

# 基于语料库的国际硕士生作文中词块的对比研究

马笑麒

布里斯托大学教育学院, 英国 布里斯托

收稿日期: 2025年5月12日; 录用日期: 2025年7月7日; 发布日期: 2025年7月21日

## 摘要

在第二语言写作中, 词块(lexical bundles)的重要性日益受到重视, 已有研究探讨了二语写作中词块产生的特征与问题。然而, 针对英语母语学生在使用词块过程中的发展特征与准确性问题的研究仍较为有限。本研究收集了29名国际学生在其研究生阶段撰写的课程作业, 构建了一个包含超过8万个词的子语料库, 旨在分析这些就读教育专业的国际学生在写作时所使用词块的模式。研究在语料库中提取高频词块, 并依据Biber等人提出的分类体系及母语者写作中的目标词块, 将其划分为六种结构类型和三种功能类型, 进而对其频率与分布特征进行对比分析。研究结果显示: 在词块的使用数量方面, 国际学生整体使用偏少; 但从使用的多样性来看, 其词块的丰富性更高。在结构类型上, 国际学生在使用以介词短语构成的词块(如含有of-短语的介词结构)和以名词短语构成的词块(如含有后置修饰语的名词结构)方面的频率明显低于母语学生。本研究有助于提升二语学习者对词块规范使用的重视, 同时也为今后的写作教学中的目标设定与实践提供了有益启示。

## 关键词

第二语言写作, 基于语料库的研究, 词块, 对比研究

# A Corpus-Based Comparative Study of Lexical Bundles in Assignments Written by Master's Level International Students

Xiaoqi Ma

School of Education, University of Bristol, Bristol, England

Received: May 12<sup>th</sup>, 2025; accepted: Jul. 7<sup>th</sup>, 2025; published: Jul. 21<sup>st</sup>, 2025

## Abstract

Lexical bundles in second language writing have been increasingly emphasized, researchers have explored the characteristics and problems of lexical bundles output in second language writing. However, few studies have addressed the developmental characteristics and accuracy issues of lexical bundles in the use of lexical bundles by native English speaker students. This study included a corpus of 29 students' assignments writing during their postgraduate studies, creating a sub-corpus of over 80,000 tokens in order to explore the patterns of lexical bundles among the same cohort of international students majoring in education when they were writing their assignments. After extracting the high-frequency lexical bundles from the corpus, this study classified these bundles into six structural and three functional types based on Biber *et al.*'s taxonomy of bundles and the target bundles in native speakers' assignments writing, and comparatively analyzed their frequency and distributional features. The results show that international students under-use lexical bundles in terms of quantity, but the richness of lexical bundles is higher. In terms of structure, international students use PP-based bundles (prepositional phrase with of-phrase fragment and other prepositional phrase fragments) and NP-based bundles (noun phrase with of-phrase fragment and noun phrase with other post-modifier fragments) less frequently than native students. This study is conducive to increasing the importance of second language learners' correct use of lexical bundles, and it is also informative for the goal-setting and practicing process of writing teaching in the future.

## Keywords

L2 Writing, Corpus-Based Study, Lexical Bundle, Comparative Study

Copyright © 2025 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

## 1. 引言

在全球化背景下, 英语已成为一种国际学术语言[1], 在学术交流和知识传播中发挥着决定性作用。语言在高等教育和科学研究中的重要性体现在多个方面, 例如以英语为工作语言的国际学术期刊数量、以英语授课的专业课程中第二语言或外语学习者的数量, 以及非母语学者需要用英语进行大部分学术研究和学术交流的需求。随着越来越多的学生选择出国留学, 写作问题日益突出, 尤其是在海外留学生群体中表现得尤为明显。非母语写作者在如何用英语准确而有效地完成写作任务方面面临着各种困难, 这引发了关于如何教授非母语者正确使用语言以及正确完成写作任务的讨论。

许多研究的发现仅强调了公式化语言在话语中的重要性; 大量研究指出, 词汇搭配等短语语言单位的再现应成为英语教学中的一个重要组成部分。然而, 很少有研究列举出对特定学生群体具有教学价值的具体词汇搭配, 尤其缺乏对非母语者在日常写作中使用词汇搭配情况的详细调查。

本研究旨在从上述方面进一步探讨海外学生课后写作任务中的词汇搭配, 主要研究目标是考察母语为英语的作者与国际学生在写作中使用的典型词汇搭配的模式、意义和功能; 鉴于面向文本的词汇搭配是组织文本结构、构建话语衔接与连贯的手段, 是构建文章时不可忽视的意义单位, 本研究的目的是首先识别英语母语作者写作中最常使用的面向文本的词汇搭配, 并找出不同母语背景写作者在其结构和功能特征上的差异, 进而深入探讨英语母语作者与非英语母语作者在使用面向文本的词汇搭配时, 在组织

