

# 基于语料库的中英学术英语写作话语标记比较研究

高胜梅

山东建筑大学外国语学院, 山东 济南

收稿日期: 2025年6月5日; 录用日期: 2025年7月14日; 发布日期: 2025年7月23日

## 摘要

本研究基于语料库语言学方法, 对比分析中英学术英语写作的话语标记差异, 旨在揭示中国学习者在国际学术语境下面临的挑战及其成因。研究语料选自宁波诺丁汉大学中国学术英语口语和书面语语料库(CAWSE)与英国学术英语语料库(BAWE)。分析表明, 中英学习者在衔接词使用上存在显著差异。可知, 中国学习者倾向于依赖显性逻辑连接与直接因果关系表达, 而英国学习者更擅长运用复杂衔接词与隐性逻辑结构。研究结果为学术英语教学提供了跨文化视角, 建议通过强化词汇多样性训练、句法灵活性与衔接策略指导, 提升中国学习者的学术写作能力。

## 关键词

学术英语写作, 中国学习者, 英国学习者, 语料库分析, 跨文化交际, 话语标记

## A Comparative Study of Discourse Markers in Chinese and English Academic Writing Based on Corpus Data

Shengmei Gao

School of Foreign Languages, Shandong Jianzhu University, Jinan Shandong

Received: Jun. 5<sup>th</sup>, 2025; accepted: Jul. 14<sup>th</sup>, 2025; published: Jul. 23<sup>rd</sup>, 2025

## Abstract

This study employs corpus linguistics methods to compare and analyze differences in discourse markers between Chinese and British academic English writing, aiming to reveal the challenges faced by Chinese learners in international academic contexts and their underlying causes. The corpus data were selected from the Chinese Academic Writing and Speaking Corpus (CAWSE) at the

University of Nottingham Ningbo China and the British Academic Writing Corpus (BAWE). The analysis indicates significant differences in the use of cohesive devices between Chinese and British learners. It is evident that Chinese learners tend to rely on explicit logical connections and direct causal relationships, while British learners are more adept at using complex cohesive devices and implicit logical structures. The findings provide a cross-cultural perspective for academic English instruction, suggesting that enhancing vocabulary diversity training, syntactic flexibility, and cohesive strategy guidance can improve Chinese learners' academic writing skills.

## Keywords

Academic English Writing, Chinese Learners, British Learners, Corpus Analysis, Cross-Cultural Communication, Discourse Markers

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

随着全球化背景下国际学术交流的深入, 学术英语写作能力已成为非母语学习者参与国际科研合作的关键门槛。在学术语篇中, 话语标记(如 *however, therefore, in addition, in contrast*)扮演着至关重要的角色, 它们如同“路标”, 引导读者理解论证的逻辑脉络、观点间的衔接与转折, 是构建清晰、连贯、具有说服力学术文本的核心要素(Fraser, 1999) [1]。熟练掌握并恰当运用这些话语标记, 对于希望在国际学术环境中有效沟通并取得优异成绩的学生而言, 不可或缺。

然而, 母语迁移与文化思维模式的差异, 常导致学习者在语篇层面的组织策略, 尤其是话语标记的选择与使用上, 面临系统性挑战(Hyland, 2002) [2]。中国学习者尤其受到汉语与英语在逻辑表达习惯和语篇衔接方式上深层差异的影响, 其学术英语写作常被指出存在逻辑衔接不够自然、话语标记使用单一或不当等问题, 有时被诟病为带有“直译化”痕迹或“逻辑推进不够流畅”(Wang & Li, 2020) [3]。这些在话语标记运用上的困难, 直接影响文本的连贯性和学术表达的严谨性。

尽管现有研究已关注中英学术写作的某些局部特征, 但聚焦于话语标记这一核心语篇组织手段, 并基于大规模可比语料库对中英学习者进行系统实证对比的研究仍属相对空白。深入探究中英学习者在学术写作中如何使用话语标记构建逻辑关系、实现语篇连贯, 对于理解其语篇能力差异至关重要。

鉴于此, 本研究题目确定为《基于语料库的中英学术英语写作话语标记比较研究》。研究将运用语料库语言学方法(主要借助 WordSmith 及 Python 工具), 系统对比分析中国学习者学术英语语料库(CAWSE)与英国本族语学习者学术英语语料库(BAWE)中话语标记的使用特征。

研究旨在解决以下问题: (1) 中英学术英语写作有哪些话语标记差异? (2) 这些差异是否与文化背景或语言习得路径相关? (3) 如何基于对比结果为教学实践提供针对性建议?

