparently been quite successful" (Swales, 2004, p. 226), and has been used as a theoretical basis for literally thousands of studies of academic and professional genres (Cheng, 2015). Thus, there is a very strong body of literature on move analysis of RAs. Focusing on the moves in RA introductions, numerous studies have reported the

applicability of the CARS model across a wide range of academic disciplines (e.g., Anthony, 1999; Lim, 2012; Ozturk, 2007).

Because the CARS model was specifically developed for move analysis of RA introductions, such introductions were the focus of early studies. Due to its practicality and utility, however, the move analysis exemplified in the CARS model has been extended and applied to other sections of RAs, which follow the IMRD/C (Introduction, Methods, Results, Discussion, Conclusion) format (e.g., Basturkmen, 2009; Holmes, 1997; Lim, 2006; Lin & Evans, 2012), including abstracts (Hyland, 2000; Martín, 2003; Pho, 2008).

Furthermore, this line of research has expanded to encompass the whole body (i.e., all sections) of RAs since Nwogu's study (1997), in which the rhetorical moves and their constituent elements (i.e., steps) of medical research papers were examined by applying move analysis. Similar undertakings have been carried out for RAs in various fields (Kanoksilapatham, 2005; Shi & Wannaruk, 2014; Stoller & Robinson, 2013; Wannaruk & Amnuai, 2015).

These past studies have provided EAP researchers and practitioners with a wealth of information about how the rhetorical and organizational structures of RAs are constructed within a specific discipline. As Hyland (2013) notes, "this kind of patterning has yielded useful information about the ways texts are constructed and the rhetorical contexts in which such patterns are used, as well as providing valuable input for genre-based teaching" (p. 2). As such, move analysis has been widely adopted and ingrained in EAP teaching practice (Cotos, 2014).

## 2.2 Corpus Use in Genre-based Writing Pedagogy

With advances in computer technology in the latter half of the twentieth century, corpus use has become indispensable to move analysis, and consequently, it has further advanced the field of genre research and pedagogy (Paltridge, 2014; Upton & Cohen, 2009). There are two corpus-based approaches to the analysis and teaching of discourse structure: top-down and bottom-up (Biber, Connor, & Upton, 2007). The top-down approach focuses on macro-textual features such as moves. Thus, the analytical framework for the particular communicative functions (i.e., discourse units) is prepared first, and the analysis is then conducted based on the predetermined discourse unit types. For example, Cotos, Link, and Huffman (2016) used top-down corpus analysis for the moves in discussion and conclusion sections across disciplines, and introduced a pedagogical model that could be applied to materials and tasks in a postgraduate writing course.

On the other hand, the bottom-up approach focuses on linguistic features such as lexico-grammatical patterns. That is, the corpus investigation of vocabulary or grammar and grouping of its functions in texts comes first, and the discourse unit types are then identified based on linguistic criteria or groupings (e.g., Biber, Connor, Upton, & Jones, 2007). Previously, the efficiency and practicality of the bottom-up approach adopted by corpus linguistics was questioned because it focused too much on examining lexico-grammatical patterns with very little attention paid to wider discourse levels (Flowerdew, 2002; Swales, 2002).

Over the years, this research gap, whereby wider discourse levels have been overlooked, has narrowed substantially (Boulton, Carter-Thomas, & Rowley-Jolivet, 2012), and in genre-based writing instruction, it is now considered to be the combination of a top-down approach to genre analysis and a bottom-up approach to corpus analysis that further helps raise learner awareness of rhetorical functions and linguistic features (Charles, 2007; Cortes, 2007; Cotos, 2009; Flowerdew, 2015b). Accordingly, Henry and Roseberry (2009) argue that the notion of the “move register,” or the list of lexico-grammatical features found in each move, is highly useful from a teaching perspective.

Genre-based pedagogy is centered around rhetorical consciousness-raising (Hyland, 2004; Tardy, 2009), and technology can better address discipline-specificity and individualization of genre-based writing instruction (Cotos, 2014). Thus, in recent years, this line of corpus-based and genre-based writing instruction has been further improved with the aid of technology (Birch-Bécaas & Cooke, 2012; C.-F. Chang & Kuo, 2011; Henry, 2007). For example, Sun (2007) created the web-based Scholarly Writing Template (SWT) to support L2 research writing. In addition to templates for moves and concordance searchers, the SWT provides users with an integrated writing environment, in which they actually draft their research papers while referring to moves and corresponding linguistic features using concordancers. Overall, positive outcomes and learners’ perceptions of consciousness-raising for moves and lexico-grammar features have been reported from use of these technology-enhanced genre pedagogical tools.

While these previous tools have annotated move boundaries manually, Anthony and Lashkia (2003) introduced machine learning techniques to move analysis and developed a software tool, the Mover, that automatically presents learners with the move structures of RA abstracts. The results of their study suggested that the Mover could be used to help users, particularly L2 scientists and engineers, better understand and construct research papers in different fields and disciplines. Another recent development in the automated move-based technologies is the Research Writing Tutor (RWT), developed by Cotos (2014, 2016) and her colleagues (Cotos, Huffman, & Link, 2015). RWT is a pedagogical tool for teaching and learning disciplinary RA writing, which aims to advance genre-based writing instruction to the next level by incorporating an automated writing evaluation program (i.e., RWT identifies move and step boundaries automatically and uses them as criteria for automated feedback). RWT gives several types of feedback on the rhetorical structures of the user's own writing, shows published writing in concordance lines, and provides instructional scaffolds with a short video lecture and an accompanying slideshow presentation on particular rhetorical concepts. With its innovative features utilizing move and step constructs to the maximum, RWT is, without doubt, the current state of the art in technology-enhanced genre writing pedagogy.

### 2.3 Lexical Bundles as Building Blocks of Discourse

Both the Mover and RWT use n-grams, which are contiguous sequences of words (or sequences of n lexical items), as a means to identify move and step boundaries. For example, Cotos,

Huffman, and Link (2015) retrieved high probability n-grams representative of each move and step in various disciplines and used them for the RWT technology. This suggests that n-grams can be used to identify linguistic realizations of moves and steps, as they can detect commonly occurring keywords, phrases, and discourse markers (Anthony & Lashkia, 2003).

Among the n-grams, those that appear more frequently (in a range of texts) than expected by chance are referred to as lexical bundles (Biber, Johansson, Leech, Conrad, & Finegan, 1999). Lexical bundles are the most frequent recurrent sequences of a fixed number of words (multi-word units). As such, they can be extracted from large corpora based on their frequency, in the form of n-grams (e.g., three-word sequences: 3-grams; four-word sequences: 4-grams). In studies of lexical bundles, 4-grams are typically used because, compared with other n-grams, they have a variety of structures and functions for analysis (Cortes, 2004; Hyland, 2008).

It should also be noted here that, while 4-grams that reach a particular frequency threshold in a corpus are likely to correlate with psycholinguistically-real units (e.g., Tremblay, Derwing, Libben, & Westbury, 2011), the precise nature of their storage or processing has not been established (see Myles & Cordier, 2016; Siyanova-Chanturia & Martinez, 2015 for detailed discussion). For this reason, lexical bundles should not be confused with psychologically valid units (i.e., formulaic sequences), as they are a corpus linguistic construct.

A great deal of research on lexical bundles indicates that, as “important building blocks of discourse in spoken and written registers” (Biber & Barbieri, 2007, p. 263), lexical bundles represent

characteristics of different registers and genres (Biber et al., 2004, 1999; Conrad & Biber, 2004). Thus, Cortes (2013, p. 36) has recently pointed out that moves and lexical bundles have more similarities than differences, in that both comprise “building blocks” of discourse. This notion has led Cortes (2013) to introduce an approach that combines moves and lexical bundles. Cortes first extracted lexical bundles in a bottom-up manner and classified them according to moves and steps in RA introductions. Although a similar approach has been suggested and conducted recently (Durrant, 2015; Durrant & Mathews-Aydinli, 2011; Flowerdew, 2009; Le & Harrington, 2015), Cortes’s study is the first to explicitly integrate moves with lexical bundles in RAs. Using a one-million word corpus of RA introductions from 13 disciplines (1,372 papers in total), Cortes demonstrated that, with this approach, it is possible to investigate the characteristics of lexical bundles in the moves and sections of RA introductions.

