

中外计算机科学学者英语学术语篇四词词块使用的对比研究

王一淳

浙江工商大学外国语学院, 浙江 杭州

收稿日期: 2025年11月9日; 录用日期: 2025年12月12日; 发布日期: 2025年12月24日

摘要

本研究通过对比分析中国学者与本族语学者在计算机科学领域英语学术论文中四词词块的使用情况, 探讨了中外学者在学术写作中词块的差异。研究基于Beijing CARE语料库中的计算机科学子库, 通过R语言提取并筛选出四词词块, 进行卡方检验分析。研究发现, 中国学者在四词词块的使用数量和频率上明显高于本族语学者, 且词块的多样性程度较高。中外学者使用的高频词块呈现共性与独特性并存的特点, 尤其是中国学者使用了大量特有的词块, 这种现象与语言迁移和本族文化影响密切相关。研究发现对中国英语学习者的词块习得和学术英语写作教学有一定参考价值。

关键词

学术语篇, 四词词块, 中外学者对比

Four-Word Lexical Bundles in the Academic Writing of Chinese and International Computer Scientists: A Comparative Study

Yichun Wang

School of Foreign Languages, Zhejiang Gongshang University, Hangzhou Zhejiang

Received: November 9, 2025; accepted: December 12, 2025; published: December 24, 2025

Abstract

This study investigates the differences in the use of four-word lexical bundles in English academic papers in the field of computer science between Chinese scholars and native English-speaking scholars

文章引用: 王一淳. 中外计算机科学学者英语学术语篇四词词块使用的对比研究[J]. 现代语言学, 2025, 13(12): 539-543. DOI: 10.12677/ml.2025.13121294

through a comparative analysis. Based on the computer science sub-corpus of the Beijing CARE corpus, four-word lexical bundles were extracted and screened using R, and a chi-square test was conducted for analysis. The findings reveal that Chinese scholars use a significantly higher number and frequency of four-word lexical bundles compared to native English-speaking scholars, with a greater diversity of such bundles. The high-frequency lexical bundles used by Chinese and international scholars exhibit both commonalities and unique characteristics. Notably, Chinese scholars employ a large number of distinctive lexical bundles, a phenomenon closely related to language transfer and the influence of their native culture. The findings of this study offer certain reference value for the acquisition of lexical bundles by Chinese English learners and the teaching of academic English writing.

Keywords

Academic Discourse, Four-Word Lexical Bundles, Comparison between Chinese and International Scholars

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

词块(lexical bundles)是语言及其使用中高频出现的多词单位,是体现语言流利性、地道性和准确性的重要标志之一[1] [2]。研究论文作为学术体裁,是传递学科知识创造的媒介[3],也是学术交流的工具[4] [5]。该类语篇通常比较简短且具有鲜明的主题焦点,展示规约化的立场选择与论辩策略。而学术语篇中词块的分布频数、结构以及语用功能存在鲜明的语域、体裁以及学科差异[6]-[10],也是区分学术专家和新手作者的关键指标之一[11]。

基于语料库的研究表明,在词块的结构和功能上,本族语者与非本族语学习者之间存在较大差异[12]。在英语作为通用语(English as a Lingua Franca, ELF)和跨文化修辞(Intercultural Rhetoric)视阈下,这些差异可能反映了非本族语学者独特的修辞策略与身份构建方式[13]。此外,国内研究表明,我国英语学习者未能充分掌握词块,过度使用、使用不足和使用不当等倾向并存[14]-[16]。研究学习者,尤其是作为高级英语学习者的中国学者词块使用特征具有重要意义,而目前国内专门针对计算机科学学者学术写作的研究较少,因此本研究通过对比中外计算机科学期刊英文论文语料库,考察中国学者在期刊论文中词块使用总体情况,并尝试从跨文化修辞视角解释差异背后的动因,以期为学术英语写作与教学提供启示。

2. 研究设计

2.1. 研究问题

本研究选择了计算机科学专业中国学者和本族语学者在英语学术期刊论文中所使用的四词词块进行对比分析,旨在回答以下问题:

- (1) 在同一学科期刊论文中,中外学者所使用的词块总体数量有何差异?
- (2) 在同一学科期刊论文中,中外学者使用的共有与专有词块有何差异?

