English for Specific Purposes 58 (2020) 138–154

Contents lists available at ScienceDirect

English for Specific Purposes

journal homepage: http://ees.elsevier.com/esp/default.asp

Promoting discipline-specific genre competence with corpus-based genre analysis activities

Jihua Dong [a,*], Xiaofei Lu [b]

[a] The School of Foreign Languages and Literature, Shandong University, 5 Hongjialou, Jinan, Shandong, 250000, China
[b] Department of Applied Linguistics, The Pennsylvania State University, 234 Sparks Building, University Park, PA, 16802, USA

ARTICLE INFO

Article history:
Available online 4 March 2020

Keywords:
Academic writing
Corpus-based instruction
Data triangulation
Genre-based instruction
Rhetorical structures
English for research publication purposes

1. Introduction

An essential goal of academic writing pedagogy is to help novice academic writers develop sufficient knowledge and skills to cope with the requirements and challenges of the diverse genres of writing that they are expected to engage with. Genre-based writing instruction has proven a highly useful pedagogical approach towards achieving this goal (Bhatia, 1991, 1993; Cortes, 2013; Hyland, 2003, 2007; Paltridge, 2014; Swales, 1990, 2004). The genre-based approach focuses on making explicit the rhetorical structures of texts of specific genres, the communicative purposes of different rhetorical moves within a text, and the linguistic resources used to construct the text (Johns, 1997). A host of empirical studies have implemented this approach in teaching different genres of academic writing (e.g., Cheng, 2007; Henry & Roseberry, 1998; Yayli, 2011) and offered converging evidence that by providing learners with knowledge of "how texts in target genres are structured and why they are written the ways they are" (Hyland, 2003, p. 26), this helps improve both learners' rhetorical awareness and their writing skills.

A critical component of genre knowledge highlighted in genre-based writing instruction is "the specific language associated with each move" (Henry & Roseberry, 1998, p. 155) in texts of specific genres. Intuitively, the systematic identification

* Corresponding author.
E-mail addresses: jdon104@aucklanduni.ac.nz (J. Dong), xxl13@psu.edu (X. Lu).

https://doi.org/10.1016/j.esp.2020.01.005
0889-4906/

and description of the linguistic features associated with different rhetorical moves stand to benefit from corpus linguistic analytical procedures. Indeed, not only has corpus-based instruction that engages learners with various corpus searching and analysis activities been considered to have "unprecedented potential for language learning and teaching" (Liu & Jiang, 2009, p. 68) in general, but it has also been considered to be highly conducive to "rhetorical consciousness-raising" (Lee & Swales, 2006, p. 72). Research into integrating corpus-based and genre-based approaches to writing pedagogy is emerging (e.g., Charles, 2007; Chen & Flowerdew, 2018; Cotos, Link, & Huffman, 2017) but still limited. There is a clear need for more research on how the integration of these two approaches may be effectively implemented in academic writing courses in different contexts.

The current study sets out to explore the potential of integrating corpus-based and genre-based approaches to teaching rhetorical structures in a discipline-specific English as a Foreign Language (EFL) academic writing course. The genre-based approach has been argued to be especially helpful in second language (L2) writing instruction, as L2 learners face additional challenges in developing genre competence (Hyland, 2007). Meanwhile, there is a dearth of research that has applied this approach to help young researchers who are novice EFL academic writers meet their discipline-specific writing needs. The study thus aims to contribute both to the emerging body of research into integrating corpus-based and genre-based approaches to academic writing pedagogy and to addressing the discipline-specific writing needs of novice EFL academic writers.

2. Corpus-based writing instruction

The past two decades have witnessed an increasing number of pedagogical applications of corpora and corpus analytical tools in writing instruction (e.g., Charles, 2007, 2012; Friginal, 2013; Gilmore, 2009; Kennedy & Miceli, 2017; Naismith, 2017). Such applications generally draw upon the constructivist theory of learning, which posits that humans construct knowledge and the meaning of entities through experience (Fosnot, 1996; Piaget, 1977) and frequently take the form of data-driven learning (DDL), an approach that promotes learning through learner exploration of patterns of language use in authentic language data (Boulton & Cobb, 2017; Conrad, 1999).

Previous applications of corpus-based writing instruction have reported various benefits of guided searches for target linguistic items or lexico-grammatical patterns in general-purpose or specialized corpora, followed by learner analyses of their contextual uses in the extracted concordance lines (e.g., Charles, 2012; Gilmore, 2009; Liu & Jiang, 2009). For example, Liu and Jiang (2009) found that using the British National Corpus in this way helped enhance L2 English learners' lexico-grammatical knowledge and their ability to use that knowledge to resolve language issues in their writing assignments. Similarly, Kennedy and Miceli (2017) reported that using the Contemporary Written Italian Corpus in such a way enabled L2 Italian learners to engage in independent language learning and identify useful patterns to refine their autobiographical writing. Poole (2016) used specialized corpora of blog posts and press releases in an undergraduate L2 writing course, and found that guided searches and analyses of keywords helped his students "notice, analyze, interpret and discuss the meaningful and purposeful linguistic and rhetorical variation present in the texts" (p. 99). Some studies integrated corpus compilation as part of the corpus-based writing instruction (Charles, 2012, 2014; Lee & Swales, 2006; Smith, 2011). For example, Smith (2011) reported that his EFL students responded positively to the experience of building and searching their own specialized corpora as part of their final course project. Charles (2012, 2014) incorporated corpus compilation and search activities throughout a discipline-specific EAP course and reported that her students found these activities highly helpful in writing discipline-specific texts. Meanwhile, some studies reported negative learner reactions to this approach, due often to inadequate training or guidance (e.g., Chambers, 2007) or to individual differences in such factors as "academic experience, search purposes, and writing tasks" (Chang, 2014, p. 243). In terms of scope, it has been noted that corpus-based instruction has not yet been widely practiced beyond teaching vocabulary or the lexico-grammar and that it is yet to reach the full "pedagogical landscape" (Römer, 2010, p.18).

While the need to combine corpus-based and genre-based approaches to teaching academic writing has long been noted (Charles, 2007), pedagogical studies that attempt to combine these approaches in practice are limited to date (Chen & Flowerdew, 2018; Cotos et al., 2017; Eriksson, 2012). Charles (2007) recommended reconciling top-down and bottom-up approaches to academic writing instruction and combining interactive genre-based discourse analysis and student corpus-investigation to examine the rhetorical and linguistic construction of texts. In a workshop for ESL doctoral students aimed at promoting their awareness of the rhetorical functions of lexical bundles, Eriksson (2012) asked students to examine lexical bundles specified by the instructor for a certain rhetorical function in a corpus of published texts and use such bundles in their own writing. While the study showcased a potentially useful approach, it did not document the impact of the approach. Chen and Flowerdew (2018) conducted a large-scale project with 473 graduate student writers that included multiple workshops on analyzing rhetorical goals and their linguistic realizations in discipline-specific corpora with concordance tools, followed by a self-directed corpus-building and analysis project that asked the participants to compare their own writing to the expert corpus they compiled. While their participants responded highly positively to this process, the impacts of the approach were again not assessed. Cotos et al. (2017) examined the effect of the use of the Research Writing Tutor, a web-based interface that allowed students to search a corpus annotated with rhetorical moves and that was able to generate rhetorical feedback on the students' writing, on novice writers' genre learning and writing improvement. Their results showed that the interface fostered novice writers' "exploration and application of genre conventions" (p. 104) and enhanced the development of their genre knowledge.

Inspired by and building upon the emerging body of research on combining corpus-based and genre-based approaches to academic writing pedagogy, this study sets out to examine the effectiveness of integrating the two approaches in teaching the rhetorical structures of engineering research article (RA) introductions to EFL engineering graduate students. The focus on a specific discipline and genre is informed by previous findings that the teaching of rhetorical structures is most effective when tailored to a specific target genre and discipline (e.g., Biber & Conrad, 2009; Henry & Roseberry, 1998; Lu, Casal, & Liu, 2020; Lu, Yoon, & Kisselev, 2018). The particular focus on engineering RAs is also informed by the increasing need for young engineering researchers in China to publish in international journals in English, as well as the lack of corpus-based writing instruction research that directly seeks to address this specific need. Drawing upon previous insights into the effectiveness of self-compiled specialized corpora, the instructor and students of the course collaboratively constructed a corpus of engineering RA introductions, which was subsequently annotated for rhetorical moves and steps. We document the implementation of guided corpus-based genre analysis activities and seek to address the following two research questions:

1. Do guided corpus-based genre analysis activities using a self-compiled specialized corpus contribute to improvement of EFL learners' genre knowledge and genre-based writing skills?
2. How do EFL learners perceive their experiences with the corpus-based genre analysis activities?

3. Methodology

3.1. Course and participants

The participants of this study were 30 electrical engineering (EE) or mechanical engineering (ME) master's students enrolled in a discipline-specific academic writing course at a highly ranked research university in China. The main objective of the course was to improve the students' competence in writing RAs in English for publication in peer-reviewed international journals. The course met three times (2 h each time) a week for 20 weeks. Participant age ranged from 21 to 26 (mean = 23.6). All participants had passed the College English Test Band 6 (CET-6)[1] prior to enrolling in the course and could thus be considered to be at an advanced level of English proficiency. Most participants were involved in one or more research projects supervised by their degree advisors while participating in the study. It was also one of their degree requirements to have at least one publication in a peer-reviewed journal. The participants were thus highly motivated to improve their skills in writing RAs in English.

