

中美应用语言学博士论文中词块使用对比研究

尹俊杰, 左欣

长春工业大学, 吉林 长春 130102

摘要 : 本研究通过分别构建中美应用语言学博士论文摘要及结论的语料库, 从结构和功能两个层面对两个语料库的四词词块进行对比分析, 探讨中美应用语言学博士论文中四词词块的使用差异。结果表明, 两国应用语言学博士使用的总体四词词块频数并无较大差异, 但从结构层面和功能层面对四词词块进行对比, 中美应用语言学博士使用的四词词块比例存在显著差异。在词块的结构层面, 中国应用语言学博士在介词词块使用方面存在过度使用的情况, 而 it 做先行主语的序列使用不足; 在词块的功能层面, 中国应用语言学博士使用语篇型词块过多, 而参与型词块使用相对不足。本研究丰富了中美应用语言学博士在学术写作中四词词块使用差异的认识, 对学术论文写作具有借鉴意义。

关键词 : 词块; 语料库; 学术写作; 应用语言学

A Comparative Study on the Use of Lexical Bundles in Chinese and American Doctoral Dissertations in Applied Linguistics

Yin Junjie, Zuo Xin

Changchun University of Technology, Changchun, Jilin, 130102

Abstract : This study explores the variance in the utilization of four-word lexical bundles within doctoral dissertations on applied linguistics from China and the United States. It does so by compiling a corpus consisting of abstracts and conclusions from dissertations in applied linguistics by Chinese and American scholars, and then conducting a comparative study of these four-word lexical sequences at structural and functional levels. The findings indicate that there is no notable difference in the overall usage frequency of these lexical chunks between the two groups of doctoral candidates. However, when examining the usage from structural and functional perspectives, a significant disparity emerges in the proportion of four-word lexical bundles employed by Chinese and American doctoral students in applied linguistics. Structurally, Chinese doctoral students in applied linguistics tend to overuse prepositional lexical bundles and underutilize "it + verb/adjective phrase fragments." Functionally, they tend to overuse text-oriented lexical bundles and underuse those that are participant-oriented. This study enhances our comprehension of the disparities in academic writing style between Chinese and American doctoral students in the field of applied linguistics, offering insights for the improvement of academic writing skills.

Keywords : lexical bundles; corpus; academic writing; applied linguistics

引言

词块 (lexical bundles) 或称为 “词汇块” “固定搭配” “预制短语” 等, 是指在语言使用中频繁使用的一组词汇, 这些词汇通常一起出现, 并且在意义上相互关联, 形成一个固定的表达单元。词块可以是一个词, 也可以是多个词的组合, 通常作为一个整体储存在人们的大脑中, 使用时, 词块可直接从大脑中提取, 无需通过语法规则生成和分析, 且它们在语言中具有特定的功能和意义。学者出于不同的学术理论背景, 研究目的与角度, 用来表达词块的术语也不同, 例如 chunks(Sinclair, 1991)、lexical bundles(Biber et al., 1999^[1]; Cortes, 2004)、clusters(Hyland, 2008a^[2]) 等。本研究选择采用 Biber 的定义方式, 将它称为 “lexical bundles”。

通过语料库分析, 从结构和功能两个维度对比了词块的使用情况, 结果显示两国学习者之间存在显著差异 (Altenberg&Granger, 2001; Chen&Baker, 2010)^[3]。国内研究发现, 国内学习者在理解和掌握词块的使用特点方面存在不足, 他们在使用词块时既存在使用不足的问题, 也存在过度使用的情况 (徐昉, 2011^[4]; 胡元江、石海漫、季萍, 2017^[5]; 张会平, 2020^[6])。对比研究中美应用语言学博士在学术论文中的词块使用特征对我国的英语学术写作国际标准化具有重要意义, 目前国内关于词块的研究多以硕士论文为研究目标, 且多数研究以论文中的正文部分为主, 以博士论文摘要及结论部分为研究目标的研究较少, 因此本研究意图通过自建并对比两国应用语言

作者简介:

尹俊杰 (2000—), 男, 汉族人, 吉林白山人, 硕士, 研究方向: 外国语言学及应用语言学;
左欣 (1978—), 女, 汉, 长春人, 副教授, 硕士, 研究方向: 语料库语言学、应用语言学。

