

中英作家英语语言学论文中四词词汇块的比较研究

Jinjin Zhang

独立学者 山东滕州

【摘要】词块是由两个或两个以上连续或不连续的词语组合而成，是一种能够表达一定意义并具有完整结构的语言现象。同时，词块也是学术论文的重要组成部分，是衡量学术论文语言能力的重要指标。本研究以英语作者撰写的英文论文和中国研究生撰写的硕士学位论文为对象，以 Biber 等人的词块结构分类为框架，从结构层面分析了四字词块的特征。研究发现：（1）中国英语语言学研究研究生使用四词块的频率高于英语作者，且中国英语语言学研究研究生在写作中存在过度使用四词块的倾向。（2）在结构层面，中国英语语言学研究研究生更倾向于使用介词四词块和动词四词块，而英语语言学作者则倾向于使用介词四词块和名词四词块。中国英语语言学研究研究生撰写的硕士论文与英语作者撰写的期刊论文在被动动词四词块的使用上存在显著差异。本研究旨在帮助学生理解并提高对词块使用的能力。同时本研究也为词块的学习与教学提供了可用的数据和实践意义。

【关键词】四词块；语料库；硕士论文；英语语言学；英语教学；比较研究

【收稿日期】2025 年 4 月 13 日

【出刊日期】2025 年 6 月 17 日

【DOI】10.12208/j.lls.20250001

A comparative study of four-word lexical chunks in English linguistic thesis by Chinese and English writers

Jinjin Zhang

Independent Researcher, Tengzhou, Shandong.

【Abstract】 The lexical chunk, a combination of two or more continuous or inconspicuous words, is a language phenomenon that can express a certain meaning and have a complete structure. Meanwhile, the lexical chunk is an essential ingredient of academic articles and an important indicator of language ability in academic articles. This study adopts English papers written by English writers and master theses written by Chinese postgraduates as the object. Using Biber et al's structural classification of lexical chunks as the frame, this study analyzes the characteristics of four-word lexical chunks at the structural levels. It was found that (1) Chinese postgraduates in English Linguistics used four-word lexical chunks more frequently than English writers. Moreover, Chinese postgraduates in English Linguistics tended to overuse four-word lexical chunks in writing. (2) At the structural level, Chinese postgraduates in English Linguistics preferred prepositional four-word lexical chunks and verbal four-word lexical chunks. English writers in English Linguistics were inclined to use prepositional four-word lexical chunks and noun four-word lexical chunks. Significant differences were found in the use of passive verbal four-word lexical chunks between master theses written by Chinese postgraduates in English Linguistics and journal papers written by English writers in English Linguistics. This study facilitates students' understanding and ability in lexical chunk use, meanwhile, this study provides available data and practical implications for lexical chunk learning and teaching.

【Keywords】 Four-word lexical chunks; Corpus; MA theses; English linguistics; English teaching; Comparative study

注：本文于 2022 年发表在 The Educational Review 期刊 6 卷 10 期，为其授权翻译版本。

1 引言

本章介绍研究的背景和目的。

1.1 研究背景

在第二语言习得或语料库语言学领域,词块是研究的焦点。众所周知,词块对语言理解和表达有着重要的贡献。Hyland 认为,词块不仅对学术话语的形成至关重要,而且是区分不同学科书面文本的重要手段。(海兰,2008)。Wray 还指出,第二语言学习者在真正掌握目标语言之前,需要熟悉目标语言中的一些常见搭配和短语。(谨慎,2002)。此外,许多研究者已经证明了在实际教学中应用词块的重要性。词块教学法已引起广泛关注,并被证明是一种促进外语学习的有效方法。因此,词块能力已被视为衡量语言学习者语言水平的重要指标之一,也是区分专家和新手的标准。

1.2 研究目标

本研究旨在对中国英语语言学研究研究生撰写的硕士论文与英语语言学期刊论文中四词组块的使用情况进行比较分析。本研究旨在回答以下问题:

(1)从中国英语语言学研究研究生撰写的硕士学位论文和英文作者撰写的学术论文中提取的四字词块有哪些特点?

(2)中国研究生撰写的英语语言学硕士论文中提取的四字词块与英国作者撰写的英语语言学期刊论文中提取的四字词块在结构层面上是否存在显著差异?

(3)中国研究生英语语言学硕士论文中提取的四字词块与英语作者撰写的英语语言学期刊论文中提取的四字词块在功能层面上是否存在显著的功能差异?