3.2. Instructional design

The study reported here was conducted in the first 10 weeks of the course. This part of the course adopted an integrated corpus- and genre-based approach to teaching the rhetorical structures of engineering RA introductions.[2] The reason for focusing on the introduction section of RAs in the current study was based on the critical role it plays in positioning and framing the research (Ozturk, 2007; Samraj, 2002) and on the availability of a well-established rhetorical structure framework (Swales, 1990; 2004). The core design of this part of the course consisted of the compilation of a specialized corpus of engineering RA introductions and annotation of the corpus with rhetorical moves and steps, followed by a series of guided corpus-based genre analysis activities. Each student also wrote the introduction section of an RA based on a completed or ongoing research project.

3.3. Corpus compilation

In week one, the instructor and students collaboratively compiled a specialized corpus of engineering RA introductions. To ensure the relevance of the samples to the participants and the manageability of the corpus size (Ädel, 2010), each participant contributed the introduction sections of five RAs recommended by his or her degree advisor. Altogether, the students submitted the introduction sections of 150 RAs in electrical and mechanical engineering, with a total of 93,756 tokens.

All RAs were published in leading engineering journals indexed in the Science Citation Index (SCI). A search on Google Scholar showed that they had each been cited 60–300 times, indicating a good level of impact. The overwhelming majority of the RAs were co-authored. We did not use the authors' L1 status as an inclusion criterion in the corpus compilation process, as the rigorous peer review and editing procedures of leading journals could be relied on to ensure that the RAs they publish conform to their rhetorical and linguistic expectations (Lillis & Curry, 2013). Nevertheless, we note that an analysis of the author names, affiliations and bios suggested that 138 (92%) RAs included one or more L1 English authors.

It was encouraging to note that all students managed to obtain their supervisors' support in the RA selection process. In fact, a third of them received a reading list with many more than five English RAs in top-tier international journals from their

[1] CET-6 has been shown to be equivalent to the B1 level in the Common European Framework of Reference for Languages in terms of vocabulary range (Tang, Pritchard, & Shi, 2012).

[2] The other ten weeks of the course focused on the rhetorical structures and linguistic features of other RA sections.

supervisors. This high level of supervisor support indicates the attention paid and the value attached to English RA writing in graduate education at this university, in line with the University's publication requirement for graduate students mentioned in Section 3.1. We return to a discussion of supervisor support in this regard as a potential avenue for collaboration between EAP/ESP instructors and expert informants in Section 5.

3.4. Rhetorical function tagging

To maximize the potential of the corpus for teaching rhetorical structures, we annotated it for rhetorical functions following the Create a Research Space (CARS) model (Swales, 2004), which has been extensively used for analyzing the rhetorical structures of RA introductions (e.g., Del Saz-Rubio, 2011; Ozturk, 2007; Samraj, 2002). The CARS model consists of three rhetorical moves (i.e., sections of text that serve a specific communicative function), each of which in turn has several steps (i.e., smaller components of the communicative function realized by the move), as summarized in Table 1.

Rhetorical function tagging took place in week two. As rhetorical move tagging was a relatively new concept to the students, the instructor first introduced them to the concept and tagset with several tagged samples. She then demonstrated the use of AntMover (Anthony, 2003), a freely available computer program designed to annotate the rhetorical function of each sentence in a text. AntMover comes with the following taxonomy of rhetorical functions: *claiming centrality, making topic generalizations, indicating a gap, announcing present study, announcing principal findings,* and *evaluation of research.* Anthony and Lashikia (2003) evaluated the performance of AntMover for annotating the rhetorical functions of sentences in computer science journal articles, and reported an accuracy of 70% on its best estimates and 90% when the top two best estimates were considered. This level of accuracy suggested that the automatic annotation could serve as a useful starting point. Therefore, following the demonstration, each student first tagged the five samples he or she submitted automatically using AntMover, without much technical difficulty. However, while the taxonomy in AntMover aligns well with the CARS model, it does not cover all the steps in the latter, resulting in some sentences in the corpus being unanalyzed. Subsequently, the students were instructed to manually check, correct and reformat (using the tagset in Table 1) the tags in their five samples and to manually annotate sentences that were left unannotated. The instructor demonstrated the procedure for these tasks. Almost all students encountered some difficulty in determining the precise rhetorical functions of some sentences. The instructor advised them to refer to the tagset and tagged samples for definitions and examples, to consider the top two candidate tags suggested by AntMover, and to consult their peers or the instructor in challenging cases. Next, each student checked the samples tagged by a peer and resolved any discrepancies through discussion. Finally, samples with five or more discrepancies were brought to the instructor, who checked those samples and resolved the discrepancies with the students. While English was the primary language used in the course, Chinese was used as well when confusion arose. In particular, in explaining the concept of rhetorical move tagging and demonstrating the tagging procedures, both English and Chinese were used to ensure that all students clearly understood them. Figure 1 shows a screenshot of the rhetorical move tags produced by AntMover. The tagged files were saved as separate files using the Save function of the File menu of the software, and the tags in the files were subsequently manually checked and edited by the students using the tagset presented in Table 1.

3.5. Guided corpus search

In weeks three to eight, the students engaged in a series of corpus-based genre analysis activities. These activities were informed by Hammond et al.'s (1992) three-phase model of genre-based pedagogy, which starts with modelling, followed by joint negotiation of text by the teacher and learners, ending with learners' independent construction of texts. This model has

Table 1
Rhetorical Moves and Steps in Research Article Introductions (adapted from Swales (2004)).

Move/Step	Tag	Description
Move 1		Establishing a territory
Step 1	M1S1	Claiming centrality
Step 2	M1S2	Making topic generalizations
Step 3	M1S3	Reviewing items of previous research
Move 2		Establishing a niche
Step 1a	M2S1a	Indicating a gap
Step 1b	M2S1b	Adding to what is known
Step 2	M2S2	Presenting positive justification
Move 3		Presenting the present work
Step 1	M3S1	Announcing present research descriptively and/or purposively
Step 2	M3S2	Presenting research questions or hypotheses
Step 3	M3S3	Definitional clarifications
Step 4	M3S4	Summarizing methods
Step 5	M3S5	Announcing principal outcomes
Step 6	M3S6	Stating the value of the present research
Step 7	M3S7	Outlining RA structure

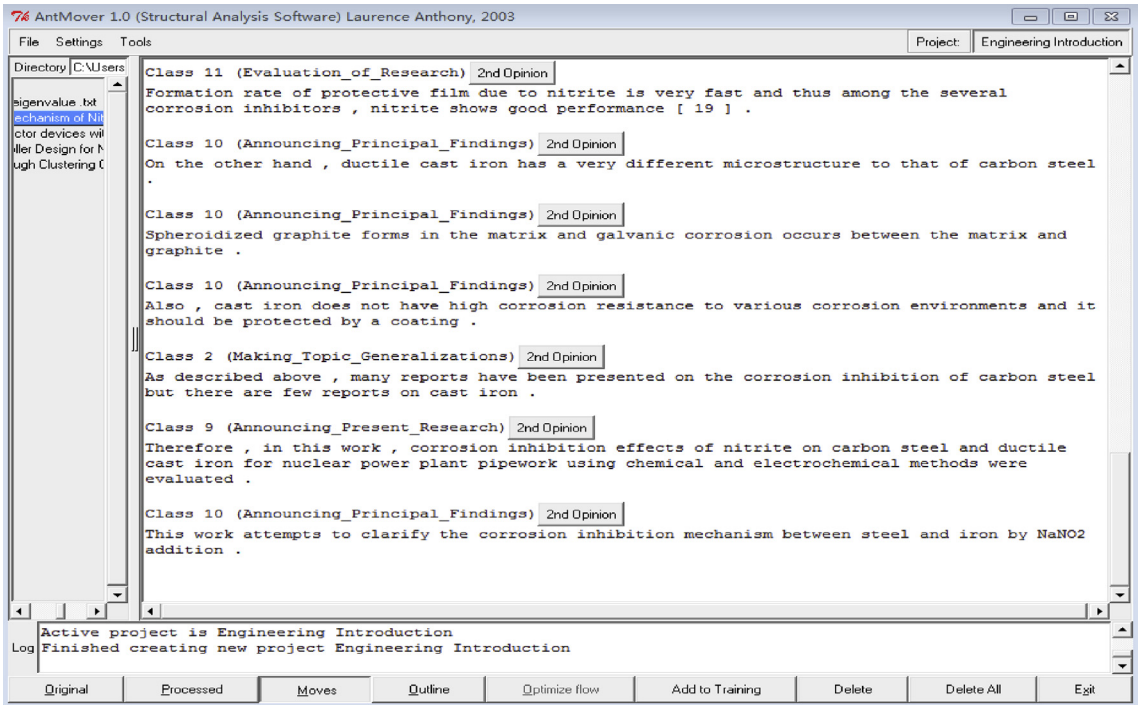

Figure 1. A screenshot of the rhetorical move tags produced by AntMover.

been recommended especially for learners beginning to familiarize themselves with genre practices (Hyland, 2003) and has been empirically shown to enhance learners' genre awareness and genre-specific writing skills (Yayli, 2011).