#### 2.4 Rationale for Developing an RA Writing Support Tool

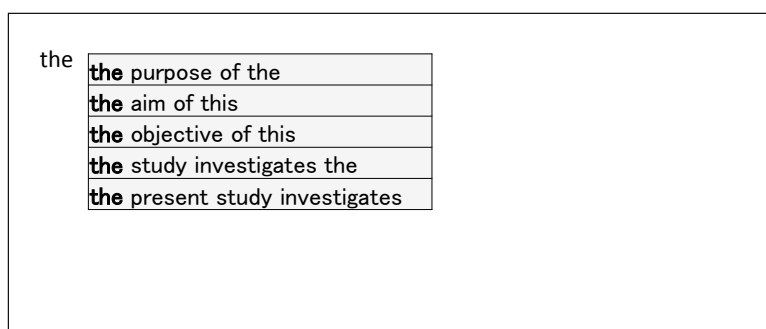
As reviewed above, there exists a robust association between moves and lexical bundles as building blocks used by the discourse community, and integrating these two constructs has gained considerable attention in recent years. Thus, while other approaches exist, we chose to use the bundle-move connection approach (Cortes, 2013) for their common characteristics as building blocks of discourse, as outlined in the literature review. To further explore this new approach and apply it to technology-based genre writing pedagogy, in this study, we aimed to develop a web-based writing

support tool for research articles, which suggests the most frequently-used lexical bundles in a move within all RA sections.

The new RA writing support tool provides immediate suggestions for the most frequent lexical bundles in a specific move with an auto-complete feature in a text area, as commonly implemented in Internet search engines such as Google auto-complete, that enables users to choose appropriate lexical bundles to achieve a specific discourse purpose. For example, if a user types the word “the” in a certain move in the abstract, as it is typed in a text area, the tool will suggest the most frequent lexical bundles conventionally used in the abstract, starting with “the.” These might include, for example, “the purpose of the,” “the aim of this,” and “the objective of this,” as shown in *Figure 1* below. It is this very feature that differentiates our new tool from other existing writing support tools, and we believe that this tool will make a unique and valuable contribution to the development of technology-enhanced genre L2 writing pedagogy for the following two reasons.

First, a writing support feature with auto-complete user interface while writing has yet to be implemented in technology-enhanced genre pedagogical tools. One writing support tool, WriteAway (<http://writeaway.nlpweb.org>), for example, provides lexical and grammatical suggestions with phrase examples based on an academic corpus while typing a certain word. The rationale behind this design is that “learner writers sorely need concrete writing suggestions, right in the context of writing. Learners could be more effectively assisted, if CALL tools provide such suggestions as learners write away” (J. Chang & Chang, 2015, p. 106). Precisely the same reasoning applies to our new

pedagogical tool, which is designed to facilitate the use and learning of move-specific bundles in the RA genre. With its integration of moves and lexical bundles and its intuitive auto-complete feature, the new tool may reduce the cognitive and metacognitive load involved in writing and be used to complement current genre-based writing pedagogy by providing support right at the time when it is most needed.



*Figure 1.* A prototype image of the writing support tool and its auto-complete feature. As a user types in a word, it auto-suggests the most frequent lexical bundles in a specific move, which start with the typed word.

Second, the advantage of a writing support feature with the auto-complete user interface is that move-specific lexical bundles are much more easily accessible and retrievable than any other materials. Retrieving an effective lexical bundle in a certain move/step is often time-consuming and frustrating in that it is simply difficult for users to spot the most appropriate phrase on a list or in a book such as the Academic Phrasebank (<http://www.phrasebank.manchester.ac.uk/>). The same is true for teachers. For example, referring to lists of formulaic sequences developed with a

corpus-driven approach, Vincent (2013) argues:

Studies aiming to identify the most common phrases in academic English naturally tend to provide long lists of formulaic phrases extracted from corpora representing academic English. Such lists . . . may not, however, be of immediate use to the EAP practitioner who wants to locate potentially useful phrases in a specific text. (p. 44)

On the other hand, with a writing support tool that provides auto-complete suggestions for move-specific lexical bundles, the users can locate lexical bundles relevant to their immediate needs without difficulty, which has the potential to broaden the scope of data-driven learning (DDL) research by introducing an alternative approach to a conventional DDL approach (i.e., referring to concordance lines). DDL refers to direct applications of corpora in which learners themselves acquire hands-on experience of using a corpus for learning purposes, often with guided tasks or materials (see Mizumoto & Chujo, 2016 for different types of DDL). Its overall effectiveness (i.e., resulting in positive outcomes) for language learning has been reported in meta-analyses (Boulton & Cobb, 2017; Mizumoto & Chujo, 2015). In L2 academic writing studies, corpus consultation as a reference source has been found effective, among other reported benefits (Yoon, 2011), in correcting writing errors (Gaskell & Cobb, 2004) and raising awareness about the writing conventions of a specific discourse community (J.-Y. Chang, 2014; Friginal, 2013), particularly by using a specialized RA corpus (Lee & Swales, 2006).

As noticing is a necessary condition for DDL to work (Flowerdew, 2015a), the degree of noticing for discipline-specific linguistic and discourse patterns may be higher; this presumably results in better understanding of building blocks, if learners can directly access the most frequent move-specific lexical bundles on the spot when required. For this reason, the new RA writing support tool could serve as a complement to the provision of a list of common lexical bundles or the use of a conventional DDL approach (i.e., referring to concordance lines) in rhetorical consciousness-raising activities.

## 2.5 Purpose of This Study

The purpose of this study was to develop a data-driven and theory-based practical writing support tool by combining rhetorical moves and lexical bundles. We first put the conceptual framework of this approach to the test by empirically examining the correspondence between all the RA sections, moves, and disciplinary-specific lexical bundles of a large corpus of RAs in one research field: applied linguistics. Based on the proof-of-concept evidence, we then devised a web-based RA writing support tool, which materializes the combination of moves and lexical bundles. The developed tool was then pilot-tested on L2 writers in Japan in order to evaluate its usability, effectiveness, and impact.

### **3. Procedures**

#### **3.1 Corpus**

While the study by Cortes (2013) has certainly ushered in a new research and pedagogical area by proposing an approach to combine moves and lexical bundles, it focused only on Introductions, and as Cortes notes, “(t)he relationship between lexical bundles and moves needs to be further developed not only in Introductions but also in other sections of RAs” (p. 42). Furthermore, Cortes compared lexical bundles in RA introduction moves across 13 disciplines to demonstrate that the concepts of both moves and lexical bundles are pertinent to a wide range of academic disciplines. This comparison is useful for gaining an overview by revealing generic move/step-bound lexical bundles, which may be of use for disciplinarily heterogeneous students in a writing course; however, it goes against the nature of moves and lexical bundles because, as the literature review suggests, different discipline communities require different discourse and language conventions. Thus, in this study, we decided to focus on one specific discipline, applied linguistics, and thoroughly investigate the moves and lexical bundles within that discipline as a first step toward incorporating various academic disciplines. We chose applied linguistics as IMRD move models have been suggested for this discipline in previous studies (e.g., Pho, 2013; Wannaruk & Amnuai, 2015) as well as due to the authors’ interests and areas of competence.

To create a corpus, we selected articles from 10 high-profile peer-reviewed international journals in the field of applied linguistics (numbers in parentheses indicate the publication year):

*Applied Linguistics* (2006–2015), *ELT Journal* (2007–2015), *English for Specific Purposes* (2007–2015), *Journal of Second Language Writing* (2008–2015), *Language Learning* (2010–2015), *Language Teaching Research* (2010–2015), *Modern Language Journal* (2012–2015), *Studies in Second Language Acquisition* (2010–2015), *System* (2013–2015), and *TESOL Quarterly* (2008–2015). As genre is known to evolve over time (Guinda, 2015), we collected the latest RAs from the most recent issues of the journals in 2015 and then moved backwards to older issues, until we had accumulated 100 articles from each journal (1,000 RAs in total). The articles were chosen based on the criteria that (a) they were empirical/experimental studies (i.e., other document types such as review articles and book reviews were not included) and (b) they followed the IMRD/C (Introduction, Methods, Results, Discussion, Conclusion) format with abstracts at the beginning of the article. Information other than the text (i.e., tables, table titles, figures, figure legends and captions, references, appendices, biographical information, and author affiliations) was not included in the corpus. The size of the RA corpus we compiled was far larger (token = 8,152,477), as a corpus of a single academic discipline, than those of similar previous studies. It was therefore considered sufficiently large to extract lexical bundles from the moves within each RA section.

