

# ChatGPT 辅助留学生作文语体纠偏实践效能研究

康雪

王曦悦

余兰心

祁峰\*

华东师范大学, 中国

## 摘要

本文聚焦国际中文教学中留学生语体能力发展滞后于语法能力的问题, 探讨 ChatGPT 等大语言模型辅助语体纠偏的实践效能。研究从 HSK 动态作文语料库中抽取 2000 份作文样本, 分析发现留学生书面语口语化偏误主要集中于单双音节对应律、长短律及时空律, 其中单双音节对应律偏误发生率达 49.6%。通过对比 ChatGPT、文心一言和讯飞星火三种大语言模型的表现, 选定 ChatGPT 作为实验工具, 对 31 名汉语水平为 HSK 4-6 级的华东师范大学留学生进行实验。结果表明, ChatGPT 辅助训练能够降低汉语作为第二语言被试的书面语口语化偏误, 对华裔被试效果虽不显著但也有一定效果。研究初步支持大语言模型作为语体教学辅助工具的可行性, 并为国际中文教学在智能化背景下的实践提供了可参考的实证依据。

## 关键词

ChatGPT, 语体教学, 人工智能, 大语言模型

## 1 引言

在国际中文教学领域, 语体能力是影响留学生汉语交际质量的关键因素, 留学生语体能力的发展大多滞后于其语法能力的发展 (司红霞, 2025), 这一现象在中高级水平学习者中尤为显著。李建芳 (1998) 指出, 语体能力表现为识别语境类型的能力和选用得体的语言材料及表达方式的能力。第二语言学习者在实际语言运用场景中因对语体语法知识和语体意识的匮乏, 难以根据交际场合、对象和媒介的差异精准选择恰当的语体形式, 进而影响其语言交际的得体性和有效性。这种语体能力与语言水平的不匹配, 容易成为制约留学生汉语能力进阶的瓶颈 (路云菲, 2025)。

造成这一现象的主要原因在于, 当前国际中文教学中缺乏高效的语体能力指导。从学术研究层面来看, 学界对于语体本体的研究虽已成果丰硕, 但在语体的界定和分类等核心问题上尚未达成共识, 多种理论观点并存, 这使得在实际语体教学应用中难以确立明确统一的标准。从教学实践角度出发, 国际中文教师在写作教学过程中因受课时限制、师生比失衡等因素影响, 对学生作文纠偏的压力较大 (翟艳、苏英霞, 2022:414), 且国际中文教材中的语体表现往往欠佳 (殷韵怡, 2022), 进而缺乏有效路径对留学生的语体知识与意识进行针对性指导。

\* 通讯作者。联系电邮: [fqi@hanyu.ecnu.edu.cn](mailto:fqi@hanyu.ecnu.edu.cn)

随着人工智能技术的迭代,大语言模型(Large Language Model, LLM)为破解这一困境提供了可能。大语言模型基于海量语料的训练特性,能够捕捉语体形式层面的共现规律。在此背景下,研究尝试借助大语言模型协助留学生进行语体纠偏,并对其深化学习者对语体理解、提高语体使用准确性并进一步提供高效助学服务的效能进行实证性检验,进而尝试为解决当前国际中文教师在语体教学中的现实困境提供参考。

## 2 文献综述

### 2.1 汉语语体语法研究现状

20世纪70年代,朱德熙(1985)、胡明扬(1993)在推进现代汉语语法系统化研究进程中,从“防乱”的角度出发,强调在语法研究中区分语体的必要性,由此引发学界对“语体”与“语法”关系的深度关注;同时明确指出书面语语法研究和口语语法研究应该分开进行,不能混为一谈。基于其对语体语法观念的认识,冯胜利(2003、2005、2006)提出了“书面正式语体”的概念,并结合韵律语法研究成果深入探讨书面正式语体的特征和模式、运行机制和语法属性等;此后正式提出了“语体语法”的概念及其基本属性问题——语体和语法词汇相互依存(冯胜利,2010)。在此基础上,冯胜利、施春宏(2018)对语体语法的基本原理、单位层级和语体系统进行了更为深入的理论建构与阐释。

在国际中文教学中,丁金国(1997)正式提出了“对外汉语教学中语体意识培养”这一课题,随后指出必须将“语体意识”的培育置于对外汉语教学的核心地位(丁金国,1999)。此后,学界众多学者积极响应,持续推动语体研究与教学实践的融合,如吴丽君(2004)、李丽丽(2005)、翟汛(2007)等学者聚焦对外汉语教学目标,主张以培养学生口语与书面语交际能力为核心,强化语体差异教学;郭颖雯(2002)、李泉(2003)等则提出构建口语与书面语独立语法体系的理论构想,为国际中文教学中的语体教学改革提供了新思路。

既有研究在汉语语体理论发展与教学应用层面形成了两个重要共识:其一,语体与语法不可分割,语体是语法实现的载体,语法规则需依托具体语体才能体现其交际价值;其二,语体能力是语言交际能力的基石,国际中文教学需突破传统“重语法轻语体”的局限。司红霞(2025)指出,语体能力在综合汉语学习和运用技能中十分重要,但当前语体理论在国际中文教育教学实践中的指导深度与针对性均显不足,现存教学模式、大纲及语言运用教学等方面存在问题,凸显了强化语体教学在国际中文教育中的重要性与紧迫性。因此,在数字化教育背景下,如何实现语体特征的自动化识别与个性化训练,以及如何借其进行系统的语体教学,都有待日后的学术研究填补该类研究的空缺。

### 2.2 大模型时代的语体研究

自ChatGPT发布以来,大语言模型凭借突破性的技术表现迅速成为人工智能领域的焦点,引发各行业广泛关注。在教育研究与实践领域,大语言模型在教育场景中的应用潜力与影响成为了学术界和教育工作者重点探讨的议题。

生成式语言模型通过分析上文已出现词语的概率分布,来推断后续最有可能出现的词汇。(袁毓林,2024),这一基于统计学习的文本生成机制与人类自然语言的生成过程存在本质差异(陆俭明,2025)。就语体语法而言,周义凯、刘海涛(2024)比较了汉语里3种语体文本与ChatGPT生成文本的句长分布,发现ChatGPT基本符合现代汉语书面语中句子单位的长度

分布和层级关系的规律,但其生成文本区分语体的能力还没有达到人类的水平;与其看法类似,毕玉德、闫丹辉(2024)指出模型缺乏对语体背后文化认知的理解,使模型的语体能力仅停留在纠正常见语法错误方面,对于文化隐喻适配等深层语体问题上可能失效;周立炜、饶高琦(2024)基于对大模型的横向对比,指出当前各类模型所生成文本的非正式表达能力普遍存在局限,未来可依托多样化语体语料资源的建设与技术迭代,进一步提升模型的语体适配与拓展能力。

综上,当前大语言模型在语体处理上仍以形式共现与概率预测为主,缺乏对语体社会文化功能及交际意图的深层理解。因此,其在语体相关任务中的适用范畴主要集中于对文本中形式化程度较高的表层语体特征进行识别、抽取与诊断性分析。需要指出的是,在大语言模型迭代升级速度持续加快的时代语境下,学术界围绕模型参与的语体分析及语体教学相关议题的探讨仍显不足,相关研究尚存较大拓展与深化空间。