Following previous calls for detailed teacher guidance in implementing corpus-based instruction (e.g., Pérez-Paredes, Sánchez-Tornel, Calero, & Jiménez, 2011), the instructor offered clear guidance with respect to the focus and procedures of the activities. Based on previous findings on the close connection between certain linguistic features and specific rhetorical moves (e.g., Cortes, 2013; Durrant & Mathews-Aydınlı, 2011), the instructor engaged the students in two types of corpus search and analysis activities with AntConc (Anthony, 2008) (see Appendix A). In the first type, students searched for specific rhetorical moves, e.g., M1S1 (Claiming centrality), and analyzed the concordance lines to identify linguistic features that occurred frequently in those moves. In the second type, students searched for specific linguistic expressions, e.g., *the results indicate* and *tend to*, and analyzed the concordance lines to determine if they tend to associate with particular rhetorical moves. For this latter type, students went back to the expressions they identified in the rhetorical function tagging activities described in Section 3.4; the instructor also recommended a set of expressions that Cortes (2013) found to associate with different rhetorical moves in her analysis of a corpus of RA introductions. The students were instructed to use queries that would cover the inflectional forms of the search terms. For example, *seem* to* was used to ensure that occurrences of *seem to, seems to,* and *seemed to* would all be extracted. These activities were designed with the general goal to help students become "more observant readers of the discoursal conventions" (Swales & Lindemann, 2002, p. 118) and the specific goal to help them notice, understand and acquire the linguistic realizations of different rhetorical moves in RA introductions. Figures 2 and 3 show screenshots of corpus searches by rhetorical move and by linguistic expression, respectively.

3.6. Data collection and analysis

The effectiveness of the corpus-based genre analysis activities and learner perception of these activities was assessed using data triangulation across pre- and post-instruction questionnaires, interviews, students' reflective journals, and student writing. The pre-instruction questionnaire (see Appendix B) was administered in week one to gather information about students' prior knowledge, experience of academic writing and corpus use, and expectations for the course. The post-instruction questionnaire (see Appendix C), adapted from Charles (2012), was administered in week 10 to investigate the students' evaluation of the effectiveness of the corpus-based genre analysis activities and their perception of their experiences with these activities. The two questionnaires were written in both English and Chinese, and the instructor orally clarified any confusion that arose when the students were filling them out. To triangulate the quantitative findings from the questionnaires, we collected reflective journals written by all 30 participants on their evaluation and perception of the course and the corpus activities in weeks three, six, and nine, and conducted semi-structured interviews with six randomly selected participants in week 10. The participants were offered the option to write their journals and be interviewed in either English

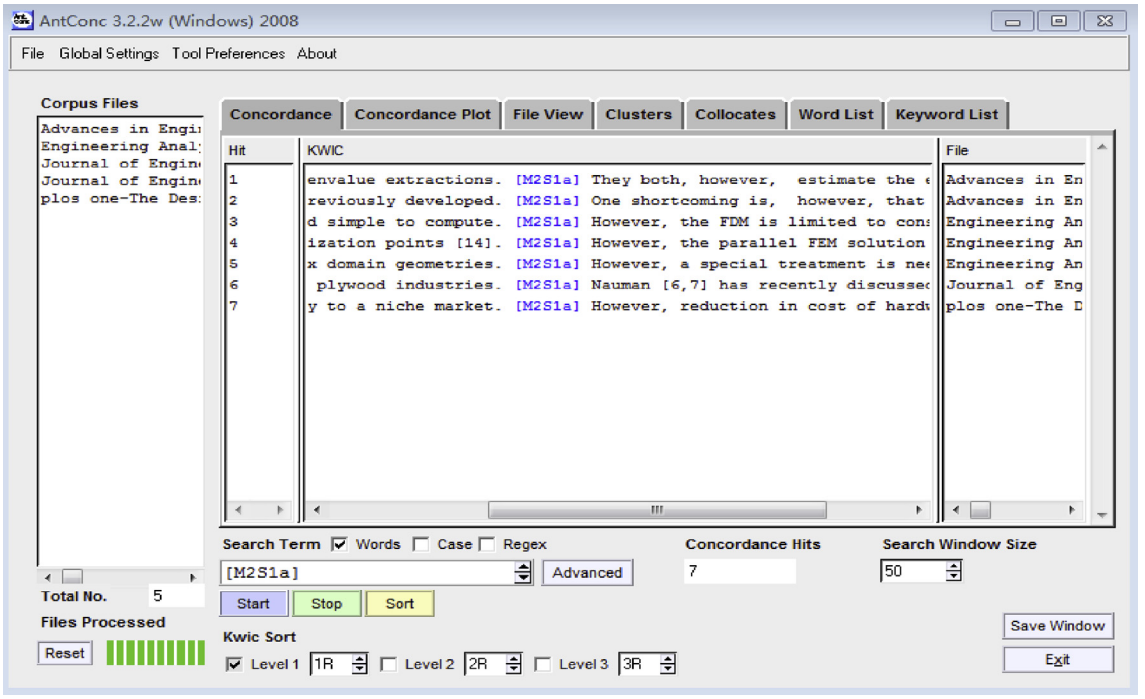

Figure 2. A screenshot of a corpus search by rhetorical move using AntConc.

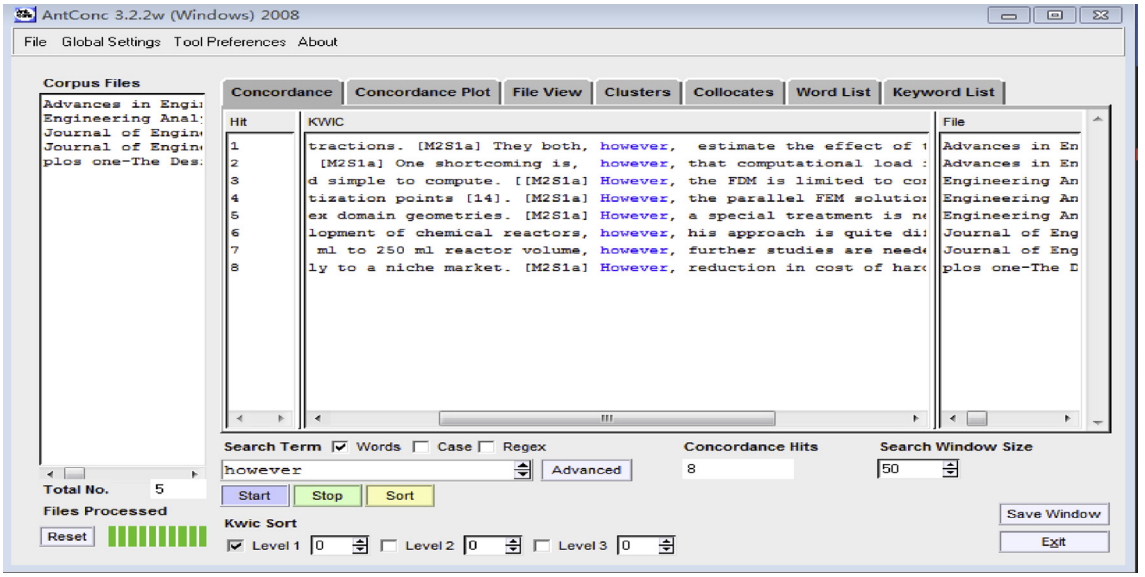

Figure 3. A screenshot of a corpus search by linguistic expression using AntConc.

or Chinese. All participants chose to write the reflective journals in Chinese, with the exception of two, who wrote them in English. All six interviewees chose to be interviewed in Chinese. The interviews were audio-recorded and subsequently transcribed. Finally, each participant was required to independently write in English and submit the introduction section of an RA based on a completed or ongoing project in week 10. The RA introductions were collected and analyzed to complement self-report data on the effectiveness of the corpus-based genre analysis activities. Kanoksilapatham (2015) noted that RA introductions across different engineering sub-disciplines may exhibit substantial variation in the use of rhetorical moves. We therefore investigated whether such variation existed in the rhetorical structures of the RA introductions produced by students majoring in the two different engineering sub-disciplines involved in the study. In addition, to obtain a sense of the

extent to which student-produced RA introductions conform to conventionalized disciplinary genre practices, we compared the distribution of rhetorical moves in student- and expert-produced RA introductions within the same sub-discipline. Given the small number of student-produced RA introductions in the two sub-disciplines, however, our comparison was descriptive only.

When coding the interview transcripts and journal entries for students' comments on and perceptions of the corpus-based genre analysis activities, we first read a number of samples to identify an initial set of themes and then refined and categorized the themes as we went through each sample line by line. The themes identified through this procedure included *benefits in terms of rhetorical structure understanding, positive features of the corpus activities, feedback from academic advisors, positive experience with corpus use, creative uses of the corpus,* and *problems and challenges*. Based on these themes, we used the MAXQDA software to code the samples and analyze the coded data.

