Form and Function in Native and EFL Learners' Academic Writing: What is the Difference?

Tsung-Ming Wu (National Taiwan Normal University, Taiwan)

An indicator of professional writing is the employment of lexical bundles, or a sequence of word combinations (Ädel & Erman, 2012). According to the definition, lexical bundles are recurrent word combinations and they "show a statistical tendency to co-occur" (Biber et al., 1999, p. 989). Previous research on lexical bundles has shown that native or professional writers tended to employ more lexical bundles than non-native or student writers (Ädel & Erman, 2012; Chen & Baker, 2010; Lin, 2011) and that their bundle use was different from non-native or student writers (Ädel & Erman, 2012; Chen & Baker, 2010; Cortes, 2004; Lin, 2011; Wei & Lei, 2011). However, earlier studies tended to adopt texts of different genres or of different disciplines for comparison, which may raise the issue of comparability. Another issue lies in the inconsistency of their comparing criteria; either their corpora size was inconsistent, with one corpus being much larger than the other, or their dispersion criterion differed among the corpora.

To address the comparability problem, the current study examined the employment of lexical bundles between the texts of the same academic genre, MA theses, by both native academic writers and the advanced EFL academic writers. The frequency threshold was set 40 times/ million words and the dispersion rate, 20% (i.e. 6 texts). This study attempted to explore the most frequently employed four-word bundles in the two corpora and to explore the form and function difference in the writing between the advanced EFL writers and native writers. It is assumed that if EFL academic writers are informed of a more professional use of lexical bundles, their writing skills are likely to improve.

A corpus of EFL MA theses and a corpus of native MA theses, which amounted to 1.3 million words, were therefore constructed for the current study. The EFL MA thesis corpus was comprised of 30 MA theses in the field of applied linguistics by Taiwanese writers. A total of 0.7 million words were included. The native MA thesis corpus, on the other hand, was made up of 30 MA theses in the field of applied linguistics. The native MA theses were collected from the graduate programs in the United States and the United Kingdom. To ensure the nativeness of the authors, their names and acknowledgements were referred to. The total number of words in this corpus was 0.6 million.

While previous studies have shown that native writers tended to employ more lexical bundles both in terms of the number and in terms of the type in their language production, the present study demonstrated the opposite pattern: EFL writers employed more lexical bundles both in number (EFL writers: native writers= 5,052: 1,576) and in type (EFL writers: native writers= 93: 42) than native writers as shown in Table 1. Overall, EFL writers demonstrated more uses of lexical bundles both in number and type. However, when the shared bundles between the two corpora were examined, EFL writers displayed an underuse trend

(23 underused bundles out of the 29 shared bundles found in the two corpora).

Table 1 Corpus results between native speaker theses (NS theses) and non-native theses (NNS theses)

Corpus	Text	Corpus size	Freq. Criterion	Dispersion Criterion	Types before adjustment	Types after refinement	Tokens before refinement	Tokens after refinement
NS theses	30	617,497	29	6	50	42	1,965	1,576
NNS theses	30	713,469	25	6	114	93	6,127	5,052

The present study also analyzed the structure and function difference between the two constructed corpora. Lin (2011)'s structural framework, a modification of Biber et al. (1999)'s structural framework, was adopted for the analysis of the bundle structure. Hyland (2008)'s functional framework was adopted for the analysis of the bundle function.

An examination of the bundle structures revealed that even though EFL writers employ more Noun Phrase (NP) bundles and Prepositional Phrase (PP) bundles, which were often perceived as productive academic bundles, in terms of the number and the type, the percentage of NP and PP bundles in their writing was not as high as native writers', as shown in Table 2. To exhibit the relative usage of the NP and PP bundles by native professional writers, a NS reference was compared. The NS reference was compiled by Lin (2011), which compiled 200 research articles from prestigious journals in applied linguistics.

When the function types of the bundles were investigated, EFL learners were found to employ lower percentage of the participant-oriented bundles despite the more varied bundle types and more bundles in number were found in the EFL learner corpus, as shown in Table 3.

An analysis of the structure type and function type of lexical bundles revealed that with different perspectives, the results differed. From the lens of percentage, native writers displayed similar bundle usage to professional writers, while in the raw frequency level, EFL writers seemed to have better command of bundle use, if more bundle uses represented higher language proficiency.