### 2.3 大模型辅助二语写作研究

大语言模型的问世对于第二语言教育领域产生了具有变革性的深远影响,写作教学是受其影响最突出的核心领域之一。然而,在具体的外语教学实践场景中,如何科学、合理地发挥大模型的工具价值、实现其与写作教学的有效融合,始终是学术界持续关注、深入探讨的重点与难点议题。

现有研究普遍聚焦于大语言模型作为自动化写作评估与反馈提供者的角色。Jining Han、Mimi Li(2024)探讨了大语言模型支持下的间接纠正反馈等反馈方式,使学生能够更清晰地意识到自己的错误所在,进而提高纠偏的效率。任伟等(2024)进一步验证了模型在错误识别方面与教师具有较高一致性,反馈数量和效率更高,因此可作为有效的辅助评估工具;在此基础上,大模型提供的写作反馈在结构、内容和语言三个维度上分布均衡,且常使用积极鼓励的话语策略,增强了反馈的可接受性,更好地适配了学习者的认知与情感需求(Kai Guo、Deliang Wang, 2024)。学习者对模型提供的词汇替代与文本润色功能关注度最高(张荔等, 2026)。

在国际中文教育领域,李韵等(2025)构建了将教师与生成式人工智能协同反馈的写作教学模式,认为该模式能够提高反馈效率、丰富学习体验、保持学习动机。类似地,马瑞凌(2026)基于大模型搭建了专业化的作文智能评改系统,构建了“人智共评”的模式,实证发现学习者的写作成绩有了一定提高,同时大模型鼓励性反馈对学习者的情感态度起到了正向调节作用。然而,不可忽视的是,大语言模型支持的写作反馈也存在一定问题:高安娜(2024)指出大模型反馈所使用的语言难度易与学习者的汉语水平脱节,并且难以结合学生学情。就大模型自身的作文纠偏能力,刘瑜等(2024)进行了测验,发现大模型虽具备较强的纠错能力,但易将学习者原有表达修改得更为书面化、正式化,也常对无错误的表达进行不必要的修改,进而可能干扰学习者自身的语言表达逻辑。

综上所述,学界当前达成了一定共识:大模型可有效承担第二语言写作评阅与反馈角色。但模型反馈研究更多聚焦于结构、内容及语言中的词汇、错别字等方面,语体方面稍弱,尤其语体语法纠偏有关的研究较匮乏。基于以上研究现状,结合大模型易将学习者作文书面化的特点,本研究拟将大语言模型的技术潜力同国际中文教育的语体教学需求关联起来,探索ChatGPT作为辅助工具的创新路径,旨在促进人工智能技术在教育领域的创新应用,为“数智化”时代背景下国际中文教育语体教学的高质量发展提供理论依据与实践方案。具体而言,本研究聚焦于三个问题:

(1) 从语距定律视角看,在中高级汉语学习者作文中语体偏误的分布呈现何种特征与规律?

(2) 在统一的评测任务与标准下,ChatGPT、文心一言、讯飞星火三种大语言模型在识别该类偏误方面表现如何?其识别准确率是否存在显著差异?

(3) 以 ChatGPT 作为辅助工具进行干预后,学习者的书面语口语化偏误是否显著减少?其效果在不同语言背景学习者群体间是否存在差异?

### 3 研究设计

#### 3.1 实验对象与分组

实验共招募 31 名汉语水平为 HSK 4-6 级的华东师范大学留学生参与实验,被试现国籍均为亚洲国家,年龄集中在 19-30 岁,属于青年语言学习者。其中,华裔被试 7 名,第一语言均为汉语,具有海外长期生活经历。实验前期,研究团队系统采集被试个人信息,并以第一语言背景是否汉语作为分组依据,以充分比较 ChatGPT 的作文书面性反馈与纠偏效能在华裔学习者和以汉语作为第二语言学习者(以下简称“汉语二语学习者”)两类群体中的异同。

#### 3.2 实验材料说明

##### 3.2.1 实验流程总述

为系统探究 ChatGPT 在语体纠偏中的应用效能,本研究遵循“偏误分析——模型筛选——训练验证”的整体思路展开,研究首先基于 HSK 动态作文语料库对留学生书面语口语化偏误进行抽样标注与统计分布分析,确定发生频次最高的一项偏误项为实验材料;进而通过对比测试筛选出纠偏效果较好的大模型作为适用工具;最终通过设计人机交互训练任务对留学生进行作文纠偏训练与效果检验,完成从规律总结到实践验证的闭环。

##### 3.2.2 留学生书面语常见偏误选择

为针对性测试大语言模型对留学生作文书面语口语化偏误的纠偏与反馈能力,研究团队选取 HSK 动态作文语料库作为留学生作文语料来源,以议论性作文文本为研究样本,系统分析留学生在书面表达中高频出现的语体偏误类型,并以此作为后续实证性研究中针对大语言模型书面语口语化纠偏效能检验的任务项目。

###### (1) 抽样总体情况

研究团队从 HSK 动态作文语料库中随机抽取 2000 份作文样本,覆盖 8 个作文题目,分别是《吸烟对个人健康和公众利益的影响》、《如何看待“安乐死”》、《如何解决“代沟问题”》、《如何面对挫折》、《我对男女分班的看法》、《我看流行歌曲》、《父母是孩子的第一任老师》、《绿色食品与饥饿》。

这些作文的作者来自日本、韩国等 38 个国家和地区,其母语背景分布如下:东南亚 592 人,东亚 939 人,中亚及南亚 36 人,欧洲 276 人,美洲 101 人,大洋洲 56 人。

## (2) 抽样标注方法

鉴于书面语与口语的界限在学界尚未形成统一界定标准,本研究采用冯胜利(2020)总结的口语非正式语体的语距定律作为本研究考察留学生书面语口语化现象的切入点和参考标准。依照此标准来判断留学生作文中出现的错误是否属于书面语口语化偏误。标注过程中,每篇作文测试者将以书面语口语化偏误条目的出现次数计数。

## (3) 标注者一致性检验

在正式标注开始前,为确保标注者理解与执行标准的统一,研究团队共同制定了详细的《语体偏误标注指南》,明确了“单双音节对应律”“长短律”等七类语距定律类偏误的操作性定义与边界案例。随后,由全部5位标注者(均为具有语言学背景的研究人员)在互不交流的情况下开始标注。标注完成后,采用组内相关系数(Intraclass Correlation Coefficient, ICC)量化评估标注者间信度,初步计算结果显示,ICC值为0.72,根据Koo & Li(2016)的标准,该值表明一致性“良好”,标注结果具有较高的可靠性。

## (4) 抽样数据结果说明

经统计,7项口语化语距定律<sup>2</sup>在样本中均有体现:“长短律<sup>3</sup>”偏误472次、“悬差律<sup>4</sup>”10次、“单双音节对应律<sup>5</sup>”992次、“融合律<sup>6</sup>”11次、“单调律<sup>7</sup>”18次、“时空律<sup>8</sup>”151次、“成分追加律<sup>9</sup>”11次。

数据显示,留学生书面语口语化偏误主要集中于单双音节对应律、长短律及时空律,其中单双音节对应律偏误发生率达49.6%。基于此,本实验选定单双音节对应律偏误作为检验大语言模型书面语口语化纠偏能力的任务项。

### 3.2.3 大语言模型选用与提示设计优化

为筛选出最适用于留学生语体纠偏辅助教学的大语言模型,本研究以留学生书面语口语化偏误中发生率最高的单双音节对应律偏误为研究对象,对GPT-3.5(以下简称ChatGPT)、ERNIE-4.0(以下简称文心一言)和Spark-V3.5(以下简称讯飞星火)三种大语言模型的表现开展对比分析。