### 3.2 Tagging Moves

The corpus of 1,000 applied linguistics RAs was tagged based on the moves in all sections (i.e., Abstract, Introduction, Methods, Results, Discussion, and Conclusion). To tag the moves, we used

the move framework for applied linguistics RAs developed by Pho (2013). We chose this framework because it is extensively based on previous move analysis studies, and covers all RA sections and moves/steps within them, focusing on applied linguistics. As comprehensive as Pho's move framework was, however, it seemed necessary to modify the classification in two RA sections (see the asterisks in Table 1). First, Pho's framework included only two moves ("Describing the data and data collection procedure" and "Describing the data analysis procedure") in the Method section, while previous RA move analysis studies have conventionally had more than two moves for this section (e.g., Cotos, Huffman, & Link, 2017; Kanoksilapatham, 2005; Wannaruk & Amnuai, 2015). Thus, in our study, we divided the Method section into four moves: (1) Describing the sample, (2) Describing research instruments, (3) Describing the procedures, and (4) Describing the data analysis procedure. Second, although Pho's move framework (2013) includes the moves in the Conclusion section as part of the Discussion-Conclusions category, it is known that RAs in applied linguistics have, to a certain degree, independent Conclusions sections, and sometimes include Pedagogical Implications as well (Yang & Allison, 2003). At the same time, an RA may end with the Discussion section, encompassing rhetorical structures equivalent to the Conclusion section without any section headings. Considering the diversity of closing sections (i.e., those sections following the Results sections in RAs) and the possibility of tagging moves in both the Discussion and Conclusion sections, we deemed it best to create an independent Conclusion section and its corresponding moves. Accordingly, we added three moves under the Conclusion section by referring to Yang and Allison

(2003): (1) Summarizing the study, (2) Evaluating the study, and (3) Deductions from the research.

Table 1 shows the final move framework used to tag moves in the current study after making the above modifications. In total, 25 moves were included in six potential RA sections.

Following this coding scheme of RA moves, the first author of this article created a move tagging sample file. Eight research assistants, who were graduate students of foreign language education and one of whom is the second author of this article, began tagging RAs individually (i.e., one person tagged one article) after a training session. In tagging, they used an Excel spreadsheet in which move tags were embedded. The files were later converted into XML format. When faced with a problem in tagging, particularly with unclear move categories, the research assistant conferred with the second author of the article and any ambiguity was resolved through discussion. Care was taken to avoid any overlapping move tags.

After tagging 1,000 RAs, the first author of this article randomly selected 50 articles and tagged them once more. The interrater agreement between the two coders was 91%. In addition to its large corpus size, the applied linguistics RA corpus used in this study is arguably the largest corpus to date with move tags.

Table 1

## Move Framework Used for Tagging Moves (\* indicates added in the current study)

Sections	Moves	Steps
Abstract	[01] Introduction (Establishes context of the paper)	<ul style="list-style-type: none"> <li>• Arguing for topic prominence (Claiming centrality)</li> <li>• Making topic generalizations</li> <li>• Defining terms, objects, or processes</li> <li>• Identifying a gap in current knowledge</li> <li>• Justifying the research study</li> </ul>
	[02] Presenting the research	<ul style="list-style-type: none"> <li>• Stating the purpose directly</li> </ul>
	[03] Describing the methodology	<ul style="list-style-type: none"> <li>• Describing the participants</li> <li>• Describing the instruments or equipment</li> <li>• Describing the procedure and conditions</li> </ul>
	[04] Summarizing the findings	<ul style="list-style-type: none"> <li>• Describing the main features or properties of the solution or product</li> </ul>
	[05] Discussing the research (Interprets or extends results beyond the scope of the paper, draws inferences, points to applications, or wider applications.)	<ul style="list-style-type: none"> <li>• Deducing conclusions from results</li> <li>• Evaluating value of the research</li> <li>• Presenting recommendations</li> </ul>
Introduction	[06] Establishing a territory (Announcing the importance of the field)	<ul style="list-style-type: none"> <li>• Claiming the centrality of the topic</li> <li>• Making topic generalizations</li> <li>• Summarizing existing studies (Reviewing items of previous research)</li> <li>• Drawing inferences from previous studies</li> <li>• Reference to main research problems</li> </ul>
	[07] Establishing a niche (Preparing for the present study)	<ul style="list-style-type: none"> <li>• Indicating a gap</li> <li>• Adding to what is known</li> <li>• Presenting positive justification</li> <li>• Raising a question</li> </ul>
	[08] Presenting the present work (Introducing the present study)	<ul style="list-style-type: none"> <li>• Announcing present research descriptively and/or purposively</li> <li>• Stating purpose(s)</li> <li>• Presenting research questions or hypotheses</li> <li>• Definitional clarifications</li> <li>• Reference to main research procedure (Summarizing methods)</li> <li>• Predicting results (Announcing principal outcomes)</li> <li>• Stating the value of the present research</li> <li>• Indicating RA (Research Article) structure</li> </ul>
	[09*] Describing the sample [10*] Describing research instruments [11] Describing the procedures [12] Describing data analysis procedure	
Results (or Results & Discussion)	[13] Preparing for the presentation of results	<ul style="list-style-type: none"> <li>• (Re)stating data collection and analysis procedure</li> <li>• Restating research questions or hypotheses</li> </ul>
	[14] Reporting specific / individual results	<ul style="list-style-type: none"> <li>• Location of results</li> <li>• Reporting most important findings</li> <li>• Substantiating (or invalidating) results</li> <li>• Indicating non-consistent observations</li> </ul>
	[15] Commenting on results	<ul style="list-style-type: none"> <li>• Interpreting results</li> </ul>
	[16] Summarizing results	<ul style="list-style-type: none"> <li>• Presenting integrated results on the basis of a number of specific results</li> </ul>
Discussion (or Conclusion)	[17] Preparing for the presentation of the discussion section	<ul style="list-style-type: none"> <li>• Giving background information (Restate the aims, objectives, procedural information, theories, and research questions)</li> </ul>
	[18] Highlighting overall research outcome	<ul style="list-style-type: none"> <li>• Reporting results (Expected or unexpected outcome)</li> </ul>
	[19] Discussing the findings of the study	<ul style="list-style-type: none"> <li>• Interpreting / discussing results</li> <li>• Indicating significance of the outcome</li> <li>• Comparing results with a hypothesis</li> <li>• Comparing results with literature</li> <li>• Exemplifying</li> </ul>
	[20] Drawing conclusions of the study Stating research conclusions	
	[21] Evaluating the study	<ul style="list-style-type: none"> <li>• Indicating significance / advantage</li> <li>• Indicating limitations</li> <li>• Evaluating methodology</li> </ul>
	[22] Deductions from the research	<ul style="list-style-type: none"> <li>• Recommending further research</li> <li>• Making suggestions / drawing (pedagogic) implications</li> </ul>
	[23] Summarizing the study	<ul style="list-style-type: none"> <li>• Providing summary</li> </ul>
Conclusion (or Pedagogical implications)	[24*] Evaluating the study	<ul style="list-style-type: none"> <li>• Indicating significance / advantage</li> <li>• Indicating limitations</li> <li>• Evaluating methodology</li> </ul>
	[25*] Deductions from the research	<ul style="list-style-type: none"> <li>• Recommending further research</li> <li>• Making suggestions / drawing (pedagogic) implications</li> </ul>

### 3.3 Examining the Correspondence Between Moves and Lexical Bundles

Most previous studies combining moves and lexical features have used predetermined groupings of organizational features without questioning the adequacy of such groupings (Cortes, 2013; Flowerdew, 2015b; Le & Harrington, 2015). Because the categorization of organizational features (i.e., rhetorical moves in this study) has significant consequences for the subsequent analyses (Durrant, 2017), it should be addressed in a systematic and objective manner. For this reason, we examined the correspondence between moves and lexical bundles to ensure that the move scheme employed was appropriate for extracting lexical bundles within each move and adequate for use in the development of a writing support tool.

Using the tagged corpus, move-specific lexical bundles were extracted with CasualConc version 2.0.2 (Imao, 2016). CasualConc is a concordance program for Mac OS. It can generate KWIC (keyword in context) concordance lines, word clusters, collocation analysis, and word count. We focused on 4-grams in the current study because they have been the most used bundle type in previous studies, and, as such, their structures and functions are comparable across studies. Consequently, a list of 4-grams and their relative frequency, adjusted by the number of words in each move, was obtained for all 25 move categories from the 1,000 RAs in the field of applied linguistics. We then performed correspondence analysis using a matrix of the 250 most frequent lexical bundles  $\times$  25 moves (see Appendix S1 in the Supporting Information online). Correspondence analysis is a

statistical technique designed to explore the associations in the data included in a frequency table (also referred to as a contingency table, a cross-tabulation, or a two-way table). With correspondence analysis, the relationships between rows and columns (in the case of this study, the 250 most frequent lexical bundles and 25 moves) can be simultaneously examined in reduced latent dimensions (Hair, Black, Babin, Anderson, & Tatham, 2006). As the outcome from correspondence analysis is graphically displayed in a two-dimensional map in which proximity represents similarity, it is possible to visually and intuitively interpret the associations between the row and column categories (Baayen, 2008). We thus used a graphical representation of the results of correspondence analysis to check the matching of moves and lexical bundles. It should be noted here that the number of lexical bundles included in the analysis (i.e., the 250 most frequent lexical bundles) was decided because, in order to better examine the profiles of moves and lexical bundles, we chose to have a larger matrix, which was 10 times larger than the 25 move categories. At the same time, we confirmed, in an exploratory way, that the results of correspondence analysis, or the relative positions of moves in the two-dimensional plot, did not change drastically when a smaller number of lexical bundles were used (e.g., 200, 150, 100, and 50 most frequent bundles).