## 2. 文献综述

学术英语写作的语言特征研究始于 20 世纪末, Biber 等(1999)通过多维分析法揭示了学术文本的正式性与信息密度[4]。Hyland (2002)提出“元话语”理论, 强调作者与读者的互动策略在学术写作中的重要性[2]。Connor (1996)的对比修辞理论指出, 英语学术写作偏好线性逻辑与显性论证, 而东方语言常隐含间接表达[5]。反观国内学者则从微观层面考察中国学习者写作的不足或从教学出发探求解法。其中,

一部分学者针对中国学习者的学术英语写作问题开展了多维度研究,从语言特征、篇章结构和文化差异等维度探讨了中英学术写作的差异。在模糊限制语方面,于强福(2019)通过对比中国理工科硕士论文与英语母语期刊摘要,发现中国学生使用模糊限制语的频率显著低于英语母语者,且类型单一,尤其缺乏范围变动型和间接缓和型,母语迁移和学术规范意识薄弱是主要原因[6]。词块使用方面,陈启敏等(2023)对比硕博论文摘要发现,博士生四词词块使用频率高于硕士生,且更倾向于介词结构和从句,但文本衔接功能仍弱于国际期刊,凸显高阶学术写作训练的不足[7]。王晶晶(2017)从自我指称语角度切入,发现中国学者倾向使用抽象主体以降低作者显现度,而英语母语者更多使用第一人称复数构建互动性[8]。而刘应亮(2023)进一步指出,中国学习者在评价性 *that* 从句中较少使用消极意义转述动词,存在误用现象,反映了学术身份构建能力的欠缺[9]。文化因素对写作风格的影响亦被广泛讨论。李建红(2014)强调,儒家文化中的集体主义倾向导致中国学者在学术写作中回避个人立场,过度依赖母语文献翻译[10],这一现象在孙庆舒(2015)的文本借用研究中得到验证,即中国研究生在英语写作中高频复制文献内容,尤其是汉语文献的未标注翻译,体现学科知识和引用规范的双重匮乏[11]。在教学策略方面,朱效惠(2018)基于活动理论设计项目式合作学习,证明多模态输入和同伴反馈能有效提升学术写作能力[12],但跨文化对比的教学干预研究仍较少。然而,现有研究多基于小规模语料且角度单一。尽管国内外的研究都涉及了中英学术写作特征,但在话语标记语料库研究仍稀缺。

国内上述研究虽语言特征、篇章结构和文化差异等角度揭示了中国学习者的写作特征,但仍存在局限,比如语料规模较小或学科覆盖单一,结论的泛化性不足;缺乏与英语母语者的系统对比,难以区分语言习得问题与文化迁移效应;对衔接策略的分析多基于人工标注,未充分结合语料库工具量化高频特征。本研究通过整合 CAWSE 与 BAWE 两大跨文化语料库(共计 1072 篇),采用定量与定性混合方法,对比中英学习者衔接策略上的差异。研究不仅弥补了国内现有成果在跨文化对比与大规模数据分析上的不足,还结合 Halliday 的衔接理论,深入解析语言差异背后的文化逻辑动因,为学术英语教学提供更具针对性的启示。

### 3. 研究方法

#### 3.1. 语料库

本研究的语料库由宁波诺丁汉大学中国学术英语口语和书面语语料库(CAWSE)和英国学术英语语料库(BAWE)组成。宁波诺丁汉大学中国学术英语口语和书面语语料库(China Academic Written and Spoken English Corpus)由宁波诺丁汉大学主导建设,旨在分析中国学生在学术英语环境中的话语标记使用特征,尤其关注非母语学习者在口语和书面表达中的挑战。该语料库包含中国学生的学术英语口语(如课堂展示、小组讨论)和书面语(如论文、报告)样本。覆盖多学科(如人文、理工、商科),样本来自本科生及研究生。英国学术英语语料库(British Academic Written English Corpus)由英国牛津大学、华威大学等机构联合开发,收录英国本土学生的优秀学术写作样本,旨在建立学术英语的参考标准,整个库包含约 3000 篇高质量学术写作(本科和硕士阶段),覆盖 4 大学科领域(人文、社科、理工、生命科学)。本研究选取 CAWSE 中艺术与科学(A&S)的 367 个文档以及 BAWE 中艺术与人文(AH)的 705 个文档,进行分析。所有文本都已进行格式处理,去除了题目以及参考文献留下了纯文本。