为增强研究设计的科学性与样本代表性,研究团队首先从HSK动态作文语料库中系统选取30篇经书面语口语化标注的留学生作文,构建为小样本提示(Few-shot Prompting)的示例集。在正式评测前,研究团队通过对话框向三个模型进行示例展示,引导其在当前会话中完成上下文学习(In-Context Learning),使其理解单双音节对应律偏误的识别标准与修改格式。

随后,在同一会话页面以16篇独立作文作为测试样本,对各模型的识别性能进行评估。

表 1. 大语言模型偏误识别准确度对比

	TP <sup>10</sup>	FP <sup>11</sup>	FN <sup>12</sup>	准确率	召回率	F1
文心一言	64	120	111	0.348	0.366	0.366
讯飞星火	83	96	92	0.464	0.474	0.469
ChatGPT	102	58	73	0.638	0.583	0.609

为评估模型准确度,本研究从真阳性(TP)、准确率、召回率与F1分数等多项指标展开综合分析。结果显示,ChatGPT在各项指标上均显著优于其他模型,其TP为102个,准确率达0.638,召回率为0.583,F1分数为0.609;讯飞星火居中,文心一言则表现相对较弱。综上,

ChatGPT 在识别准确率与综合性能上均表现最优,说明其在该类偏误识别任务中具备更高的可靠性与实际应用潜力。基于此,本实验选定 ChatGPT 作为识别留学生作文中书面语口语化偏误(单双音节对应律偏误)的大语言模型工具,在原示例展示和评估的会话进行。

### 3.3 实验方案

#### 3.3.1 实验流程

实验开展前,研究团队完成对被试信息的采集。实验过程中,被试首先在一周时间内完成命题作文《就业危机》(以下称为“作文1”)。提交后,被试将依据主试提供的指令词,利用 ChatGPT 对作文中单双音节对应律偏误进行识别与修改。在此过程中,被试通过 ChatGPT 进行主动的作文修改与语体学习(交互指令词如下图所示);同时,主试团队将对作文1中的单双音节对应律偏误进行标注。每篇文章均由5名主试独立标注,对于同一处偏误,若有超过3名标注者共同标出,则计为该篇文章的一处有效偏误。标注者间信度通过 ICC 系数进行量化,结果为 0.76,表明信度良好。学习阶段结束后,被试在后一周内完成第二篇命题作文《我看流行音乐》(以下称为“作文2”)。以上两篇作文题目均选自 HSK 动态作文语料库中的议论文题目。作文2在一周后交由同一主试团队采用相同流程进行标注,最终通过对比两篇作文的偏误数据得出结论。

图 1. 交互指令词示意图

#### 交互指令词设计

- 指令1: 你好,我是一名留学生,现在在中国留学学习汉语。需要你帮我找出我的作文中出现的书面语口语化偏误中的单双音节对应律偏误。我写了一篇作文,作文要求如下。(具体作文要求)如果你读懂了上述要求,请回答“我已知晓,请进行下一步。”
- 指令2: 好的。下面是我的作文。(此处留学生粘贴自己的作文段落之一)请结合作文要求,对我的作文进行针对性纠偏,并帮我提供改正建议。在纠偏过程中注意要使用以下格式进行回答。(具体格式)
- 指令3: 我对“××××”不理解,你能为我解释一下吗?
- 指令4: 我学好了知识,现在想进行单双音节对应律的相关练习。请为我生成两个练习,且不生成答案。
- 指令5: 请告知我答案。

本实验周期共计两周,所观察到的训练效果主要反映 ChatGPT 辅助训练对留学生书面语口语化偏误的短期干预成效。后续研究可进一步延长实验周期,开展追踪调查,以检验训练效果的持续性与稳定性。

#### 3.3.2 学习者反馈收集

为全面评估学习者对 ChatGPT 作为语体学习辅助工具的体验与认知,本研究在实验结束后开展了结构式访谈与问卷调查,旨在从学习者视角补充验证定量分析结果,并为进一步优化人机协同教学模式提供参考依据。

研究采用 Braun & Clarke (2006) 提出的主题分析法,对问卷文本的质性分析遵循自下而上的归纳流程,主要包含三个阶段:首先进行初始编码,编码者通过反复阅读文本将相关内容抽象为关键概念并建立编码;其次进行主题归类,编码者通过比较和整合初始编码,从中提炼出更高级别的主题类别并厘清其关系;最后总结核心主题,在所有类别中识别出能够统领和解释整体现象的核心主题,从而形成最终的编码框架(如表2所示)。

表 2. 主题编码框架表

编码框架				
核心主题	内涵	主题归类	定义	对应初始编码示例
A. 效能感知 与学习赋能	学习者肯定ChatGPT在提升学习效率、语言能力及多场景辅助方面的积极作用	A1效率与便捷性	学习者高度评价ChatGPT在节省时间、快速响应、简化流程方面的价值	省时省力、很快就提供出答案
		A2多功能集成支持	学习者认可其在写作、语法、翻译、举例、课件制作等多方面的辅助能力	改善词汇用法、改正语法错误、帮文章的意思更顺畅更有逻辑
		A3语言能力提升感知	学习者认为使用后写作更规范、表达更流畅、语体意识增强	表达更流畅、语体转换有效
B. 局限认知 与风险意识	学习者认识到ChatGPT在技术准确性、文化语境理解、情感交互等方面的不足	B1技术可靠性局限	学习者指出其可能出现错误、识别不够精准、答案需进一步验证	偶尔会出现错误、识别不够精准、需自行判断
		B2文化语境理解缺失	学习者强调其在文化背景、历史内涵、语境适配方面的不足	缺乏文化背景讲解、语境适配差、难以解释新潮用语
		B3情感与深度交互不足	学习者认为其缺乏情感交互、深层逻辑解释与人文关怀	没有情感、解释不深入、缺乏创造性
C. 使用定位 与协同观念	学习者明确ChatGPT的辅助工具属性，强调人机协同、教师引导与自主思考并重的使用哲学	C1辅助工具定位明确	学习者明确将其定位为“辅助工具”，强调其补充而非替代作用	只能作为辅助工具、不能替代教师
		C2批判使用与风险警示	学习者警惕过度依赖可能导致的思维惰性、自主判断力下降	不能过度依赖、警惕思维惰性、需自己思考
		C3人机协同教学观	学习者主张与教师引导、传统教学方法相结合，形成互补	结合传统教学、教师引导不可替代、需与真人互动结合
D. 改进期望 与优化建议	学习者从功能、场景、实验设计等角度提出具体改进建议，反映其期待更加智能化、人性化、情境化的学习支持	D1功能优化需求	学习者希望增加水平适配、文化融入、逻辑分析、多模态等功能	按水平推荐、增加文化背景
		D2场景与交互优化建议	学习者建议区分实验与日常场景、优化指令设计、增加练习环节	增加错误点练习、区分实验与日常场景、优化指令设计
		D3实验设计反馈	学习者对实验主题、时长、任务设计、沟通方式提出改进意见	实验时间延长、主题和时间安排优化、加强提问与沟通