话语结构和促进文本连贯方面的异同。

## 2. 文献综述

本研究将聚焦于母语为英语的作者所写作业中使用的词汇搭配,以及与之不属于同一语言背景的海外学生和国际学生在其硕士学习阶段英语写作中的词汇搭配使用情况。

目前,大多数关于中国英语学习者的词块(*lexical chunking*)研究主要集中在本科阶段(包括英语专业和非英语专业)写作方面[2];而有关中国英语使用者学术写作的词块研究则主要集中于硕士论文和英文期刊发表文章,研究语料多集中于摘要和引言部分[3]。随着海外留学生数量的增加,尽管其中一些已成为较为成熟的英语学习者,但在英语写作方面,其写作水平仍未达到以英语为母语的论文作者的水平。因此,本研究将以非英语母语的海外学生所写作业为主要研究语料,尝试探讨国际学生英语写作中词汇搭配使用的特点。

鉴于词块在语言本体、语言加工和语言学习中的重要性,二语学习者需要掌握大量的词块,才能达到接近母语者的语言水平[4]。只有在掌握了母语者经常使用的多词组合,并且更重要的是这些表达被目标语社群所接受的前提下,学习者才能实现有效和流利的交流[5]。否则,如果缺乏这些固定的程序化表达,将不可避免地给人留下不地道、不自然的印象。因此,二语研究者特别强调应识别母语者最常使用的词汇组合,并将其教给二语学习者[5][6]。

基于学习者语料库的共现分析显示,母语者与非母语者在搭配使用上存在差异,表现为非母语者的搭配使用不足、过度使用和误用等情况。多项学习者语料库研究发现,学习者在使用低频但紧密结合的词汇搭配时存在使用不足的现象[7]而过度依赖少量高频搭配[8],即 Nesselhauf 所称的“最简单的搭配类别”(如 *have a chance, have a look, have time, have a problem, take care of*) (Nesselhauf, 2005)。Nesselhauf [9]在对德语背景英语学习者写作中 2000 个动词-名词搭配进行调查时发现,其中有相当一部分使用不当,最常见的错误类型是动词、名词和介词的选择不当(如 \**make homework* 替代 *do homework*, \**life quality* 替代 *quality of life*) [10]。

另一方面,许多研究者使用 N-gram 分析和词串识别方法研究学习者语言的短语特征,发现学习者在某些类型的词序列使用上明显多于母语者[11]。这种现象可能与学习者词汇能力有限有关(Granger & Paquot, 2008)。此外,学习者在词序列的结构、功能和使用方式方面与母语者也表现出更大的差异[7][11]。

词汇组合等短语序列在口语和书面语的语言产出和语言学习中都起着重要作用。Pawley 和 Syder [7]指出,语言的流利和地道表达在很大程度上依赖于对语言中成块和程序化表达的掌握。同时,一些研究者也认为,能熟练使用词汇组合等短语结构,是语言使用者在特定领域中语言熟练度的重要表现,如学术写作[12]。虽然前述基于学习者语料库的短语学研究大多聚焦于一般层次的英语学习者和第二语言学习者,但近年来已有越来越多的研究关注具有一定学术能力的非母语作者在学术词汇序列使用方面的表现[3][11][13][14]。事实上,大多数国际学生在出国前已接受过英语语言训练。本研究正是聚焦于非母语作者在英文写作中的表现。

以往研究比较了母语者和二语学习者在词汇串使用上的差异[15]-[17]。Shin (2019)从母语和非母语学生的大学作文中提取词汇串,发现非母语者的使用较不多样,且更倾向于使用动词密集型和人称化表达。

## 3. 方法论

### 3.1. 研究问题

本研究旨在识别国际学生与以英语为母语者在使用词块方面存在的问题。根据研究目标,以下研究问题帮助明确研究重点:

- (1) 国际学生与英语母语者在作业写作中最常用的四词词块和三次词块有哪些?
- (2) 国际学生与英语母语者在词块使用上的结构性差异有哪些?

