I Think We Should…:
Investigating Lexical Bundle Use in the Speech of English Learners Across Proficiency Levels

Hengbin Yan, Guangdong University of Foreign Studies, Guangzhou, China

ABSTRACT

High-frequency recurrent word combinations known as lexical bundles are an essential component in the second language development. However, existing research on second language lexical bundle use has focused on writing proficiency, while oral proficiency has not received adequate attention. This study adopts a corpus-driven approach to the investigation of the speech of second language learners, comparing lexical bundle use across proficiency levels in several areas of interest including frequency, functional distribution and bundle fixedness. Results show that low-proficiency students tend to use significantly more context-dependent bundles than high-proficiency students, but do not differ in overall lexical bundle use. The patterning of lexical bundle use in non-native speech exhibits features that are typical in the register of classroom teaching. Additionally, the frequency and functional distributions of non-native speech share many similarities with those of non-native writing. Implications of the author’s findings are discussed in relation to previous studies.

KEYWORDS

Formulaic Language, Learner Corpora, Lexical Bundles, Second Language Development

1. INTRODUCTION

Researchers in Second Language Acquisition (SLA) have long been interested in high-frequency recurrent sequences of words known as lexical bundles (Biber et al., 1999, 2004). Compared with arbitrary word sequences, lexical bundles are distinguished by being: 1) extremely common; 2) usually not idiomatic in meaning and not perceptually salient and 3) usually structurally incomplete, but instead bridging structural units (Biber, 2009a). Lexical bundles are ubiquitous in the language use of both native speakers (NS) and non-native speakers (NNS) and have demonstrated considerable semantic and pragmatic utility (Conklin & Schmitt, 2008). Research in SLA has shown that lexical bundles can have great influence on language development in both early and more advanced stages (Ellis et al., 2011; Wray, 2002).

Since Biber et al. (1999, 2004), mainstream approaches to the study of lexical bundles in SLA have adopted corpus-driven approaches to automatically identify and track the development of lexical bundles in learner corpora. Earlier studies on second language bundle use have focused on contrastive studies on native vs. non-native corpora (Ädel & Erman, 2012; Chen & Baker, 2010; Hyland, 2008a; Römer, 2009). Recent years have seen a rise in interest in comparative research on learner corpora across different proficiency levels (Appel & Wood, 2016; Chen & Baker, 2016; Staples et al., 2013), which has mostly centered on NNS academic writing, presumably because the majority of scored learner data currently available to researchers are in the form of writing in

DOI: 10.4018/IJTIAL.2019070105

Copyright © 2019, IGI Global. Copying or distributing in print or electronic forms without written permission of IGI Global is prohibited.
academic contexts. On the other hand, NNS speech, while acknowledged as a mode complimentary to and with characteristics drastically different from writing (Biber et al., 2004), has not been duly investigated across proficiency levels.

To fill this gap, we explore in this study the lexical bundle used in the speech of English-as-a-Foreign-Language (EFL) students across proficiency levels. Utilizing a corpus of scored NNS speech taken from a large-scale, officially administered speaking test in China, we investigate the frequency, functions and degrees of bundle fixedness to reveal overall distribution and potential patterning in learner speech across different proficiency levels. Through such empirical investigations, this study has the potential to enrich our understanding of the language features of EFL learners at different developmental stages of language learning and may prove valuable to improving language instruction through targeted teaching of various forms of formulaic language.

2. LITERATURE REVIEW

2.1. Spoken Bundles

The majority of comparative studies on the lexical use in SLA have been conducted on academic writing (e.g. Chen & Baker, 2010; Chen & Baker, 2016; Hyland, 2008). Of the relatively few recent studies that explored spoken lexical bundles, most have focused on contrasting native and native bundle use. Shirato and Stapleton (2007) compared the lexical bundle use of a NNS speech corpus with an established NS corpus, and found significant differences in the overuse and underuse of NNS vocabulary that deviate from NS norms including discourse markers and modal items. Sánchez Hernández (2013) examined 4-word lexical bundles in corpora of university students, both native and nonnative. In addition to structural and functional differences, the study revealed a larger number of lexical bundles in non-native speech. Comparing two non-native subcorpora during different stages of interaction, it was also found that classroom instruction can lead to increased use of lexical bundles (although no statistical tests have been performed).

One potential problem with the above approaches employing NS and NNS corpora is that such corpora are often collected in different contexts/registers with numerous uncontrollable variables, leading to reduced corpus comparability and difficulty in accounting for the observed differences (Gablasova et al., 2017a). From a developmental perspective, these approaches are also unable to provide insights into how NNS students of different proficiencies differ in the use of lexical bundles.

2.2. Proficiency Levels

To answer the above questions in learner language development, several studies have explored non-native language use across different proficiency levels. As proficiency information is not always readily obtainable, some have resorted to indirect variables to differentiate proficiency (e.g. age, grade, nativeness), while others have used candidate responses to high-stakes exams to investigate distinguishing features in learner language performance (Hawkey & Barker, 2004).

Previous studies on the relationships between proficiency levels and frequency of lexical bundle use have presented complex and often conflicting results. Chen and Baker (2010) reported that the percentage of bundle use in academic writing increases as proficiency levels advance. However, the token and type frequencies in this study have not been normalized and no statistical tests have been performed for the comparison. Similarly, Ådel and Erman (2012) found that NNS undergraduate used fewer types of lexical bundles than NS counterparts (token frequencies, however, were not reported). Huang (2015) reported that advanced learners (senior English majors) used two and a half times more bundles (both tokens and types) than less advanced counterparts (junior English majors).

However, there are also contrary findings in a number of studies. De Cock (2000) failed to confirm the hypothesis of “learner individual bricks and NS prefabricated sections” and reported higher frequencies of lexical bundle use by learners than NS speakers in interview speech. In writing,
EFL Learning Beyond the Wall with MALL: College Students’ Perceptions
Mohialdeen Alotumi (2020). Enhancements and Limitations to ICT-Based Informal Language Learning: Emerging Research and Opportunities (pp. 138-160).
www.igi-global.com/chapter/efl-learning-beyond-the-wall-with-mall/242380?camid=4v1a