Lexical Bundles

The term "lexical bundles" was defined as "recurrent expressions, regardless of their idiomaticity, and regardless of their structural status". As is well documented, lexical bundles not only contribute to fluent linguistic production but also form essential building blocks of discourse. A good command of lexical bundles could be indicative of a proficient and professional academic writer and is thus considered a pivotal skill for student writers, especially EFL student writers, for achieving sustainable growth of writing competence. Appropriate use of lexical bundles in academic writing helps writers from an academic community demonstrate their research writing ability.

1. Lexical Bundles

As Biber et al. [1] posit, lexical bundles are the most frequent, recurrent, multiword sequences in a register, which are defined "strictly on the basis of frequency" (p. 399) rather than intuitive criteria. Even though the identification of lexical bundles is solely based on frequency without considering structural or functional features, Biber and associates think that these multiword sequences are "interpretable in both structural and functional terms" (p. 399).

The first structural classification was proposed by Biber et al. [2], in which the prevalent lexical bundles were grouped into fourteen structural categories in conversation and twelve categories in academic prose. Concerning structural analysis of lexical bundles, their framework has been used as a major reference.

Functionally, two taxonomies are widely adopted. Biber et al. [1] distinguished three main functions: (1) stance expressions for displaying "attitudes or assessments of certainty", (2) discourse organizers that "reflect relationships between prior and coming discourse", and (3) referential bundles that "make direct reference to physical abstract or single out some particular attribute of the entity as especially important." (ibid., p. 384). Inspired by Biber et al. [1], Hyland [3] proposed his functional taxonomy of lexical bundles, including (1) research-oriented bundles that "help writers to structure their activities and experiences of the real world", (2) text-oriented bundles that are "concerned with the organization of the text and its meaning as a message or argument", and (3) participant-oriented bundles that are "focused on the reader or writer of the text" (ibid., p. 13).

2. Studies of the Structural and Functional Analyses of Lexical Bundles

Comparative studies of lexical bundles are largely carried out along three dimensions. They are lexical bundle use across registers, across disciplines, and across writer groups. We review these studies below.

2.1. Lexical Bundle Use across Registers

One of the important variables influencing the use of lexical bundles is register variation. Based on the comparison of two types of registers, i.e., conversation and academic prose, Biber et al. [2] found that, in terms of structure, bundles were more clausal in conversation but more phrasal in academic prose. In their studies of university classroom teaching and university textbooks, Biber et al. [1] concluded that the use of lexical bundles in classroom teaching reflects a mixture of characteristics typical of both conversation and academic prose. Similarly, Nesi and Basturkmen [4] found that academic lectures are also featured by combined use of oral and literate bundles. In terms of function, classroom teaching combines functional characteristics of both spoken (by using stance and discourse organizing bundles) and written registers (by using referential bundles) [1]. Biber and Barbieri [5] further examined lexical bundles in a broader range of spoken and written university registers. They concluded that both spoken/written register differences and communicative purposes influence the use of lexical bundles.

2.2. Lexical Bundles across Disciplines

Discipline is also a crucial variable influencing the use of lexical bundles [6][7][8][3]. Of these studies, lexical bundles in soft science and hard science are often compared. In terms of structure, two main structural types are found in history, including noun phrases and prepositional phrases, whereas more structural types are found in biology [6]. It is further found that social science texts make use of a large number of bundles with an embedded *of* phrase to identify the logical relations in the argument. By contrast, hard science texts make more use of formulaic passive constructions and *anticipatory it* patterns to disguise the personal role of writers in the interpretation of data [3]. In terms of function, soft-science texts use more text-oriented and participant-oriented bundles, whereas hard science texts are dominated by research-oriented bundles [3][9]. Disciplinary variation is also explored in student writing [7]. The results suggest that research-oriented bundles are used for assertion of importance in soft science but for physical descriptions in hard science. Stance-oriented bundles are used to evaluate the importance of the topic in soft science but to state findings in hard science. Furthermore, soft science writing is characterized by text-oriented bundles indicating relationships or differences. Hard science writing, by contrast, contains text-oriented bundles that guide readers' attention to data presented in figures and tables. Some studies have further investigated the relation between lexical bundles and rhetorical moves in a given discipline [10][11][12]. Previous studies mostly compare lexical bundles from a macrodiscipline level, such as soft/hard science distinctions and humanities/natural sciences distinctions. It would be more helpful to investigate disciplinary-specific lexical bundles, which may help writers express stances more appropriately in their research community.