2 文献综述

本章主要介绍词块的相关研究。

2.1 国内外词块研究

在本研究中,词块指的是来自不同领域的众多词实体。不同的语言学家从不同的语言学视角对词块给出了不同的定义。Firth 是第一个提出这种语言现象的人,他称之为“扩展集合”。Syder 称之为“形式-意义配对”和“词汇化的句子词干”。Nattinger 和 DeCarrico 将词集定义为“词汇短语”。(纳廷格和 DeCarrico,2000)。一些中国语言学家称之为“词块”、“词束”等。Sinclair 基于长期的语料库语言学研究

调查,提出了“惯用原则”,认为词块是词汇综合集合的发展和延伸。(Sinclair,1991)。不同的研究者可能只从太阳的角度来研究词块,因此很难给词块一个固定的定义。Biber 等人是第一个研究词集的人。他在其语法书中将其命名为“词束”。(Biber et al.,1998)。Altenberg 对词块的发展也做出了贡献。他将连续的词集合称为“多词序列”,而不连续的词集合不包含在“多词序列”中。“多词序列”是一个具有不同结构的综合集合。“多词序列”可以是句子、短语和不完整的短语的集合。Skehan 提出了语言习得理论。他认为语言学习有三个必要的阶段:词汇化-语法化-词汇化(斯基汉,1999年)。因此,词块具有表达各种预制意义的特点,并且以未修改的形式存储。Mitchell 认为,词不同于搭配,后者属于更高级抽象的语法范畴。

许多中国语言学家也对词块的研究做出了贡献。杨永昌分析了语言“板块”结构的语用功能和分类,并认为学生在词汇学习中应该学习词块,因为学习词块可以帮助学生避免一些语言使用错误,增强书面表达的信心,提高口语和写作的流利程度。肖凤山认为,词块属于一种词搭配、一种词组或词块。他认为,在英语词汇教学中运用词块教学法可以提高学生的学习动机,促进学生积极参与课堂。吴静和王荣德认为,词块教学适合中国学生的语言学习特点,因为词块是记忆词汇的理想模型。蒲建忠认为,语言学习者在词块连接和搭配方面经常犯错误。语言学习者大脑中词块的缺乏,导致词块的连接和搭配能力较弱。

3 理论框架

本章主要从结构角度介绍词块的分类,并阐述选择四字词块的理由。

3.1 词块的结构分类

Biber 等将词块分为“人称代词+动词短语片段”;“it+动词短语或形容词短语片段”;“被动动词+代词短语片段”等12种类型。这种分类方法已经沿用多年,几乎所有研究都采用这种方法来分析词块的结构。由于缺乏基础理论,研究者发现从语料库中提取出来的词块很难进行分类和描述。因此,传统的结构分类方法有语言理论的基础,操作起来也更方便。表1给出了从结构角度对词块的详细分类。

3.2 选择四词词块的原因

Cortes 指出,大多数四字词块都包含三字词块,例如“由于”包括“结果”(Cortes, 2004)。此外,四字词块的频率高于五字词块的频率,因此,四字词

块在书面语中出现的频率高于其他词块。四字词块也更容易从语料库中提取,能够更精确地展现其结构和功能,是最适合研究的长度。

表 1 词块的结构分类

结构	示例
Noun phrase with of fragment	The end of the, the nature of the
Noun phrase with other post-modifier fragment	The way in which, the relationship between the
Prepositional phrase with embedded of-phrase fragment	At a result of, in terms of the
Other prepositional phrase fragment	On the other hand, at the same time, by the fact that
Passive verb +prepositional phrase fragment	Is based on the, referred to as the
Anticipatory it + verb phrase/ adjective phrase	It is important to, it is possible to, it can be seen
Copula be + noun/ adjective phrase	Is the same as, is a matter of, is one of the, is due to the
(verb phrase) that- clause fragment	Should be noted that, that there is a, studies have shown that
(verb/ adjective) to- clause fragment	Are likely to be, has been shown to, to ensure that the
Adverbial clause fragment	As shown in figure, as we have seen
Pronoun/ noun phrase +be	This is not the, there was no significant
others	As well as the, may or may not, as shown in figure

4 研究设计

本章介绍了本研究的目的和方法,并阐述了数据收集和分析的过程。

4.1 研究方法

本研究采用定量与定性相结合的方法,对四字词块的使用情况进行比较分析。研究对象为英语作者撰写的英语语言学期刊论文和中国研究生撰写的英语语言学硕士论文中的四字词块。本研究以英语作者撰写的英语语言学期刊论文中的四字词块为标准,从使用频率、功能和结构三个角度分析了四字词块使用的异同。