R version 3.3.2 (R Core Team, 2016) was used to conduct the analysis. In order to ensure reproducibility and transparency in the data analysis (Larson-Hall & Plonsky, 2015; Marsden, Mackey, & Plonsky, 2016), all the data and R codes used in this study are shared on the IRIS repository (<https://www.iris-database.org/iris/app/home/detail?id=york:932233>).

### 3.4 Pilot Testing

After confirming the appropriateness of the move scheme and retrieved lexical bundles, we developed a web-based RA writing support tool. The developed tool was then pilot-tested with eight L2 writers in Japan: six undergraduate students majoring in English who used the tool to write their Bachelor's theses in applied linguistics and two graduate students writing their M.A. theses in applied linguistics. They participated in this small pilot study on a voluntary basis.

The participants' proficiency, based on their self-report of several proficiency test scores, ranged from B2 (Independent User) to C1 (Proficient User) in the Common European Framework of Reference (CEFR) for Languages: learning, teaching, assessment (Council of Europe, 2001). Given their intermediate level of proficiency in English, the participants were considered appropriate as a target population sample (i.e., novice or non-native English researchers). None of them had written an RA prior to their thesis writing. As their theses followed the Abstract and IMRD/C patterns, and consulting the developed tool as a reference resource would greatly help them with their linguistic difficulties, the participants' needs matched the purpose of the current pilot testing. No training was given to the participants except for explaining the concept of the tool.

The six undergraduates (5 females, 1 male) provided feedback on an open-ended response sheet after using the developed tool for two months. The two graduate students (both females) were asked, for five months, to keep an online log of the search terms, open-ended comments, and

feedback each time they used the tool while writing their manuscripts.

## 4. Results

### 4.1 Move and Lexical Bundle Correspondence

The results of correspondence analysis using a matrix of the 250 most frequent lexical bundles  $\times$  25 moves are visually displayed in two dimensions (*Figure 2*). In this figure, 250 lexical bundles and 25 moves are placed according to their values for two dimensions obtained from correspondence analysis (as lexical bundles overlap and their visibility is low in *Figure 2*, see Appendix S2 in the Supporting Information online for dimension scores). In the figure, moves close to each other are similar (i.e., the correlation is stronger), while those distant from each other are dissimilar (i.e., the correlation is weaker). This also applies to lexical bundles. Furthermore, associations between moves and lexical bundles can be simultaneously inspected, as their relative proximity represents stronger associations between them. That is, lexical bundles close to a certain move indicate that they are highly associated with that specific move.

The figure reveals that lexical bundles that are often used to realize the intended rhetorical purposes of a given move are placed close to that move. For example, in the Abstract section, Move 02: “Presenting the research” and Move 01: “Introduction” are distributed on the right-hand side of the figure (i.e., positive values on Dimension 1), with lexical bundles, such as *this study investigated the*, *this study examined the*, *investigated the effects of*, *in a second language*, *the purpose of this*, and

*little is known about*, that are more highly associated with these moves. This is undoubtedly because these bundles are vital for setting the research scene for research papers in applied linguistics. The left-hand side of the figure (i.e., negative values in Dimension 2) is occupied by moves other than the first three moves of the Abstract (the moves placed in the right-hand side of the figure) and three moves in the Introduction (the center of the figure). This implies that the first dimension is composed of the first three moves in the Abstract and others, indicating the strong influence of the first three moves in the Abstract on the relationships between the moves and lexical bundles.



In Dimension 2 (i.e., the vertical axis), two moves in the Abstract (Move 05: “Discussing the research” and Move 01: “Summarizing the findings”) and moves in the Discussion and Conclusion sections lie from the top to the center of the figure (i.e., positive values on Dimension 2), whereas moves in the Method and Results sections are plotted from the bottom to the center of the figure (i.e., negative values on Dimension 2). It can be speculated from these results that Dimension 2 reflects the differences in the interpretive and descriptive natures of various sections in RAs in applied linguistics. The positioning of those moves is also reflected in the distribution of lexical bundles. Bundles such as *the article concludes with*, *discussed in terms of*, and *implications of these findings*, which are necessary to discuss the findings (in the Abstract), are placed at the top of the figure, while bundles at the bottom half of the figure include *ranged from # to*, *participants were asked to*, and *at the time of*, which are necessary for presenting empirical or experimental procedures.

Many of the most frequent lexical bundles can be observed in the middle, left-hand side of the figure, where the Introduction, Discussion, and Conclusion sections gather. This indicates that moves in these sections are realized using a wide range and variety of lexical bundles; although some appear more likely in a specific move, these are ubiquitous bundles, which are used extensively across different moves in different sections.

While the lexical bundles characteristic of moves in the Abstract and Method obviously account for the profiling of moves and lexical bundles in *Figure 2*, it is worth noting that the relative data points of the moves in different sections of applied linguistics RAs do not change even if (a) a

smaller number of lexical bundles are included (e.g., 200, 150, 100, and 50 most frequent bundles) in the correspondence analysis or (b) moves in the Abstract section are deleted in the correspondence analysis. In addition, the effect size (the measure of the strength of association) was large, Cramer's  $V$  [95%CI] = 0.29 [0.28, 0.30], considering the size of the matrix (Volker, 2006).

These results confirm that the association of lexical bundles and moves is robust, and it is possible to let the Abstract and IMRD/C patterns emerge themselves as a result, which in turn suggests that the conceptual framework of bundle-move connection, the move scheme, and the coding of moves in this study were all empirically tenable. As the current approach proved effective, we took a step further, realizing the concept of bundle-move connection with a web-based RA writing support tool.

#### 4.2 Developing the Web-based RA Writing Support Tool

Based on the results of the proof-of-concept, we developed a web-based RA writing support tool. *Figure 3* displays a screenshot of the online application “AWSuM: Academic Word Suggestion Machine.” It is accessible online free of charge (<http://langtest.jp/awsum/>).

We designed the support tool to allow us to fully exploit the potential of the association of lexical bundles and moves. First, users can choose a discipline, section, and move within each section from the top pull-down menu (see *Figure 3*). The feature allowing selection of a discipline was added for later use, when other disciplines are included. The ability to select a section and move

are an integral part of this tool because users need to pay attention to the correspondence between moves and lexical bundles. The option “ALL” is also available for all these selections, allowing users to expand the search area. For these options, both top-down and bottom-up searchers are available for users, depending on their immediate needs. On the right edge of the screen, the most frequently selected lexical bundles are displayed in frequency order, in accordance with the selected section and move, so users can gain hints about what lexical bundles are often used in a specific section or move.

Then, as a user types a word in the box, the tool dynamically and instantaneously suggests a list of the most frequent bundles, using PHP and MySQL in the web server. The keyword function specifies the reference words on the left (1 to 5 words). For example, if a user wants to know what lexical bundles are used after “the,” the keyword is 1; likewise, “the purpose” is 2, “the purpose of” is 3, “the purpose of this” is 4, and “the purpose of this study” is 5. In this way, the more that users increase the number of reference words for the phrase for which they wish to acquire suggestions, the more the context will be limited, which allows users to check in more detail what kind of words follow. Users can also set the number of n-grams (3R, 4R, or 5R) to be auto-suggested to the right of the keywords. This function allows users to be presented with lexical bundles of different numbers of words.

The tool auto-suggests not only what Cortes (2013) termed “trigger” bundles, which are very easy to identify and connect to a move and even to a step in a move, but also “non-trigger” bundles

(i.e., called ubiquitous bundles in this study; see the results of move and lexical bundle correspondence above). Non-trigger bundles have very important discursal functions to which L2 writers need to be exposed for later use in their own writing.

**Discipline**  
Applied Linguistics

**Section**  
abstract

**Move**  
02\_presenting\_research

**Keyword**  
1 2 3 4 5 3R 4R 5R ☒ Auto-suggest

of this \* was

- of this \* was to investigate the
- of this \* was to explore the
- of this \* was to determine whether
- of this \* was to examine the
- of this \* was to examine whether
- of this \* was to validate the
- of this \* was to explore in
- of this \* was to identify and
- of this \* was to investigate a
- of this \* was the point at

off 1L 2L

the aim

the objective

major purpose

the acquisition

(s) used for \*

this study was

on this research was

of this investigation was

of this feature was

**Concordance**

"of this \* was" Search

The aim	<b>of this study was</b>	therefore to investigate the effectiveness of a
The primary aim	<b>of this study was</b>	to assess the impact of a curricular intervent
The goal	<b>of this study was</b>	to determine the overall effects of pronunciat
The purpose	<b>of this study was</b>	to determine whether it is possible to disting
The goal	<b>of this study was</b>	to determine whether learner beliefs regardin

**Google search**

"of this \* was" Search

**Most frequent 4-grams in abstract 02\_presenting\_research**

- as a foreign language
- this article reports on
- as a second language
- english as a second
- english as a foreign
- the extent to which
- this study investigated the
- this study examines the
- of english as a
- article reports on a
- this study investigates the
- this study examined the
- the purpose of this
- reports on a study
- of this study was
- learners of english as
- this study was to
- this article explores the
- of this article is
- this article is to
- this article examines the
- investigated the effects of
- this paper reports on

Figure 3. A screenshot of the web-based RA writing support tool “AWSuM: Academic Word Suggestion Machine.”