2.3. Lexical Bundles across Writer Groups

The third influential factor regarding the use of lexical bundles is the background of different writers, such as between L1 English and L2 English writers [13][14][15][16][17][18], between student writers across different proficiency levels [19][20], or between expert writers and novice writers [6][21].

Most of the research indicated structural and functional differences between L1 and L2 writings. In terms of structure, for instance, L1 Swedish student writers are found to use a higher number of *anticipatory it* (e.g., *it is difficult to*) and *attended this* (e.g., *in this essay I*) constructions than L1 English student writers [13]. It is also found that L1 English writers produce more verb phrase (with a passive verb) lexical bundles, whereas L1 Persian writers use more noun phrase bundles [15]. However, L1 Chinese writers, including both student writers and expert writers, use more verb patterns, whereas L1 English writers use a slightly more extensive range of noun sequences and prepositional sequences [16][17]. In terms of function, L1 English writers are found to use a higher proportion of stance bundles than Swedish writers [13] but a smaller proportion than L1 Chinese writers [17]. Chinese writers are also found to use lexical bundles of description, transition and structure more frequently than English writers, whereas English writers employ more quantification and framing bundles than Chinese writers [16]. Persian writers overused statistical markers compared to English writers [15]. Other studies, however, reported no significant differences between L1 and L2 writing. For instance, Chen and Baker [14] found that lexical bundles in L1 and L2 student writing are surprisingly similar. This finding is consistent with Shin [18], who found that both L1 and L2 student writers heavily use clausal bundles.

Comparisons have also been made between student writings of different proficiency levels [19][20][21] and between student writers and expert writers [6][3][14].

Regarding student writings of different proficiency levels, previous research has demonstrated a mixture of divergent and even contrasting results. Whereas lower proficiency student writings are reported to feature a higher number of NP-based lexical bundles [20], Chen and Baker [21] found that the lowest level has the lowest proportion of NP-based bundles. Similarly, Vo [20] reported a higher frequency of stance bundles in lower-level writing, whereas Staples et al. [19] found different proficiency groups have a similar distribution of stance bundles and discourse organizing bundles. Such diversity in research results may be largely due to the different criteria for determining the proficiency level of different students.

Römer [22] and Chen and Baker [14] conducted a three-way comparison: L1 English expert writer versus student writers of both L1- and L2-English backgrounds. It was argued that novice/expert distinction is more important than L1/L2 distinction based on the findings that few differences existed between the L1 and L2 student writers. It was found, though, that many lexical bundles frequently used by expert writers are rarely found in student writing [6][3][14][22], whereas student writing features more VP-based bundles [14] and bundles commonly found in the spoken register [23]. Nonetheless, the findings, useful as they are, may not reflect the whole picture of novice writers' discourse features. Many previous studies focus on how bundles identified in expert writing are used by student writers. Such comparisons generate insightful findings but provide insufficient understanding of lexical bundles that are unique to student writing. In addition, most studies on student writers focus on writings by undergraduate writers, including, for example, argumentative essays [17][18][24], research

papers [6], and writing examination papers [20]. Very little attention has been paid to the use of lexical bundles in MA student thesis writing. One of the few existing relevant studies was Hyland [3], which compared published article bundles to those identified in Master theses and PhD dissertations. However, he treated MA theses in his corpus as highly proficient texts and explained the feathers from the perspective of genre variation rather than novice/expert distinction. Therefore, it offers limited pedagogical guidance for student writers in developing sustainable linguistic resources to express their stance in more mature way. Another relevant study is by Pan and Liu [25], who compared L1-L2 differences in bundles in masters' theses and research articles. Although their findings indicated that both L1 background and the level of expertise affect the bundle employment, Pan and Liu [25] did not compare the student bundles directly with expert bundles and they mainly focused on comparison between L1-L2 students and between L1-L2 experts. Despite the fact that their research was among one of few attempts to investigate how postgraduate students employ lexical bundles, comparing MA student writing to expert writing can provide useful information on expert writers' linguistic choices.

Therefore, the current study seeks to focus on this understudied writer group by comparing the use of lexical bundles between Chinese English-major MA theses and expert writers' published articles. Informed by the previous literature, research articles published in leading international journals such as those covered in the SSCI can generally be considered as samples of expert writing. MA theses can be viewed as unique pieces of student writing at the level between argumentative essays/course papers and published research articles. They are written by apprentice academic writers who are under the pressure to display their extensive knowledge in one discipline as well as the ability to conduct independent research appropriately. It is hoped that the present study will contribute to existing knowledge of bundle research on MA writers, especially on Chinese EFL learners. The study aims to provide further insights into pedagogical implications for teaching academic writing.