本研究以 Biber 等人提出的词块语法结构和 Hyland 提出的语篇分析功能范畴为框架,探讨四字词块的使用特点。四字词块样本选自英语语言学期刊论文(英语作者撰写)和中国英语语言学研究生

4.2 数据收集

本研究包含两个样本:语料库 1 和语料库 2。语料库 1 包含从 300 篇英语语言学期刊论文中随机抽取的 30 篇论文。语料库 2 包含从中国英语语言学研究

生撰写的硕士论文中随机抽取的论文。本研究选取的论文发表在 SSCI 收录的、国际上有较大影响力(影响因子>2)的英语语言学期刊上,包括: Applied Linguistics、Bilingualism-language and Cognition、International Journal of Bilingual Education and Bilingualism、Journal of Applied Linguistics 和 Studies of Second Language Acquisition。从这五个期刊中选取的 300 篇论文,发表年份从 2016 年至 2021 年不等。并且,这些论文的第一作者必须是英语母语人士。

4.3 研究工具

本研究使用 AntConc3.5.8 提取四字词组块。AntConc3.5.8 是一款免费的单语语料库分析和研究工具,由 Laurence Anthony 于 2002 年开发。它包含“Concordance”、“Concordance Plot”、“File View”、“Clusters”(或称“N-Gram”)、“Collocates”、“Word Lists”、“Keyword List”等索引工具以及其他功能。本研究采用“N-Gram”索引工具提取四字词组块。语料库 1 和语料库 2 中的四字词块。

5 结果与讨论

结构视角探讨四字词块的特点。首先,本章从整体上分析了英语作者撰写的英语语言学期刊论文和中国研究生撰写的英语语言学硕士学位论文中四字词块使用的特点。然后,在结构层面探讨了四字

词块使用的异同。

5.1 词汇束的总体分布

借助 AntConc3.5.8 对英语语言学领域英文作者撰写的期刊论文和中国英语语言学专业研究生撰写的硕士论文中的四字词块进行了抽取。数据显示, 中国英语语言学专业研究生使用四字词块 40 个, 出现频率为 542; 英语语言学专业英文作者使用四字词块 40 个, 出现频率为 264。结果表明, 两个语料库中均有一定数量的四字词块被使用。但两个语料库中四字词块的使用总数和频率仍然存在差异, 并且可以发现, 语料库 1 中的四字词块的数量和频率都高于语料库 2。本研究证明中国学生在写作中存在过度使用词块的倾向, 这与以前对词块的研究相一致。

英语作者撰写的英语语言学期刊论文和中国英语语言学研究研究生撰写的硕士论文在四字词块使用频率上存在差异。但本研究的重点是从功能、结构角度找出英语作者撰写的英语语言学期刊论文和中国英语语言学研究研究生撰写的硕士论文在四字词块使用上是否存在显著差异。本论文继续找出词块的特点。

本论文整理出了中国英语语言学研究研究生和英语作者撰写的英语语言学硕士论文中四字词块的使用情况, 结果如表 2 所示。借助 SPSS, 显著性差异的结果如表 3 所示。

结果显示了中国研究生和英语作者在英语语言学专业写作时使用的四字词块。40 个四字词块中有 13 个是中国研究生和英语作者在英语语言学专业写作时同时使用的。然而, 结果也表明, 只有不到 50% 的四字词块是中国研究生和英语作者在英语语言学专业写作时同时使用的。中国研究生和英语作者在写作中使用四字词块的偏好有所不同。

表 2 还显示, 有些四字词块被中国的英语语言学研究研究生和英语作者同时使用。虽然这些四字词块被中国的英语语言学研究研究生和英语作者同时使用, 但是这些四字词块在两个语料库中的出现频率不同。大多数四字词块在语料库 1 中出现的频率高于语料库 2。此外, 还有一些四字词块在语料库 2 中出现的频率高于语料库 1, 这意味着有些四字词块是中国的英语语言学研究研究生经常使用的, 而英语作者并不经常使用的。

表 2 四字词块的使用

		CP	EW
全部的		208	108
1	at the same time	39	7
2	as well as the	36	19
3	on the other hand	25	13
4	on the basis of	16	7
5	in the field of	13	9
6	can be used to	12	10
7	is one of the	12	4
8	in the present study	11	4
9	in terms of the	10	5
10	as a result of	9	9
11	can be seen in	9	10
12	on the one hand	9	5
13	at the end of	7	6