### 3.2. 语料库构建

为了实现研究目标, 本文收集了某英国大学国际学生课程作业来构建语料库。所有写作样本均来自某英国大学教育学专业的研究生。他们在出国留学前基本都接受过系统的英语学习, 每学期课程结束时, 需完成课程作业。本文选择了他们在布里斯托学习期间所完成且得分在 50~70 分之间的所有作业。在此基础上建立了一个新的语料库, 共计 29 篇文章, 词汇总量为 80,558 个词项。

### 3.3. 词块识别

词块分析的第一步是建立一份英语母语者使用的高频词块清单, 这些词块对学术英语的学习与教学具有重要价值。在 Cortes 关于历史与生物学领域中, 已发表论文与学生写作中词块对比的研究中, 她提出的方法是将英语母语语料库中最频繁的词块识别为“目标词块”(target bundles), 并进一步考察非母语语料库中这些词块的使用情况[12]。本文将采用类似方法, 对非母语者语料库中“目标词块”的使用情况进行考察。因此, 本研究中, “目标词块”一词也将按照 Cortes 的方法使用[12]。英语母语语料库“Louvain Corpus of Native English Essays (LOCNESS)”将作为本研究的目标词块来源。

根据 Biber 等人(2000)对词块的定义本文所指的词块为高频且重复出现的词语序列[18]。由于二词词块数量庞大, 且多为固定搭配, 本文并不将其纳入词块研究范围。Simpson-Vlach 和 Ellis 研究表明, 三词词块在教学中具有显著价值, 他们通过频率标准和心理验证方法, 得出了一份具有教学意义的学术语言词块清单, 其中前 50 个核心词块多为三词结构[19]。Biber 在对大学课堂对话与教材中的词块研究中, 为了控制研究范围, 主要关注四词词块[20]。他认为三词词块仍然数量过多, 而五词或六词词块出现频率较低, 且多为四词词块的延展形式。此外, Hyland 亦在学术写作中研究了四词词块。因此, 为了获取更全面的词块列表, 本文将词块长度设置为三到四个词, 即 N-gram 中的  $3 \leq N \leq 4$  [21]。

### 3.4. 频数标准

词块(Lexical bundles)最初被定义为在某一语篇领域中高频重复出现的词语序列[18]。频率标准既界定了基于语料库的词汇-句法模式的研究对象, 也反映了词语序列作为预制块被储存和使用的程度, 高频序列更可能被存储并作为程序化语言提取使用。

在本研究中, 学生的作文在收集与转码之后, 手动删除了如标题和注释等无效信息。随后将所有文本分段, 并使用 AntConc (Anthony, 2006)软件观察其索引行, 以分析目标词汇串的结构特征。从语料库中提取出最常用的三词串和四词串列表。为了确保最终入选的词汇串确实是因为在多个文本中被频繁使用, 而非因个体偏好或任务依赖等偶然因素, 筛选词汇串时遵循以下四项原则:

1. **保守的频率下限:** 至少出现 3 次(即每百万词中 40 次) [22];
2. **广泛的分布度:** 至少出现在 5 篇文章中, 且涉及多个写作任务;
3. **较高的互信息值(MI):**  $MI \geq 3$ , 代表词汇在特定语境中一起出现的概率;
4. **独立性:** 不与其他短语重叠或组成五词及以上的长串(例如: “they do not want” 和 “do not want to”)。

基于上述标准, 最终从国际学生语料中筛选出 162 个高频词汇串, 其中四词串为 14 个。

### 3.5. 特征分析

本研究通过比较国际学生语料库与母语者语料库的词汇串, 在频率、结构、多样性和功能等方面进

行分析。由于两组语料在文本规模上存在差异,即词量和文本数量不同,因此不适合直接进行对比分析。为此,研究通过比较以下指标来说明不同年级学生所写语篇在词汇层面的多样性:**Token**: 词汇总出现次数;**Type**: 不同词块的数量;**标准类型/词频比(Type/Token Ratio, TTR)**: 用于衡量语篇词汇多样性。TTR 值越高,说明重复使用的词块越少,语篇在词汇层面的多样性越高。进一步分析中,按照词块分类的主流方法,本研究借鉴以往研究将词汇串分为六类结构类别:1. 基于名词短语(NP-based bundles);2. 基于介词短语(PP-based bundles);3. 基于动词短语(VP-based bundles);4. 从句片段(Clause fragments);5. 修饰语(Modifiers);6. 连接类词汇串(Linking bundles) [18] [23]。

## 4. 研究结果

### 4.1. 最常用的四词与三词词块

本节将重点分析国际学生与英语母语者在课程作业中最常使用的四词词块与三词词块。从表 1 中可以看出,在四词词块方面,母语者语料库(LOCNESS)共包含 31 个不同的四词词块,总出现频率为 333 次。而国际学生语料库共包含 14 个不同的四词词块,总出现频率为 167 次。在三词词块方面,LOCNESS 包含 299 个不同的三词词块,总出现频率为 4330 次;而国际学生语料库包含 162 个三词词块,总出现频率为 2160 次。

**Table 1.** Comparison of commonly used three-word bundles and four-word bundles

**表 1.** 常用三词词块和四次词块对比

	LOCNESS		International Students	
	Three-word bundles	Four-word bundles	Three-word bundles	Four-word bundles
Types	299	31	162	14
Tokens	4330	333	2160	167