学博士论文摘要与结论部分语料库，借用语料库研究工具提取并分析词块，对比研究中美应用语言学博士在论文中词块使用的结构及功能特征，帮助国内英语学习者撰写更符合国际学术规范的论文摘要与结论。

一、文献回顾

词块理论的起源可以追溯到组块这一术语。在20世纪70年代，Becker(1975)首次提出了“预制语块”(prefabricated chunk)的概念，是指人们将一系列相互关联的词汇作为一个单元记忆和学习。Becker提出，人类的记忆中储存着大量的预制语块，这些语块由多个词汇构成，并且携带的信息量超过了单个词汇。这些预制语块存在于记忆中，随着人们频繁地使用它们，使得人们对这些语块的熟悉度增加，从而延长了它们在记忆中的保留时间。预制语块的概念提出之后，词块理论开始在语言学界受到关注，并逐渐发展成为一个研究领域。在众多研究中，词块的划分通常聚焦于三词、四词、五词和六词的组合。特别是四词词块，在书面文本中出现的频率最高。相较于五词词块，四词词块的出现率更为显著；与三词词块相比，四词词块在结构和功能上展现出更高的明确性。基于这些观察，本研究选择四词词块作为分析对象。

在词汇学家对词块的分类方法中，Biber(1999)在词块结构层面的分类方法被认为较为全面，且较能被大多数学者接纳并采用。Biber在学术语篇的结构层面对词块进行了细致的分类，共划分为12种类型，涵盖了从名词短语、介词短语到动词短语等多个语法结构，包括名词短语加上of短语片段、名词短语加上其他修辞语片段、介词短语内置of短语片段、其他介词短语片段、被动动词短语片段、系动词be加上名词或形容词短语片段、动词短语加上that从句、动词或形容词加上to短语片段、先行词it加上动词或形容词短语片段、副词短语片段、代词或名词短语加上be短语片段以及其他类型。Chen&Baker(2010)在Biber研究的基础上，进一步将词块结构归纳为四大主要类别：基于介词词组的词块、基于名词词组的词块、基于动词词组的词块和其他类别。Biber的词块分类框架不仅囊括了词块的结构属性，还涵盖了词块在语境中的作用，使得分类工作更为详尽和精确，这种分类方式有助于研究者更系统地分析词块在学术写作中的结构特征，因此本文在词块结构层面的研究采用Biber(1999)的分类方法。

目前对词块功能层面的划分存在两种主流方法，两种方法均围绕概念、语篇和人际三大语言元功能理论展开。一种为Biber提出，将词块划分为指示标记，立场标记，与文本组织三种；另一种也比较相似，由Hyland提出，将词块划分为研究型词块，参与型词块和文本型词块。研究型词块用于描述研究相关的内容，如介绍实验过程或结果，它能够协助作者构建或复原现实世界中的事件，传达研究的主题和内容，例如“the character of the”“of the current study”等；参与型词块，此类词块关注文本的作者或读者，倾向于作者与读者两者间的互动，大多用于表达作者的态度或评价，试图辅助读者代入作者视角理解文章，例如“it is important to”“as we can see”等；语篇型词块多用于组织语篇，包括用于承接上下文的过度词块，构想基础框架的结构词块等，例如“in the next part”“in addition to the”等。词块按照功能分为三大类后可细分为12小类，研究型词块

包括时间、位置、过程、量化、描述和话题；语篇型词块包括过渡、结果、文本结构和框架；参与型词块包括立场和融入类型的词块。Hyland的分类方法涵盖了词块的各种功能，且更加适用于学术文本的研究，本研究主要研究对象为学术语篇，因此本文在词块功能方面的研究采用Hyland(2008a)的分类方法。

二、研究设计

(一) 研究问题

- (1) 中美应用语言学博士论文中词块语法结构类型使用有何异同？
- (2) 中美应用语言学博士论文中词块语用功能类型使用有何异同？
- (3) 中美应用语言学博士论文中词块结构和功能类型使用不同的原因？