2.2. 语料来源

本研究所使用的语料来自 Beijing CARE (Beijing Collection of Academic Research Essays)语料库,该

库由北京航空航天大学语料库语言学团队创建,旨在为学术英语的研究提供大量数据支持。语料库的建立始于 2015 年,并持续更新。该语料库分为中国学者子库和西方学者子库,包含了来自 23 个学科领域的学术论文,总计收录了 6155 篇论文,库容约为 3000 万词。语料库中的文本主要选自 2000 至 2017 年间中国和西方学者在国际期刊上发表的英语论文,特别关注研究性论文,排除综述性文章。期刊的选取依据包括期刊的学科代表性和影响因子,每个学科的文獻通过从 150 种国际期刊中随机抽取符合标准的论文来组成。Beijing CARE 语料库的主要目的是为学术英语研究、跨文化与跨学科话语对比研究提供数据支持,同时也为学术写作和教学提供实例数据[17]。

根据上述研究问题,笔者聚焦于 Beijing CARE 中的两个子库,即计算机专业中国学者库(CHCS)以及计算机专业本族语学者库(WCS),分别包含论文 78、147 篇,形符数分别为 468661、992876。为避免 WCS 库中语料的数量影响分析结果,笔者对该子库的文本进行了分层抽样,选取其中的 74 篇论文,形符数为 472151。

2.3. 词块提取与处理

本研究使用标准阈值每百万次出现 40 次以上,且不少于 10%文本范围标准来提取四词词块,确保提取出最代表性的词块。笔者利用 R 语言编写了相关代码,在 CHCS 库与 WCS 库中分别提取到了 73 种与 36 种四词词块。随后对所提取的词块进行人工核对和剔除(主要是重叠词块),保证统计结果的准确性。

完成词块处理与分类后,本研究采用卡方检验对各个词块进行比较分析,探究中外学者英语学术语篇中四词词块的使用差异。

3. 结果与讨论

3.1. 词块总体特征

经过筛选后,CHCS 与 WCS 中的四词词块数量分别为 58 和 33,中国学者使用的四词词块数量高于本族语学者。且词块的使用频率也明显高于本族语学者,如 as shown in fig 和 is shown in fig。两库中相同词块共有 19 个,相当于 CHCS 中的 33%和 WCS 中的 57%。只出现在 CHCS 中的词块有 33 个,仅出现在 WCS 中的词块有 14 个。本研究采用卡方检验测量两库词块的显著性,在 72 个被检测的词块中,仅有 8 个词块(in this paper we, the size of the, the performance of the, is based on the, with respect to the, the total number of, and the number of, at the same time)无显著性。其中四个名词词块(the size of, the performance of, the total number of, and the number of)往往用于描述某一客体的特性;is based on the, with respect to the, at the same time 则一般用于语篇内的衔接;尽管 in this paper we 这一词块并无显著性,但在 CHCS 中出现的频次高达 111 次,在 WCS 中仅有 43 次。

总体而言,中外计算机科学领域学者的词块使用表现为共性和独特性并存。这与潘璠[18]对中外机械专业学者进行的词块研究相呼应。中国学者能够使用常见的四词词块,如 on the other hand, it should be noted, as well as the, in terms of the 等来组织论文语言,但作为二语学习者,受限其写作水平,词块使用的种类及频率明显多于本族语学者。同时,中外学者的论文写作都有各自的特有词块存在。

3.2. 词块差异对比

对进行卡方检验后的两个词表进行筛选,观察具有显著性词块,以下部分分析了中外学者使用的共有词块与特有词块的差异。

3.2.1. 共有词块对比

在中外学者的共有词块中, as shown in fig ($X^2 = 10.098, p = 0.001$)和 is shown in fig ($X^2 = 6.070, p =$

0.01)两个词块尽管显著性并不强,但分别在 CHCS 中出现了 138、121 次,而在 WCS 中则仅有 41、40 次。这一结果与 Staples 等[19]的研究结论一致。缺少写作经验的作者在构建语篇时往往会更依赖程式化表达;另一方面,也能看出中国学者虽然在二语写作中受到语言水平的限制,但会积极采用图表来建构更为直观的表达方式。此外,本文认为中国学者高频使用“as shown in fig”等指引性词块,并非单纯的语言冗余,而可能是一种语用策略。在跨文化学术交流中,非本族语作者往往倾向于使用更多的路标式词块来降低读者的认知负荷,确保信息传递的准确性。

