

Lexical Bundles in Conversation across English varieties: A Core-Periphery approach

Zeping Huang

Hong Kong Baptist University

There has been no dearth of research on lexical bundles (i.e., recurrent multi-word sequences) from the perspective of register variation or disciplinary variation in academic contexts since the term was first coined by Biber et al. (1999). Yet, comparative research on lexical bundles in the context of world Englishes is relatively scarce. In light of this research lacuna, this study adopted Nelson's (2014) a "core-periphery" approach to investigate the core and peripheral lexical bundles in conversation, using data from the International Corpora of English (ICE) for four varieties of English: British, Canadian, Singapore, and Hong Kong English. Two research questions were addressed as follows:

1. What is the distribution of the core and peripheral bundles in conversation across the four varieties?
2. Can the core and peripheral bundles reveal any (dis)similarity in linguistic patterns among the four varieties in conversation? If so, what are the similarities and the differences?

To identify the core and the peripheral bundles, a list of three-word bundles with a minimum frequency of 25 per million words was retrieved from each corpus, making four lists in total. If a bundle occurs four times in the combined list, it means the bundle exists in all four corpora, and it is therefore an absolute core bundle (4-overlap). If a bundle occurs only once in the combine list, it means the bundle exists in one corpus only, and it is therefore an absolute peripheral bundle (non-overlap). Acronyms (e.g., *BBC*, *PhD*, or *NUS*), proper nouns (e.g., *Hong Kong*) were excluded, as they are content-based lexis rather than discourse building devices. Sound fillers such as *um*, *uh* and *mmm* were also excluded for analysis since they are not lexis but rather filled pauses. Lexical bundles with significant frequency in each variety were identified by the log-likelihood test (Rayson & Garside, 2000).

Through examining 3-word bundles overlapping in the four varieties, the quantitative and qualitative analyses unveiled (dis)similarities in the use of bundles across the four English varieties. The overlap and non-overlap comparisons enabled us to identify bundles with significantly high frequencies in each regional variety, suggesting speakers of each variety under scrutiny have their preferred patterns in constructing utterances, organizing discourse, sustaining conversational routine, and maintaining the flow of speech in face-to-face conversation. Our finding also shows a consistent correlation between the use of lexical bundles and the evolutionary stages of each English variety as defined in Schneider's (2007) Dynamic Model: the more advanced a variety is, the more similar its use is to the source variety British English in face-to-face conversation. Canadian English shared the most common ground with British English. As a nascent variety, Hong Kong English differed most from British English while its Asian counterpart Singapore English fell between Canadian English and Hong Kong English in the continuum of the evolutionary cycle. The "core-periphery" approach has helped producing generalizable insights into quantitative comparison, meanwhile allowed more fine-grained qualitative analysis to elucidate area of difference in the use of lexical bundles that have been overlooked by previous research.

Keywords: Core-periphery approach, lexical bundles, conversation, English varieties, Dynamic Model of New Englishes

References

- Biber, Douglas, Johansson Stig, Leech Geoffrey, Conrad Susan, and Finegan Edward. 1999. *The Longman grammar of spoken and written English*. London: Longman.
- Nelson, Gerald. 2014. "The circle of English: An exploration of the "core" and "periphery" of world Englishes". In E. Green & C. Meyer eds. *The variability of current world Englishes*. De Gruyter: Berlin/Boston, 99-119.
- Rayson, Paul, and Garside, Roger. 2000. "Comparing Corpora Using Frequency Profiling". *Proceedings of the workshop on comparing corpora held in conjunction with the 38th annual meeting of the Association for Computational Linguistics (ACL 2000)*. Hong Kong.
- Schneider, Edgar. 2007. *Postcolonial English: Varieties around the World*. Cambridge: Cambridge University Press.