为确保研究过程严谨、分析结果可靠，本研究在编码阶段采取了以下策略对分析的信度与效度进行检验与控制。在正式编码前，研究者与经过培训的编码员共同对编码框架进行讨论与校准。随后，从全部 31 份访谈转录稿中随机抽取 6 份（约 20%）作为信度检验样本，由两位编码员分别进行独立编码。编码完成后，对两份编码结果进行科恩卡帕系数计算。得到 Kappa 值为 0.84，根据 Landis 与 Koch（1977）的判断标准，该结果高于 0.80，表明编码的一致性较高，编码框架具有较高的客观性与稳定性。对于编码不一致的个别条目，两位编码员进行了复核与讨论，依据原始文本与编码框架达成共识。

## 4 实验结果

### 4.1 数据统计方式

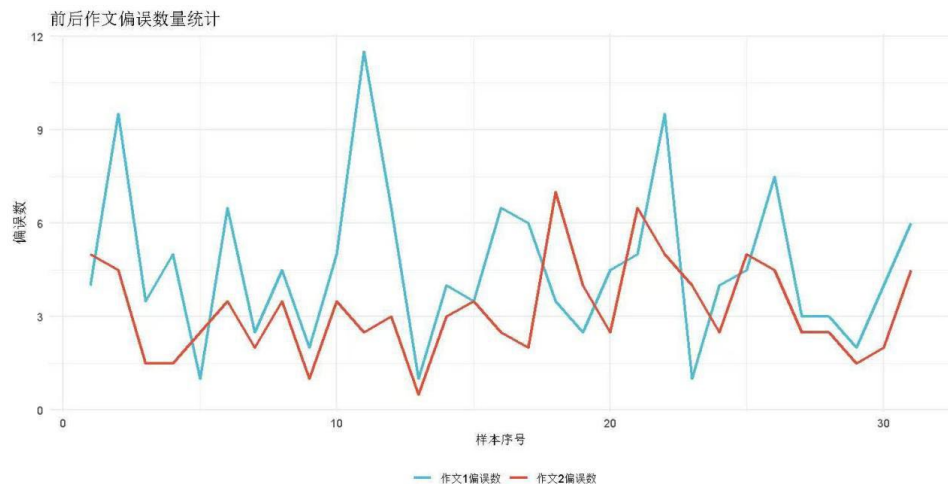
本研究采用多维度数据分析策略：首先，计算每位被试第二次写作相较于第一次的错误变化量，分别统计“华裔被试”“汉语二语（非华裔）被试”与“全体”三个组的平均错误变化量，并整合为统一的数据框架。其次，通过箱线图呈现各组错误变化量的分布特征，直观展示数据离散程度与集中趋势。进一步地，计算训练前平均错误数、训练后平均变化量及其占原始错误数的百分比，以可视化图表呈现，并对比分析两组在“平均变化量”和“增长比例”上的差异。同时，通过自定义函数对每组数据进行分类计数与比例计算，呈现不同群体中写作错误数“正增长”与“负增长”的人数比例。

鉴于华裔被试组样本量较小（ $n=7$ ），为增强统计检验的稳健性并验证两组学习者的进步表现是否存在显著差异，本研究采用非参数检验方法进行效果评估：使用配对样本 Wilcoxon 检验分别对汉语二语被试组和华裔被试组进行组内前后测比较，并通过独立样本 Mann-Whitney U 检验对两组的进步幅度进行组间比较，以综合评估训练的有效性。

## 4.2 数据呈现与分析

### 4.2.1 前后增长偏误量变化

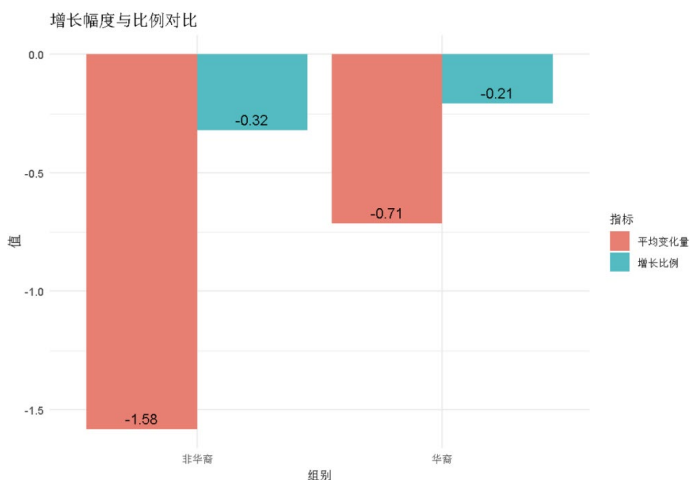
图 2. 被试前后作文偏误数量统计



从图 2 中可直观看到全体被试在两次写作中的偏误数整体呈现下降趋势，不同样本的偏误减少幅度存在个体差异。需要说明的是，本实验中华裔被试仅 7 人，汉语二语被试 24 人，样本结构存在一定不平衡。小样本群体易受个体差异影响，组内波动较大，本文在后续数据分析中不再对样本结构差异进行反复强调，但在解读华裔被试组数据时需考虑其可能对统计检验效力产生的影响。

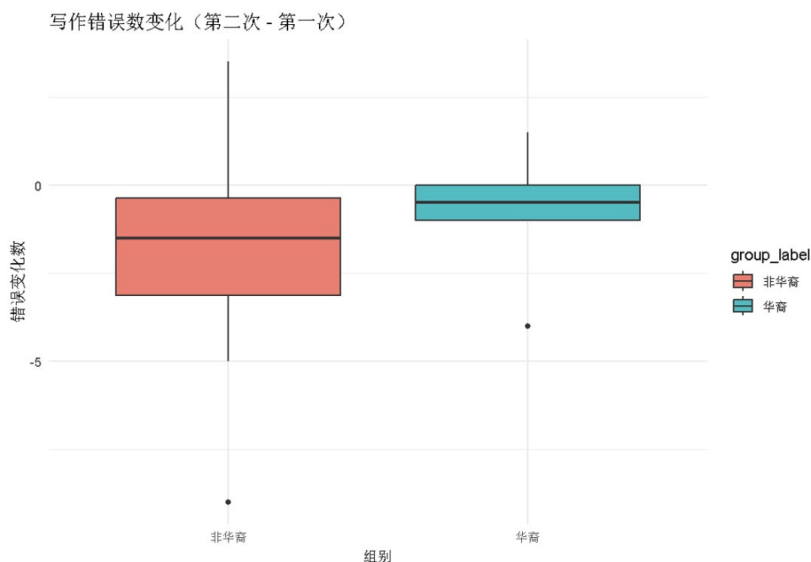
从具体数据来看，作文 1 中全体被试平均偏误数为 4.60 个，其中华裔被试平均偏误数为 4.29 个，汉语二语被试平均偏误数为 4.66 个，可见汉语二语被试的汉语语体意识相对低于华裔被试，书面作文中口语化偏误更多。作文 2 中，全体被试平均偏误数降低至 3.21 个，华裔被试平均偏误数为 3.57 个，汉语二语被试平均偏误数为 3.09 个，此时汉语二语被试偏误减少数量高于华裔被试。

图 3. 华裔被试与汉语作为第二语言被试作文偏误增长幅度与比例对比



从图 3 中可清晰直观地看到华裔被试与汉语二语被试在偏误平均变化量和增长比例上的差异。全体被试作文 2 较作文 1 平均减少偏误 1.39 个，华裔被试平均减少偏误 0.71 个，汉语二语被试平均减少 1.58 个。无论是绝对变化量还是相对变化比例，汉语二语被试组均显著优于华裔被试组。