4. Results and discussion

This section reports and discusses the results pertaining to the two research questions. To address research question one, we examine quantitative results from the pre- and post-instruction questionnaires and qualitative results from the reflective journals, interviews and writing samples to understand the effectiveness of the corpus-based genre analysis activities in promoting learners' genre knowledge and genre-based writing skills. To address research question two, we examine quantitative results from the post-instruction questionnaire and qualitative results from the reflective journals and interviews to understand the students' perceptions of their experiences with the corpus-based genre analysis activities.

4.1. Effects on students' genre knowledge and genre-based writing skills

4.1.1. Results from the questionnaires

The students' responses to the pre-instruction questionnaire indicated that they had very limited knowledge of the rhetorical structures and linguistic features of RA introductions and little experience of corpus use prior to the course. Very few students reported familiarity with writing RAs in English (6.7% or two students selected "Somewhat agree") or understanding of the rhetorical structures (6.7% or two students selected "Somewhat agree") or linguistic features of RA introductions (3.3% or one student selected "Somewhat agree"). No student reported prior experience with corpus use. Only one student had previously written RA introductions in English.

Table 2 summarizes the results from the post-instruction questionnaire pertaining to the participants' evaluation of the effectiveness of the course's focus on the rhetorical structures of RA introductions (Items 1–3) and the corpus-based genre analysis activities (Items 5–10) in promoting their genre knowledge and genre-based writing skills. With respect to the course in general, the overwhelming majority of the students agreed (either somewhat or strongly) that the course's rhetorical function tagging activities helped improve their RA writing skills (96.7%, 29 students) and enhanced their knowledge of the rhetorical structures (93.3%, 28 students) and linguistic features of RA introductions (90%, 27 students). These positive evaluations of the effectiveness of genre-based instruction echo Yayli's (2011) findings on the positive role it plays in enhancing learners' genre awareness and genre-specific writing skills.

Similarly, most students agreed (either somewhat or strongly) that the corpus activities helped improve their RA writing skills (93.3%, 28 students), enhanced their knowledge of the rhetorical structures (90%, 27 students) and linguistic features of RA introductions (86.7%, 26 students), and helped them generalize the patterns of linguistic features (73.3%, 22 students) and master the use of specific linguistic expressions in RA introductions (73.3%, 22 students). These evaluations not only confirm previous findings on the usefulness of corpus-based writing instruction in general (Charles, 2012; Naismith, 2017) but more importantly highlight the effectiveness of integrating it into genre-based writing instruction.

In response to the open-ended item on other benefits of the academic writing course (Item 4), 90% (27) of the students stated that the course enabled them to read RAs more efficiently. One student noted the following in his free-text response:

Table 2
Participant Evaluation of the Effectiveness of the Course's Rhetorical Function Tagging Activities and Corpus Activities.

Item on the post-instruction questionnaire	Strongly disagree	Somewhat disagree	Neither agree nor disagree	Somewhat agree	Strongly agree
1. Course helped improve RA writing skills	0.0%	0.0%	3.3%	30.0%	66.7%
2. Course enhanced knowledge of the rhetorical structures of RA introductions	0.0%	3.3%	3.3%	40.0%	53.3%
3. Course enhanced knowledge of the linguistic features of RA introductions	0.0%	3.3%	6.7%	40.0%	50.0%
5. Corpus activities helped improve RA writing skills	0.0%	3.3%	3.3%	43.3%	50.0%
6.Corpus activities enhanced knowledge of the rhetorical structures of RA introductions	0.0%	3.3%	6.7%	50.0%	40.0%
7. Corpus activities enhanced knowledge of linguistic features of RA introductions	0.0%	3.3%	3.3%	33.3%	53.3%
8. Corpus activities helped generalize patterns of linguistic features	0.0%	3.3%	23.3%	33.3%	40.0%
9. Corpus activities helped master the use of specific linguistic expressions	3.3%	6.7%	16.7%	30%	43.3%
10. Corpus activities helped understand research topics	3.3%	13.3%	33.3%	26.7%	23.3%

> "I used to read research articles word by word, and often felt frustrated or lost during the reading. From this course, I have learned to focus on the structure of an article and pay attention to different types of information in different sections … I think I now read research articles more purposefully and selectively".

While the relationship between academic writing and reading is beyond the scope of the current study, it is certainly worth exploring how the corpus-based genre analysis approach may enhance both at the same time in future research.

4.1.2. Results from students' reflective journals and post-instruction interviews

4.1.2.1. Benefits in terms of rhetorical structure understanding. The analysis of the students' reflective journals provided additional evidence of the pedagogical effectiveness of the corpus activities. Evidence of an improved understanding of the rhetorical structures of the RA introduction was abundant in the reflective journals, with many comments similar to the following one:

> "Before this course, I had no idea about the organization of the introduction section. Now I understand how to frame different types of information in corresponding rhetorical moves in the introduction."

Many students also noted in the reflective journals that the corpus activities helped improve their understanding of the linguistic features characteristic of different rhetorical moves in RA introductions, as illustrated by the comment below:

> "I have read some English articles prior to this course, but it never occurred to me that there is such a connection between linguistic markers and rhetorical moves. Now I know how to use certain linguistic expressions to frame certain rhetorical moves in the introduction section."

Indeed, the corpus activities revealed strong associations between various linguistic markers and rhetorical moves, such as *plays an important role* and M1S1 (Claiming centrality), and *however, little knowledge* and M2S1a (Indicating a gap), echoing findings from previous studies on linguistic realizations of rhetorical moves (Cortes, 2013; Samraj, 2005).

4.1.2.2. Positive features of the corpus activities. Analysis of the reflective journals and post-instruction interviews also offered insights into the features of the corpus activities that may have contributed to their pedagogical effectiveness. First, many students indicated that the hands-on nature of the corpus search and analysis activities enhanced their level of engagement with the learning process, as illustrated by the following comment in a student's reflective journal:

> "In the corpus activities, I searched the corpus with the assigned query terms, read the concordance lines, and tried to generalize the usage patterns of the linguistic devices."

It was clear that the corpus activities entailed more active involvement of the students in the learning process than passive reception that is typical of textbook- or teacher-centered instruction (Lantolf, 2009). The active interaction with genre knowledge enacted by the corpus activities corroborates the role of corpora in facilitating knowledge construction reported previously (Gardner & Davies, 2007; Lin, 2016).

Second, the majority of the interviewees (5 out of 6) explicitly noted that the inductive learning processes created by the corpus activities helped enhance their understanding of the linguistic patterns associated with different rhetorical moves, as illustrated by the following comments:

> "When searching for a move, I can see all the instances of the same move listed in one window. This allows me to see the common linguistic patterns used to express that move."

> "The concordance lines list all the contexts and moves a linguistic item occurs in. This helps me to see clearly how to use that item".

These and other similar comments suggest that searching the corpus by move and by linguistic item and analyzing the concordance lines retrieved helped them notice and understand the relationship between certain linguistic features and rhetorical moves. These results are in line with Kennedy and Miceli's (2017) finding that reading concordance lines allowed students to generalize patterns and enhanced their use of those patterns. The results also provide additional empirical support for Ballance's (2017) call to encourage beginning or infrequent corpus users to take advantage of the benefits of relevant authentic examples that concordances offer.

One particular advantage of the corpus-based instruction is that the specialized corpus provided pedagogical materials that were directly relevant to the subject content (Cargill, O'Connor, & Li, 2012; Hyland, 2000), exposing students to the rhetorical and linguistic practices of RA introductions within their own sub-disciplines. The corpus search and analysis activities instantiated an inductive learning process that not only allowed students to generalize patterns but also alleviated "the need for teacher mastery of the details of the content" (Cargill et al., 2012, p.67), although the corpus-based course design posed additional demands on instructor expertise and workload compared to traditional course design.

Finally, some students noted that the corpus activities served as a useful way to help consolidate the knowledge they had acquired in class, as illustrated by the following comment in a student's reflective journal:

> "The class demo and my initial observation gave me an idea of the structures of RAs, but I don't think I understood everything completely. The corpus search activities provided many examples of the moves and linguistic patterns discussed in class. Reading through those examples really helped deepen my understanding of them."

4.1.2.3. Feedback from academic advisors. Somewhat unexpectedly, a few students voluntarily noted comments from their academic advisors on their writing, as illustrated by the comment below, although we did not include advisor evaluation of the participants' writing as part of our data collection plan:

> "When I showed my supervisor the introduction section I wrote for a research article, she praised me and said that it was really well-structured".

Favorable responses such as these from the discipline experts can be regarded as further indirect evidence of the success of the corpus-based instruction of genre knowledge. Given the course's goal to help students write more effective RA introductions in specific disciplines, such feedback from discipline experts was heartening for both the students and the instructor and may constitute a productive source of data in future research along similar lines.

4.1.3. Rhetorical moves in student-produced RA introductions

The RA introductions produced by the students were analyzed for rhetorical moves and their linguistic realizations. The 30 samples ranged from 141 words to 776 words, with a total of 11,997 words (mean = 399.9, standard deviation = 139.1). Initially, 30% (i.e., 9) of the samples were independently coded for rhetorical moves by the first author and an experienced academic English teacher at the research site. Inter-annotator reliability, evaluated using Cohen's Kappa, reached .95. All coding discrepancies for these samples were resolved through discussion. The other 70% (i.e., 21) of the samples were then coded for rhetorical moves by the first author only.