本族语学者使用频次明显较高的词块有 can be used to ($X^2 = 74.056, p = 0.007e - 15$), in the case of ($X^2 = 37.869, p = 0.007e - 7$), 其频次分别为 99、47,中国学者的使用频次则分别为 48、20。前者是一种典型的被动结构,本族语学者对被动语态的使用更为熟练,能自然地运用该词块来表达概念的客观性和中立性。中国学者虽然也使用被动语态,但其使用可能受到母语迁移和对复杂英语语法形式的掌握程度的影响,对类似表达的依赖性较低。英语写作风格通常遵循清晰的逻辑链条,并通过条件性表达(如 in the case of)对具体情况进行详细说明。本族语学者受英语文化影响,倾向于在学术写作中大量使用结构化的词块来构建逻辑框架,而中国学者则可能受汉语文化中整体性和意境性的表达习惯影响,更倾向于通过上下文或隐含关系来传递信息,而非依赖固定词块。

3.2.2. 特有词块对比

在中国学者所使用的特有词块中,有 3 个词块都出现了人称代词 we: we can see that ($X^2 = 32.300, p = 0.001e - 5$), we assume that the ($X^2 = 13.143, p = 0.00003$), in this section we ($X^2 = 7.727, p = 0.005$)。此外, in the following we ($X^2 = 11.569, p = 0.0006$)这一词块虽然是共有词块,但使用频率在 CHCS 高达 98 次,而在 WCS 中仅有 28 次。这一现象主要归因于中国文化强调个体与群体间相互依存,倾向集体主义,而这种集体意识在写作中也得以显现。同时,不同文化背景的作者在建立作者与读者之间的关系时也存在修辞偏好差异。本族语学者倾向于使用名词化和被动结构来构建客观、非人称的“科学真理”形象;而中国学者大量使用包含第一人称复数“we”的词块,体现了一种介入式的修辞立场。这一差异也印证了 Connor [20]关于写作风格不仅受母语迁移影响,更受特定学术社群文化认同建构影响的观点。

在本族语学者所使用的特有词块中,更多使用了名词后修饰结构词块,如 the extent to which ($X^2 = 59.456, p = 0.001e - 11$)。指出,中外作者在学术语篇中名词后修饰结构词块的差异可以归因于汉英两种语言中关系从句的句法不同。英语语言限定性关系从句,是中心词前置型,其基本构式为“RRCs-de-名词”。汉语限定性关系从句是中心词后置型,其构式为“名词-that-RRCs”。句法迁移是语言迁移的一种,在中国作者学术英语写作中是一种普遍的补偿策略,随着语言能力和学术水平的提升,写作者的迁移程度会有所降低[21]。

4. 结论与启示

以 Beijing CARE 的两个子库作为语料,本文使用 R 语言对中外计算机科学学者四词词块的使用进行了对比研究。研究发现,相对于英语母语学者,中国学者在学术写作中更依赖于四词词块,其词块多样性程度略高。中外学者高频词块呈现出共性与独特性共存现状,最为突出的是中国学者大量特有词块的使用,这种现象主要来源于中国学者受语言迁移和本族文化的影响。

因此,英语学术写作教学不应仅停留在纠正语法的层面,而应引入语料库辅助的教学手段,提升学生对学科特有词块的敏感度。基于本研究发现,笔者提出以下针对计算机科学学术英语的具体教学建议:

第一,开展“语境共现”的数据驱动学习。教师可引导学生利用语料库相关软件,检索本研究提取的高频词块,设计“语境重构”任务,让学生观察该词块在 WCS 库中的左右两侧搭配词,从而掌握比单纯记忆词块更深层的句法韵律。