注: CP=中国研究生 EW=英文作家

表 3 是中国英语语言学研究研究生 ($\chi^2=7.635$, $df=1$, $p>.05$) 和英语作者 ($\chi^2=7.635$, $df=1$, $p>.05$) 四字词

块使用情况的卡方检验。结果表明, 英文作者撰写的英语语言学期刊论文与中国英语语言学研究研究生撰

写的硕士学位论文在使用频率上存在显著差异。这也表明中国英语语言学研究使用的四字词块种类比英语作者更多。此外,中国英语语言学研究在写作中使用四字词块的频率更高。中国英语语言学研究研究生和英语作者使用四字词块的频率都不到一半。这一结果表明,中国英语语言学研究已经认识到词块在写作中的重要性,并认识到使用四词词块是硕士论文的必要条件。语言学习者在硕士论文

中已经具备使用各种四词词块的意识。

5.2 四词词块的结构

第三章介绍了 Biber 等人提出的词块结构分类方法。他根据词块的结构差异,将词块分为 12 类。本研究采用 Biber 等人提出的词块结构分类方法,排序后的四字词块如表 4 所示。为了更清楚地表达词块结构差异,本研究还对四字词块的结构分布进行了卡方检验,结果如表 5 所示。

表 3 四词词块总体分布的卡方检验

	CP	EW	value	df	sig.
数字	48	43	7.635	1	.006
频率	542	264	7.635	1	.006

注: CP=中国研究生 EW=英文作者; a.0 个单元格 (0.0%) 的期望计数小于 5。最小期望计数为 31.14; b.0 个单元格 (0.0%) 的期望计数小于 5。最小期望计数为 31.14

表 4 四字词块在 12 种结构类型的分布

	N	CP 的 百分比	F	北	EW 的 百分比	F
总计/种类	48		542	43		264
1 n + of	2	3%	16	2	4.50%	12
2 n + other	-	-	-	2	4.50%	12
3 p + of	17	36.90%	200	15	32%	85
4 P + other	8	21%	114	6	14.70%	39
5 p. v.+ p	6	9%	49	8	19.60%	52
6 it + v./ adj.	3	4.20%	23	1	2.60%	7
7 be + n./adj/	1	2.20%	12	2	3%	8
8 v + that-clause	2	4.60%	25	2	4.50%	12
9 v./adj.+ to-clause	4	7.30%	40	-	-	-
10 adv.- clause	1	1.60%	9	-	-	-
11 p/ n + be	-	-	-	-	-	-
12 其他表达方式	4	10%	54	5	14%	37

注: n+of=名词短语带 of 片段; n+other=名词短语带其他后置修饰语片段; p+of=介词短语带 of 片段; p+other=其他介词短语片段; pv+p=被动动词+介词短语片段; it+v./adj.=先行 it+动词短语/形容词短语; be+n./adj/=系词 be/名词/形容词短语; v+that-clause=动词短语 that-clause 片段; v./adj.+to-clause=动词/形容词 to-clause 片段; adv.-clause=状语从句片段; p/n+be=代词/名词短语+be; 注: N=数量 F=频率 CP=中国研究生 EW=英语作家

从表 4 中,我们可以看出这些四字词块的结构分布情况,以及每个结构类别中四字词块的数量。此外,该表还显示,语料库 1 中的四字词块涵盖了 12 种结构类型中的 10 种,语料库 2 中的四字词块

涵盖了 12 种结构类型中有 9 种类型,说明中国研究生采用的四字词块结构类型范围比较广。

不难得出,从结构上看,中国英语语言学研究研究生和英语作者都倾向于在写作中大量使用四字词

块。中国研究生撰写的英语语言学硕士论文和英语作者撰写的英语语言学期刊论文中介词四字词块的比例相同。此外,介词四字词块在学术论文中起着至关重要的作用。此外,英语语言学的英语作者在写作中更喜欢使用名词四字词块,占23%。名词四字词块在母语作者的期刊论文中的比例高于中国研究生硕士论文中的比例。中国英语语言学研究生使用名词四字词块的比例为13%。研究发现,动词性词块在对话中出现频率较高,而名词性词块通常出现在正式写作中。结果表明,英语语言学专业的英语作者在写作中倾向于使用正式的四字词块。相比之下,中国英语语言学专业的研究生比英语语言学专业的英语作者使用更多的动词性四字词块。这一现象表明,从中国研究生撰写的硕士论文中提取的四字词块比从英语作者撰写的英语语言学期刊论文中提取的四字词块更具口语化。