图 4. 作文偏误数量变化箱线图



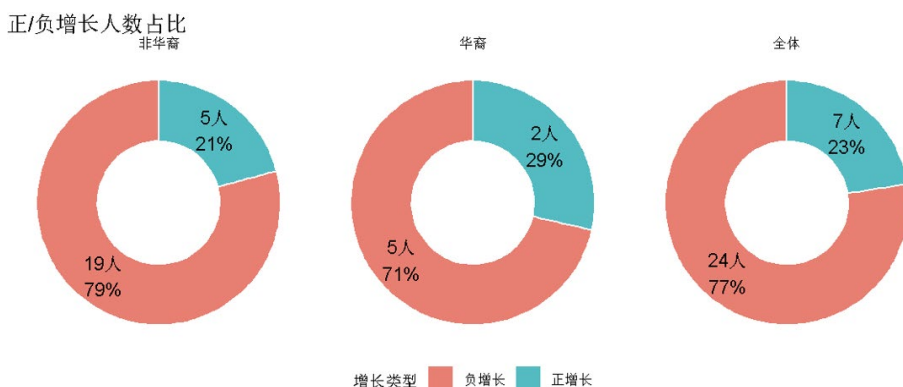
结合图 4 作文偏误数量变化箱线图可见，由于样本量有限，数据分布相对集中，华裔被试组中位数趋近于 0，部分数据为正值，显示该组整体变化不明显，甚至存在个别退步现象。

汉语二语被试初始口语化偏误基数较大，在与 ChatGPT 交互学习过程中，因其对书面语规范的认知起点较低，提升空间较大，所以能从反馈中获得更显著的改进。而华裔被试本身具有一定的汉语基础，部分口语化表达可能源于习惯，较难通过短期交互学习改变，且可能存在对模型反馈的接受度与应用效率不足的问题。

因此，与 ChatGPT 交互学习对汉语二语被试减少书面语口语化偏误有更明显的效果，能显著提升其作文的语体意识，减少书面语口语化偏误问题；华裔被试虽有进步，但幅度较小，在利用大语言模型改进书面语口语化偏误方面效果不如汉语二语被试。

#### 4.2.2 正负增长比率

图 5. 作文 2 较作文 1 偏误数量增长人数占比环形图



“正负增长比率”指在 ChatGPT 辅助训练后，留学生第二次作文相比第一次在书面语口语化偏误数量上发生不同方向变化的人数比例。其中，“正增长”表示偏误数量增加，反映写作表现退步；“负增长”表示偏误数量减少，反映写作表现进步。该比率用于直观呈现训练效果在不同学习者群体中的分布情况，揭示干预措施效果的普遍性与个体差异。

图 5 展示了作文 2 较作文 1 偏误数量增长人数占比情况，汉语二语被试组约 80% 的学生为负增长，说明该组大部分学生在与 ChatGPT 交互学习后作文语体意识进步显著；华裔被试组正增长占比相对更高，表明较多该组被试在第二次写作中表现退步；从整体趋势看，负增长占主导，证实 ChatGPT 辅助训练对减少偏误具有总体有效性。

#### 4.2.3 显著性检验

为评估训练效果，本研究采用配对样本 Wilcoxon 检验分析各组前后测偏误数的变化，并借助独立样本 Mann-Whitney U 检验比较两组间的进步幅度差异，从而对训练成效进行综合检验。

表 3. 显著性检验数据表

组别	均值差	P 值 <sup>13</sup>	统计量 (V)	效应量 (r) <sup>14</sup>	显著性
全体被试	-1.387	0.003	367.5	0.537	显著
华裔被试	-0.714	0.222	19	0.461	不显著
汉语二语被试	-1.583	0.003	228	0.602	显著

对于样本数据采用非参数的 Wilcoxon 符号秩检验进行前后测对比。结果显示，全体被试在训练后的写作错误数显著降低 ( $V=367.5$ ,  $p=0.003$ )，华裔被试组从测试 1 到测试 2 平均减少 0.71 个错误，但未达到统计学显著水平 ( $V=19$ ,  $p=0.222$ )；而汉语二语被试组 ( $n=24$ ) 平均减少 1.58 个错误，进步显著 ( $V=228$ ,  $p=0.003$ )。Mann-Whitney U 检验显示两组的进步幅度并无显著差异 ( $W=69$ ,  $p=0.769$ )，且华裔被试组的效应量  $r=0.461$  (中等效应) 表明可能存在实际进步趋势。

综上，大语言模型辅助学习对减少留学生书面语口语化偏误具有积极意义，尤其对汉语二语学习者效果较明显；对华裔学习者的作用虽未达统计显著，但初步数据支持其潜在有效性，后续可在保证样本量的基础上进一步探讨其适用性与优化路径。

### 4.3 讨论：ChatGPT 辅助语体纠偏的效能分析

#### 4.3.1 训练效果的整体有效性及群体差异

本研究发现，经过 ChatGPT 辅助训练后，全体被试的书面语口语化偏误数量显著减少，这一结果初步验证了大语言模型作为语体教学辅助工具的可行性。从正负增长比率来看，约 80% 的汉语二语学习者在训练后偏误数下降，表明 ChatGPT 辅助训练对该群体具有相对的适用性。

训练效果在不同语言背景学习者群体间呈现出一定差异：

##### (1) 汉语二语被试组的显著进步

汉语二语被试组在训练后偏误数平均减少 1.58 个，效应量  $r=0.602$ ，达到大效应水平。这一群体初始偏误基数较大，对书面语规范的认知起点较低，在与 ChatGPT 交互过程中获得了较大的提升空间。

## (2) 华裔被试组的非显著进步趋势

华裔被试组虽平均偏误数从 4.29 个降至 3.57 个，但未达到统计学显著水平。然而，效应量  $r=0.461$  表明存在中等效应的实际进步趋势。

### 4.3.2 组间对比及成因分析

从偏误数量的变化分布来看，汉语二语被试组的进步幅度明显大于华裔被试组。可以推知，ChatGPT 辅助训练更适合于那些对书面语规范尚未形成稳定认知的学习者，使他们能够从模型提供的即时反馈中获得明确的修改方向，并在后续写作中有意识地规避同类偏误。

相比之下，华裔被试组的学习机制可能更为复杂。对于已经具备一定语体意识的学习者，ChatGPT 提供的表层纠偏反馈可能不足以撼动其固有的语言习惯；他们需要更具深度和解释性的反馈，才能真正实现语体能力的进阶。同时，被试内部的语言能力差异也可能对个体学习效果产生影响——虽华裔被试第一语言均为汉语，但其海外生活经历长短、家庭语言环境等因素可能存在差异，这些变量在本研究中未能有效控制。

在解读本研究发现时，需考虑两方面的潜在影响。其一，本研究未设置无干预对照组，后测偏误减少可能部分源于重复写作带来的练习效应，难以完全归因于 ChatGPT 干预。其二，前后测作文题目《就业危机》与《我看流行音乐》虽同属议论文，但未经严格的难度等价性检验，题目差异也可能对结果产生一定影响。后续研究应通过设置对照组、平衡题目难度等方式，进一步明确大语言模型辅助的净效能。

