

英语写作修改中数据驱动学习的应用

杨黥阳

北京科技大学外国语学院 北京 100083

[摘要] 本文针对基于数据驱动学习的书面纠错反馈法和问卷、回溯访谈等回顾方法进行阐述和研究。研究结果证明, 学生使用基于语料库的纠错方法后, 在搭配和短语错误、选词错误和英语词汇学习方面有显著提升, 从而提升英语写作水平。本研究助推二语学习者和语料库之间的互动相关研究, 同时为教师在布置和设计英语写作等教育教学任务中使用数据驱动学习的潜在作用方面提供有价值的参考和帮助。

[关键词] 数据驱动学习; 语料库; 纠错反馈; 英语写作

[中图分类号] H319 **[文献标识码]** A **[文章编号]** 1893-9431 (2026)-0024-98 **[收稿日期]** 2025-08-11

一、引言

半个多世纪以来, 语料库作为文本或语音记录的集合, 一直被学术研究者用于计算语言学、词典学、写作和其他语言领域^[1]。从教学的角度来看, 有效利用语料库、从语料库的输出开发形成教学材料, 以间接方式向学习者提供真实的范例, 或者通过学习者实际操作查询语料库直接获得语言的正确表达^[1-3], 这两种方法的实现都可以归类为数据驱动学习 (DDL)。数据驱动学习得到了教育工作者和研究人员的极大关注, 并为语言学习者提供了大量的语言输入材料, 即通过教学材料或语料库查询, 使他们能够无需咨询更精通的语言使用者而获得学习语言模式和解决语言问题。通过这种方式, 学生可以借助老师的指导从语料库中接收语言输入, 通过实践查询归纳总结获得语言知识^[4-5], 并独立发现和应用语言的正确表达形式。尽管业界对数据驱动学习应用的研究兴趣日益增进, 但

现有的许多研究仍然仅以结果为导向, 在二语写作修改过程中查阅语料库和决策的过程中, 鲜少有人关注学习者与语料库的互动。本研究旨在克服上述不足, 结合回顾性数据作为支撑来评估数据驱动学习在写作修改中的有效性。

二、数据驱动学习和书面纠错反馈

(一) 基于数据驱动学习的纠错反馈修改效能

自2005年以来, 数据驱动学习研究经历了一个快速增长期, 语料库在语言课堂中的教学功能包括提供真实的语言范例, 促进写作和口语过程, 以及给学习者带来越来越多的学习乐趣和参与感。通过语料库检索, 语料库会给出所检索的关键词 (即KWIC, Key word in context) 在真实英语语境下的使用情况, 使学习者沉浸在真实的语言环境中^[1, 6]。

作者在对数据驱动学习的深度应用和教学研究中发现, 基于语料库的写作修改在提

高词汇习得和书面表达方面十分有效。首先，基于语料库的纠错方法通过为学习者提供海量且真实的地道语言样本，能够有效支撑其语言输出实践，不仅增强了学习材料的可信度，也为语言习得提供了可靠的参照标准。其次，语料库的输出功能对二语学习者具有多重促进作用。该方法在一方面有助于学习者深化对目标词汇的认知理解，另一方面能够有效修正学习者的典型语言错误。更重要的是，通过接触大量真实语境中的词汇使用范例，学习者能够更准确地掌握词汇在句子结构中的运用规律，从而全面提升书面表达的准确性和流畅性^[7]。此外，双语语料库既能显著降低学习者的认知负荷，又通过清晰的跨语言对比呈现，在帮助学习者掌握词语

搭配规律方面的表现尤为突出，这种双重视角的学习支持为词汇的深度习得提供了有效路径^[8]。

（二）基于语料库的纠错反馈分类

在二语写作中使用数据驱动学习的一种常见方式是将语料库的使用与书面纠错反馈(Written corrective feedback, WCF)相结合，语言学习者在使用语料库时可以从中受益。基于语料库的纠错方法作为书面反馈的一部分，对学习者书面文本中语言问题提供反馈，帮助语言学习者提高写作准确性并通过修改过程促进语言习得的方法。基于数据驱动学习，纠错反馈类型包括直接反馈和间接反馈，见表1。