Table 3 summarizes the rhetorical moves of the RA introductions produced by student and expert writers in EE and ME. Notably, for each of the three moves, at least one step was used in all samples, and some steps that have been characterized as obligatory were used in the overwhelming majority of the samples. At the same time, one or two specific steps of each move were found in the overwhelming majority of the samples.

For Move 1 (Establishing a territory), M1S1 (Claiming centrality) and M1S3 (Reviewing items of previous research) were found in 90% and 96.7% of the samples respectively, indicating the students' awareness of the importance to claim centrality of their own research topic and to refer back to relevant previous research. As shown in Table 3, expert-produced RA introductions also commonly contained these two steps, albeit not as frequently. In terms of sub-disciplinary variation, a larger proportion of student and expert EE writers used M1S1, while a larger proportion of student and expert ME writers used M1S3.

In Example (1), the EE student establishes a territory by claiming an important role for the technology of localization (the topic of the RA) in wireless sensor networks, using the expression *a significant factor and one of the supporting technologies.* In Example (2), the ME student reviews previous studies on the same topic as the RA, namely, *the dynamics and effects of water droplet energy and impact forces.*

(1) The technology of localization, by which the position of blind node can be acquired, is *a significant factor* and *one of the supporting technologies* in wireless sensor networks. [M1S1, EE]

Table 3
Number and Percentage of RA Introductions Containing Different Moves.

Move/steps	Expert EE (N = 75)		Expert ME (N = 75)		Student EE (N = 15)		Student ME (N = 15)	
	Number	Percentage (%)	Number	Percentage %	Number	Percentage %	Number	Percentage %
Move 1	75	100.00	75	100.00	15	100.00	15	100.00
M1S1	36	48.00	34	45.33	14	93.33	13	86.67
M1S2	5	6.67	13	17.33	5	33.33	12	80.00
M1S3	54	72.00	65	86.67	14	93.33	15	100.00
Move 2	46	61.33	56	74.67	13	86.67	14	93.33
M2S1a	46	61.33	55	73.33	12	80.00	14	93.33
M2S1b	2	2.67	5	6.67	0	0.00	0	0.00
M2S2	3	4.00	2	2.67	4	26.67	3	20.00
Move 3	75	100.00	53	70.67	15	100.00	15	100.00
M3S1	32	42.67	53	70.67	11	73.33	14	93.33
M3S2	3	4.00	3	4.00	9	60.00	11	73.33
M3S3	5	6.67	2	2.67	0	0.00	0	0.00
M3S4	59	78.67	38	50.67	10	66.67	6	40.00
M3S5	5	6.67	3	4.00	0	0.00	0	0.00
M3S6	24	32.00	20	26.67	11	73.33	9	60.00
M3S7	47	62.67	29	38.67	2	13.33	1	6.67

(2) The dynamics and effects of water droplet energy and impact forces have been investigated for a wide range of applications. Dadiao and Wallender (1985) and Stillmunkes and James (1982) found that as droplet kinetic energy increased, soil sealing also increased. [M1S3, ME]

For Move 2 (Establishing a niche), M2S1a (Indicating a gap) was found in 86.7% of the student-produced introductions, higher than expert-produced introductions. More student and expert ME writers used M2S1a than EE writers. In contrast, more student and expert EE writers used M2S2 (Presenting positive justification) than ME writers. M2S1b (Adding to what is known), which was used by a few expert ME (6.67%) and EE (2.67%) writers, was absent in student-produced RA introductions. In Example (3), the ME student employs *however* and *there is little knowledge of* as linguistic markers to indicate a gap in previous research.

(3) *However, there is little knowledge of* the relationship between MW dielectric properties and pH value of raw milk during storage, which can be used as an indicator of spoilage. [M2S1a, ME]

For Move 3 (Presenting the present work), M3S1 (Announcing present research descriptively and/or purposively) appeared in 83.3% of the samples. More student and expert ME writers used M3S1 than their EE counterparts, with student writers using it more than expert writers. Cargill and O'Connor (2013) considered M3S1 an obligatory step in the introduction section of RAs and called for student attention to this step when writing RA introductions. The high proportion of student-produced RA introductions that contained this step was thus a good indication of the students' awareness of the importance of this step. Meanwhile, the absence of this step in the other 16.7% of student-produced RA introductions deserves further instructional attention.

Similarly, M3S2 (Presenting research questions or hypothesis) was used by a notably larger proportion of student writers than expert writers, and a larger proportion of ME writers than EE writers. Examples (4) and (5) present two cases in which student writers present the research objectives or questions.

(4) T*he aim of this work* was to investigate the performance of an intra-row weed knife path control system. The specific objectives were (1) to prove the in-field feasibility of a complete GPS-based system for intra-row weed knife control; (2) to assess that the intra-row weed knife control system with the GPS can accomplish the task accurately and precisely under standard field conditions in the A&F technology university. [M3S1, ME]

(5) This study will answer *the following questions:* (1) what are the relationship among PWM frequency, PWM duty cycle, liquid pressure and nozzle flow; (2) how to build the flow models for Teejet AITXA 8002, 8003 and 8004 nozzles? [M3S2, EE]

For M3S6 (Stating the value of the present research), a larger proportion of student writers used this step than expert writers, and more EE writers used this step than ME writers. Example (6) presents an instance of this step produced by an EE student.

(6) Evaluation of the influence of various factors provided insight into the functions of hot air heating methods and provided a means of optimizing these treatments. [M3S6, EE]

In contrast, M3S4 (Summarizing methods) was used by a larger proportion of expert writers than student writers. More EE writers used this step than ME writers. Example (7) illustrates how an EE student explicated the method of his/her study.

(7) The wattmeter method measures the power used by the transducer and is utilised to measure the output acoustic power, mechanical loss energy, and the electro-acoustic efficiency of the system. [M3S4, EE]

M3S7 (Outlining RA structure) was also used by a larger proportion of expert writers than student writers. While only 10% of student-produced introductions contained this step, all instances of this step were realized with explicit linguistic markers. In Example (8), the EE student signals the organization of the RA with *The paper is organized as follows*. This type of usage shows that a few students were already skillful in using a rhetorical map to guide readers through their discussion. One possible reason for the low use of this step by the students, however, could be that many of them had not yet fully planned the other sections of their RAs by the time they submitted the introductions.

(8) The paper is organized as follows. In the first part the theoretical background of the dynamic spline formulation will be recalled. In the second part, the human spine model is presented and the corresponding equations of motions are deduced. In the last part, the model has been validated in terms of modal behavior and included into a whole body model in order to be compared to previous works. [M3S7, EE]

M3S5 (Announcing principal outcomes) was absent in student-produced introductions in both sub-disciplines. This absence echoes the low usage of this step by expert EE (6.67%) and ME (4%) writers. Additionally, some students who did not yet have research outcomes to report may have opted to present a specific hypothesis in M3S2, as illustrated in Example (9).

(9) The hypothesis of this study was that PWM technology can be used to control solenoid valve at high liquid pressure (50–100 psi), and nozzles in the same model have the similar flow characteristics. [M3S2, EE]

Overall, the analysis revealed that most students were able to integrate steps of each of the three rhetorical moves into their RA introductions, and to use appropriate linguistic markers to realize these steps. In general, similar patterns of sub-disciplinary variation in the distribution of rhetorical moves were found in student- and expert-produced RA introductions. Meanwhile, some steps were used by a larger proportion of student writers, such as M3S1, M3S2, and M3S6, while some steps were used by a larger proportion of expert writers, such as M3S4 and M3S7. A few steps were completely absent in student-produced RA introductions, including M2S1b, M3S3, and M3S5. However, these three steps were found in a very low proportion of expert-produced introductions and could thus be deemed to be optional (Kanoksilapatham, 2015). Further analysis is necessary to establish whether their complete absence is cause for pedagogical concern or a natural consequence of the nature of the RAs being produced by the students.

4.2. Student perceptions of their experiences with the corpus activities

4.2.1. Results from the post-instruction questionnaire

Table 4 summarizes the results of the items (Items 11–18) on the post-instruction questionnaire pertaining to students' perceptions of their experiences with the corpus activities. The majority of students agreed (either somewhat or strongly) that it was easy to use AntMover and AntConc and to find solutions to the tasks (i.e., identifying linguistic patterns and markers associated with different rhetorical moves) by searching the corpus (73.3%, 22 students). As shown by the positive responses from the students, the corpus activities, which were developed based upon Charles (2012) and Kennedy and Miceli (2017), were relatively straightforward for the learners to interact with. The ease of use of these activities certainly contributed to their positive pedagogical effects discussed above.

Additionally, most students agreed (somewhat or strongly) that the corpus activities were fun (83.3%, 25 students), and helped raise their interest (76.7%, 23 students) and confidence (80%, 24 students) to write RA introductions. The majority of the students also agreed (somewhat or strongly) that they would use corpora in future academic writing (76.7% 23 students) and recommend corpus activities to their peers (70%, 21 students). These results align with the positive experiences with corpus activities and sustained interest in future use of corpus tools reported in previous studies (Cheng, 2007; Naismith, 2017; Yoon & Hirvela, 2004). Nevertheless, further research is necessary to ascertain the long-term effectiveness of such corpus activities, particularly with respect to the students' sustained use of corpus tools and specialized corpora to continue improving their writing.