(二) 语料来源

本研究采用自建语料库，包括中国应用语言学博士论文摘要与结论语料库(A Corpus of Abstracts and Conclusions of Chinese Doctoral Dissertations简称为CCDD)和美国应用语言学博士论文摘要与结论语料库(A Corpus of Abstracts and Conclusions of American Doctoral Dissertations简称为CADD)。CCDD文本提取主要通过中国博士学位论文全文数据库进行随机语料选取，通过阅读其内容，确认其研究内容为应用语言学方向，且作者母语为汉语，选取2018–2023年“外国语言学及应用语言学”学科专业中的50篇博士学位论文，最终随机抽取30篇符合要求的论文。CADD的文本取自Proquest学位论文数据库，设置搜索词条为“applied linguistics”进行检索，人工阅读其文献信息，剔除不属于应用语言学领域的论文，确定论文作者姓名具有英语本族语国家姓名特征且作者隶属单位为美国高校，选取2018–2023年论文50篇，最终随机抽取30篇论文。CCDD总库容为100571词，平均每篇3352词；CADD总库容为99376词，平均每篇3313词。在学术论文研究中，摘要通常提供研究的概述，包括研究的目的、方法、结果和结论。通过阅读摘要，研究者可以迅速了解论文的核心内容，而结论部分总结了研究的主要发现和意义，包含更深入的分析与更全面的研究结果。本研究剔除论文的正文部分，保留摘要和结论部分，可避免词块受作者个人写作风格和文章主题影响，影响整体研究结果。在分别确定30篇中美应用语言学博士论文后，对文本进行集中处理，删除其中的题目、目录、表格、附录、参考文献以及正文，只留下摘要和结论部分，并将这些内容保存为TXT格式的文件，以此构建语料库。

(三) 研究步骤

- (1) 词块提取。设置词丛(N-Gram)大小为4，使用Antconc提取四词词块，参照Hyland(2008)设定的频率阈值，即每百万词出现20次。在提取词块时，应确保这些词块至少出现在语料库中

10%的文本里，以减少词块提取过程中受作者个人写作风格影响。

(2) 词块分类与对比分析。参照 Biber et al.(1999) 和 Hyland (2008a) 的分类框架，将提取的四词词块分别进行结构层面与功能层面的分类。使用 Chi-square Calculator 对比词块频数间的差异，试图分析造成此差异的原因。

三、结果与分析

(一) 词块频数对比

表1 CCDD和CADD词块频数信息

	CCDD		CADD		LL	P
	原始频数	标准化频数	原始频数	标准化频数		
类符数	100	994	77	775	+2.73	0.099
形符数	741	7368	698	7024	+0.82	0.366

注：标准化频数指的是每百万词中的出现次数；LL代表对数似然比；“-”表示中国学者较少使用；P代表显著性值。
 $P > 0.05$: 差异不显著； $P \leq 0.05$: 差异显著； $P < 0.001$: 差异非常显著；***表示显著值 $P < 0.001$ 。

本项研究采用基于语料库的方法，从 CCDD 语料库中提取了 100 个词块，从 CADD 语料库中提取了 77 个词块。表 1 显示，目前中国应用语言学博士使用的四词词块在种类上和使用频数上与美国应用语言学博士所使用的四词词块皆无显著性差异 ($LL = +2.73$, $p = 0.099 > 0.05$) ($LL = +0.82$, $p = 0.366 > 0.05$)。根据徐昉 (2012) 研究指出，随着英语学习水平的提高，中国英语学习者在词块使用上逐步接近国际水准，表 1 的研究成果显示，词块的使用频率与使用者的语言能力密切相关，中国的应用语言学博士在使用词块方面正逐渐接近美国应用语言学博士的水平。