### 4.3.3 学习者反馈分析

基于对 31 名汉语水平为 HSK 4-6 级的留学生的问卷文本进行主题分析，本研究系统梳理了其在效能感知、局限认知、使用哲学与改进期望四个核心维度上的反馈，如图表所示。

表 4. 学习者反馈群体分布

核心主题	主题归类	提及人数	华裔	非华裔	总人数中占比	华裔提及占比	非华裔提及占比
A. 效能感知	A1 效率便捷	17	3	14	54.84%	42.86%	58.33%
	A2 多功能支持	4	0	4	12.90%	0.00%	16.67%
	A3 能力提升	4	0	4	12.90%	0.00%	16.67%
A类提及人数总数据		20	3	17	64.52%	42.86%	70.83%
B. 局限认知	B1 技术可靠性	8	4	4	25.81%	57.14%	16.67%
	B2 文化语境缺失	7	2	5	22.58%	28.57%	20.83%
	B3 情感交互不足	1	0	1	3.23%	0.00%	4.17%
B类提及人数总数据		15	4	11	48.39%	57.14%	45.83%
C. 使用哲学	C1 辅助工具定位	4	1	3	12.90%	14.29%	12.50%
	C2 批判使用意识	4	1	3	12.90%	14.29%	12.50%
	C3 人机协同观念	1	0	1	3.23%	0.00%	4.17%
C类提及人数总数据		6	2	4	19.35%	28.57%	16.67%
D. 改进期望	D1 功能优化	4	0	4	12.90%	0.00%	16.67%
	D2 场景交互优化	3	0	3	9.68%	0.00%	12.50%
	D3 实验设计反馈	8	2	6	25.81%	28.57%	25.00%
D类提及人数总数据		14	2	12	45.16%	28.57%	50.00%

具体来看，在各项主题的提及人数上，“效率便捷”（A1）以 17 次提及形成显著峰值，表明学习者对 ChatGPT 辅助语体学习最直观、最强烈的感受集中于即时响应与操作低门槛；在写作场景中，ChatGPT 能够快速识别偏误、即时提供修改建议，这一点相较于传统课堂反馈周期更具优势。

“技术可靠性”(B1)与“实验设计反馈”(D3)同获8次提及。B1集中体现了学习者对模型输出准确性的审视态度。究其原因,大语言模型基于概率生成文本的本质使其存在“幻觉”风险,在二语写作场景中可能生成看似合理却不符合汉语语体规范或文化语境的表达,进而误导学习者(陆俭明,2025)。因此,模型输出的可信度问题不仅是当前体验的短板,更需在后续应用中通过教师监督、提示工程与人工复核机制加以系统应对。D3则指向学习者对交互流程、任务适配性等实验设计层面的具体反思。

上述分析结果表明,ChatGPT等大语言模型的应用需要建立在系统化的“人机协同”框架之下。这一框架既需要语言学家与教育者共同构建专业化语料库,以提升模型反馈的准确性与文化适配性;也应在教学实践中明确技术工具的辅助定位,由教师发挥教学引领作用,学生践行自主学习理念,技术工具提供智能支持,从而构建三者深度融合、相互促进的教育闭环。

#### 4.3.4 研究发现对语体教学的启示

本研究聚焦于单双音节对应律偏误——这一在留学生书面语口语化偏误中占比最高的类型(49.6%)。实验结果表明,通过ChatGPT辅助训练,学习者的该类偏误显著减少。这一发现为国际中文教学提供了两点启示:

第一,对于形式化程度高、规律性强的语体偏误类型,大语言模型可以作为有效的辅助教学工具。单双音节对应律具有较为明确的转换规则,模型能够通过上下文学习掌握这些规律,并为学习者提供准确的修改建议。

第二,模型的辅助效能与学习者的语言背景密切相关。对于汉语二语学习者,模型反馈能够显著提升其语体意识;而对于华裔学习者,模型的辅助作用可能有限,需要结合教师的深度讲解与针对性训练,才能实现语体能力的实质性提升。

#### 4.3.5 技术层面反思及未来方向

##### (1) 模型筛选阶段的局限性

本研究在筛选模型时,使用了规模有限的示例集进行提示示例学习和评估。这种小样本设置可能对模型性能评估的稳定性带来一定挑战,性能指标存在波动。因此,本研究得出的ChatGPT优于其他模型的结论,是在本次特定实验设置下观察到的结果。未来研究可通过扩大示例库与测试集的规模、采用交叉验证等方法,增强评估结果的稳定性和普适性。

##### (2) 模型基线能力的未知性

本研究在模型筛选阶段,直接使用了经过提示示例学习后的性能进行横向比较,未能预先测试并记录各模型在零样本条件下的原始识别能力。因此,虽然数据显示ChatGPT表现最佳,但缺乏训练前后的直接对比,难以严格量化提示示例学习对模型性能的具体提升幅度,也无法完全排除模型本身原有能力差异对结果的影响。未来研究应在提示学习前增设零样本测试环节,获取模型基线数据,从而更精准地评估提示策略的有效性。

## 5 结语

本研究系针对大语言模型辅助国际中文教育留学生书面语口语化偏误纠正效能的探索性研究,通过实证分析,初步探讨了大语言模型在汉语语体学习中的辅助作用,印证了李韵等(2025)既有研究关于大模型写作反馈支持功能的结论。结果显示,ChatGPT在单双音节对应

律偏误识别方面具有一定优势；在模型辅助纠偏的支持下，汉语二语被试偏误减少显著，而华裔被试虽平均错误数下降，但受样本量所限未达到统计显著水平。

本研究通过非参数检验验证了模型辅助训练在整体及汉语二语学习者中的有效性，为语体教学的智能化路径提供了初步实证参考。在此基础上，研究尝试验证前人研究中提出的“智能诊断—自主探究”的教学环节，由大语言模型承担诊断与反馈支持角色，学习者在模型辅助下自主完成偏误识别、修改与练习，教师则可在后续教学中有机介入、深化引导。这一路径在降低教师批改负荷的同时，增强反馈的即时性与学生的参与感，从而为国际中文教学的数字化转型探索可行路径。

本研究还存在一定局限。就技术而言，学习者反馈大模型生成内容容易包含幻觉，与陆俭明（2025）结论相似。因此，在投入教学使用与开展实验前，应通过大规模汉语语料进行数据预处理和问答对训练，以提高输出内容精准度；就研究范畴而言，实验仅聚焦单双音节对应律这一种偏误类型，尚未覆盖其他语体偏误类型。这些研究缺陷有待后续研究进一步弥补。基于研究发现，未来研究可在以下方向实现突破：一是可以开发融合语言学知识库的专用模型，通过注入汉语语体特征规则来提升反馈准确性；二是在具体教学应用中针对初级、中级、高级学习者设计差异化的语体训练模块；最后，可以开展跨学科合作，将计算语言学的算法优势与二语习得的认知理论相结合，使教师、智慧工具和学习者三者协同运作。

综上，本研究证实大语言模型可作为语体教学的有效辅助工具，但仍需对模型本身进行专门化训练，使之更好地服务于汉语学习者的语言学习，进一步促进国际中文教育提质增效，实现数字化转型。