#### 3.2. 数据分析

本研究使用语料库语言学软件 WordSmith、LancsBox 和 Python 对两个语料库进行了对比分析,以确定中国和英国学习者学术写作中的话语标记特点。之所以选择以上三个软件是因为三者结合能满足研究需求并便于可视化。采用以上 WordSmith 以及 python 软件衔接词定量指标进行了研究。

## 4. 研究成果

衔接词是构建文本逻辑连贯性的核心手段。Halliday & Hasan (1976)提出衔接理论,将衔接分为指称、替代、省略、连接和词汇衔接五类[13]。其中,连接词(如加合、转折、因果词)通过显性标记逻辑关系,直接影响读者对论证结构的理解(见表 1)。

**Table 1.** Halliday & Hasan's classification of connectives

**表 1.** Halliday & Hasan 的衔接词分类

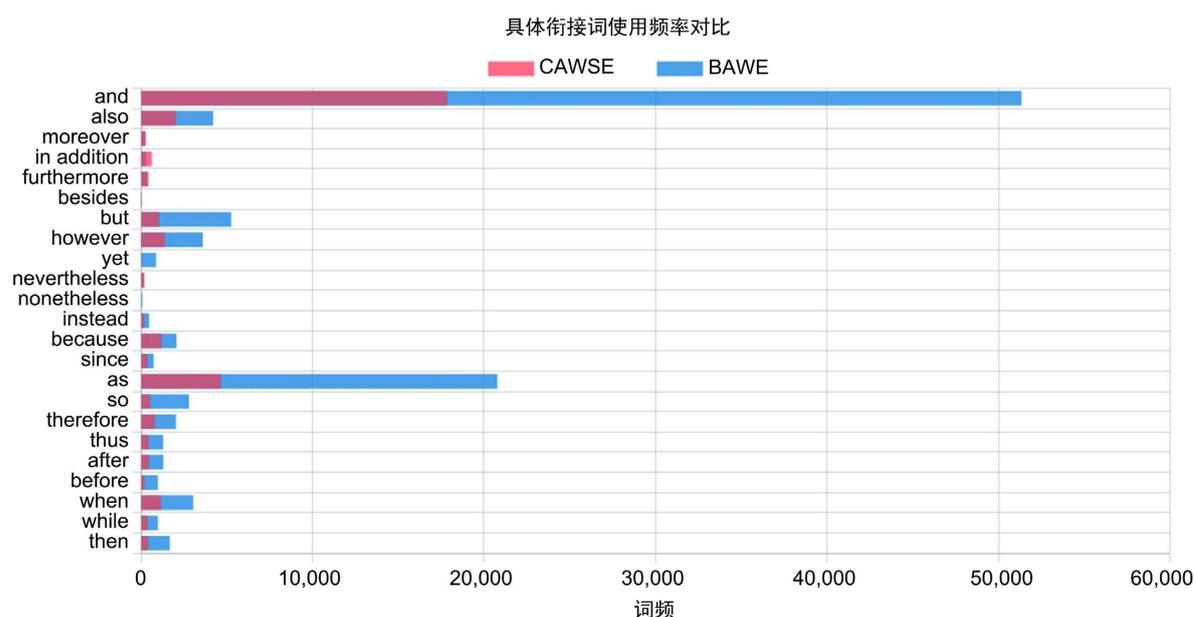
类别	连接词
additive 加合	and, also, moreover, in addition, furthermore, besides, likewise, similarly, equally, additionally, alternatively, on the other hand, too, as well, not only... but also...
adversative 转折	but, however, yet, nevertheless, nonetheless, instead, on the contrary, although, even though, despite, in spite of, regardless of, whereas, while, conversely
causal 因果	because, since, as, so, therefore, thus, consequently, as a result, hence
temporal 时间	after, before, when, while, then, meanwhile, subsequently, later, immediately, finally, first, second, next, last

**Table 2.** Total word frequency of transition words and STTR in the Chinese and English corpus

**表 2.** 衔接词在中英语料库中的总词频以及 STTR

Category	CAWSE		BAWE	
	Total Frequency	STTR	Total Frequency	STTR
additive	22,471	1498.067	58,898	3926.533
adversative	4663	310.867	14,500	966.667
causal	8678	964.222	30,649	3405.444
temporal	4424	316.000	12,748	910.571