4.2.2. Results from the reflective journals and post-instruction interviews

4.2.2.1. Positive experience with corpus use. Analysis of the post-instruction interviews revealed further evidence of the students' perception that the corpus activities made learning interesting and increased their confidence in writing research articles, as illustrated by the comments below:

> "Using the corpus makes learning interesting and not boring. I enjoy using the corpus to learn about language use."

> "I am working on a research project, and my supervisor is very happy with my study, and he wants me to aim for an SCI-indexed journal. I was quite scared at this thought, as publishing in English seems to be too challenging for me, not to mention in SCI-indexed journals. After this course, I have learned how to write research articles in English, and I am confident that I can write up the study I am working on in English now."

Table 4
Participant Perceptions of their Experiences with the Corpus Activities.

Item on the post-instruction questionnaire	Strongly disagree	Somewhat disagree	Neither agree nor disagree	Somewhat agree	Strongly agree
11. It was easy to use the Antmover software	0.0%	3.3%	23.3%	46.7%	26.7%
12. It was easy to use the AntConc software	0.0%	0.0%	16.7%	40.0%	43.3%
13. It was easy to find solutions to the tasks by searching the corpus	0.0%	6.7%	20.0%	43.3%	30.0%
14. Corpus activities raised interest in writing RA introductions	0.0%	3.3%	20.0%	40.0%	36.7%
15. Corpus activities raised confidence about writing RA introductions	0.0%	3.3%	16.7%	40.0%	40.0%
16. It was fun to search the corpus	0.0%	3.3%	13.3%	40.0%	43.3%
17. Would use corpora in future academic writing	0.0%	13.3%	10.0%	56.7%	20.0%
18. Would recommend corpus activities to others	0.0%	20.0%	10.0%	43.3%	26.7%

4.2.2.2. Creative uses of the corpus. Some interviewees also mentioned creative uses of the corpus to find discipline-specific information beyond the task requirements. For example, one interviewee noted the following:

> "I often search the corpus with AntConc to see how certain terms in my discipline are used by other scholars. AntConc allows me to see how the terms are used in different contexts and to compare how different writers talk about similar concepts in different ways."

These creative uses show that the students have come to take advantage of the corpus for "lexicogrammatically fine-tuning" (Lee & Swales, 2006, p.57) their understanding and use of discipline-specific language. The benefits these creative uses bring to the students also help explain their interest in using corpus tools in the future.

4.2.2.3. Problems and challenges. Analysis of the reflective journals and post-instruction interviews also revealed several problems or challenges that the students experienced in interacting with the corpus. Some students noted the longer than expected time spent on reading concordance lines, especially early on, as illustrated by the following comments in students' reflective journals:

> "When I first retrieved the concordance lines from the corpus, I easily ended up reading everything in those lines, and felt quite at a loss. Later, I started referring back to the task requirements to remind myself what I should be looking for specifically."

> "I need to stare at the screen to see what the possible questions are. I find this has taken up a lot of time."

The difficulty to decode concordance lines was also previously reported by Charles (2012) and Liu and Jiang (2009) and points to the need for more explicit instructor guidance at the onset of the corpus activities. The concordance lines were presented in the keyword in context format, with the query expression in the center, and the students were not expected to read the lines from the beginning to end (or left to right), but to look for shared linguistic patterns surrounding the query expression across the lines. In the current study, we trained the students to start with the query expression highlighted in the center, to sort the concordance lines on the left and/or right of the query in various ways, to attempt to find linguistic patterns around the query expression across the concordance lines, and to discuss the patterns they found with their peers. Our training proved useful to many students, but more hands-on training and group work may further enhance the students' efficiency in reading and analyzing concordance lines. Additionally, computational tools that can both perform rhetorical move annotation and retrieve linguistic patterns associated with specific rhetorical moves/steps will certainly be of tremendous value in helping learners efficiently obtain key information on form-function mappings from the corpus.

More importantly, some students noted the difficulty in obtaining a full understanding of the rhetorical structures and linguistic features of the target genre through analyzing the concordance lines alone. One interviewee mentioned the following:

> "The search results only show some fragments of the texts, and it is hard for me to get a holistic understanding of the content knowledge just by analyzing them".

This challenge resonates with previous observations of the fragmentary nature of what concordance lines supply (Swales, 2002) and the "disproportionate focus on individual items and a concomitant neglect of phenomena" (Charles, 2011, p. 48) in looking at concordance lines alone. It highlights the need to combine corpus search and analysis activities with other genre analysis activities to ensure a more holistic view of the rhetorical structures and linguistic features of the target genres, such as the rhetorical move annotation activity used in this study.

Finally, an analysis of the interview transcripts and reflective journals of those students who responded negatively to the post-instruction questionnaire indicated that many of them had not yet made adequate progress in their research projects to be in a position to start writing an RA. This may have resulted in a lower level of motivation to engage with the corpus activities than those students who were already in a position to write an RA (see, e.g., Charles, 2012). Future research may explore ways to better motivate such students to engage in corpus-based genre analysis activities.

5. Conclusion

This study contributes evidence of the potential of corpus-based genre analysis activities for enhancing EFL learners' discipline-specific genre knowledge and genre-based writing skills. By participating in the compilation and rhetorical move annotation of a specialized corpus of RA introductions in engineering, learners gained opportunities to not only access but also rhetorically analyze authentic materials directly related to their own research interests.

The frequent and systematic exposure to discipline-specific discourses of the target genre facilitated by subsequent corpus search and analysis activities helped promote learners' "academic discoursal consciousness" (Belcher & Braine, 1995, p. xv) and enrich their "discipline-based experiences" (Hirvela, 1997, p. 83). More specifically, these hands-on activities created useful inductive learning processes that helped them generalize linguistic patterns associated with different rhetorical moves and consolidate the genre knowledge they acquired in the writing course. The writing samples produced by the students showcased their enhanced ability to employ appropriate linguistic devices to realize different rhetorical moves. Many learners also found the corpus activities to be easy to follow, interesting, and confidence-boosting.

The success in encouraging students to solicit supervisors' recommendation in identifying RAs for the specialized corpus and the supervisors' feedback on student-produced RA introductions suggests that it may be highly fruitful to enlist the collaboration of expert informants in future EAP/ESP course teaching. Notwithstanding the resistance or reluctance previously reported (Braine, 2001), collaboration from expert informants can help ensure the relevance of the pedagogical focus to disciplinary needs and assess the effectiveness of the pedagogy in disciplinary learning contexts.

Meanwhile, our results revealed the importance of an "appropriate level of teacher guidance or pedagogical mediation" (McEnery & Xiao, 2011, p. 371) to ensure that learners are not overwhelmed by the vast amount of information present in the concordance lines and that they effectively focus on the analytical concerns of the corpus activities. Our results also showed that using the types of corpus search and analysis activities implemented in the current study alone could result in a fragmented understanding of the rhetorical structures of the target genre, and it is necessary to combine them with other types of activities that take a more holistic approach to analyzing genre. These observations have useful implications for training language teachers with limited knowledge of and experience with corpus tools (Tono, Satake, & Miura, 2014; Tribble, 2015) to implement such corpus-based genre analysis activities.

The current study has several limitations, some of which can be addressed in future research. First, the participants of the current study were at an advanced level of English proficiency and were highly motivated to develop their discipline-specific RA writing skills. It would be useful to see whether and what adjustments to the instructional approach may be necessary to achieve the pedagogical effectiveness and positive learner experiences observed in this study on other EFL learners with different levels of language proficiency and/or motivation. In general, however, the types of materials and activities used in the current study may not be appropriate for learners with low English proficiency, as they have high demands in both content and language knowledge.

Second, given the time and resource constraints, the specialized corpus compiled in this course was relatively small in size. The corpus compilation and annotation activities integrated in this study may empower learners to build their own specialized corpora and use them in data-driven learning in the future. Meanwhile, efforts in creating larger-scale, functionally tagged corpora of academic texts for specific disciplines and target genres will certainly prove highly beneficial to corpus-based genre pedagogy (e.g., Cotos et al., 2017; Dong & Buckingham, 2018).

Third, the current study focused on the instructional implementation of the corpus-based genre analysis activities and learners' evaluation and perception of these activities. While the student-produced writing samples provided some evidence of their developing genre knowledge and genre-based writing skills, we did not include a control group for comparisons with alternative instructional approaches, nor did we systematically track the learners' development of genre knowledge and genre-based writing skills over time. Follow-up interviews with the participants at a later point (e.g., six months after the end of the course) could have revealed information on whether they attempted to expand the specialized corpus and/or continued to use the corpus tools and activities to improve their writing skills. In our future research, we plan to integrate these design aspects to more rigorously assess the long-term learning outcomes that the corpus-based genre analysis approach can produce.