It is obvious that these non-trigger bundles are not as salient as trigger bundles. As a result, they will pose a major challenge if L2 writers cannot use them at their disposal in their writing. For this reason, the tool auto-suggests both non-trigger and trigger bundles in order to further support L2 writers.

In addition to these features that utilize the bundle-move connection while writing RAs, we have implemented other functions to allow the tool to be used for DDL activities and as a useful reference resource both inside and outside the classroom. First, it is possible to display 1 to 3 high frequency words (1L-3L) that come to the left of the searched keyword. This feature enables users to examine collocations. Second, users can perform wildcard searches to search for strings containing any word in a particular position. It is thus possible to search what words are more likely to be used in a position marked by asterisks in a search string such as “it is \* to \* that.”

Furthermore, users can refer to the concordance lines (keyword in context or KWIC) to investigate how a word or phrase is used within the original articles. We also incorporated the Google search feature, with which users can search words and phrases using the Google Custom Search Engine (see Geluso, 2013 for details of Google as a corpus and concordancer). With this custom search, the specified domain is higher educational institutions in countries where English is the first language, which means it can be used to confirm words and phrases from academic texts other than those covered by the move-tagged corpus included in the developed tool.

As described above, the newly developed web-based RA writing support tool features a combination of lexical bundles and moves. It is an innovative, interactive tool with an auto-complete

user interface. Aside from being a technology-enhanced genre pedagogical tool, it is also equipped with other features, which make it an ideal online reference resource both inside and outside the classroom. We have also prepared a user's manual, which provides complete details about the features implemented in the new tool (see Appendix S3 in the Supporting Information online).

#### 4.3 User Feedback

After the tool was developed, it was piloted with eight L2 writers for two to five months. The open-ended comments from all eight participants at the end of pilot testing came to a total of 284 entries, which were manually analyzed and classified for thematic categories. This analysis was conducted by the first and second authors of this article, according to the following categories formulated from the literature on evaluating learners' responses to DDL tools (Mizumoto, Chujo, & Yokota, 2016; Yoon, 2011): (1) Noticing of lexico-grammatical patterns (Discovery), (2) Confirmation of lexico-grammatical patterns (Reference), (3) Benefits unique to this tool, (4) Greater confidence and autonomy in L2 RA writing, and (5) Suggestions for improvement. Using the Kappa coefficient, the inter-rater reliability for the coding was confirmed to be 0.83 (a very good level of agreement). Where disagreement occurred, a decision was made by consensus.

Those five categories, their frequency, and the excerpts characteristic of each category are provided in Table 2. The open-ended user comments suggest that the tool was useful for discovering and confirming lexico-grammatical patterns in RA moves in a specific discipline. In addition, the

benefits unique to this tool were also reported, which most likely led to the development of greater confidence and autonomy in the users' L2 RA writing. These findings suggest the beneficial effects of the web-based RA writing support tool, which we set out to achieve in developing it. At the same time, users' feedback pointed to the difficulties they faced in making the most of using the tool (see "Suggestions for improvement" in Table 2). These difficulties will be addressed in our future studies.

Taken together, although limited in terms of the number of participants, this pilot study provided valuable information about users' perceptions of the usability, effectiveness, and impact of the new tool and about how we could further improve it.

Table 2

User Feedback on the Newly Developed Tool (Categories, Frequency, and Excerpts)

(1) Noticing of lexico-grammatical patterns (Discovery)	67 / 284 entries (23.6%)
<ul style="list-style-type: none"> <li>- I was able to learn the appropriate verb for a given noun. [e.g., excerpt # =&gt; excerpt # <b>illustrates</b>]</li> <li>- I noticed a number of verbs suitable for a certain context. [e.g., these studies have =&gt; these studies have <b>shown</b>]</li> <li>- I was able to learn the appropriate preposition for a given noun. [e.g., the Cronbach's alpha =&gt; the Cronbach's alpha <b>for</b>]</li> <li>- I was able to learn the specific nouns following a given adjective by setting the left-hand range. [e.g., has been a controversial =&gt; has been a controversial <b>topic</b>.]</li> <li>- I was able to learn how technical terms are used. [e.g., Bonferroni =&gt; Bonferroni multiple comparison]</li> </ul>	
(2) Confirmation of lexico-grammatical patterns (Reference)	67 / 284 entries (23.6%)
<ul style="list-style-type: none"> <li>- I was able to confirm that the phrase I came up with or wanted to use is frequently used. [e.g., procedure was followed]</li> <li>- I was able to confirm phrases for which I was unsure of their usage. [e.g., is defined <b>as</b>]</li> <li>- I was able to check if a phrase I had used in a different genre could be used in research articles. [e.g., the reason why]</li> <li>- I was able to check how an unfamiliar word like a synonym is used in research articles.</li> <li>- I was able to check how sentences could be constructed by viewing more suggestions.</li> </ul>	
(3) Benefits unique to this tool	58 / 284 entries (20.4%)
<ul style="list-style-type: none"> <li>- Setting the section and the move helped me to find more appropriate phrases in the context. [e.g., university students <b>who were majoring</b>]</li> <li>- When I did not know how to begin a sentence, I looked at the "most frequent n-grams" in the move.</li> <li>- I came across a more appropriate phrase in the "most-frequent n-grams" when I set the section and move and tried to search for a different phrase.</li> <li>- This tool suggests frequently-used phrases for a given move, which made me realize how important those phrases are.</li> <li>- By fixing the left-hand range, I was able to find a more appropriate phrase in the context.</li> </ul>	
(4) Greater confidence and autonomy in L2 RA writing	39 / 284 entries (13.7%)
<ul style="list-style-type: none"> <li>- I can write my paper with confidence because the tool suggests appropriate academic phrases.</li> <li>- I can write my paper with a more professional tone, compared with writing without any help.</li> <li>- As I use the tool, I may be able to learn many structural patterns and use them in my writing.</li> <li>- The more I used the tool, the more smoothly I was able to write. I want to keep using the tool.</li> <li>- I was able to feel assured that the phrases I had never used in my writing were actually used in other published papers.</li> </ul>	
(5) Suggestions for improvement	53 / 284 entries (18.7%)
<ul style="list-style-type: none"> <li>- I felt I was not able to search for phrases well. It may take time to get used to using the tool adequately.</li> <li>- It would be better if more example sentences were suggested because I did not know how to begin a sentence.</li> <li>- Lexical bundles should be arranged and suggested according to their meanings.</li> <li>- It would be better if a dictionary or thesaurus were included in the tool.</li> <li>- While the concordance lines were useful, I was not sure which phrases would be most appropriate.</li> </ul>	

## 5. Discussion and Conclusions

Corpus-based approaches to the analysis and teaching of discourse structure have increasingly become the established norm in language teaching and learning in the 21st century. One of the most advanced applications of such corpus-based approaches is their use in technology-enhanced genre L2 writing pedagogical tools (Cotos, 2014). Against this background, the current study was designed to further the genre-based technology by developing a data-driven and theory-based practical RA writing support tool, which combines rhetorical moves and lexical bundles. The results of the study suggest that the bundle-move connection approach (Cortes, 2013) was sufficiently robust to implement in the web-based RA writing support tool. We then translated our findings to the design and development of a pedagogical tool. Positive preliminary user feedback on its usefulness was obtained in the pilot study. Given the saying “the proof of the pudding is in the eating,” it was thus confirmed that the developed online RA writing support tool may play a facilitating role in technology-enhanced genre writing teaching and learning.