### 衔接词词频对比 (CAWSE vs BAWE)



**Figure 1.** Specific word frequency comparison of transition words in Chinese and English corpus

**图 1.** 衔接词在中英语料库中的具体词频对比

表 2 显示, BAWE 的总词频显著高于 CAWSE (各类型平均高出 2.6 倍), 其中 additive 类差距最大 (58,898 vs. 22,471)。尽管中英学习者对衔接词类型的频率排序一致 (additive > causal > adversative > temporal), 这种表面相似性掩盖了深层的语义功能分化。STTR (标准化类符形符比) 指标显示 BAWE 的词汇多样性是 CAWSE 的 2.6 倍最大差距出现在 causal 类 (3405.444 vs. 964.222), 最小在 temporal 类 (910.571 vs. 316.000)。由此可知, 学生在写作中运用的衔接词要比中国学生多, 写作的逻辑更紧密。而出现如此差异的原因可能在于中国学习者将衔接词视为逻辑显化工具, 而英语母语者将其作为学术思辨载体。为进一步探讨具体衔接词使用差异以及原因, 我们进一步比对了两个语料库中具体衔接词的使用频率。

由图 1 可知, 中国学生使用频率最高的是 “and” “also” “besides” 等简单并列词, 而英国学生用词较复杂且正式, 如 “moreover” 和 “furthermore”。且转折类和时间类衔接词使用频率更高, 体现更强的逻辑严谨性和结构复杂性。

衔接词是实现语篇逻辑性、连贯性与说服力的核心工具。两者在衔接词中使用差异较大, 这反映在中国学生的写作中更依赖直接的逻辑连接方式, 语言风格较为简洁, 而英语学术写作需避免重复并追求语言丰富性。例 1 和例 2 都是进一步阐述事实, 但很明显, 例 2 furthermore 的使用进一步加强了上下句的联系, 符合学术论文写作追求客观的要求, 更进一步加强了逻辑上的衔接。

例 1: The pragmatic view of human rights is often applauded in that it offers little in the way of metaphysical justification for its ideas and instead focuses on agreement between nations on morality. (CAWSE)

例 2: North American colonists had little moral qualms when enslaving Africans. Furthermore, their lack of moral consciousness was a result of the accepted belief in the inferiority of Africans because of their skin-colour. (BAWE)

另一方面转折词绝非简单的“对比信号”, 而是学术写作者驾驭复杂思想、导航读者认知、融入学科共同体的核心能力, 其缺失将导致文本沦为观点罗列, 丧失学术对话生命力。通过图 1 可知, 在转折类词中, 中国学生较依赖 “but” “yet” 等基础转折词, 而英国学生除了这些基础转折词以外, 还使用了像 “nevertheless” 一类的复杂转折词。虽然二者都是属于转折词, 但是在学术英语写作中呈现出来的效果却天差地别, 如例 3 和例 4 所示。

例 3: It seems that the criticism lacks convincing but the support is reliable as it based on a large amount of researches. (CAWSE)

例 4: Even though the characters may be separated spiritually or physically, they are nevertheless bonded through these shared realities. (BAWE)

例 3 用了简单的 “but”, 直接对比前后两个观点, 批评缺乏说服力, 而支持可靠。例 4 的结构更复杂, 用 “nevertheless” 带出转折, 强调尽管有分离, 他们依然有联系。这种结构可能更突出矛盾双方共存的状态, 而不仅仅是简单对比。可知, 例 4 的让步结构更突出矛盾的复杂性。分离与联系并存, 而例 3 的对比相对简单, 这显示转折词的选择影响论述的深度。

究其根本, 以上差异可能与思维模式有关, 中国学生较少接受 “让步 - 反驳” 逻辑训练, 而英语学术写作强调辩证性。在衔接词分析中, 我们可以看出中国学生用以显性因果关系为主的词较多, 如 “because” “so”。而英国学生在此之上使用的隐性因果词较多 (“since” “as”), 再次之上还添加复杂句式。这可能因为在中文里, 语法结构偏线性, 多出现像 “因为……所以” 的固定搭配, 而英语从句更灵活, 需逻辑嵌套能力。在时间类词的使用中, 我们可以发现中国学生用 “after” 以及 “before” 较多, 而英国学生的时间词多用于阐述 “历程”, 可见英国学生学术写作需清晰呈现时间顺序。