#### 4.1.1. 目标词块(TBs)产出频率

首先,采用斯皮尔曼相关分析来探究学生与母语作者在目标词块(TBs)产出频率之间的关系。结果显示,在四词词块中,国际学生与母语者在 TBs 的使用频率之间没有显著相关性,但在三词词块的使用上,二者之间存在显著相关性:相关系数为 0.354 ( $P = 0.022$ )。这表明在三词词块的使用上,国际学生作业中 TBs 的产出与频率在很大程度上可以通过母语者学术写作中的 TBs 产出与频率进行预测。

#### 4.1.2. 目标词块(TBs)产出分布

随后,为了进一步检验两个语料库在 TBs 结构和功能分布上的差异,研究根据前一章所介绍的分体系,将所有 TBs 划分为六种结构类型(介词短语型 pp-based、名词短语型 np-based、动词短语型 vp-based、句子片段、连接词块和修饰词块)和三种功能类型(指称表达、篇章组织和立场表达),并对两个语料库中所有词块的产出率进行了比较。结果显示,在词块类型与频率方面,两个语料库之间存在显著差异。英语母语学生在三词与四词词块的使用数量上均高于国际学生。

此外,对于两个语料库中均有出现的词块,母语者的使用频率也普遍高于国际学生。例如,在三词词块中,母语者最常用的词块出现了 88 次,而国际学生语料库中仅为 43 次;在四词词块中,母语者最常用的词块出现了 49 次,而国际学生为 27 次。

另外,从收集到的数据(表 2)可以看出,国际学生四词词块的类型-标记比(TTR)为 10.17,而母语者为 9.30。从表 3 来看,三词词块的 TTR 值国际学生为 8.37,而母语者为 6.90。 $TTR(\text{type-token ratio}) = \text{类}$

型数/标记数  $\times 100\%$ ，是衡量语料词汇密度的指标。从公式可以看出，TTR 值越高，语料中的词汇丰富度与多样性越高。因此，相较于母语者，国际学生在课程作业写作中所使用的词汇表现出更高的丰富度与多样性。这个结果表明，国际学生具备良好的英语运用能力，善于使用不同的词块表达相同的意义，例如 *moreover* (此外)、*furthermore* (进一步)和 *besides* (同样地)等。

**Table 2.** Comparison of four-word bundles

**表 2.** 四词词块对比

Four-word bundles	LOCNESS	Students
Types	31	17
Tokens	333	167
Ration	9.30	10.17

**Table 3.** Comparison of three-word bundles

**表 3.** 三次词块对比

Three-word bundles	LOCNESS	Students
Types	299	181
Tokens	4330	2160
Ration	6.90	8.37

## 4.2. 四词词块与三词词块结构分类的比较

### 4.2.1. 四词词块的分类比较

本节将从结构的角度出发，探讨两个语料库之间的差异。根据 Biber 的分类标准，在剔除部分无意义的词块后[18]。表 4 则展示了两个语料库中不同类别四词词块的比例。

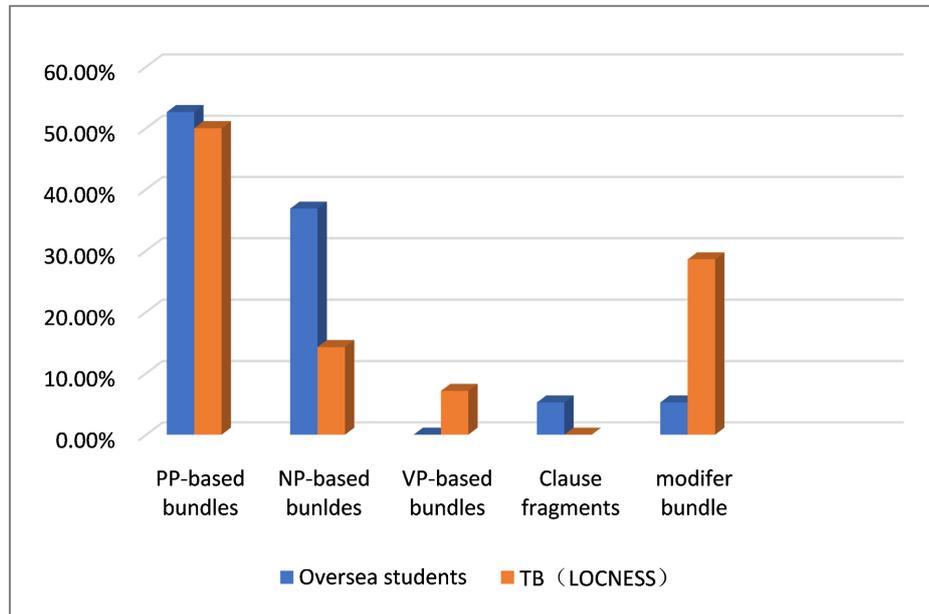
**Table 4.** Differences in the structure of four-word bundles