With regard to the tool we developed, writing specialists may argue that providing lexical bundles “on-the-go” interferes with the composition process and is rather prescriptive. With regard to the first concern, the auto-suggest feature can be turned off by removing the check mark. As for the second concern (i.e., its prescriptive nature), this criticism has often been directed at DDL as well. In the case of DDL, it is claimed that consulting concordance lines increase learners’ consciousness of

descriptive rather than prescriptive language (Chambers, 2005) and helps them develop their own descriptive frameworks (Gavioli & Aston, 2001). The participants in the current pilot study seemed to have gained similar perspectives in that they cited noticing and confirmation of lexico-grammatical patterns (almost half of the open-ended comments) as empowering features of the developed tool. Thus, noticing, hypothesis formulation, and hypothesis testing are attainable with the tool, in the same way as conventional DDL approaches offer. Therefore, although it may seem prescriptive, the tool can greatly scaffold L2 writers who have difficulty acquiring knowledge of multi-word units (Boers et al., 2016), let alone how to use them appropriately. It does this by suggesting the most frequent lexical bundles in a given move, thus catering to the user's immediate needs. In this sense, the tool converts many of the advantages of linguistic approaches to genre-based pedagogy (Hyland, 2004) into tangible forms; it is needs-based, supportive, empowering, and consciousness-raising. From these perspectives, we feel confident that the developed tool will comprise a very practical, promising means for preparing L2 writers to enter the discourse community.

Let us now discuss the strengths and limitations of the bundle-move connection approach as well as the strengths and weaknesses of the developed tool. First, the strength of the bundle-move connection approach is obviously its high level of salience. As echoed in the EAP literature (e.g., Charles, 2007; Flowerdew, 2015b; Henry & Roseberry, 2009), the combination of a top-down and bottom-up approach is what makes genre-based pedagogy more successful in terms of raising

learners' awareness of rhetorical patterns and conventions. Accordingly, the bundle-move connection approach explicitly focuses on one of the most prevalent linguistic features, lexical bundles. With this approach, learners can see how communicative purposes are realized using lexical bundles, which result in better learning of these bundles in genre-based writing pedagogy.

The limitation of the bundle-move connection approach, on the other hand, is that there are also other linguistic features, such as metadiscourse (Hyland & Tse, 2004), in which L2 writers need to become proficient in order to make a persuasive argument. As such, it should be kept in mind that the bundle-move connection approach is not a panacea for all aspects of L2 academic writing fluency.

Next, the strength of the developed tool is that it can be used in both a top-down and bottom-up approach. In the top-down approach, users pay more attention to rhetorical functions, while in the bottom-up approach, they can focus on lexical bundles and other linguistic features. Specifically, in a top-down approach, users specify the section and move within which to inspect the use of lexical bundles. However, as users do not always need to specify sections or moves, in the bottom-up approach, the focus can be primarily on form. Furthermore, the auto-complete user interface of the newly developed tool can free learners from the burden of going through the concordance lines while drafting a manuscript. It is worth noting that, when necessary, examining the concordance lines is of course possible using the tool's concordance feature.

The weakness of the tool is that, despite its level of sophistication and convenience, it does not

help users to put their thoughts together or show them how to begin a sentence. This limitation also applies to all other existing tools (e.g., Birch-Bécaas & Cooke, 2012). Of course, the most frequent lexical bundles are suggested in a list form (on the right hand of the screen) as one of the tool's features. However, they are most useful for RA sections/moves with the distinct use of lexical bundles or “trigger” bundles (Cortes, 2013), such as Abstract and Method sections (as can be clearly seen in *Figure 2*). For this reason, the list of lexical bundles shown in the tool may not be helpful in presenting an extensive discussion in the Introduction or Discussion sections. That is because those sections are highly argumentative, but ubiquitous bundles, which are used extensively across different moves in different sections, such as *on the other hand*, *in the case of*, *as well as the*, *in terms of the*, *on the basis of*, *in the context of* occupy the top of the lexical bundles list. As a few participants in the pilot study made the similar comment, “I did not know how to begin a sentence,” it would be desirable to provide explicit teaching sessions (e.g., Cotos et al., 2016; Flowerdew, 2015b) in addition to training for successful use of the tool. In fact, the top-down and deductive teaching of genre and the provision of a training session to use the tool is crucial, in that the DDL literature has repeatedly pointed out that training is a prerequisite for students to benefit from corpus consultation (Boulton, 2010). In this sense, we agree with Conrad (2000) that corpus use in the classroom should be incorporated into existing pedagogy with the realization that the tools would not be fully utilized if presented in isolation.

The frequency-based suggestion of lexical bundles is another limitation of the tool. The tool

auto-suggests the most frequent lexical bundles in a certain move in an RA, but, as some researchers have pointed out, retrieving bundles based on frequency alone could be improved by utilizing other indices or methods (e.g., Simpson-Vlach & Ellis, 2010; Vincent, 2013). We will therefore try to improve the usability of the online writing support tool to appeal to as many researchers and practitioners as possible who are interested in making use of genre-based move-bundles in their research and practice.

Further studies will be necessary to test the utility of the newly developed web-based RA writing support tool, which has the potential to broaden the existing agenda of EAP practitioners. In particular, we could conduct a quasi-experimental study with a pre/post-test design (along with a delayed test) by assigning control/experimental groups, tracking tool uses, and analyzing writing to see if users do, in fact, use the predicted bundles more frequently in the relevant moves and use them appropriately in an overall more stylistically appropriate end product.

Cortes (2013, p. 42), for example, suggests a DDL activity, in which academic writing instructors introduce a list of frequent lexical bundles in a particular move or step, and students directly consult the corpus for the use of lexical bundles in a specific move in their discipline. Instead of providing a list of common lexical bundles, with this web application, teachers and learners can directly access move-specific bundles, presumably resulting in better noticing and understanding of the building blocks often used in a specific discourse community. The effectiveness of the tool could then be tested in a large-scale DDL study in which the results can be compared with those of

previous studies on teaching lexical bundles (Cortes, 2006).

Another possible research venue for testing the usefulness of this tool would be in a study wherein L2 writers test the tool on an existing paper or section of a paper that they would like to improve, using bundles that frequently connect with moves or steps. The changes in writing quality could then be analyzed from various perspectives.

Future studies could also explore two aspects that were not addressed in the current study. First, although we intentionally did not tag steps in moves due to the very large corpus size, it may be possible to investigate the bundle-move connection in more detail by including steps in the analysis. Second, this study, again intentionally, focused on just one discipline (i.e., applied linguistics), but future studies can be conducted that target multiple academic fields to compare disciplinary differences in bundle-move connection. To tag moves and steps from different disciplinary fields, an automated tagging system is currently under development. As the main value of the bundle-move connection approach can be found more in its practical application, we are also considering the possibility of using the move-step framework of the Research Writing Tutor (Cotos, 2014; Cotos et al., 2015); this program has an established move-step framework for 30 different disciplines and automated tagging system for these moves and steps, thus allowing us to avoid reinventing the wheel.

As it stands, the developed web-based RA writing support tool, featuring the combination of lexical bundles and moves, can assist researchers and genre-based EAP practitioners in need of an

alternative RA writing support/teaching tool. As such, it could potentially contribute to the growing trend of technology-enhanced genre writing pedagogy. With its auto-complete feature, the tool will be appealing to researchers and practitioners in the modern, technologically advanced RA writing arena. By the same token, although the tool has been designed for genre-based classroom consciousness-raising classroom activities, it can also be used by novice or non-native English researchers as a beacon that guides them through the long and arduous journey of RA writing outside the classroom.

### **Supporting Information**

Additional Supporting Information may be found in the online version of this article at the publisher's website:

Appendix S1. A Matrix of the 250 Most Frequent Lexical Bundles and 25 moves.

Appendix S2. Result of the Correspondence Analysis.

Appendix S3. AWSuM User's Manual.

## References

- Ackermann, K., & Chen, Y.-H. H. (2013). Developing the Academic Collocation List (ACL) - A corpus-driven and expert-judged approach. *Journal of English for Academic Purposes*, 12, 235–247. doi:10.1016/j.jeap.2013.08.002
- Anthony, L. (1999). Writing research article introductions in software engineering: How accurate is a standard model? *IEEE Transactions on Professional Communication*, 42, 38–46. doi:10.1109/47.749366
- Anthony, L., & Lashkia, G. V. (2003). Mover: A machine learning tool to assist in the reading and writing of technical papers. *IEEE Transactions on Professional Communication*, 46, 185–193. doi:10.1109/TPC.2003.816789
- Baayen, R. H. (2008). *Analyzing linguistic data: A practical introduction to statistics using R*. Cambridge, England: Cambridge University Press.
- Basturkmen, H. (2009). Commenting on results in published research articles and masters dissertations in language teaching. *Journal of English for Academic Purposes*, 8, 241–251. doi:10.1016/j.jeap.2009.07.001
- Bhatia, V. K. (1993). *Analysing genre: Language use in professional settings*. London, England: Longman.
- Biber, D., & Barbieri, F. (2007). Lexical bundles in university spoken and written registers. *English for Specific Purposes*, 26, 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. doi:10.1075/scl.28
- Biber, D., Connor, U., Upton, T. A., & Jones, J. K. (2007). Vocabulary-based discourse units in biology research articles. In D. Biber, U. Connor, & T. A. Upton (Eds.), *Discourse on the move: Using corpus analysis to describe discourse structure* (pp. 175–212). Amsterdam, the Netherlands: John Benjamins. doi:10.1075/scl.28.10jon
- Biber, D., Connor, U., Upton, T. A., & Kanoksilapatham, B. (2007). Introduction to move analysis. In D. Biber, U. Connor, & T. A. Upton (Eds.), *Discourse on the move: Using corpus analysis to describe discourse structure* (pp. 23–41). Amsterdam, the Netherlands: John Benjamins.
- Biber, D., & Conrad, S. (2009). *Register, genre, and style*. Cambridge, England: Cambridge University Press.
- Biber, D., Conrad, S., & Cortes, V. (2004). If you look at. . . : Lexical bundles in university teaching and textbooks. *Applied Linguistics*, 25, 371–405. doi:10.1093/applin/25.3.371
- Biber, D., Johansson, S., Leech, G., Conrad, S., & Finegan, E. (1999). *Longman grammar of spoken and written English*. London, England: Longman.
- Birch-Bécaas, S., & Cooke, R. (2012). Raising collective awareness of rhetorical strategies: Using an online writing tool to demonstrate discourse moves in the ESP classroom. In A. Boulton, S. Carter-Thomas, & E. Rowley-Jolivet (Eds.), *Corpus-informed research and learning in ESP: Issues and applications* (pp. 239–260). Amsterdam/Philadelphia: John Benjamins.

