Allameh Tabataba'i University

Faculty of Persian Literature and Foreign Languages

Department of English Language and Literature

A Contrastive Analysis of Lexical Bundles in Applied Linguistics Theses Written in English by Persian and English Students

A Thesis Submitted in Partial Fulfillment of the Requirements for the Degree of Master of

Arts (M.A.) in Teaching English as a Foreign Language (TEFL)

Advisor: Dr. Shahin Vaezi

Reader: Dr. Hossein Mollanazar

By

Qasem Barkhi

Tehran, Iran

June 2010

فرم گرد آوری اطلاعات پایان نامه

کتابخانه مرکزی دانشگاه علامه طباطبائی

عنوان: بررسی مقابله ای خوشه های واژگانی در پایان نامه های زبانشناسی کاربردی نوشته شده توسط دانشجویان ایرانی و انگلیسی **A Contrastive Analysis of Lexical Bundles in Applied Linguistics Theses Written in English by Iranian and English Students**
نویسنده/محقق: قاسم برخی
مترجم: خیر
استاد مشاور: جناب آقای دکترحسین ملانظر   **استاد راهنما:** سرکارخانم دکتر شهین واعظی **استاد داور:** سرکارخانم دکتر فهیمه معرفت
کتابنامه: **واژه نامه:**
نوع پایان نامه: بنیادی ☐   توسعه ای ☐   کاربردی ☑
مقطع تحصیلی: کارشناسی ارشد   **سال تحصیلی:** 1389
محل تحصیل: تهران   **نام دانشگاه:** علامه طباطبائی   **دانشکده:**ادبیات فارسی و زبان های خارجی
تعداد صفحات: **گروه آموزشی:**زبان و ادبیات انگلیسی
کلید واژه ها به زبان فارسی: 1.خرد گونه 2. زبان پیش ساخته 3 . خوشه های واژگانی
کلید واژه ها به زبان انگلیسی: 1. Sub-genre 2. Formulaic Language 3. Lexical Bundles

چکیده

الف: موضوع و طرح مساله (اهمیت موضوع و هدف):

از آنجا که تحقیقات اخیر دال بر نقش خوشه های واژگانی در عملکرد بهتر در محیطهای دانشگاهی و در بافتهای گفتاری و نوشتاری از لحاظ سازماندهی کلام و انتقال بهتر و واضح تر مفهوم دارند، محقق بر آن داشته شد که مقایسه ای بین دانشجویان ایرانی و انگلیسی انجام دهد تا به بررسی این نکته بپردازد که آیا فراوانی استفاده از این اجزای بسیار مهم کلامی و نحوه کاربرد آنها در پایان نامه های دانشجویان ایرانی و انگلیسی به هم تفاوت دارد و اینکه آیا این تفاوت معنا دار است یا خیر تا مشاهده کنیم که آیا نحوه آموزش و منابع استفاده شده برای تدریس مهارت نوشتن در محیطهای دانشگاهی کارا بوده اند یا خیر. در وهله بعدی این سوال مطرح بود که آیا خرد گونه های پایان نامه های دانشجویان انگلیسی از قبیل مقدمه، روش شناسی، و نتیجه گیری از لحاظ فراوانی استفاده از خوشه های واژگانی و نحوه کاربرد آنها تفاوت معنا داری وجود دارد یا خیر که به درک بهتر ما از این سه خردگونه می توانست کمک شایانی بکند.

ب:مبانی نظری شامل مرور مختصری از منابع ،چارچوب نظری و پرسش ها و فرضیه ها:

پرسش های تحقیق:

1- آیا تفاوت معناداری بین پایان نامه های زبانشناسی کاربردی نوشته شده به زبان انگلیسی توسط دانشجویان ایرانی و انگلیسی از لحاظ فراوانی کاربرد ونوع کاربرد خوشه های واژگانی وجود دارد؟

2- آیا تفاوت معنا داری بین سه خرد گونه پایان نامه (مقدمه، روش شناسی، و نتیجه گیری) از لحاظ فراوانی کاربرد و نوع کاربرد خوشه های واژگانی وجود دارد؟

فرضیه های تحقیق:

1. تفاوت معناداری بین پایان نامه های زبانشناسی کاربردی نوشته شده به زبان انگلیسی توسط دانشجویان ایرانی و انگلیسی از لحاظ فراوانی کاربرد ونوع کاربرد خوشه های واژگانی وجود ندارد.

2- تفاوت معنا داری بین سه خردگونه پایان نامه (مقدمه، روش شناسی، و نتیجه گیری) از لحاظ فراوانی کاربرد و نوع کاربرد خوشه های واژگانی وجود ندارد.