**表 4.** 四词词块结构差异

Structural types	LOCNESS		Students	
PP-based bundles	10	52.63%	7	50%
NP-based bundles	7	36.84%	2	14.28%
VP-based bundles	0	0%	1	7.14%
Clause fragments	1	5.26%	0	0
Modifier bundle	1	5.26%	4	28.57%
Total	19		14	

如图所示，母语为英语的学生在使用中占比最高的类型是介词短语、以介词短语(pp-based bundles)和名词短语(np-based bundles)为基础的词块。这些高频词块都是论文写作中典型的结构。

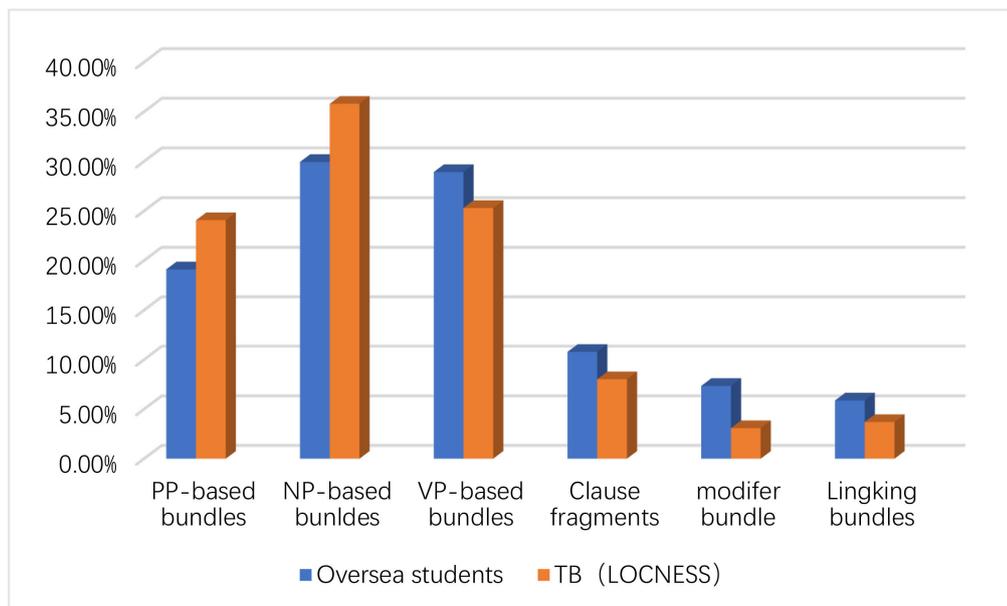
通过对四词词块的比较分析(见图 1)，可以看出国际学生在介词短语的使用上与母语学生基本相同，但在名词短语的使用上，母语学生明显高于国际学生。而国际学生则更倾向于使用由 *it is* 引导的从句类修饰词块，如 *it is necessary*、*it is important* 等。



**Figure 1.** Comparative analysis of four-word bundles  
**图 1.** 四词词块比较分析

#### 4.2.2. 三词词块分类比较

通过使用 AntConc 软件对数据进行分析, 表 5 显示了国际学生和以英语为母语的学生在三词词块应用方面的情况。如图 2 所示, 在最常用的三词词块中国际学生与以英语为母语的学生使用比例几乎相同。该结果与四词词块的使用情况略有不同, 这种差异可能源于四词词块样本量较小。从图中可以看出, 除了介词短语(pp-based bundles)和名词短语(NP-based bundles)使用频率低于以英语为母语的学生外, 国际学生在其他类型的词块使用频率普遍高于母语者。



**Figure 2.** Comparative analysis of three-word bundles  
**图 2.** 三词词块比较分析

**Table 5.** Comparative analysis of three-word bundles  
**表 5.** 三次词块比较分析

Structural types	LOCNESSS		Students	
PP-based bundles	39	19.11%	39	24.07%
NP-based bundles	57	29.94%	58	35.8%
VP-based bundles	59	28.92%	41	25.30%
Clause fragments	22	10.78%	13	8.02%
Modifier bundle	15	7.35%	5	3.08%
Linking bundles	12	5.88%	6	3.70%
	204		162	