Table 2 Lexical bundles categorized according to structure types

Category	NS	NNS	*NS reference
NP	14 (33.33%)	26 (27.96%)	33.77%
PP	17 (40.48%)	32 (34.41%)	43.05%
VP	6 (14.29%)	21 (20.43%)	20.53%
Others	5 (11.90%)	16 (17.20%)	2.65%
Total	42	93	

Table 3 Lexical bundles categorized according to function types

Function	NS	%	NNS	%	*NS reference
1. Research-oriented bundles	17	39.53%	43	45.74%	37.70%
2. Text-oriented bundles	18	41.86%	41	43.62%	46.40%
3. Participant-oriented bundles	8	18.60%	10	10.64%	15.90%

An examination of the highly overused bundles (e.g., *on the other hand*) among the shared bundles illustrated some learner usage problems. Also, the examination of the highly underused bundles (e.g., *it is possible that*, *as well as the*, and *the fact that the*) suggested that teaching and learner style might play a role in the use of lexical bundles. The lexical bundles with significant difference were shown in Table 4. Based on the findings, some pedagogical implications are discussed for English learners and instructors.

Table 4 Lexical Bundles Employed with Significant Difference between NS and NNS

Rank	Bundles	NNS Freq	NS Freq	LL	Sig value
1	on the other hand + the	317	58	15.73	****
2	in the present study + the	302	68	6.31	*
3	the results of the	149	88	-21.05	****
4	at the end of + the	108	67	-18.22	****
5	at the same time	62	39	-10.98	***
6	at the beginning of + the	62	49	-22.08	****
7	the extent to which	56	28	-3.91	*
8	in the form of	55	45	-21.52	****
9	there was no significant difference	51	41	-19.03	****
10	the results of this	50	42	-20.97	****
11	as well as the	48	61	-51.76	****
12	on the basis of	46	28	-7.26	**
13	the fact that the	43	51	-40.23	****
14	in the field of	43	27	-7.57	**
15	i don t know	43	27	-7.57	**
16	the meaning of the	42	30	-11.19	***
17	the purpose of the	41	38	-22.03	****
18	to the fact that	40	27	-8.94	**
19	it is possible that	39	52	-46.28	****
20	no significant difference between	39	34	-18.00	****
21	significant difference between the	39	28	-10.55	**
22	in the case of	37	35	-20.88	****
23	can be seen in	34	27	-12.27	***
24	it should be noted	31	25	-11.67	***
25	there was a significant	30	28	-16.39	****

Note: LL= log likelihood; += NNS overuse; -= NS overuse;
LL>3.84 → *p<.05; LL>6.63 → **p<.01; LL>10.83 → ***p<.001; LL>15.13 → ****p<.0001

References

Ädel, A. (2006). *Metadiscourse in L1 and L2 English* (Vol. 24). John Benjamins Publishing.

Ädel, A., & Erman, B. (2012). Recurrent word combinations in academic writing by native and non-native speakers of English: A lexical bundles approach. *English for Specific Purposes*, *31*(2), 81-92.

Biber, D., Johansson, S., Leech, G., Conrad, S., Finegan, E., & Quirk, R. (1999). *Longman grammar of spoken and written English* (Vol. 2). MIT Press.

Biber, D., Conrad, S., & Cortes, V. (2004). If you look at...: Lexical bundles in university teaching and textbooks. *Applied linguistics*, *25*(3), 371-405.

Biber, D., & Barbieri, F. (2007). Lexical bundles in university spoken and written registers. *English for specific purposes*, *26*(3), 263-286.

Chen, Y. H., & Baker, P. (2010). Lexical bundles in L1 and L2 academic writing. *Language Learning & Technology*, *14*(2), 30-49.

Cortes, V. (2002). Lexical bundles in freshman composition. *Using corpora to explore linguistic variation*, 131-145.

Cortes, V. (2004). Lexical bundles in published and student disciplinary writing: Examples from history and biology. *English for specific purposes*, *23*(4), 397-423.

Field, Y., & Oi, Y. L. M. (1992). A comparison of internal conjunctive cohesion in the English essay writing of Cantonese speakers and native speakers of English. *RELC journal*, *23*(1), 15-28.

Hyland, K. (2008). Academic clusters: Text patterning in published and postgraduate writing. *International Journal of Applied Linguistics*, *18*(1), 41-62.

Hyland, K. (2008). As can be seen: Lexical bundles and disciplinary variation. *English for specific purposes*, *27*(1), 4-21.

Milton, J. C., & Tsang, E. S. C. (1993). A corpus-based study of logical connectors in EFL students' writing: directions for future research. In *Studies in lexis. Proceedings of a seminar on lexis organized by the Language Centre of the HKUST, Hong Kong (Language Centre, HKUST, Hong Kong, 1993)*. Lin, Y.H. (2011). A Corpus-based Analysis of the Use of Lexical Bundles in English Academic Writing. (Master), National Taiwan Normal University, Taipei.

Römer, U.(2009).English in academia: Does nativeness matter. *Anglistik: International Journal of English Studies*, *20*(2), 89-100.

Wei, Y., & Lei, L. (2011). Lexical bundles in the academic writing of advanced Chinese EFL learners. *RELC journal*, *42*(2), 155-166.