表1.基于语料库的纠错反馈类型和范围

纠错反馈的分类	描述
直接电子反馈	教师指出错误，并提供指示如何修改错误的语料库超链接
间接反馈	
1. 指示和定位	教师在文中强调或突出错误。
2. 元语言反馈	教师在文字页边空白处标注(例如，WC =单词选择；art = article)或者用语法描述标记错误。

书面反馈的不同类型在实施过程中对师生双方的时间投入和认知负荷存在显著差异。基于语料库的间接反馈与传统书面间接反馈在形式上具有相似性，由教师识别并定位错误位置，但不直接提供正确的表达，而是要

求学习者根据反馈内容自主检索语料库以获取恰当的表达方式并完成修正。相比之下，基于语料库的直接反馈与传统直接书面反馈存在本质区别，传统直接反馈由教师直接提供修正后的准确表达^[9]，而基于语料库

的直接反馈要求学习者从教师预先筛选的语料库数据中自主提取合适的语言形式,这种反馈机制类似于电子反馈(electronic feedback),其实现包括定位错误和提供指向相关语料库检索结果的超链接^[9-10]。在此过程中,教师需要预先发起语料库查阅,并为学习者生成包含针对性语料示例的超链接,从而引导其完成错误修正,这种反馈方式既保留了直接反馈的指导性,又通过语料库的实证数据增强了学习者的自主探究能力。

另一方面,基于语料库的间接书面纠正反馈则不同于直接反馈,其核心特征在于要求学习者在整合和应用语料库数据前,必须主动进行语料库检索。通过应用研究发现,仅标示错误位置而不提供正确答案的间接反馈方式,在降低书面语言错误率方面具有显著效果。无论采用何种具体形式,基于语料库的间接反馈在解决语言共现问题、词汇语法错误、成分省略或冗余添加等典型语言问题方面均展现出独特的教学价值。这种反馈机制通过促使学习者主动参与语言探究过程,强化了学生与语料库的互动,不仅提升了错误识别能力,更培养了自主语言学习策略。

(三) 错误标注数量

在基于语料库的写作修改研究中,错误标注的数量和范围需根据研究目标和教学情境进行科学设计。在给予间接纠正反馈时通常采取三种主要标注策略:全面标注法、选择性标注法和聚焦标注法。全面标注主张对所有可识别的错误类型进行系统标注。该标注法能为学生和教师提供完整的错误分布

图谱,但可能增加教师的时间投入和学习者的认知负荷。相比之下,根据研究目的或教学目的,选择性标注法则更具针对性。一些倾向使用选择性标注的学者偏向于广泛关注不同错误类型,其种类可高达10个以上。值得注意的是,无论错误类型的多少,每篇文章的错误标注数量应控制在10-15处,在保证研究信度的同时确保学生在基于语料库进行错误修改时的操作可行性。而更为精细化的聚焦标注通常将标注数量控制在2-5个关键错误,但不区分错误类型。该方法具有高度选择性和聚焦性,适合于探究更精细化的写作修改过程。

在修改要求方面,每篇文章的实际修正量应维持在3-10个错误之间,这一区间既保证了研究数据的有效性,又兼顾了学习者的修改能力。当前研究在错误类型覆盖方面已形成较完善的体系,特别是在对比分析语料库检索对语法/词汇错误或局部/整体错误的修正效果方面积累了丰富成果。就错误类型研究的深度而言,研究应在错误类型上进行深度挖掘,针对某一特定错误类型开展系统深入的纵向追踪。这种研究取向能够充分揭示不同类型错误的修正机制差异。在研究维度上不仅聚焦于错误修正的结果评估(如错误率变化),且对修正过程的动态进行深入挖掘。