Finally, while our analysis revealed some variation in the distribution of rhetorical moves in RA introductions across the two sub-disciplines of engineering, we did not focus on such variation in the corpus-based instruction. In light of such variation (see also Kanoksilapatham, 2015), future research may pay more attention to sub-disciplinary variation in teaching rhetorical structures in disciplinary writing and test the effectiveness of corpus-based instruction in promoting learners' awareness of such variation.

Acknowledgments

We would like to express our thanks to the editor and anonymous reviewers for their valuable comments on this paper. The work was supported by a grant from the National Social Science Foundation of China (No. 17CYY053) to the first author.

Appendix A. Corpus-based search query samples

Task 1: Identify linguistic patterns that commonly occur in M1S1 (Claiming centrality).

1. Open Antconc.
2. Load the tagged files in AntConc by clicking on 'Open File(s) … ' or 'Open Dir … ' in the 'File' menu.
3. Type in "M1S1" in the main search box at the bottom of the search interface.
4. Read and analyze the concordance lines. Use the 'Kwic Sort' function to rearrange the concordance lines with the options 1R and 2R (i.e., first and second context word to the right of the search word).
5. Check if there are linguistic patterns that occur frequently in this move. If yes, write them down.
6. Discuss the linguistic patterns you found with your group members. Note their comments on the usefulness of the patterns you identified and any additional useful patterns that they identified.

Task 2

1. Type in "seem* to" in the main search box at the bottom of the search interface. Take note of the moves that contain this linguistic marker.
2. In a similar fashion, search the useful linguistic patterns you recorded in Task 1 above and note the moves each of them frequently occurs in. Record your results in the table below. For example, if you searched "seem* to", write down "seem* to" in the linguistics markers column and the moves that it frequently appears in the moves column.

Linguistic markers	Moves

3. Discuss your findings with your group members. Determine if there is any specific relationship between one or more linguistic markers and one or more moves.

Appendix B. Pre-Instruction Questionnaire

Thank you for taking the time to fill out this questionnaire. Your responses will help us understand your prior knowledge and experience of English research article (RA) introduction writing and corpus use. Your responses will be treated with complete confidentiality and will be used for research purposes only. For each item, please circle the letter that best corresponds to your response.

1. I know how to write the introduction section of an RA.
   A. Strongly disagree B. Somewhat disagree C. Neither agree nor disagree
   D. Somewhat agree E. Strongly agree
2. I am familiar with the rhetorical structures of RA introductions.
   A. Strongly disagree B. Somewhat disagree C. Neither agree nor disagree
   D. Somewhat agree E. Strongly agree
3. I am familiar with the linguistic features of RA introductions.
   A. Strongly disagree B. Somewhat disagree C. Neither agree nor disagree
   D. Somewhat agree E. Strongly agree
4. I have used a corpus in a course or for other learning purposes before.
   A. Strongly disagree B. Somewhat disagree C. Neither agree nor disagree
   D. Somewhat agree E. Strongly agree
5. I have written the introduction section for one or more RAs prior to this course.
   A. Strongly disagree B. Somewhat disagree C. Neither agree nor disagree
   D. Somewhat agree E. Strongly agree
6. How many published RAs have you read in the past six months?
   A. 0 B. 1-5 C. 6-10 D. 11-20 E. 21 or more

Appendix C. Post-Instruction Questionnaire

Thank you for taking the time to fill out this questionnaire. Your responses will help us understand your experience with the course. Your responses will be treated with complete confidentiality and will be used for research purposes only. For each item, please circle the letter that best corresponds to your response.

1. The course's focus on the rhetorical structures of RA introductions helped improve my RA writing skills.
   A. Strongly disagree B. Somewhat disagree C. Neither agree nor disagree
   D. Somewhat agree E. Strongly agree
2. This course's focus on the rhetorical structures of RA introductions helped me learn the rhetorical structures of RA introductions.
   A. Strongly disagree B. Somewhat disagree C. Neither agree nor disagree
   D. Somewhat agree E. Strongly agree
3. This course's focus on the rhetorical structures of RA introductions helped me learn the linguistic features of RA introductions.
   A. Strongly disagree B. Somewhat disagree C. Neither agree nor disagree
   D. Somewhat agree E. Strongly agree
4. What else did this course help you improve?
   A. Reading skills B. Communication skills C. Corpus use skills
   D. Other: please specify_____
5. The corpus activities helped improve my RA writing skills.
   A. Strongly disagree B. Somewhat disagree C. Neither agree nor disagree
   D. Somewhat agree E. Strongly agree
6. The corpus activities enhanced my knowledge of the rhetorical structures of RA introductions.
   A. Strongly disagree B. Somewhat disagree C. Neither agree nor disagree
   D. Somewhat agree E. Strongly agree
7. The corpus activities enhanced my knowledge of the linguistic features of RA introductions.
   A. Strongly disagree B. Somewhat disagree C. Neither agree nor disagree
   D. Somewhat agree E. Strongly agree
8. The corpus activities helped me generalize the patterns of different linguistic features of RA introductions.
   A. Strongly disagree B. Somewhat disagree C. Neither agree nor disagree
   D. Somewhat agree E. Strongly agree
9. The corpus activities helped me master the use of specific linguistic expressions.
   A. Strongly disagree B. Somewhat disagree C. Neither agree nor disagree
   D. Somewhat agree E. Strongly agree
10. The corpus activities helped improve my understanding of my own research topics.
    A. Strongly disagree B. Somewhat disagree C. Neither agree nor disagree
    D. Somewhat agree E. Strongly agree
11. It was easy to use the AntMover software.
    A. Strongly disagree B. Somewhat disagree C. Neither agree nor disagree
    D. Somewhat agree E. Strongly agree
12. It was easy to use the AntConc software.
    A. Strongly disagree B. Somewhat disagree C. Neither agree nor disagree
    D. Somewhat agree E. Strongly agree
13. It was easy to find answers to the tasks by searching the corpus.
    A. Strongly disagree B. Somewhat disagree C. Neither agree nor disagree
    D. Somewhat agree E. Strongly agree
14. The use of the corpus made me feel more interested in writing RA introductions.
    A. Strongly disagree B. Somewhat disagree C. Neither agree nor disagree
    D. Somewhat agree E. Strongly agree
15. The use of the corpus made me feel more confident about writing RA introductions.
    A. Strongly disagree B. Somewhat disagree C. Neither agree nor disagree
    D. Somewhat agree E. Strongly agree
16. It was fun to search the corpus.
    A. Strongly disagree B. Somewhat disagree C. Neither agree nor disagree
    D. Somewhat agree E. Strongly agree
17. I would use corpora to help with my RA writing in the future.
    A. Strongly disagree B. Somewhat disagree C. Neither agree nor disagree
    D. Somewhat agree E. Strongly agree
18. I would recommend corpus compilation and search activities to other English language learners who want to improve their RA writing skills.
    A. Strongly disagree B. Somewhat disagree C. Neither agree nor disagree
    D. Somewhat agree E. Strongly agree

References

Ädel, A. (2010). Using corpora to teach academic writing: Challenges for the direct approach. In M. Campoy-Cubillo, B. Belles-Fortuno, & L. Gea-Valor (Eds.), *Corpus-based approaches to English language teaching* (pp. 39-55). New York: Continuum.
Anthony, L. (2003). *AntMover (version 1.0) [computer software]*. Tokyo, Japan: Waseda University. Available from: http://www.laurenceanthony.net/software.
Anthony, L. (2008). *AntConc (version 3.2.2) [computer software]*. Tokyo, Japan: Waseda University. Available from: http://www.antlab.sci.waseda.ac.jp.
Anthony, L., & Lashikia, G. V. (2003). Automatic identification of organizational structure in writing using machine learning. In *Paper presented at the sixth international conference on languages for specific purposes. Leiden, Netherlands*.