## 5. 讨论

本研究的发现如下。首先，国际学生在写作中使用介词短语类(PP-based bundles)和名词短语类(NP-based bundles)词块的频率低于母语者。这一结果与 Hyland [21] 的研究一致，他认为写作中最常见的结构类型是名词短语和介词短语。Hyland 指出，写作中“名词短语+介词短语”的结构能传达多种含义，如数量(如 *a great deal of*、*a great number of*)、地点(如 *the end of the*、*the beginning of the*)等[22]。这类结构被 Biber 等人称为“极具生成力的结构框架”(extremely productive frames) [20]。然而，相较于母语写作者，国际学生显然并未意识到这些名词表达在议论文写作中的重要作用。此外，国际学生在写作中更倾向于使用动词短语类(VP-based bundles)词块，这一结果与 Biber 等人提出的规律一致[24]。他们指出，在语言习得的早期阶段，学习者广泛使用从句类词块，这些词块可划分为动词短语和句子起始结构。然而，随着语言能力的不断提升，学习者逐渐过渡到更多使用短语类词块，即名词短语和介词短语。因此，尽管国际学生已经接受了一定的语言训练，但其英语能力与母语者仍存在差距，这也导致他们在完成写作任务时更倾向于使用 VP-based 词块而非 NP-based 和 PP-based 词块。因此，正如 Biber 等人所建议的那样，在研究写作技能的发展时，应将词块的使用情况纳入考虑[25]。

### 频率特征

总体而言，学生所产出的高频词块数量少于母语者。此外，母语学术语篇中的高频词块与学生写作中出现的高频词块之间存在不匹配的情况：许多国际学生常用的高频词块并未出现在目标词块列表中，而国际学生语料库中这些目标词块的产出频率也远低于应有水平。在 204 个 TBs 中，四词词块的高频使用( $\geq 6$ 次)仅有 1 位学生达到，而三词词块的高频使用则增加到了 29 个。更严重的是，在两个语料库中，几乎有一半的母语学术语篇中最常用的词块在学生文本中从未出现过。这些“零产出”词块表明，学生在被动语态的使用上可能存在困难。例如，学生几乎没有产出包含被动语态的学术 TBs (如“*should be made as shown in the figure*”)以及其他包含“*It*”类动名词结构的 TBs (如“*it should be*”，“*it is possible*”)。这些都是学生在写作中存在的薄弱环节，值得在今后的写作反馈与教学实践中予以重点关注。

根据词频效应[12] [25]，在母语写作中高频出现的目标词块(Target Bundles, TB)更容易被二语学习者在输入过程中注意到、习得并加以巩固，而那些在母语写作中出现频率较低的目标词块由于在学习材料中出现次数较少，则更难以习得。因此，从理论上讲，既然 TB 是在母语写作中出现频率最高的词块，学生在写作中也理应更频繁地使用这些词块，然而这一现象却并未出现在海外学生的实际写作中。

为深入探究国际学生在某些词块(TB)使用不足的原因，需综合考虑语言输入环境、教学导向、写作

策略及语言认知等多重因素。首先,学生在写作中高频使用的TB(如the development of the、is one of the)往往结构简洁、符合语法规则,构成词汇也多为早期英语学习阶段所接触的常用词。这类表达被认为是“性价比高”的语言资源,学生可以在认知负担较小的情况下完成表达任务。这种依赖熟悉表达的现象,可借助提出的“泰迪熊原则”(teddy-bear principle)进行解释,即学生更倾向于使用已掌握、可快速调用的语言单位[25]。

其次,部分国际学生在非正式访谈中透露,他们在写作时会主动回避一些使用频率较低或语法结构复杂的表达方式,主要出于对犯错的担忧。这种“风险规避型策略”可能源自以往标准化英语考试训练中形成的写作习惯:在限定时间内产出尽可能多且尽量无误的内容。这一策略会无形中抑制学生尝试使用更复杂、但功能性更强的TB,进而影响写作中词块的数量与多样性。

再者,一些TB的缺失可能与学生对其语用功能的认知不足有关[12]。如in the present study或as shown in the figure等学术衔接类表达,在语篇组织中具有显著作用,但学生可能未意识到它们在学术写作中的功能,因而并未有意使用。此外,由于这类TB通常在正式学术文本中才高频出现,学生在课堂输入或课外阅读中较少接触,因此缺乏模仿和产出的机会。这也可归因于语言输入环境与学习资源之间的差异。

从语言结构角度分析,某些结构(如带后置修饰语的名词短语,如such)在目标语写作中较少出现,也可能反映出源语迁移的影响[15][26]。例如,部分学生母语习惯将修饰成分置于名词之前,而英语中则常用后置修饰。这种结构层面的差异可能使学生在使用相关TB时感到困难或不自然,从而倾向于规避。