پ: روش تحقیق شامل تعریف مفاهیم، روش تحقیق، جامعه مورد تحقیق ،نمونه گیری و روش های نمونه گیری، ابزار اندازه گیری، نحوه ی اجرای آن، شیوه ی گرد آوری و تجزیه و تحلیل داده ها:

جامعه آماری این تحقیق دربرگیرنده پایان نامه های رشته زبانشناسی کاربردی بودند که به زبان انگلیسی توسط دانشجویان ایرانی محصل در دانشگاههای قابل دسترس در تهران و پایان نامه های دانشجویان انگلیسی دانلود شده از اینترنت بود. پایان نامه های انگلیسی بر اساس بیوگرافی دانشجویان و دانشگاه آنها انتخاب شد. در قسمت دوم تحقیق برای بررسی تفاوت خردگونه ای فقط از پایان نامه های دانشجویان انگلیسی استفاده شد. برای شمارش و استخراج خوشه های واژگانی از نرم افزار (کی اف ان گرام) نسخه سال 2007 استفاده شد و برای تحلیل نوع کاربردی آنها از چهارچوب کاربردی سال 2008 هایلند استفاده شد. برای تشخیص تفاوت معنادار برای مقایسه پایان نامه های

دانشجویان ایرانی و دانشجویان انگلیسی از آزمون آماری تی تست استفاده شد و برای مقایسه خردگونه های پایان نامه ها از آزمون آنوای یک طرفه و شفی استفاده شد.

ت: یافته های تحقیق: نتایج نشان داد که تفاوت معناداری بین دانشجویان ایرانی وانگلیسی از لحاظ فراوانی کاربرد و نوع کاربرد خوشه های واژگانی وجود دارد. دانشجویان انگلیسی از خوشه های واژگانی بیشتری استفاده کرده بودند و خوشه های واژگانی مشارکت محور در پایان نامه های دانشجویان انگلیسی بیشترو خوشه های واژگانی تحقیق محور در پایان نامه های دانشجویان ایرانی بیشتر استفاده شده بود. در بین خردگونه های پایان نامه های دانشجویان انگلیسی نیز تفاوت معنا داری وجود داشت و فراوانی خوشه های واژگانی در روش شناسی بیشتر از دو خردگونه دیگر بود. نوع کاربردی تحقیق محور در روش شناسی بیشتر و نوع مشارکت محور و متن محور در قسمت نتیجه گیری بیشتر استفاده شده بود.

ث: نتیجه گیری و پیشنهادات: بر اساس یافته های تحقیق می توان نتیجه گرفت که دانشجویان ایرانی در استفاده از خوشه های واژگانی و نحوه کاربرد آنها از نسبت به دانشجویان ایرانی مهارت کمتری دارند و می بایست در مطالب درسی برای آموزش مهارت نوشتن به این اجزای مهم کلامی توجه بیشتری مبذول شود. تفاوت بین خردگونه ها نیز دال بر این نکته است که با آگاه ساختن دانشجویان به این تفاوتها و بررسی خوشه های واژگانی مختص هر خردگونه و تدریس آنها در کلاسهای آموزش مهارت نوشتن ممکن است باعث تسهیل وتسریع مهارت نوشتن در محیطهای دانشگاهی شود.

صحت اطلاعات مندرج در این فرم بر اساس محتوای پایان نامه و ضوابط مندرج در فرم را گواهی می نمایم.

نام استاد راهنما: سرکارخانم دکتر شهین واعظی

سمت علمی: استادیار

امضاء استاد راهنما:

رئیس کتابخانه:

Acknowledgments

I gratefully acknowledge the help of the following people, without whose help and support this thesis would have not been written:

First of all, my sincere thanks go to Dr. Vaezi and Dr. Mollanazar, my advisor and reader respectively, for carefully reading my thesis and giving insightful comments and guidelines.

Second of all, I really want to seize the chance to express my deep gratitude and appreciation to my professors in M.A. program including Dr. Khatib for helping us to form an informed view on teaching and learning theories, Dr. Tajeddin for familiarizing us with the fundamental and new concepts in discourse, Dr. Mostafaei for her Syllabus Design and Materials Development classes, Dr. Marefat for her really informative Testing and Research classes, and Dr. Fahim for his Linguistics class.

I would also like to thank my professors in B.A. program including Dr. Ghaemi, Dr. Akef, and Dr. Behroozi for their encouragement, inspiration, and help to find my way and keep on going forward.

I would also like to gratefully express my gratitude to my dear friends F. Tabandeh for sharing his thoughts and ideas and R. Dashtestani for helping me with statistical procedures.

Last but not least, my deepest thanks go to my parents for their unconditional love and support and for their endless sacrifice and encouragement.