(二) 词块结构对比

表2 CCDD和CADD词块结构频数信息

结构	结构类型	CCDD		CADD		LL	P
		原始频数	标准化频数	原始频数	标准化频数		
介词词块	介词短语 + 内置 of 短语片段	165	1641	120	1208	+6.61	0.010
	其他介词短语片段	142	1412	114	1147	+2.74	0.098
	小计	307	3053	234	2355	+9.03	0.002
名词词块	名词短语 + 其他后置修辞语	102	1014	99	996	+0.02	0.899
	名词短语 + of 短语	174	1730	210	2113	-3.82	0.051
	小计	276	2744	309	3109	-2.28	0.132
动词词块	被动动词 + 介词短语	32	318	8	81	+15.13	***
	副词短语片段	62	616	44	443	+2.86	0.092
	动词 (be)+ 名词短语 / 形容词短语	18	179	27	272	-1.92	0.167
	(动词 / 形容词) + to 短语片段	13	129	7	71	-1.76	0.189
	先行词 it + 动词短语 / 形容词短语片段	33	328	70	704	-14.04	***
	小计	158	1571	155	1560	+0.00	0.949

表2展示了不同结构类型词块的原始频数、标准化频数、对数似然率以及显著性值。研究结果表明，中国应用语言学博士在四词介词词块的使用频数上显著超过了美国应用语言学博士 ($LL = +9.03$, $p = 0.002 < 0.05$)，且中国学者在构建学术语篇时更依赖介词短语 + 内置 of 短语片段的介词词块，卫乃兴 (2007)^[1] 的研究指出中国学者在构建学术语篇时，由于汉语语言习惯和学术传统的影响，导致介词使用在英语写作中得到迁移，这种迁移表现在中国应用语言学博士会过多的使用表示逻辑关系的介词词块构建文章，例如 “in terms of the” “in the process of” 等^[7]。中国应用语言学博士对于介词词块的偏好可能源于汉语中的介词短语对于语篇组织逻辑性的重要性，例如汉语中的“关于”、“由于”等介词短语在英语中对应“about”、“due to”，此类介词短语在学术写作中多用于表达关系和原因等，且符合汉语语言习惯和学术传统。使用介词词块，能够帮助作者以一种客观的方式引出研究背景、原因和结果等，且帮助作者实现信息重组和结构组织，增强学术论文内容逻辑严密性，例如 “in the context of” “on the basis of” 等介词词块，可以清晰地为学术写作界定研究背景和基础，用于构建表达准确且逻辑严密的学术语篇。中国应用语言学博士使用被动语态词块的频数显著高于美国应用语言学博士 ($LL = +15.13$, $p = *** < 0.05$)，而动词词块中先行词 it 引导的动词短语 / 形容词短语片段的频数显著低于美国应用语言学博士 ($LL = -14.04$, $p = *** < 0.05$)。造成这种现象的原因可能是语言习惯因素和文化背景因素导致，美国学术会议上，演讲者会更加频繁地使用 “it is widely acknowledged” 引入一个观点，所以对于美国应用语言学博士，这种表达习惯更为熟悉，而中国应用语言学博士可能接触到该类表达方式的机会相对较少；王传艳 (2013) 研究发现被动语态的使用可以使文章更加客观，排除个人情感，强调作品本身，可能相较于美国应用语言学博士倾向主观直接明确地表达，中国应用语言学博士更加倾向委婉的方式来表达类似的观点^[9]。

(三) 词块功能对比

表3 CCDD和CADD词块功能频数信息

功能	功能类型	CCDD		CADD		LL	P
		原始频数	标准化频数	原始频数	标准化频数		
语篇型词块	文本结构	129	1283	90	906	+6.52	0.011
	框架	45	438	33	332	+1.71	0.192
	过渡	75	746	32	322	+17.27	***
	结果	46	457	65	654	+10.08	0.002
	小计	295	3918	220	2214	+11.12	***
研究型词块	量化	19	189	18	181	+0.02	0.898
	描述	190	1889	159	1560	+2.40	0.122
	时间 / 位置	132	1313	126	1268	+0.08	0.781
	过程	17	169	44	443	-12.70	***
	小计	358	3560	347	3492	+0.07	0.799
参与型词块	融入	42	418	72	725	-8.35	0.004
	立场	46	457	59	594	-1.77	0.184
	小计	88	875	131	1318	-9.02	0.003