References

1. Biber, D.; Conrad, S.; Cortes, V. If you look at...: Lexical Bundles in University Teaching and Textbooks. Appl. Linguist. 2004, 25, 371–405.

2. Biber, D.; Johansson, S.; Leech, G.; Conrad, S.; Finegan, E. Longman Grammar of Spoken and Written English; Pearson Education: Harlow, UK, 1999.

3. Hyland, K. As can be seen: Lexical bundles and disciplinary variation. Engl. Specif. Purp. 2008, 27, 4–21.

4. Nesi, H.; Basturkmen, H. Lexical bundles and discourse signalling in academic lectures. Int. J. Corpus Linguist. 2006, 11, 283–304.

5. Biber, D.; Barbieri, F. Lexical bundles in university spoken and written registers. Engl. Specif. Purp. 2007, 26, 263–286.

6. Cortes, V. Lexical bundles in published and student disciplinary writing: Examples from history and biology. Engl. Specif. Purp. 2004, 23, 397–423.

7. Durrant, P. Lexical Bundles and Disciplinary Variation in University Students' Writing: Mapping the Territories. Appl. Linguist. 2017, 38, 165–193.

8. Hyland, K. Academic clusters: Text patterning in published and postgraduate writing. Int. J. Appl. Linguist. 2008, 18, 41–62.

9. Omidian, T.; Shahriari, H.; Siyanova-Chanturia, A. A cross-disciplinary investigation of multi-word expressions in the moves of research article abstracts. J. Engl. Acad. Purp. 2018, 36, 1–14.

10. Abdollahpour, Z.; Gholami, J. Embodiment of rhetorical moves in lexical bundles in abstracts of the medical sciences. S. Afr. Linguist. Appl. Lang. Stud. 2019, 37, 339–360.

11. Cortes, V. The purpose of this study is to: Connecting lexical bundles and moves in research article introductions. J. Engl. Acad. Purp. 2013, 12, 33–43.

12. Qi, H.; Pan, F. Lexical bundle variation across moves in abstracts of medical research articles. S. Afr. Linguist. Appl. Lang. Stud. 2020, 38, 109–128.

13. Ädel, A.; Erman, B. Recurrent word combinations in academic writing by native and non-native speakers of English: A lexical bundles approach. Engl. Specif. Purp. 2012, 31, 81–92.

14. Chen, Y.H.; Baker, P. Lexical bundles in L1 and L2 academic writing. Lang. Learn. Technol. 2010, 14, 30–49.

15. Esfandiari, R.; Barbary, F. A contrastive corpus-driven study of lexical bundles between English writers and Persian writers in psychology research articles. J. Engl. Acad. Purp. 2017, 29, 21–42.

16. Pan, F.; Reppen, R.; Biber, D. Comparing patterns of L1 versus L2 English academic professionals: Lexical bundles in Telecommunications research journals. J. Engl. Acad. Purp. 2016, 21, 60–71.

17. Bychkovska, T.; Lee, J.J. At the same time: Lexical bundles in L1 and L2 university student argumentative writing. J. Engl. Acad. Purp. 2017, 30, 38–52.

18. Shin, Y.K. Do native writers always have a head start over nonnative writers? The use of lexical bundles in college students' essays. J. Engl. Acad. Purp. 2019, 40, 1–14.

19. Staples, S.; Egbert, J.; Biber, D.; McClair, A. Formulaic sequences and EAP writing development: Lexical bundles in the TOEFL iBT writing section. J. Engl. Acad. Purp. 2013, 12, 214–225.

20. Vo, S. Use of lexical features in non-native academic writing. J. Second Lang. Writ. 2019, 44, 1–12.

21. Chen, Y.H.; Baker, P. Investigating criterial discourse features across second language development: Lexical bundles in rated learner essays, CEFR B1, B2 and C1. Appl. Linguist. 2016, 37, 849–880.

22. Römer, U. English in academia: Does nativeness matter? Anglistik Int. J. Engl. Stud. 2009, 20, 89–100.

23. Wang, Y. As Hill seems to suggest: Variability in formulaic sequences with interpersonal functions in L1 novice and expert academic writing. J. Engl. Acad. Purp. 2018, 33, 12–23.

24. Jaworska, S.; Krummes, C.; Ensslin, A. Formulaic sequences in native and non-native argumentative writing in German. Int. J. Corpus Linguist. 2015, 20, 500–525.

25. Pan, F.; Liu, C. Comparing L1-L2 differences in lexical bundles in student and expert writing. S. Afr. Linguist. Appl. Lang. Stud. 2019, 37, 142–157.