Abstract

In recent decades, the focus of register analysis has been on the study of word clusters and their functions in different written and spoken genres through corpus-based studies. Lexical bundles are one kind of multi-word expressions which appear more frequently than by chance which seem to contribute to shaping meaning and introducing coherence in specific contexts more vividly. As the frequent use of these discourse elements in language production has been asserted as a sign of language proficiency, this study embarked on investigating the question whether there is a gap between Persian and English M.A. students majoring in ELT in terms of frequency of form and functional types of lexical bundles used in their theses. The corpus of 847000 words consisted of 40 theses, 20 of which written by Persian students. Another question was also raised to inspect the frequency of form and functional type of bundles in three sub-genres of thesis, namely introduction, methodology, and conclusion in English students' theses. The 189000 word corpus was consisted of 60 'introduction', 'methodology', and 'conclusion' parts. The findings of the first part show the higher frequency of use of lexical bundles by English students and different ratio of functional types in the Persian and English students' theses. In sub-genres, the highest frequency of form was witnessed in 'Methodology' part and different ratios of functional types of bundles were witnessed in all three sub-genres. Some implications are drawn for teaching writing in academic settings.

Table of Contents

Chapter One: Introduction

Chapter 2: Review of the Literature

Chapter 3: Methodology

Chapter 4: Results and Discussion

Chapter 5: Conclusion, Implications, and Suggestions for Further Research

References

List of Tables

Table 4.29. Scheffe test to locate the difference in P-O bundles of three sub-genre corpora 79

List of Figures

List of abbreviations

ANOVA: Analysis of Variance

EAL1: English as L1 (first language)

PAL1: Persian as L1 (first language)

Introduction

1.1. Overview

In the current chapter, attempts have been made to first have a concise introduction to the significant attention the study of lexical structures has been gained in different domains of study, e.g. in register, genre-based and discourse analyses, emphasizing the fact that much of such research has centered on single words rather than strings of words, namely lexical bundles. Then, the gap to be filled between a battery of current research focusing on bundles used by English natives and required research on the use of such bundles by English non-natives has been accentuated. Afterwards, the research questions and null hypotheses to address the gap have been stated. Finally, the most noteworthy terms frequently referred to in this study have been defined and briefly elaborated on.

1.2. Introduction

The systematic investigation of the lexical patterns and structures in different text types goes back to early register analysis. Register analysis with an exclusive focus on the words and sentences later gave way to discourse analysis with focus on stretches of sentences unveiled in real language use (Widdowson, 2007). Later on, genre analysis gained ground focusing on the linguistic characteristics of communicative events whose members share the same communicative purpose (Swales, 1990). The study of the rhetorical structures and functional moves of written academic genres have been prevalent in the past fifteen years (Hyland, 1996, 1998, 1999, 2004, 2005; Hyland & Tse, 2005; Samraj, 2005; Van Bonn & Swales, 2007; Aktas & Cortes, 2008).

Most of the research in this direction has put emphasis on the study of single words or functional moves between different academic disciplines' genres and sub-genres and less on the strings of words occurring in different genres, despite the fact that multi-word expressions (also called formulaic language) play an unquestionably significant role in fluent written language production (Hyland, 2008a).

Such multi-word expressions have been studied under many names, including 'lexical phrases', 'formulas', 'routines', 'fixed expressions', and 'prefabricated patterns' (Biber, 2006). The diversity of given names shows different underlying approaches each describing the subject under study in specific ways, and so they offer different perspectives on the use of multi-word sequences. As an example, some studies describe multi-word sequences that are idiomatic (e.g., idioms like *in a nutshell*), whereas other studies focus on sequences that are non-idiomatic but perceptually salient (e.g., *if you know what I mean*). Even though a bulk of research has been conducted into the analysis of these multi-word expressions (all having complete structural units) under different names and focuses of study, non of them systematically analyzed "structurally incomplete" salient strings of

words known as "lexical bundles" which happen to occur in discourse more frequently than by chance, and help to form the text meaning and lead to better understanding of particular genres (Biber et al. 1999; Biber & Conrad, 1999; Marco, 2000; Biber & Reppen, 2002; Biber et al. 2004a, 2004b; Hyland, 2008; Durrant, 2009; Vongpumivitch, Huang & Chang, 2009).