## 5. 结论

本研究基于 CAUSE 与 BAWE 语料库的对比, 揭示了中英学术英语写作话语标记差异。研究发现, 量化分析表明, 中国学习者显著依赖简单显性逻辑连接词, 如高频使用“and”(占加合类词频 32.1%)、“because”(占因果类 68.3%)等基础词汇, 而英国学习者更倾向采用复杂衔接词(如“furthermore”在 BAWE 的使用频率为 CAUSE 的 4.2 倍)与隐性逻辑结构(如通过“since”引导的从句嵌套因果关系)。在词汇多样性上, BAWE 的标准化类符形符比(STTR)达 CAUSE 的 2.6 倍, 其中因果类词汇差异最为突出(3405.444 vs. 964.222), 凸显中国学习者衔接策略的单一性。

显性衔接是语篇的有形网络体现在语篇的表层结构上; 隐性衔接是语篇的“无形”网络[14]。本研究进一步说明在学术英语写作中中国英语学习者的差异成因可追溯至文化思维模式与语言习得路径的双重影响。一方面, 汉语的线性逻辑传统促使中国学习者偏好“因为……所以”式显性因果表达, 而英语学术写作强调辩证性结构(如“although...nevertheless”在 BAWE 占比 31%)。另一方面, 母语迁移导致中国学习者过度依赖翻译对等词(如误用“besides”替代“moreover”), 且缺乏句式嵌套能力训练(CAUSE 中分词结构表因果仅占 9%, 远低于 BAWE 的 28%)。

很明显, 文化、价值观念和思维方式等差异制约了我们对英语语篇的识别和研究[14]。学习者可以在词汇层面系统性扩充高阶衔接词库, 如以“however”替代“but”, 减少基础词依赖; 在句法层面应强化隐性逻辑表达训练(例如将直陈式“Because A, B”改写为“Given A, B emerges”); 教学方面, 应从修辞层面出发, 建议增设跨文化模块, 重点解析英语学术语篇“主张-反证-综合”的辩证推进模式。本研究的实证结果为学术英语教学提供了跨文化视角的操作框架, 未来可进一步探究学科子语域的差异化衔接策略。

## 参考文献

- [1] Fraser, B. (1988) Types of English Discourse Markers. *Acta Linguistica Hungarica*.
- [2] Hyland, K. (2002) Authority and Invisibility: Authorial Identity in Academic Writing. *Journal of Pragmatics*, 34, 1091-1112. [https://doi.org/10.1016/S0378-2166\(02\)00035-8](https://doi.org/10.1016/S0378-2166(02)00035-8)
- [3] 王力, 李洋. 基于语料库的中英学术写作词汇对比分析[J]. *现代外语*, 2020, 43(2): 234-247.
- [4] Biber, D., Johansson, S., Leech, G., Conrad, S. and Finegan, E. (1999) *Longman Grammar of Spoken and Written English*. Pearson Education.
- [5] Connor, U. (1996) *Contrastive Rhetoric: Cross-Cultural Aspects of Second-Language Writing*. Cambridge University Press. <https://doi.org/10.1017/CBO9781139524599>
- [6] 于强福. 理工科硕士学位论文英文摘要中模糊限制语缺失问题及其对策研究[J]. *教育评论*, 2019(2): 122-126.
- [7] 陈启敏, 龚扬, 陈明芳. 硕士和博士学位论文摘要中词块的对比分析[J]. *海外英语*, 2023(3): 72-74.
- [8] 王晶晶. 理工科博士生学术英语写作中的作者自我指称语研究[J]. *外语界*, 2017(2): 89-96.
- [9] 刘应亮, 杜凤双, 裴宇雨. 中国英语专业硕士论文摘要中评价性 that 从句的引述特征研究[J]. *外国语言与文化*, 2023, 37(2): 122-134.
- [10] 李建红. EFL 语境下作者身份与 EAP 写作的有效性研究[J]. *南通大学学报(社会科学版)*, 2014, 30(6): 64-68.
- [11] 孙厌舒, 王俊菊. 学术英语写作中的文本借用研究[J]. *外语与外语教学*, 2015(6): 36-41, 48.
- [12] 朱效惠, 袁欣. 活动理论视角下的学术英语写作教学研究[J]. *外语界*, 2018(1): 71-78, 87.
- [13] Halliday, M.A.K. and Hasan, R. (1976) *Cohesion in English*. Longman.
- [14] 王振福. 论衔接与连贯的显明性和隐含性[J]. *东北师大学报*, 2002(4): 92-97.