Use of formulaic bundles by non-native English graduate writers and published authors in applied linguistics

CrossMark

Jingjing Qin*

Department of English Language Education, College of Education, Yeditepe University, Istanbul 34755, Turkey

ARTICLE INFO

ABSTRACT

This corpus-based cross-sectional study examines how advanced non-native English graduate students of applied linguistics at different levels of study used target formulaic bundles in their academic papers. First, five-unit formulaic bundles were extracted from a one-million-word reference corpus composed of 128 published articles in applied linguistics. Then the use of these target bundles was examined in 136 academic papers written by 20 non-native English graduate students of applied linguistics at four levels of study and 15 published articles by English-speaking expert writers in the same field. It was found that, as the level of study increased, students used a greater number and variety of target bundles. Specifically, non-native graduate students at the higher levels of study used more bundles characteristic of academic writing (e.g., *noun phrases with post-modifier fragments* or certain *prepositional phrases*) than those at the lower levels. Furthermore, the former group used more bundles as text organizers (e.g., *explanation*, *exemplification*, and *focus*) and stance bundles than the latter group. The pedagogical implications are suggested regarding what and how to teach non-native English graduate writers regarding the use of formulaic bundles.

1. Introduction

It is generally agreed that language is formulaic in nature, whether it is spoken or written (Ellis, 1996, 2008; Granger & Meunier, 2008; Pawley & Syder, 1983; Sinclair, 1991, 2004; Wray, 2002). Studies show that formulaic language plays a crucial role in academic writing, as it contributes to 21–52.3% of written discourse (Biber, Johansson, Leech, Conrad, & Finegan, 1999; Erman & Warren, 2000). Specifically, it has been observed that advanced and fluent writing is characterized by appropriate and frequent use of formulaic language, which also helps language users maintain identity in a disciplinary community; conversely, the absence of such formulaic language may indicate writers' inexperience or lack of expertise in an academic context (Bamberg, 1983; McCully, 1985; Wray, 2002).

Despite its importance, there is no consensus on how to define and identify formulaic language (Wray, 2008; Wray & Perkins, 2000). Due to its fuzziness, a plethora of terms have been used in the literature to refer to it, such as sentence stems (Pawley & Syder, 1983), prefabs or lexical phrases (Nattinger & DeCarrico, 1992), formulaic sequences (Schmitt & Carter, 2004), chunks (Sinclair, 1991), lexical bundles (Biber et al., 1999) and others, depending on how it is perceived and operationalized by different researchers. Wray (2008) has proposed four different sub-types of formulaic language and corresponding ways of identifying them, among which frequency seems to be more objective and clear-cut. This criterion, however, is not without problems, as it unavoidably excludes less frequent formulaic language which is nevertheless important. Although frequency should not be

* Tel.: +90 5394363802.
E-mail addresses: jingjingqin79@gmail.com, qinjingjing79@hotmail.com.

used as the sole criterion for the identification of formulaic language, it is effective for identifying the list of formulaic sequences which are most often used by L2 academic writers (Ellis, Simpson-Vlach, & Maynard, 2006 cited in Wray, 2008). Within this framework, Biber et al. (1999) defined formulaic language as the most frequently recurring sequences of words in a given genre, which they called 'lexical bundles,' as identified from computerized text analysis based on specified cut-off frequencies. The words in these lexical bundles occur together more often than would be expected by chance, and they may or may not constitute a complete syntactic unit (e.g., *as a result of, the extent to which, the fact that the*). They further categorized these bundles in terms of structural forms. For example, the majority of lexical bundles in academic writing take the form of *noun phrase with post-modifier fragment* and *preposition + noun phrase fragment*. In addition, the lexical bundles were found to serve important functions in various types of academic discourse, such as stance bundles, discourse organizers, referential expressions (Biber, Conrad, & Cortes, 2003, 2004; Cortes, 2004; Hyland, 2008a).

There is a growing awareness that the unnatural, unidiomatic nature of papers written by L2 students is due to a lack or misuse of formulaic language (Granger, 1998; Howarth, 1998; Meunier & Granger, 2008). In the field of EAP and L2 writing, researchers have shown great interest in understanding how the formulaic language used by L2 writers and native English-speaking writers differs. For example, Chen and Baker (2010) conducted both structural and functional analyses of lexical bundles in academic writing by Chinese EFL university students, native English-speaking university students and native expert writers. They found that the native English-speaking expert writers used the widest range of lexical bundles, whereas the Chinese students had the smallest and also overused certain lexical bundles. Both groups of student writers underused some lexical bundles compared to expert writers. Hyland (2008a) composed a corpus from published articles in four disciplines (electrical engineering, business studies, applied linguistics, and microbiology) and identified the most frequent four-word clusters. These were then compared to the four-word clusters identified in PhD dissertations and Master theses from the same four disciplines written by Chinese-speaking university students in Hong Kong. Interestingly, the number and range of four-word clusters employed by the graduate students exceeded those used by the published writers. Using the same data as in Hyland (2008a), Hyland (2008b) also found that Master students used more clusters than published writers, probably due to the pedagogic genre of these, where students were expected to display their research skills and mastery of disciplinary knowledge. In addition, Master students, doctoral students and published writers employed different clusters, with less than half of the 50 most common clusters overlapping among the three groups.

The aforementioned studies compare lexical bundles extracted from academic journal articles to texts written by L2 writers to examine whether the two groups use the same or different bundles. Approaching the issue of the use of lexical bundles from a different perspective, however, an important question remains relatively unexplored, namely whether the use of lexical bundles by novice L1 or L2 writers increasingly approximates target constructions in an academic field as they become more experienced. One of the few existing relevant studies was conducted in an L1 academic setting by Cortes (2004). She first identified four-word lexical bundles in published academic articles in the disciplines of history and biology, and then examined the use of these bundles in the writings of English-speaking university students at three levels of study (undergraduate lower division, undergraduate upper division and graduate level) in each discipline. Students at higher levels of study in biology were found to use more target bundles, especially in the use of text organizers and stance bundles, whereas students at different levels of study in history did not show much difference. Generally, student writers from both disciplines rarely used the target bundles and, even if they used them, their functions did not match those employed in published articles. Little research, so far, has been conducted in the usage of target lexical bundles by L2 academic writers.

The current study aims to investigate how L2 student writers at different levels of study in applied linguistics use target formulaic bundles in English-medium contexts. An unresolved question in L2 writing pedagogy is whether L2 writers can master formulaic bundles simply by being immersed in the target setting or whether focused instruction is needed to teach the formulaic bundles explicitly. The current cross-sectional corpus-based study attempts to contribute to answering this question by examining whether advanced L2 graduate student writers' use of lexical bundles in a specific discipline, applied linguistics, increases in line with their levels of study to approximate to expert writers' usage. First, five-unit target bundles in published applied linguistics research articles were identified. Then academic papers written by non-native graduate students at four levels of study and research papers written by native English-speaking academics were examined regarding the use of these target bundles, both in terms of structural forms and functions. This study was guided by the following three research questions:

1. What are the most frequent five-unit bundles in research articles in applied linguistics?
   a. What are the structural forms of these bundles?
   b. What are the functions of these bundles?
2. How do non-native graduate students and expert writers use these target bundles in terms of structural forms?
3. How do non-native students and expert writers use these target bundles in terms of functions?

2. Methods

2.1. Corpora

To investigate the three research questions listed above, a reference corpus and an analysis corpus were compiled. The one-million-word reference corpus, from which the target bundles were extracted, was composed of 128 research articles