此外,本研究还发现,语言能力较高的学生更倾向于产出更多、结构更复杂的词汇短语,这与Chen & Baker及Huang等人的研究结果一致。这可能反映了他们具备更强的语篇意识与学术表达能力[15][26]。然而,这一发现与部分文献中的结果不一致,如Hu等人与Hyland指出,语言能力较弱的学习者更依赖公式化表达[6][23]。这一矛盾可能与不同研究在词块识别方法上的差异有关。本研究中对重叠短语和与写作任务高度相关的功能短语进行了排除,以避免主题偏差的影响[14]。这一操作提升了提取结果的客观性,使所反映的TB使用情况更能体现学生真实的语言表达倾向。

综上所述,国际学生在TB使用上的不足并非单一因素所致,而是教学背景、写作策略、语言输入、语用意识及母语迁移等多方面交织的结果。未来研究可结合定量分析与访谈、课堂观察等质性方法,以更全面揭示学生在写作中TB选择与回避的动因。

## 6. 总结

本研究通过收集研究生阶段学生的写作材料,从更客观的角度探讨国际学生群体在写作中使用词块的模式。与大规模的横断面研究相比,本研究在一定程度上排除了个体因素的影响,研究对象均为海外留学生,研究结果具有较强的代表性。在方法论方面,本研究对所检索出的词块从多个角度进行了验证,排除了重叠的词块,提升了所提取词块的有效性,为后续分析奠定了坚实的基础。此外,研究还发现一些国际学生过于依赖熟悉的词块表达,因此在未来的学习中,国际学生应加强对词块的理解和记忆,同时可以适当加入一些被动语态表达,使写作更具连贯性。

本研究的发现对学术英语写作的教学和实践具有以下几点启示:首先,对写作中词块的识别可以为英语教师、教材编写者以及其他英语教学工作提供具体的教学对象和工具。本研究分析了英语母语写作者所使用词块的频率、结构和功能,提供了典型目标词块的使用实例,全面展现了母语写作者在写作中如何运用词块。这些词块列表并非孤立、按频率排列的短语,而是可以作为教学内容的重要来源,为教学中多词单位的选择提供参考,帮助学习者深入理解词块在学术语篇中的实际使用。基于对母语者词块使用的系统性理解,学习者可以逐步掌握这些固定搭配的表达方式,从而更自然、更地道地在写作中

使用它们。

其次，由于学术作业写作具有范式性和规范性，写作中语言的流畅性和地道性在很大程度上依赖于对公式化或已词化句型的掌握。词块的高频使用已成为流畅写作的重要标志，对达到教师预期的写作水平至关重要。对于国际学生而言，掌握第二语言中的词块特征是习得写作规范的重要部分。为了提升作业成绩，非母语写作者及初级学者需要掌握一定数量的固定模式和表达公式。这些表达方式能帮助他们以更符合写作主题、目的、方法和目标读者的方式表达思想。这一点强调了在英语教学课程中加入相关词块教学的重要性。因此，以词块为核心的语言学研究具有重要的教学意义和实践价值。

