Lexical bundles in academic bio-data: A corpus interdisciplinary analysis

Kathy Ling Lin (Shanghai Jiao Tong University, China) and Isaac Nuokyaa-Ire Mwinlaaru (Hong Kong Polytechnic University, Hong Kong)

In the investigation of academic English, compared with the much-studied genres such as research articles and degree theses, academic bio-data is an underexplored "para-genre" or "para-text" (Genette, 1987[1991]). This "para-genre", however, is very important and closely related to an academic's identity construction (i.e., claim membership of a disciplinary discourse community while presenting his or her own professional identity). To bridge the gap and establish its structural identity (or in Bhatia's (1993) term "generic integrity"), Mwinlaaru (in press) innovatively gave an SFL-based account of its generic structure by analyzing 200 biodata written by applied linguistics scholars, with 100 each from journal articles and seminar posters. The present project is a further extension in this line to examine cross-disciplinary similarities and differences in lexical bundle use in academic bio-data, and to associate the findings from bundles analysis with their structural variations and peculiarities across disciplinary boundaries.

To fulfill the research purpose, the authors collected 300 academic bios from 15 high-ranking prestigious journals in three disciplines based on Journal Citation Database and the recommendations from our disciplinary informants. The corpus of academic bio statements includes 100 each from Applied Linguistics (AL), Industrial Engineering (IE), and Physical Sciences (PS). The source journals for the three disciplines are: for PS, *Solid* State *Nuclear Magnetic Resonance* (SSNMR), *CHEMPHYCHEM* (CHEM), *The Journal of Physical Chemistry Letters* (JPCL), *Soft Matter* (SM), and *Chemical Physics Letters* (CPL); for AL, *Discourse & Society* (D&S), *Discourse Studies* (DS), *English for Specific Purposes*(ESP), *Language & Learning* (L&L), and *Text & Talk* (T&T); and for IE, *IEEE Transactions on Engineering* (IEEETE), IIE Transactions (IIET), *Industrial Management & Data Systems* (IM&DS), *Journal of Manufacturing Systems* (JMS), and *Journal of Product Innovation Management* (JPIM). The average lengths of academic bio statements for the three disciplines are respectively 73.9 words (Applied Linguistics), 88.7 words (Industrial Engineering), and 85.6 words (Physical Sciences). Before we conducted bundles analysis, rhetorical structural theory and SFL genre-based approach have been applied to describing the typical structure of this genre and their related cross-disciplinary variations, which could help account for the differences and the featured use of lexical bundles across the three contrasting disciplines. The findings from the structural analysis show that while Applied Linguistics scholars favour a two-tier contextual structure and an inventorying style of presentation in bios, Physical Science scholars prefer a three-tier structure and a chronicling mode. Industrial Engineering scholars lie between the two, preferring a three-tier structure and a synthesis of chronicling and inventorying styles. These disciplinary variations in the

rhetorical structure of biodata are realised by different choices in phraseological patterns in the bios.

In phraseology study, 4-word bundles analysis was conducted, as "the four-word scope is the most researched length for writing studies…manageable size for manual categorization and concordance checks" (Chen & Baker, 2010: 32). AntConc 3.4.4w was used, with cut-off points determined based on our data observation and a very close reading of the rich literature on bundles analysis (the cut-off standardized frequency: 0.6 times per thousand words, the raw cut-off frequency: 4 for AL, 5 for PS and IS; distribution: 5%). The study shows interesting cross-disciplinary commonalities and differences in four-word lexical bundle use in terms of their structure and function. In our talk, we will present in detail our findings on bundles analysis, e.g., the most frequently-used lexical bundles and their categories in terms of function and structure, to see how disciplinary variations in the rhetorical structure of biodata are realised by different choices in phraseological patterns in the bios. As an illustration, we made a table to show the findings from the comparative study of lexical bundle use in AL and PS bio-data (see Table 1). More interesting findings will be presented in our talk. Research contributions and implications for ESP teaching and research will be discussed.

Table 1 A case study: use AL and PS for a comparison

Bundles (tokens, range) in AL bios (14 types, 132 tokens)	**Bundles shared by both disciplines**	**Bundles (tokens, range) in PS bios** (39 types, 342 tokens)
at the university of (40, 39), in the department of (15, 15), her research interests include (13, 13), in the school of (9, 9), research interests include the (7, 7), his research interests include (6, 6), is the author of (6, 6), **university of hong kong (6, 5), from the university of (5, 5),** is a senior lecturer (5, 5), is an associate professor (5, 5), is associate professor of (5, 5), is senior lecturer in (5, 5), **the school of English (5, 5)**	**at the university of, in the department of, from the university of**	**at the university of (38, 30),** under the supervision of (22, 19), **from the university of (21, 20),** his ph.d in (18, 18), received his ph.d (15, 15), his research interests include (11, 11), **in the department of (11, 8),** the department of chemistry (11, 9), the university of california (10, 9), a ph.d in (9, 9), degree in chemistry from (9, 8), in chemistry from the (9, 9), ph.d in chemistry (8, 8), the max planck institute (8, 7), his ph.d from (7, 7), ph.d from the (7, 7), ph.d **in #** from (7, 7), received her ph.d (7, 7), received his b.sc (7, 7), b.s degree in (6, 6), he is **currently** a (6, 6), he received his ph (6, 6), her ph.d in (6,

		6), **in # from the university (6, 6),** in the laboratory of (6, 5), of science and technology (6, 6), ph.d degree in (6, 6), **after postdoctoral research in (5, 5),** and his ph.d (5, 5), as a postdoctoral fellow (5, 5), chinese academy of sciences (5, 5), **d in # from the (5, 5)**, d in physical chemistry (5, 5), his ph.d degree (5, 5), his research interests are (5, 5), **in # he received his (5, 5)**, obtained his ph.d (5, 5), of solid state nmr (5, 5), professor of chemistry at (5, 5), received his b.s (5, 5)

References

Bhatia, Vijay K. 1993. *Analyzing genre: Language use in professional settings*. London: Longman.

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

Genette, Gerard. 1987. Introduction to the paratexts. Reprinted in translation by M. Maclean (1991). *New Literary History* 22(2). 261–272.