1.3. Statement of the Problem

Lexical bundles have obtained much attention in corpus studies based on "frequency-driven" approaches in general spoken and written discourse from one hand (Salem, 1987; Aletenberg, 1998; Biber et al. 1999) and in ESP/EAP genre-based domains of study from the other hand (Cortes, 2002, 2004; Levy, 2003; Biber et al. 2004a, 2004b; Biber, 2006; Chen, 2008; Hyland, 2007, 2008a, 2008b). Whilst most of such studies centered on the analysis of lexical bundles used in English academic corpus, few studies have focused on the effects of L1 on the use of lexical chunks (or bundles) in English as an L2 or the role of ESL students' proficiency on the use of such bundles (e.g. Levy, 2003; Prodromou, 2003a, 2003b; Lindstromberg & Boers, 2008; Tabandeh, 2009). According to Haswell (1991), the fluent use of lexical bundles contributes directly to the native-like performance of non-native writers and reflects their familiarity with particular genres, and as Adolphs & Durow (2004) mention, students' more integration to L2 environment help them acquire and use such formulaic language more naturally. In addition, while it has been shown that lexical bundles also vary in different sub-genre parts of research articles (Tabandeh, 2009), to probe such

possible variations in other main academic genres, e.g. MA theses, might prove fruitful to the study of such bundles and their functional significance in academic texts.

1.4. Significance and Purpose of the Study

On the one hand, as Hyland (2008a) suggests, research into the use of lexical bundles should reflect on the role of writers' L1 in the English as L2 writing performances. On the other hand, according to Haswell (1991), the frequent use of such bundles shows how much the L2 writer is able to speak his thought fluently and naturally in particular genres, or to what degree students are integrated into L2 environment (Adolphs & Durow; 2004). Therefore, this study seeks to investigate the possible variations between applied linguistics MA theses written in English by students with Persian as their L1 (PAL1) and students with English as their L1 (EAL1) to see whether there is any variation due to the writer's L1. Such an analysis would throw some light on the effects of L1 on L2 writing performances and the possible need to raise L2 writers' awareness of the use and the functions of lexical bundles in EAP academic settings. Moreover, the study of the use of lexical bundles and their related functions in sub-genres of MA theses, i.e. introduction, methodology, and conclusion parts, paves the way for understanding how lexical bundles are distributed frequently and functionally in one of the major genres of academic writing, namely MA theses.

1.5. Research Questions

The following research questions were posed to investigate in this study:

1. Is there any significant difference between MA theses written in English by   students with

English as their L1 (EAL1) and students with Persian as their L1 (PAL1)   in terms of the frequency of

the forms and   functional types of lexical bundles?

2. Is there any significant difference among sub-genres of MA theses (introduction, methodology,

and conclusion) in terms of frequency of use and functional types of lexical bundles?

1.6. Null-Hypotheses

Based on the above mentioned research questions, the following null-hypotheses were

formulated:

1. There is no significant variation between MA theses written in English by students with (EAL1) and

students with (PAL1) in terms of the frequency of the forms and   functional types of lexical bundles.

2. There is no significant variation among sub-genres of MA theses (introduction, methodology, and

conclusion) in terms of the frequency of forms and the functional types of lexical bundles.

1.7. Definition of the Key Terms

Sub-genre: In a genre set, a particular micro-genre that along with other related micro-genres

comprises a macro-genre or a genre-set (Samraj, 2005), like research article as a macro-genre and

abstract and introduction parts as sub-genres.

Formulaic Language: it is "a sequence, continuous or discontinuous, of words or other meaning elements, which is, or appears to be prefabricated: that is, stored and retrieved whole from memory at the time of use, rather than being subject to generation or analysis by language grammar" (Wray and Perkins, 2000, p. 2). Some examples could be *at first sight* and *kick the bucket*.

Lexical Bundles: Defined by Hyland (2008a) as "extended collocations which appear more frequent than expected by chance, helping to shape meanings in specific contexts and contributing to the coherence of the text" (p. 4). Three characteristics make lexical bundles different from other lexical chunks: they are continuous (fixed in sequence), grammatically incomplete, and identified based on their frequency (Cortes, 2004; Biber et al. 2004; Biber et al. 2006; Hyland, 2008a). Some examples could be *in this study* and *the size of the*.

1.8. Limitations and Delimitations of the Study

The limitation the researcher is faced with is the unavailability of more sophisticated software created specifically to count and analyze lexical bundles. The researcher is going to analyze the lexical bundles by "kfNgram" software, which is a semi-professional software designed specifically to count lexical bundles in texts.

The researcher decided to exclude the review of the literature from corpora due to the point that it may be more heavily affected by works of others because the use of quotations-which are not writers' own language performance- is more frequent in this part compared with other parts and the exclusion of quotations was so time-consuming. It was also decided to exclude results section from sub-genre corpora, because the results section is full of tables and the number of words is really small. Furthermore, as the number of native speakers' theses that could be converted to text file

format was limited and in some theses there was no discussion part and it was merged with conclusion, it was decided to exclude this sub-genre from the study.

Chapter Two: Review of the

Literature