从表5可以看出,名词类四词块($\chi^2=1.344$, $df=1$, $p>.05$)、介词类四词块($\chi^2=4.524$, $df=1$, $p>.05$)、动词类四词块($\chi^2=1.256$, $df=1$, $p>.05$)、形容词类四词块($\chi^2=.489$, $df=1$, $p>.05$)和小句四词块($\chi^2=.002$, $df=1$, $p>.05$)的使用不存在显著差异,但被动类动词类四词块($\chi^2=18.393$, $df=1$, $p<.05$)的使用存在显著差异。因此,从中国英语语言学研究生硕士论文中提取的大多数结构性四词块与从英语作者撰写的英语语言学期刊论文中提取的词块并无显著差异。此外,中国英语语言学研究生使用的被动动词性四词块与英语作者使用的词块存在显著差异。中国英语语言学研究生倾向于使用主动四词块,而不是被动动词性四词块。相反,英语作者更喜欢使用被动动词性四词块。鉴于此现象,中国研究生在语言学习中应内化更多被动动词性四词块,并在写作中运用它们。

表5 四词词块结构分布的卡方检验

	CP	EW	value	df	sig
n + of	16	12	1.344	1	0.246
p + other	114	39	4.524	1	0.042
p. v.+ p	49	52	18.393	1	0
it + v./ adj.	23	7	1.256	1	0.262
be + n./adj/	12	8	0.489	1	0.484
v + that- clause	25	12	0.002	1	0.966

a.0个单元格(0.0%)的期望计数小于5,最小期望计数为0.917; b.0个单元格(0.0%)的期望计数小于5,最小期望计数为50.11; c.0个单元格(0.0%)的期望计数小于5,最小期望计数为33.08; d.0个单元格(0.0%)的期望计数小于5,最小期望计数为9.83; e.0个单元格(0.0%)的期望计数小于5,最小期望计数为6.55; f.0个单元格(0.0%)的期望计数小于5,最小期望计数为12.12; 注: CP=中国研究生 EW=英语作者

综上所述,无论是在中国英语语言学研究生的硕士论文中,还是在英语作者撰写的英语语言学期刊论文中,介词四词块出现的频率都是最高的。英语语言学论文的英语作者不太倾向于使用介词和名词四词块,而中国英语语言学研究生使用动词四词块的频率高于英语作者。这些数据表明,中国研究生和英语作者在写作中都偏向于使用介词四词块,而在其他四词块的使用上略有差异。动词四词块在中国研究生的硕士论文中出现的频率和比例都较高。英语语言学论文的英语作者在写作中偏向于使用名词四词块,而不是动词四词块。

6 结论

6.1 本研究的结果

本研究中,中国英语语言学研究生在其学术写作中使用了各种各样的四字词块。此外,从中国英语语言学研究生撰写的硕士论文中提取的四字词块的使用频率高于从英语作者撰写的英语语言学期刊论文中提取的四字词块。而且,在40个四字词块中,有13个被中国英语语言学研究生和英语作者同时使用,这意味着不到一半的词块被中国英语语言学研究生和英语作者同时使用。同时,研究数据表明,中国英语语言学研究生撰写的硕士论文和英语作者

撰写的英语语言学期刊论文在四字词块使用上存在显著差异。

参考文献

- [1] Biber, D., & S, Concord. (1998). *Corpus Linguistics: Investigating Language Structure and Use*. Cambridge, Cambridge University Press.
- [2] Cortes, V. (2004). Teaching lexical bundles in the disciplines: An example from a writing intensive history class [J]. *Linguistic and Education*, 17, 391-406.
- [3] Hyland, K. (2008). Academic clusters: text patterning in published and postgraduate writing [J]. *International Journal of Applied Linguistics*, 18(1), 41-62.
- [4] Nattinger, J. & DeCarrico, J. (2000). *Lexical Phases and*

Language Teaching. Shanghai, Shanghai Foreign Language Teaching Press.

- [5] Sinclair, J. (1991). *Corpus, Concordance, Collocation*. Oxford: Oxford University Press.
- [6] Skehan, P. A. (1999). *Cognitive Approach to Language Learning*. Shanghai: Shanghai Foreign Language Teaching Press.
- [7] Wray, A. (2002). *Formulaic Language and the Lexicon*. Cambridge: CUP.

版权声明: ©2025 作者与开放获取期刊研究中心(OAJRC)所有。本文章按照知识共享署名许可条款发表。

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



OPEN ACCESS