## 注释

1. 所有语距定律详见：冯胜利、刘丽媛（2020）《语体语法的生物原理与生成机制》，《民俗典籍文字研究》第2期，第76-103+281-282页。
2. 长短律与平衡律：相对的两个节律单位如长短不一，则具口语性；相对的两个节律单位如长短一致，则具正式性。
3. 悬差律：相差一半（或以上）的两个韵律单位的组合，均具谐俗性。
4. 单双音节对应律：汉语单双音节对应之词，单音节具口语性，双音节均具正式性。
5. 融合律：两个邻接的语法单位在韵律压力下发生融合或删除，成为一个音节，均具口语属性。
6. 单调律：缩短句法单位的标准长度以满足“一句一动一调”的运作结果，均具口语的属性。
7. 时空律：时空标记越少，或具有时空要素的语言单位所含“时空素”越少，则越正式，反之亦然。
8. 成分追加律：在句末进行的句法成分追加或者复说的运作，均具口语体。
9. 模型正确捕捉到目标对象的次数。
10. 模型的误报，即错误地将负样本识别为正样本。
11. 模型的漏报，即未能成功抓取出实际存在的正样本。
12. 本研究以0.05为统计学显著性水平，即p值小于0.05时视为组内前后测差异或组间差异具有统计学意义，p值大于或等于0.05则视为未达到显著水平。
13. 本研究以0.10、0.30、0.50为效应量r的评判标准，即r值在0.10附近为小效应，0.30附近为中等效应，0.50及以上为大效应；r值越大表示效应越强，不涉及“显著”与否的判断，而是用于衡量差异的实际意义大小。

## 致谢

本研究系国家级大学生创新实践项目“基于语言大模型的留学生习作语体偏误分析与纠偏实践”(项目编号:202510269028G)的阶段性研究成果,特此感谢学校、学院提供的宝贵平台与资源支持,使本研究得以顺利开展。

## 参考文献

- Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, 3(2), 77–101.
- Guo, K., & Wang, D. (2024). To resist it or to embrace it? Examining ChatGPT's potential to support teacher feedback in EFL writing. *Education and Information Technologies*, 29(7), 8435–8463.
- Han, J., & Li, M. (2024). Exploring ChatGPT-supported teacher feedback in the EFL context. *System*, 126, 103502.
- Koo, T. K., & Li, M. Y. (2016). A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *Journal of Chiropractic Medicine*, 15(2), 155–163.
- Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, 33(1), 159–174.
- Bi, Yude (毕玉德), & Yan, Danhui (闫丹辉). (2024). 通用人工智能时代关于语言研究的思考 [Reflections on Linguistic Study in the Age of Artificial General Intelligence]. *语言文字应用 [Applied Linguistics]*, (03), 119-131. <https://doi.org/10.16499/j.cnki.1003-5397.2024.03.007>
- Ding, Jinguo (丁金国). (1997). 对外汉语教学中的语体意识 [Stylistic Consciousness in Teaching Chinese to Foreigners]. *烟台大学学报 (哲学社会科学版) [Journal of Yantai University (Philosophy and Social Science Edition)]*, (01), 89-96. <https://doi.org/10.13951/j.cnki.issn1002-3194.1997.01.015>
- Ding, Jinguo (丁金国). (1999). 再论对外汉语教学中的语体意识 [Re-exploring stylistic consciousness in teaching Chinese to foreigners]. *语言文字应用 [Applied Linguistics]*, (02), 28-32. <https://doi.org/10.16499/j.cnki.1003-5397.1999.02.006>
- Feng, Shengli (冯胜利). (2003). 书面语语法及教学的相对独立性 [An Independent Grammar for Written Chinese in Second Language Teaching]. *语言教学与研究 [Language Teaching and Linguistic Studies]*, (02), 53-63.
- Feng, Shengli (冯胜利). (2005). 汉语韵律语法研究 [Studies on Chinese Prosodic Grammar.] 北京大学出版社 [Peking University Press].
- Feng, Shengli (冯胜利). (2006). 论汉语书面正式语体的特征与教学 [On the properties and pedagogy of written Chinese]. *世界汉语教学 [Chinese Teaching in the World]*, (04), 98-106+148. <https://doi.org/10.13724/j.cnki.ctiw.2006.04.015>
- Feng, Shengli (冯胜利). (2010). 论语体的机制及其语法属性 [On mechanisms of Register System and its grammatical property]. *中国语文 [Studies of the Chinese Language]*, (05), 400-412+479.
- Feng, Shengli (冯胜利), & Shi, Chunhong (施春宏). (2018). 论语体语法的基本原理、单位层级和语体系统 [The Principles, Hierarchical Units and Register System of Register Grammar]. *世界汉语教学 [Chinese Teaching in the World]*, 32(03), 302-325. <https://doi.org/10.13724/j.cnki.ctiw.2018.03.002>
- Feng, Shengli (冯胜利), & Liu, Liyuan (刘丽媛). (2020). 语体语法的生物原理与生成机制 [On Biologic Principles and Generative Mechanisms in Stylistic-Register Grammar]. *民俗典籍文字研究 [Studies on Folkloristics Classics and Philology]*, (02), 76-103+281-282.

- Gao, Anna (高安娜). (2024). ChatGPT 辅助国际中文教学的自我叙事探究: 机遇和挑战 [Self-Narrative Inquiry in ChatGPT Assisted TCSL: Opportunities and Challenges]. (eds.) 中文教学现代化研究——2024 年专题文集 [Research on Modernization of Chinese Language Teaching: 2024 Thematic Collection] (pp.109-120). 华东师范大学国际汉语文化学院;
- Guo, Yingwen (郭颖雯). (2002). 汉语口语体口语教学语法体系的建立与量化 [On the Establishment and Specification of a Grammar System in the Oral Teaching of Chinese Language]. 汉语学习 [Chinese Language Learning], (06), 63-67.
- Hu, Mingyang (胡明扬). (1993). 语体和语法 [Register and Grammar]. 汉语学习 [Chinese Language Learning], (02), 1-4.
- Li, Jianfang (李建芳). (1998). 试析语体的习得过程 [An Analysis of the Acquisition Process of Register]. 修辞学习 [Contemporary Rhetoric], (01), 14-15. <https://doi.org/10.16027/j.cnki.cn31-2043/h.1998.01.008>
- Li, Lili (李丽丽). (2005). 在对外汉语教学中培养学生的口语交际能力 [Cultivating Overseas Students' Oral Communicative Competence in TCFL]. 云南师范大学学报 [Journal of Yunnan Normal University (Teaching & Studying Chinese as a Foreign Language Edition)], (01), 72-76. <https://doi.org/10.16802/j.cnki.ynsddw.2005.01.017>
- Li, Quan (李泉). (2003). 基于语体的对外汉语教学语法体系构建 [Building of Grammar System for Teaching Chinese as a Second Language Based on Style of Writing]. 汉语学习 [Chinese Language Learning], (03), 49-55.
- Li, Yun (李韵), Han, Jining (韩继宁) & Li, Yutong (李雨瞳). (2025). 人机双师协同写作反馈在汉语二语教学中的应用及其效果评估 [Application and evaluation of human-AI dual-teacher collaborative feedback in Chinese second language writing]. 语言教学与研究 [Language Teaching and Linguistic Studies], (02), 25-37.
- Liu, Yu (刘瑜), Tan, Tan (谭坦) & Chen, Qing (陈青). (2024). ChatGPT 在汉语二语写作中的纠正反馈能力考察 [ChatGPT Corrective Feedback Ability for Chinese L2 Writing]. 国际汉语 [Chinese Language in the World], (00), 83-98+164-165.
- Lu, Jianming (陆俭明). (2025). 大语言模型的“语言”跟自然语言性质迥然不同 [The “Language” of Large Language Models Differs Fundamentally from Natural Language]. 语言战略研究 [Chinese Journal of Language Policy and Planning], 10(01), 1.
- Lu, Yunfei (路云菲). (2025). 高级阶段留学生汉语语体转换能力调查研究 [Research on the Ability of Advanced Stage International Students to Change Chinese Language Registers] (硕士学位论文, 河北师范大学) [Unpublished Master's Thesis, Hebei Normal University]. 硕士 <https://doi.org/10.27110/d.cnki.ghsfu.2025.000679>
- Ma, Ruiling (马瑞凌). (2026). 大语言模型赋能中文二语作文智能评改的应用与评价 [Application and Efficacy Assessment for Large Language Models Enabled Intelligent Evaluation & Correction of L2 Chinese Writing]. 语言教学与研究 [Language Teaching and Linguistic Studies], (01), 13-23.
- Ren, Wei (任伟), Liu, Yuanbo (刘远博) & Xie, Yu (解月). (2024). 英语写作教学中 ChatGPT 与教师反馈的对比研究 [A Comparative Study of ChatGPT and Teacher Feedback in English Writing Instruction]. 外语教学理论与实践 [Foreign Language Learning Theory and Practice], (04), 30-38+60.
- Si, Hongxia (司红霞). (2025). 国际中文教育的语体教学新思考 [New Thoughts on Register Teaching for International Chinese Education]. 语言教学与研究 [Language Teaching and Linguistic Studies], (02), 14-24.
- Wu, Lijun (吴丽君). (2004). 口语词汇与书面语词汇教学研究 [Teaching of vocabulary of the spoken and written Chinese language]. 云南师范大学学报 [Journal of Yunnan Normal University (Teaching & Studying Chinese as a Foreign Language Edition)], (03), 14-19. <https://doi.org/10.16802/j.cnki.ynsddw.2004.03.004>