doi:10.1075/scl.52.10bir

Boers, F., Demecheleer, M., He, L., Deconinck, J., Stengers, H., & Eyckmans, J. (2016).

Typographic enhancement of multiword units in second language text. *International Journal of Applied Linguistics*. Advance online publication. doi:10.1111/ijal.12141

Boulton, A. (2010). Data-driven learning: Taking the computer out of the equation. *Language Learning*, 60, 534–572. doi:10.1111/j.1467-9922.2010.00566.x

Boulton, A., Carter-Thomas, S., & Rowley-Jolivet, E. (Eds.). (2012). *Corpus-informed research and learning in ESP: Issues and applications*. Amsterdam, the Netherlands: John Benjamins.

doi:10.1075/scl.52

Boulton, A., & Cobb, T. (2017). Corpus use in language learning: A meta-analysis. *Language Learning*. Advance online publication. doi:10.1111/lang.12224

Chambers, A. (2005). Integrating corpus consultation in language studies. *Language Learning & Technology*, 9, 111–125. Retrieved from <http://llt.msu.edu/vol9num2/chambers/>

Chang, C.-F., & Kuo, C.-H. (2011). A corpus-based approach to online materials development for writing research articles. *English for Specific Purposes*, 30, 222–234.

doi:10.1016/j.esp.2011.04.001

Chang, J.-Y. (2014). The use of general and specialized corpora as reference sources for academic English writing: A case study. *ReCALL*, 26, 243–259. doi:10.1017/S0958344014000056

Chang, J., & Chang, J. S. (2015). WriteAhead2: Mining lexical grammar patterns for assisted writing.

*Proceedings of NAACL-HLT 2015*, 106–110. Retrieved from

<http://www.aclweb.org/anthology/N15-3022>

Charles, M. (2007). Reconciling top-down and bottom-up approaches to graduate writing: Using a corpus to teach rhetorical functions. *Journal of English for Academic Purposes*, 6, 289–302.

doi:10.1016/j.jeap.2007.09.009

Cheng, A. (2015). Genre analysis as a pre-instructional, instructional, and teacher development framework. *Journal of English for Academic Purposes*, 19, 125–136.

doi:10.1016/j.jeap.2015.04.004

Conrad, S. (2000). Will corpus linguistics revolutionize grammar teaching in the 21st century?

*TESOL Quarterly*, 34, 548–560. doi:10.2307/3587743

Conrad, S., & Biber, D. (2004). The frequency and use of lexical bundles in conversation and academic prose. *Lexicographica: International Annual for Lexicography*, 20, 56–71.

doi:10.1515/9783484604674.56

Cortes, V. (2004). Lexical bundles in published and student disciplinary writing: Examples from history and biology. *English for Specific Purposes*, 23, 397–423. doi:10.1016/j.esp.2003.12.001

Cortes, V. (2006). Teaching lexical bundles in the disciplines: An example from a writing intensive history class. *Linguistics and Education*, 17, 391–406. doi:10.1016/j.linged.2007.02.001

Cortes, V. (2007). Exploring genre and corpora in the English for academic writing class. *The ORTESOL Journal*, 25, 8–14.

- 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, 33–43.  
doi:10.1016/j.jeap.2012.11.002
- Cotos, E. (2009). Designing an intelligent discourse evaluation tool: Theoretical, empirical, and technological considerations. In C. A. Chapelle, G. H. Jun, & I. Katz (Eds.), *Developing and evaluating language learning materials* (pp. 103–127). Ames, IA: Iowa State University.  
Retrieved from [https://www.apling.engl.iastate.edu/tsll/2008/12 IADE Elena.pdf](https://www.apling.engl.iastate.edu/tsll/2008/12%20IADE%20Elena.pdf)
- Cotos, E. (2014). *Genre-based automated writing evaluation for L2 research writing: From design to evaluation and enhancement*. New York, NY: Palgrave Macmillan.
- Cotos, E. (2016). Computer-assisted research writing in the disciplines. In S. A. Crossley & D. S. McNamara (Eds.), *Adaptive educational technologies for literacy instruction* (pp. 225–242). New York, NY: Routledge/Taylor & Francis.
- 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: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:10.1016/j.esp.2017.01.001
- Cotos, E., Link, S., & Huffman, S. R. (2016). Studying disciplinary corpora to teach the craft of

Discussion. *Writing & Pedagogy*, 8, 33–64. doi:10.1558/wap.v8i1.27661

Council of Europe. (2001). *Common European Framework of Reference for Languages: Learning, teaching, assessment*. Cambridge, England: Cambridge University Press.

Coxhead, A. (2000). A new academic word list. *TESOL Quarterly*, 34, 213–238.

doi:10.2307/3587951

Durrant, P. (2009). Investigating the viability of a collocation list for students of English for academic purposes. *English for Specific Purposes*, 28, 157–169. doi:10.1016/j.esp.2009.02.002

Durrant, P. (2017). Lexical bundles and disciplinary variation in university students' writing:

Mapping the territories. *Applied Linguistics*, 38, 165–193. doi:10.1093/applin/amv011

Durrant, P., Mathews-Aydinli, J., & Mathews-Aydinli, J. (2011). A function-first approach to identifying formulaic language in academic writing. *English for Specific Purposes*, 30, 58–72.

doi:10.1016/j.esp.2010.05.002

Flowerdew, L. (2002). Corpus-based analyses in EAP. In J. Flowerdew (Ed.), *Academic discourse* (pp. 95–114). London, England: Longman.

Flowerdew, L. (2009). Applying corpus linguistics to pedagogy: A critical evaluation. *International Journal of Corpus Linguistics*, 14, 393–417. doi:10.1075/ijcl.14.3.05flo

Flowerdew, L. (2015a). Data-driven learning and language learning theories: Whither the twain shall meet. In A. Leńko-Szymańska & A. Boulton (Eds.), *Multiple affordances of language corpora for data-driven learning* (pp. 15–36). Amsterdam/Philadelphia: John Benjamins.

doi:10.1075/scl.69.02flo

Flowerdew, L. (2015b). Using corpus-based research and online academic corpora to inform writing of the discussion section of a thesis. *Journal of English for Academic Purposes*, 20, 58–68.

doi:10.1016/j.jeap.2015.06.001

Friginal, E. (2013). Developing research report writing skills using corpora. *English for Specific Purposes*, 32, 208–220. doi:10.1016/j.esp.2013.06.001

Gardner, D., & Davies, M. (2014). A new academic vocabulary list. *Applied Linguistics*, 35, 305–327. doi:10.1093/applin/amt015

Gaskell, D., & Cobb, T. (2004). Can learners use concordance feedback for writing errors? *System*, 32, 301–319. doi:10.1016/j.system.2004.04.001

Gavioli, L., & Aston, G. (2001). Enriching reality: Language corpora in language pedagogy. *ELT Journal*, 55, 238–246. doi:10.1093/elt/55.3.238

Geluso, J. (2013). Phraseology and frequency of occurrence on the web: native speakers' perceptions of Google-informed second language writing. *Computer Assisted Language Learning*, 26, 1–14. doi:10.1080/09588221.2011.639786

Grabowski, Ł. (2015). Keywords and lexical bundles within English pharmaceutical discourse: A corpus-driven description. *English for Specific Purposes*, 38, 23–33. doi:10.1016/j.esp.2014.10.004

Guinda, C. S. (2015). Genres on the move: Currency and erosion of the genre moves construct.