## 参考文献

- [1] Hoffmann, C. (2000) The Spread of English and the Growth of Multilingualism with English in Europe. In: Cenoz, J. and Jessner, U., Eds., *English in Europe*, Multilingual Matters, 1-21. <https://doi.org/10.21832/9781800417991-002>
- [2] Du, J. (2013) The Use of Lexical Bundles by Chinese EFL English-Major Undergraduates at Different University Levels: A Corpus-Based Study of L2 Learners' Examination Essays. PhD Thesis, Chinese University of Hong Kong.
- [3] Qin, J. (2014) Use of Formulaic Bundles by Non-Native English Graduate Writers and Published Authors in Applied Linguistics. *System*, **42**, 220-231. <https://doi.org/10.1016/j.system.2013.12.003>
- [4] Ellis, N.C., Simpson-Vlach, R. and Maynard, C. (2008) Formulaic Language in Native and Second Language Speakers: Psycholinguistics, Corpus Linguistics, and TESOL. *TESOL Quarterly*, **42**, 375-396. <https://doi.org/10.1002/j.1545-7249.2008.tb00137.x>
- [5] Crossley, S. and Salsbury, T.L. (2011) The Development of Lexical Bundle Accuracy and Production in English Second Language Speakers. *IRAL-International Review of Applied Linguistics in Language Teaching*, **49**, 1-26. <https://doi.org/10.1515/iral.2011.001>
- [6] Hyland, K. (2008) Academic Clusters: Text Patterning in Published and Postgraduate Writing. *International Journal of Applied Linguistics*, **18**, 41-62. <https://doi.org/10.1111/j.1473-4192.2008.00178.x>
- [7] Durrant, P. and Schmitt, N. (2009) To What Extent Do Native and Non-Native Writers Make Use of Collocations? *IRAL-International Review of Applied Linguistics in Language Teaching*, **47**, 157-177. <https://doi.org/10.1515/iral.2009.007>
- [8] Kaszubski, P. (2006) Ideas from the Corpora. *Humanising Language Teaching Magazine*. <http://old.hltmag.co.uk/jul06/idea.htm>
- [9] Nesselhauf, N. (2005) Collocations in a Learner Corpus. John Benjamins Publishing Company. <https://doi.org/10.1075/scl.14>
- [10] Meng, Y.R., Huang, X.Y. and Fan, X.H. (2025) A Diachronic Study on the Developmental Characteristics of Lexical Bundle Competence in English Writing. *Language Education*, **13**, 110-118.
- [11] Wei, Y. and Lei, L. (2011) Lexical Bundles in the Academic Writing of Advanced Chinese EFL Learners. *RELC Journal*, **42**, 155-166. <https://doi.org/10.1177/0033688211407295>
- [12] Pawley, A. and Syder, F.H. (2014) Two Puzzles for Linguistic Theory: Nativelike Selection and Nativelike Fluency. In: *Language and Communication*, Routledge, 191-226. <https://www.taylorfrancis.com/chapters/edit/10.4324/9781315836027-14/two-puzzles-linguistic-theory-andrew-pawley-frances-hodgetts-syder>
- [13] Cortes, V. (2004) Lexical Bundles in Published and Student Disciplinary Writing: Examples from History and Biology. *English for Specific Purposes*, **23**, 397-423. <https://doi.org/10.1016/j.esp.2003.12.001>
- [14] Ädel, A. and Erman, B. (2012) Recurrent Word Combinations in Academic Writing by Native and Non-Native Speakers of English: A Lexical Bundles Approach. *English for Specific Purposes*, **31**, 81-92. <https://doi.org/10.1016/j.esp.2011.08.004>
- [15] Chen, Y.-H. and Baker, P. (2010) Lexical Bundles in L1 and L2 Academic Writing. [https://scholarspace.manoa.hawaii.edu/bitstream/10125/44213/1/14\\_02\\_chenbaker.pdf](https://scholarspace.manoa.hawaii.edu/bitstream/10125/44213/1/14_02_chenbaker.pdf)
- [16] Chen, Y.-H. and Baker, P. (2016) Investigating Criterial Discourse Features across Second Language Development: Lexical Bundles in Rated Learner Essays, CEFR B1, B2 and C1. *Applied Linguistics*, **37**, 849-880.
- [17] Esfandiari, R. and Barbary, F. (2017) A Contrastive Corpus-Driven Study of Lexical Bundles between English Writers and Persian Writers in Psychology Research Articles. *Journal of English for Academic Purposes*, **29**, 21-42. <https://doi.org/10.1016/j.jeap.2017.09.002>
- [18] Yang, Y. (2017) Lexical Bundles in Argumentative and Narrative Writings by Chinese EFL Learners. *International*

- 
- Journal of English Linguistics*, 7, 58-69. <https://doi.org/10.5539/ijel.v7n3p58>
- [19] Biber, D., Johansson, S., Leech, G., Conrad, S. and Finegan, E. (2000) Longman Grammar of Spoken and Written English. Longman. <http://tesl-ej.org/wordpress/issues/volume4/ej15/ej15r14/?wscr>
- [20] Simpson-Vlach, R. and Ellis, N.C. (2010) An Academic Formulas List: New Methods in Phraseology Research. *Applied Linguistics*, 31, 487-512. <https://doi.org/10.1093/applin/amp058>
- [21] Biber, D. (2004) If You Look at ...: Lexical Bundles in University Teaching and Textbooks. *Applied Linguistics*, 25, 371-405. <https://doi.org/10.1093/applin/25.3.371>
- [22] Hyland, K. (2008) As Can Be Seen: Lexical Bundles and Disciplinary Variation. *English for Specific Purposes*, 27, 4-21. <https://doi.org/10.1016/j.esp.2007.06.001>
- [23] Biber, D., Connor, U. and Upton, T.A. (2007) Discourse on the Move. John Benjamins Publishing Company. <https://doi.org/10.1075/scl.28>
- [24] Hu, J., Shi, H. and Ji, P. (2017) Structural and Functional Characteristics of Discourse Word Blocks for English Learners and Native Speakers. *Foreign Language Education*, 38, 57-64.
- [25] Biber, D. and Finegan, E. (2014) On the Exploitation of Computerized Corpora in Variation Studies. In: *English Corpus Linguistics*, Routledge, 204-220. <https://www.taylorfrancis.com/chapters/edit/10.4324/9781315845890-17/exploitation-computerized-corpora-variation-studies-douglas-biber-edward-finegan>
- [26] Ellis, N. (2003) Constructions, Chunking, and Connectionism: The Emergence of Second Language Structure. The Handbook of Second Language Acquisition/Blackwell Publishing. <https://doi.org/10.1002/9780470756492.ch4>