- Yin, Yunyi (殷韵怡). (2022). 对外汉语口语教材语体意识研究 [Research on the Stylistic Consciousness in Spoken Chinese Textbooks for Foreign Learners: A Case Study of Developing] (硕士学位论文, 江西科技师范大学) [Unpublished Master's Thesis, Jiangxi Science and Technology Normal University]. 硕士 <https://doi.org/10.27751/d.cnki.gjxkj.2022.000112>
- Yuan, Yulin (袁毓林). (2024). ChatGPT等大模型的语言处理机制及其理论蕴涵 [ChatGPT and Other Large Language Models: Their Language Processing Mechanisms and Their Theoretical Implications]. 外国语 (上海外国语大学学报) [Journal of Foreign Languages], 47(04), 2-14.
- Zhai, Xu (翟汛). (2007). 对外汉语口语教学的几点思考 [Reflections on Oral Chinese Teaching for Learners of Chinese as a Foreign Language]. 长江学术 [Yangtze River Academic], (02), 96-100.
- Zhang, Li (张荔), Yu, Yangyang (郁杨洋) & Pan, Zhixi (潘之欣). (2026). 大语言模型辅助二语写作反馈方式及效果研究 [Exploring the Approaches and Effects of ChatGPT-Assisted Feedback in Second Language Writing]. 外语学刊 [Foreign Language Research], (01), 77-86. <https://doi.org/10.16263/j.cnki.23-1071/h.2026.01.012>
- Zhou, Liwei (周立炜) & Rao, Gaoqi (饶高琦). (2024). 大语言模型中文语体能力评测研究 [Research on Evaluation of Chinese Stylistic Competence on Large Language Models]. 语言文字应用 [Applied Linguistics], (01), 69-82. <https://doi.org/10.16499/j.cnki.1003-5397.2024.01.007>
- Zhou, Yikai (周义凯) & Liu, Haitao (刘海涛). (2024). ChatGPT掌握现代汉语书面语的句长规律了吗? [Has ChatGPT Mastered the Sentence Length Patterns of Written Modern Chinese?]. 语言文字应用 [Applied Linguistics], (02), 124-134. <https://doi.org/10.16499/j.cnki.1003-5397.2024.02.010>
- Zhu, Dexi (朱德熙). (1985). 现代书面汉语里的虚化动词和名动词为第一届国际汉语教学讨论会而作 [Delexical Verbs and Verbal Nouns in Modern Written Chinese: Presented at the First International Conference on Chinese Language Teaching] 北京大学学报 (哲学社会科学版) [Journal of Peking University (Philosophy and Social Sciences)], (05), 3-8.
- Zhai, Yan (翟艳) & Su, Yingxia (苏英霞). (2022). 汉语技能教学 (第2版) [Chinese as a Second Language Skills Teaching (Second Edition)] 北京语言大学出版社 [Beijing Language and Culture University Press].

投稿: 2025年12月15日; 接受: 2026年3月8日; 出版: 2026年4月10日

## 作者简介

康雪, 华东师范大学国际汉语文化学院在读, 研究方向为国际中文教育, 语体教学。

王曦悦, 华东师范大学国际汉语文化学院在读, 研究方向为中文信息处理、国际中文教学、社会语言学。

余兰心, 华东师范大学统计学院在读, 研究方向为在线共形预测。

祁峰, 教授, 华东师范大学国际汉语文化学院, 研究方向为语言学、国际中文教学、语料库语言学。

## **An Empirical Evaluation of ChatGPT's Effectiveness in Assisting International Students with Genre-Style Chinese Writing Correction**

**Xue Kang**

**Xiyue Wang**

**Lanxin Yu**

**Feng Qi**

East China Normal University, China

### **Abstract**

This study investigates the practical efficacy of ChatGPT and other large language models (LLMs) in assisting genre-style correction in writing, addressing the issue that international students' genre competence often lags behind their grammatical competence in international Chinese language education. A sample of 2,000 compositions was drawn from the HSK Dynamic Composition Corpus, revealing that the colloquialization errors in students' written productions predominantly involve the laws of monosyllabic-disyllabic correspondence, length, and temporospatiality, among which violations of the monosyllabic-disyllabic correspondence law account for 49.6% of all errors. A comparative evaluation of three LLMs—ChatGPT, Ernie Bot, and iFlytek Spark—was conducted, and ChatGPT was selected as the experimental tool. An experiment was carried out with 31 Chinese learners at the HSK Level 4-6 from East China Normal University. The results indicate that ChatGPT-assisted training significantly reduces colloquialization errors among Chinese-as-a-second-language (CSL) participants. For heritage learners, although the reduction did not reach statistical significance, a positive trend was observed. This study provides preliminary evidence supporting the feasibility of LLMs as auxiliary tools for genre instruction and offers empirical insights for the intelligent transformation of international Chinese language education.

### **Keywords**

ChatGPT, genre teaching, artificial intelligence, large language model

*Xue Kang*, postgraduate student at East China Normal University, School of International Chinese Studies. Research interests: International Chinese Language Education, Stylistic Teaching.

*Xiyue Wang*, postgraduate student at East China Normal University, School of International Chinese Studies. Research interests: Chinese Information Processing, International Chinese Language Teaching, Sociolinguistics.

*Lanxin Yu*, postgraduate student at East China Normal University, School of Statistics. Research interests: Online Conformal Prediction.

***Feng Qi***, Professor, East China Normal University, School of International Chinese Studies. Research interests: Linguistics, International Chinese Language Teaching, Corpus Linguistics.