Ballance, O. J. (2017). Pedagogical models of concordance use: Correlations between concordance user preferences. *Computer Assisted Language Learning, 30*(3–4), 259-283.
Belcher, D. D., & Braine, G. (Eds.). (1995). *Academic writing in a second language: Essays on research and pedagogy*. Norwood, NJ: Ablex.
Bhatia, V. K. (1991). genre-based approach to ESP materials development. *World Englishes, 10*(2), 1-14.
Bhatia, V. K. (1993). *Analyzing genre: Language use in professional settings*. London: Longman.
Biber, D., & Conrad, S. (2009). *Register, genre, and style*. New York: Cambridge University Press.
Boulton, A., & Cobb, T. (2017). Corpus use in language learning: A meta-analysis. *Language Learning, 67*(2), 348-393.
Braine, G. (2001). When professors don't cooperate: A critical perspective on EAP research. *English for Specific Purposes, 20*(3), 293-303.
Cargill, M., & O'Connor, P. (2013). *Writing scientific research articles: Strategy and steps* (2nd ed.). Oxford: Wiley-Blackwell.
Cargill, M., O'Connor, P., & Li, Y. (2012). Educating Chinese scientists to write for international journals: Addressing the divide between science and technology education and English language teaching. *English for Specific Purposes, 31*(1), 60-69.
Chambers, A. (2007). Popularising corpus consultation by language learners and teachers. In E. Hidalgo, L. Quereda, & J. Santana (Eds.), *Corpora in the foreign language classroom* (pp. 3-16). Amsterdam: Rodopi.
Chang, J. (2014). The use of general and specialized corpora as reference sources for academic English writing: A case study. *ReCALL, 26*(2), 243-259.
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*(4), 289-302.
Charles, M. (2011). Adverbials of result: Phraseology and functions in the problem-solution pattern. *Journal of English for Academic Purposes, 10*(1), 47-60.
Charles, M. (2012). "Proper vocabulary and juicy collocations": EAP students evaluate do-it-yourself corpus-building. *English for Specific Purposes, 31*(2), 93-102.
Charles, M. (2014). Getting the corpus habit: EAP students' long-term use of personal corpora. *English for Specific Purposes, 35*, 30-40.
Chen, M., & Flowerdew, J. (2018). Introducing data-driven learning to PhD students for research writing purposes: A territory-wide project in Hong Kong. *English for Specific Purposes, 50*, 97-112.
Cheng, A. (2007). Transferring generic features and recontextualizing genre awareness: Understanding writing performance in the ESP genre-based literacy framework. *English for Specific Purposes, 26*(3), 287-307.
Conrad, S. (1999). The importance of corpus-based research for language teachers. *System, 27*(1), 1-18.
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.
Cotos, E., Link, S., & Huffman, S. R. (2017). Effects of DDL technology on genre learning. *Language, Learning and Technology, 21*(3), 104-130.
Del Saz-Rubio, M. M. (2011). A pragmatic approach to the macro-structure and metadiscoursal features of research article introductions in the field of Agricultural Sciences. *English for Specific Purposes, 30*(4), 258-271.
Dong, J., & Buckingham, L. (2018). The textual colligation of stance phraseology in cross-disciplinary academic discourse. *International Journal of Corpus Linguistics, 23*(4), 408-436.
Durrant, P., & Mathews-Aydınlı, J. (2011). A function-first approach to identifying formulaic language in academic writing. *English for Specific Purposes, 30*(1), 58-72.
Eriksson, A. (2012). Pedagogical perspectives on bundles: Teaching bundles to doctoral students of biochemistry. In J. Thomas, & A. Boulton (Eds.), *Input, process and product: Developments in teaching and language corpora* (pp. 195-211). Brno: Masaryk University Press.
Fosnot, C. T. (1996). Constructivism: A psychological theory of learning. In C. T. Fosnot (Ed.), *Constructivism: Theory, perspectives, and practice* (pp. 8-33). New York: Teachers College Press.
Friginal, E. (2013). Developing research report writing skills using corpora. *English for Specific Purposes, 32*(4), 208-220.
Gardner, D., & Davies, M. (2007). Pointing out frequent phrasal verbs: A corpus-based analysis. *TESOL Quarterly, 41*(2), 339-359.
Gilmore, A. (2009). Using online corpora to develop students' writing skills. *ELT Journal, 63*(4), 363-372.
Hammond, J., Burns, A., Joyce, H., Brosnan, D., Gerot, L., Solomon, N., & Hood, S. (1992). *English for social purposes: A handbook for teachers of adult literacy*. Sydney: National Centre of English Language Teaching Research.
Henry, A., & Roseberry, R. L. (1998). An evaluation of a genre-based approach to the teaching of EAP/ESP writing. *TESOL Quarterly, 32*(1), 147-156.
Hirvela, A. (1997). Disciplinary portfolios" and EAP writing instruction. *English for Specific Purposes, 16*(2), 83-100.
Hyland, K. (2000). *Disciplinary discourses: Social interactions in academic writing*. London: Pearson.
Hyland, K. (2003). Genre-based pedagogies: A social response to process. *Journal of Second Language Writing, 12*(1), 17-29.
Hyland, K. (2007). Genre pedagogy: Language, literacy and L2 writing instruction. *Journal of Second Language Writing, 16*(3), 148-164.
Johns, A. M. (1997). *Text, role and context: Developing academic literacies*. Cambridge: Cambridge University Press.
Kanoksilapatham, B. (2015). Distinguishing textual features characterizing structural variation in research articles across three engineering sub-discipline corpora. *English for Specific Purposes, 37*, 74-86.
Kennedy, C., & Miceli, T. (2017). Cultivating effective corpus use by language learners. *Computer Assisted Language Learning, 30*(1–2), 91-114.
Lantolf, J. P. (2009). Dynamic assessment: The dialectic integration of instruction and assessment. *Language Teaching, 42*(3), 355-368.
Lee, D., & Swales, J. (2006). A corpus-based EAP course for NNS doctoral students: Moving from available specialized corpora to self-compiled corpora. *English for Specific Purposes, 25*(1), 56-75.
Lillis, T., & Curry, M. J. (2013). *Academic writing in a global context: The politics and practices of publishing in English*. London: Routledge.
Lin, M. H. (2016). Effects of corpus-aided language learning in the EFL grammar classroom: A case study of students' learning attitudes and teachers' perceptions in Taiwan. *TESOL Quarterly, 50*(4), 871-893.
Liu, D., & Jiang, P. (2009). Using a corpus-based lexicogrammatical approach and ESL contexts to grammar instruction in EFL. *The Modern Language Journal, 93*(1), 61-78.
Lu, X., Casal, J. E., & Liu, Y. (2020). The rhetorical functions of syntactically complex sentences in social science research article introductions. *Journal of English for Academic Purposes, 44*, 1-16.
Lu, X., Yoon, J., & Kisselev, O. (2018). A phrase-frame list for social science research article introductions. *Journal of English for Academic Purposes, 36*, 76-85.
McEnery, T., & Xiao, R. (2011). What corpora can offer in language teaching and learning. In E. Henkel (Ed.), *Handbook of research in second language teaching and learning* (vol. 2, pp. 364-380)New York: Routledge.
Naismith, B. (2017). Integrating corpus tools on intensive CELTA courses. *ELT Journal, 71*(3), 273-283.
Ozturk, I. (2007). The textual organisation of research article introductions in applied linguistics: Variability within a single discipline. *English for Specific Purposes, 26*(1), 25-38.
Paltridge, B. (2014). Genre and second-language academic writing. *Language Teaching, 47*(3), 303-318.
Pérez-Paredes, P., Sánchez-Tornel, M., Calero, J. M. A., & Jiménez, P. A. (2011). Tracking learners' actual uses of corpora: Guided vs non-guided corpus consultation. *Computer Assisted Language Learning, 24*(3), 233-253.
Piaget, J. (1977). *The development of thought: Equilibration of cognitive structures*. New York: Viking.
Poole, R. (2016). A corpus-aided approach for the teaching and learning of rhetoric in an undergraduate composition course for L2 writers. *Journal of English for Academic Purposes, 21*, 99-109.
Römer, U. (2010). Using general and specialized corpora in English language teaching: Past, present and future. In M. Campoy-Cubillo (Ed.), *Corpus-based approaches to English language teaching* (pp. 18-38). London: Continuum.
Samraj, B. (2002). Introductions in research articles: Variations across disciplines. *English for Specific Purposes, 21*(1), 1-17.
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.
Smith, S. (2011). Learner construction of corpora for general English in Taiwan. *Computer Assisted Language Learning, 24*(4), 291-316.

Swales, J. M. (1990). *Genre analysis: English in academic and research settings*. Cambridge: 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: Longman.
Swales, J. M. (2004). *Research genres: Exploration and applications*. Cambridge: Cambridge University Press.
Swales, J., & Lindemann, S. (2002). Teaching the literature review to international graduate students. In A. M. Johns (Ed.), *Genre in the classroom: Multiple perspectives* (pp. 105-119). Mahwah, NJ: Lawrence Erlbaum.
Tang, J., Pritchard, N., & Shi, L. (2012). Calibrating English language courses with major international and national EFL tests via vocabulary range. *Chinese Journal of Applied Linguistics, 35*(1), 24-43.
Tono, Y., Satake, Y., & Miura, A. (2014). The effects of using corpora on revision tasks in L2 writing with coded error feedback. *ReCALL, 26*(2), 147-162.
Tribble, C. (2015). Teaching and language corpora: Perspectives from a personal journey. In A. Leńko-Szymańska, & A. Boulton (Eds.), *Multiple affordances of language corpora for data-driven learning* (pp. 37-62). Amsterdam: John Benjamins.
Yayli, D. (2011). From genre awareness to cross-genre awareness: A study in an EFL context. *Journal of English for Academic Purposes, 10*(3), 121-129.
Yoon, H., & Hirvela, A. (2004). ESL student attitudes toward corpus use in L2 writing. *Journal of Second Language Writing, 13*, 257-283.

Jihua Dong is Professor, Qilu Young Scholar, and Taishan Young Scholar in the School of Foreign Languages and Literature at Shandong University, China. Her research interests include corpus linguistics, corpus-based teaching, academic writing, and discourse analysis. She has previously published in *International Journal of Corpus Linguistics*, *Journal of English for Academic Purposes*, and *System*, among others. E-mail: jdon104@aucklanduni.ac.nz

Xiaofei Lu is Professor of Applied Linguistics and Asian Studies at The Pennsylvania State University, where he directs the graduate programs in the Department of Applied Linguistics. His research interests are in corpus linguistics, English for Academic Purposes, second language writing, and intelligent computer-assisted language learning. E-mail: xxl13@psu.edu, http://www.personal.psu.edu/xxl13.