*Journal of English for Academic Purposes*, 19, 73–87. doi:10.1016/j.jeap.2015.07.001

Hair, J. F. J., Black, W. C., Babin, B. J., Anderson, R. E., & Tatham, R. L. (2006). *Multivariate data analysis* (6th ed.). Upper Saddle River, NJ: Pearson Prentice Hall.

Hardy, J. A., & Friginal, E. (2016). Genre variation in student writing: A multi-dimensional analysis. *Journal of English for Academic Purposes*, 22, 119–131. doi:10.1016/j.jeap.2016.03.002

Henry, A. (2007). Evaluating language learners' response to web-based, data-driven, genre teaching materials. *English for Specific Purposes*, 26, 462–484. doi:10.1016/j.esp.2007.01.003

Henry, A., & Roseberry, R. L. (2009). Move registers and language teaching. *The Journal of AsiaTEFL*, 6, 101–119.

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, 321–337. doi:10.1016/S0889-4906(96)00038-5

Hyland, K. (2000). *Disciplinary discourses: Social interactions in academic writing*. London, England: Longman.

Hyland, K. (2004). *Genre and second language writing*. Ann Arbor, MI: University of Michigan Press.

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

Hyland, K. (2013). Genre and discourse analysis in language for specific purposes. In C. A. Chapelle

(Ed.), *The Encyclopedia of Applied Linguistics*. Oxford, England: Blackwell.

doi:10.1002/9781405198431.wbeal0452

Hyland, K. (2015). *Academic publishing: Issues and challenges in the construction of knowledge*. Oxford University Press.

Hyland, K. (2016). Methods and methodologies in second language writing research. *System*, 59, 116–125. doi:10.1016/j.system.2016.05.002

Hyland, K., & Tse, P. (2004). Metadiscourse in academic writing: A reappraisal. *Applied Linguistics*, 25, 156–177. doi:10.1093/applin/25.2.156

Hyon, S. (1996). Genre in three traditions: Implications for ESL. *TESOL Quarterly*, 30, 693–722. doi:10.2307/3587930

Imao, Y. (2016). CasualConc (Version 2.0.4) [Computer software]. Retrieved from <https://sites.google.com/site/casualconc/>

Kanoksilapatham, B. (2005). Rhetorical structure of biochemistry research articles. *English for Specific Purposes*, 24, 269–292. doi:10.1016/j.esp.2004.08.003

Larson-Hall, J., & Plonsky, L. (2015). Reporting and interpreting quantitative research findings: What gets reported and recommendations for the field. *Language Learning*, 65, 127–159. doi:10.1111/lang.12115

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. doi:10.1016/j.esp.2015.03.003

Lee, D., & Swales, J. M. (2006). A corpus-based EAP course for NNS doctoral students: Moving from available specialized corpora to self-compiled corpora. *English for Specific Purposes*, 25, 56–75. doi:10.1016/j.esp.2005.02.010

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

Lim, J. M.-H. (2012). How do writers establish research niches? A genre-based investigation into management researchers' rhetorical steps and linguistic mechanisms. *Journal of English for Academic Purposes*, 11, 229–245. doi:10.1016/j.jeap.2012.05.002

Lin, L., & Evans, S. (2012). Structural patterns in empirical research articles: A cross-disciplinary study. *English for Specific Purposes*, 31, 150–160. doi:10.1016/j.esp.2011.10.002

Marsden, E., Mackey, A., & Plonsky, L. (2016). The IRIS Repository: Advancing research practice and methodology. In A. Mackey & E. Marsden (Eds.), *Advancing methodology and practice: The IRIS Repository of Instruments for Research into Second Languages* (pp. 1–21). New York, NY: Routledge.

Martín, P. M. (2003). A genre analysis of English and Spanish research paper abstracts in experimental social sciences. *English for Specific Purposes*, 22, 25–43.  
doi:10.1016/S0889-4906(01)00033-3

Mizumoto, A., & Chujo, K. (2015). A meta-analysis of data-driven learning approach in the Japanese

EFL classroom. *English Corpus Studies*, 22, 1–18. Retrieved from

<http://mizumot.com/files/ecs2015.pdf>

Mizumoto, A., & Chujo, K. (2016). Who is data-driven learning for? Challenging the monolithic view of its relationship with learning styles. *System*, 61, 55–64.

doi:10.1016/j.system.2016.07.010

Mizumoto, A., Chujo, K., & Yokota, K. (2016). Development of a scale to measure learners' perceived preferences and benefits of data-driven learning. *ReCALL*, 28, 227–246.

doi:10.1017/S0958344015000208

Myles, F., & Cordier, C. (2016). Formulaic sequence (FS) cannot be an umbrella term in SLA.

*Studies in Second Language Acquisition*, 12, 1–26. doi:10.1017/S027226311600036X

Nwogu, K. N. (1997). The medical research paper: Structure and functions. *English for Specific Purposes*, 16, 119–138. doi:10.1016/S0889-4906(97)85388-4

Ozturk, I. (2007). The textual organisation of research article introductions in applied linguistics: Variability within a single discipline. *English for Specific Purposes*, 26, 25–38.

doi:10.1016/j.esp.2005.12.003

Paltridge, B. (2014). Genre and second-language academic writing. *Language Teaching*, 47, 303–318. doi:10.1017/S0261444814000068

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,

231–250. doi:10.1177/1461445607087010

Pho, P. D. (2013). *Authorial stance in research articles: Examples from applied linguistics and educational technology*. Basingstoke, England: Palgrave Macmillan.

Polio, C., & Williams, J. (2009). Teaching and testing writing. In M. H. Long & C. J. Doughty (Eds.), *The Handbook of Language Teaching* (pp. 486–517). Oxford, England: John Wiley & Sons.

R Core Team. (2016). R: A language and environment for statistical computing (Version 3.3.2) [Computer software]. Vienna, Austria. Retrieved from <http://www.r-project.org/>

Shi, H., & Wannaruk, A. (2014). Rhetorical structure of research articles in agricultural science. *English Language Teaching*, 7, 1–13. doi:10.5539/elt.v7n8p1

Simpson-Vlach, R., & Ellis, N. C. (2010). An academic formulas list: New methods in phraseology research. *Applied Linguistics*, 31, 487–512. doi:10.1093/applin/amp058

Siyanova-Chanturia, A., & Martinez, R. (2015). The Idiom principle revisited. *Applied Linguistics*, 36, 549–569. doi:10.1093/applin/amt054

Stoller, F. L., & Robinson, M. S. (2013). Chemistry journal articles: An interdisciplinary approach to move analysis with pedagogical aims. *English for Specific Purposes*, 32, 45–57. doi:10.1016/j.esp.2012.09.001

Sun, Y.-C. (2007). Learner perceptions of a concordancing tool for academic writing. *Computer Assisted Language Learning*, 20, 323–343. doi:10.1080/09588220701745791

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

England: Cambridge University Press.

Swales, J. M. (2002). Integrated and fragmented worlds: EAP materials and corpus linguistics. In J.

Flowerdew (Ed.), *Academic discourse* (pp. 150–164). London, England: Longman.

Swales, J. M. (2004). *Research genres: Explorations and applications*. Cambridge, England:

Cambridge University Press.

Tardy, C. M. (2009). *Building genre knowledge*. West Lafayette, IN: Parlor Press.

Tardy, C. M. (2012). Genre-based language teaching. In C. Chapelle (Ed.), *The Encyclopedia of*

*Applied Linguistics* (pp. 10–13). Oxford, England: Blackwell.

doi:10.1002/9781405198431.wbeal0453

Tremblay, A., Derwing, B., Libben, G., & Westbury, C. (2011). Processing advantages of lexical

bundles: Evidence from self-paced reading and sentence recall tasks. *Language Learning*, 61,

569–613. doi:10.1111/j.1467-9922.2010.00622.x

Upton, T. A., & Cohen, M. A. (2009). An approach to corpus-based discourse analysis: The move

analysis as example. *Discourse Studies*, 11, 585–605. doi:10.1177/1461445609341006

Vincent, B. (2013). Investigating academic phraseology through combinations of very frequent

words: A methodological exploration. *Journal of English for Academic Purposes*, 12, 44–56.

doi:10.1016/j.jeap.2012.11.007

Volker, M. (2006). Reporting effect size estimates in school psychology research. *Psychology in the*

*Schools*, 43, 653–672. doi:10.1002/pits

- Wannaruk, A., & Amnuai, W. (2015). A comparison of rhetorical move structure of applied linguistics research articles published in international and national Thai journals. *RELC Journal*, 47, 193–211. doi:10.1177/0033688215609230
- 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, 365–385. doi:10.1016/S0889-4906(02)00026-1
- Yoon, C. (2011). Concordancing in L2 writing class: An overview of research and issues. *Journal of English for Academic Purposes*, 10, 130–139. doi:10.1016/j.jeap.2011.03.003