第二,设计“语体转换与修辞重写”练习。针对中国学者过度使用“We can see that”等口语化或介

入性过强词块的问题,教师可设计改写练习。例如,给出包含“We propose a method that...”的句子,要求学生参考本族语学者的名词化特征,将其改写为“The proposed method aims to...”。通过对比练习,帮助学生体会从“叙事主导”向“概念主导”的修辞转变,从而更符合国际计算机科学界的行文规范。

第三,注重跨文化修辞意识的培养。在教学中,教师应明确指出中西学术思维的差异,解释为何英语学术写作倾向于“作者责任制”(writer-responsible),即通过名词后修饰结构(如 the extent to which)来精确界定概念,而非依赖读者的语境推断。

尽管本研究已经探讨了中外学者在计算机科学领域中的四词词块使用差异,但并未对词块的结构与功能做进一步分析。此外,还可以深入探讨不同学术背景、研究方向对学者写作风格和词块使用的影响,以便为学术写作教学和研究提供更加细致和广泛的理论支持。

参考文献

- [1] Pawley, A. and Syder, F.H. (1983) Two Puzzles for Linguistic Theory: Nativelike Selection and Nativelike Fluency. In: Richards, J.C. and Schmidt, R.W., *Language and Communication*, Longman, 191-226.
- [2] 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>
- [3] Hyland, K. (2009) *Academic Discourse: English in a Global Context*. Continuum.
- [4] Duszak, A. (1997) Cross-Cultural Academic Communication: A Discourse Community View. In: Duszak, A., *Culture and Styles of Academic Discourse*, Mouton de Gruyter, 11-39.
- [5] 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>
- [6] Biber, D., Conrad, S. and Cortes, V. (2004) If You Look at... Lexical Bundles in University Teaching and Textbooks. *Applied Linguistics*, 3, 371-405.
- [7] Pérez-Llantada, C. (2014) Formulaic Language in L1 and L2 Expert Academic Writing: Convergent and Divergent Usage. *Journal of English for Academic Purposes*, 14, 84-94. <https://doi.org/10.1016/j.jeap.2014.01.002>
- [8] Durrant, P. (2017) Lexical Bundles and Disciplinary Variation in University Students' Writing: Mapping the Territories. *Applied Linguistics*, 38, 165-193. <https://doi.org/10.1093/applin/amv011>
- [9] Shirazizadeh, M. and Amirfazlian, R. (2021) Lexical Bundles in Theses, Articles and Textbooks of Applied Linguistics: Investigating Intradisciplinary Uniformity and Variation. *Journal of English for Academic Purposes*, 49, Article 100946. <https://doi.org/10.1016/j.jeap.2020.100946>
- [10] 王芙蓉, 王宏俐. 基于语料库的语言学和工科学术英语词块比较研究[J]. 外语界, 2015(2): 16-24.
- [11] 李燕, 姜亚军. 语料库驱动的中外学术语篇词块变异研究[J]. 外语电化教学, 2023(2): 56-62+113.
- [12] Chen, Y. and Baker, P. (2010) Lexical Bundles in L1 and L2 Academic Writing. *Language Learning & Technology*, 14, 30-49. <https://doi.org/10.64152/10125/44213>
- [13] Mauranen, A. (2012) *Exploring ELF: Academic English Shaped by Non-Native Speakers*. Cambridge University Press.
- [14] 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(4): 51-56.
- [15] 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(4): 58-62.
- [16] 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(4): 19-25+86.
- [17] 卫乃兴, 王冰昕. 学术英语语料库 Beijing CARE 简介[J]. 语料库语言学, 2020(7): 72-77.
- [18] 潘璠. 语料库驱动的英语本族语和中国作者期刊论文词块结构和功能对比研究[J]. 外语与外语教学, 2016(4): 115-123.
- [19] Staples, S., Egbert, J., Biber, D. and McClair, A. (2013) Formulaic Sequences and EAP Writing Development: Lexical Bundles in the TOEFL iBT Writing Section. *Journal of English for Academic Purposes*, 12, 214-225. <https://doi.org/10.1016/j.jeap.2013.05.002>
- [20] Connor, U. (2011) *Intercultural Rhetoric in Second Language Writing*. University of Michigan Press.
- [21] 李燕, 姜峰. 中外学者英语学术语词块使用的共性与特性[J]. 外语教学, 2023, 44(5): 41-46.