表3表示，中国应用语言学博士语篇型词块使用过多，但是相较于美国应用语言学博士，在参与型词块的使用上还有显著不足 ($LL=-9.02$, $P=0.003<0.05$)。表3研究结果与李梦骁(2017)研究结果相同，参与型词块倾向于构建读者与作者间的互动沟通，帮助读者代入作者视角，辅助表达研究结果，而中国应用语言学博士在学术写作中没有充分利用评价型词块资源，导致学术语篇在表达立场与态度时过于保守，缺乏与读者的互动与对话^[10]。包凯(2021)研究指出，中国英语学习者在学术写作中受到的教育可能更偏向于使用固定形式化的词块构建语篇，这些词块往往与文本框架相关，较少涉及个人立场^[11]。这种写作传统可能导致中国应用语言学博士在写作时倾向于使用语篇型词块，确保文本逻辑紧密及形式规范。中国应用语言学博士使用参与型词块较少的原因也可能因为文化差异，西方学术文化中，个人立场和参与度受到高度重视，因此美国应用语言学博士在写作中倾向于使用参与型词块表达个人观点及态度，而在中国学术写作中，可能更加强调客观性，因此中国应用语言学博士可能更加倾向于使用语篇型词块构建论文框架，从而导致参与型词块的频数显著低于美国应用语言学博士。

四、结论

本项研究对比分析了中美应用语言学博士论文中四词词块的

使用情况，从结构和功能两个维度进行了深入探讨。研究发现：1.中国应用语言学博士在四词词块使用的种类和频数上与美国应用语言学博士无显著差异。2.结构层面：中国应用语言学博士过多依赖介词词块构筑语篇，而先行词it引导的动词/形容词短语片段的使用频率上显著低于美国应用语言学博士，原因可能是受汉语语言习惯和学术传统的影响，从而较多使用表示逻辑关系的介词词块组织篇章结构。3.功能层面：由于文化差异与学术写作传统差异，中国应用语言学博士过多使用语篇型词块，参与型词块使用显著不足。

本研究发现，中国应用语言学博士在四词词块使用上已经接近美国应用语言学博士词块使用标准，但是对于先行词it引导的动词/形容词短语片段和参与型词块的使用应该加强，培养国内学者在使用英语进行写作时的英语语言习惯，减少对于介词词块和名词词块的依赖度，合理运用参与型词块表达研究内容，丰富学术英语写作的写作结构。

本研究的局限性在于语料库的规模有限，所得出的研究结果可能存在一定的偏差，无法全面反映应用语言学学者在词块使用上的所有特征。后续研究可以扩大库容量，降低研究误差，从而深入分析中美应用语言学博士四词词块使用的异同。

参考文献

- [1]Biber D, Conrad S, Cortes V. If you look at …: Lexical Bundles in University Teaching and Textbooks [J]. *Applied Linguistics*, 2004, 25 (3):35.
- [2]Hyland K. Academic clusters: text patterning in published and postgraduate writing [J]. *International Journal of Applied Linguistics*, 2008, 18 (1): 41–62.
- [3]Chen Y, Baker P. Lexical Bundles in L1 and L2 Academic Writing [J]. *Language Learning and Technology*, 2010, 14 (2): 30.
- [4]徐昉. 中国学习者英语学术词块的使用及发展特征研究 [J]. 《中国外语》, 2012, 9 (4): 51–56.
- [5]胡元江, 石海漫, 季萍. 英语学习者与本族语者议论文词块的结构与功能特征——基于语料库的对比研究 [J]. 外语研究, 2017, 34 (04): 58–62.
- [6]张会平. 中国英语初学者写作词块使用特征研究 [J]. 解放军外国语学院学报, 2020, 43(04): 19–25+86+160.
- [7]Hyland K. Academic Discourse: English In A Global Context [M]. New York: Continuum, 2009.
- [8]卫乃兴. 中国学生英语口语的短语学特征研究——COLSEC语料库的词块证据分析 [J]. 现代外语, 2007, (03): 280–291+329–330.
- [9]王传艳. 被动语态在科技英语翻译中的处理 [J]. 武汉纺织大学学报, 2013, 6(01): 91–93.
- [10]李梦骁, 刘永兵. 基于语料库的中外学者学术语篇词块使用对比研究 [J]. 现代外语, 2016, 39(04): 507–515+584.
- [11]包凯. 谷歌翻译汉译英错误类型及纠错方法初探 [J]. 中国科技翻译, 2017, 30(04): 20–23.