三、语料库在写作修改中的应用

(一) 语料库的选择

作为促进二语写作的参考工具,不同类型的语料库因其设计特点和语料来源的差异

而呈现出不同的教学功能。具有代表性的四个常见单语语料库是：英式英语学术写作语料库(BAWE corpus, British Academic Written English corpus)、SKELL (Sketch Engine for Language Learning)、英国国家语料库(BNC, British National Corpus)、当代美国英语语料库(COCA, Corpus of Contemporary American English),在教学研究应用场景上它们各具特色。BAWE作为专门针对学术英语研究的语料资源,其核心在于系统呈现学术文本的语言特征。BAWE不仅为语言研究者提供了分析学术话语模式的书面范例,也能帮助二语写作者掌握特定学科领域的语言使用惯例,适合有志于提升学术写作水平的学习者参考使用。而SKELL为普通语言学习者量身定制,通过简化的检索界面降低了语料库使用的技术门槛,使其成为普通二语学习者日常写作的理想辅助工具。SKELL将庞大的语料数据转化为易于理解的语言学习资源,在保持语料真实性的同时,兼顾了语言教学的可操作性。BNC和COCA作为综合性语料库的代表,则跨年龄层的语言样本,涵盖了从正式书面语到日常口语的多种语体类型。通过系统收录各类媒体文本,全面记录英语语言变体的实际使用情况。在二语写作教学中,BNC和COCA的对比使用可以帮助学习者深入理解英式英语与美式英语在词汇选择、语法结构和表达习惯等方面的细微差异。

在选择语料库资源时,应当综合考虑学习者的语言水平、学习目标以及具体的写作

任务要求。对于学术写作训练,BAWE的专业性优势明显;而针对基础语言能力的培养,SKELL则更具实用价值;至于需要了解语言变体或更广泛的语言特征的学习者,BNC和COCA能提供更为丰富的语言输入。这种基于语料库特性的差异化应用策略,对于提升二语写作水平和教学效果均具有重要意义。

(二) 学习对象的选择和语料库使用培训

学习者的语言熟练程度是影响语料库在二语写作中应用效果的关键调节变量。现有研究在受试者选择上呈现出明显的集中趋势:多数实证研究以欧洲共同语言参考框架(CEFR)B1至A1级的中高级水平学习者为研究对象。超过一半的DDL写作研究聚焦于大学本科或研究生阶段的中级到中高级学习者。然而,值得注意的是,本研究发现数据驱动学习教学在低水平学习者(CEFR B1以下)中反而表现出更显著的效果提升。数据驱动学习的教学优势在语言熟练度较低的各类学习者群体中具有普遍适用性,这种效应不受年龄因素的调节。其潜在机制可能包括:(1)初级学习者对显性语言输入的依赖度更高;(2)语料库提供的语境化范例能有效补偿其有限的内隐语言知识;(3)数据驱动学习的发现式学习更符合初级学习者的语言认知特点。在进行后续基于语料库的教学和研究,应避免可能存在的“中高级偏向”对象选择,以期在更广泛学习者群体中发现的数据驱动学习的潜在价值^[11]。

对于语料库使用的培训时长而言,其设

置存在显著差异,其持续时间从不足两小时的短期集中培训到超过十次、每次半小时的分布式培训不等。值得注意的是,尽管培训时长存在较大变异,但采用前测-后测设计的各项研究均一致报告了积极的学习效果,这一发现具有重要的教学启示意义。这种时长设置的灵活性表明基于语料库的教学具有较强的适应性,不同培训强度均可产生积极效果。且培训效果可能更多取决于培训质量而非单纯的时间投入,也许在达到基本培训门槛后即可产生显著效果。

(三) 学习者对基于语料库的写作修改之看法

采用三角验证法,通过问卷、访谈、学习日志等回顾性数据来补充定量研究后发现,总的来说,除了定量修改结果显示的积极效应,学习者的认知评价总体也较为积极。大多数学习者认可语料库作为多功能参考工具的价值,跨语料库集成平台显著提升了写作表达的多样性和词汇习得效率。然而,一些学习者特别是学术写作群体表现出抵触情绪:学习者认为书面纠错修改纯属教师职责,且在使用语料库修改的负面体验情况下否定语料库查询的必要性。虽然总体而言大多学习对基于语料库的写作修改持积极态度,但在技术层面,学习者面临三大挑战,即,在使用语料库接口平台时的操作困难,从检索结果归纳语言规律时信心不足,以及完整修正流程耗时费力等。

研究虽证实了数据驱动学习在提升二语

写作质量方面的有效性,但在理解学习者运用语料库进行特定错误类型修正的机制方面仍存在研究空白。在认知过程、人机互动、个体差异等方面可作进一步研究和探讨。

四、结语

综上所述,通过系统深入研究数据驱动学习在二语写作教学中的应用,本文发现基于语料库的纠错方法对提升学习者的语言和写作能力具有显著作用,在词汇搭配和短语结构的错误修正方面效果尤为突出,并且能显著改善选词准确性,通过语境化示例促进词汇深度习得,提升学习者学习自主性。这种提升不仅体现在语言形式的准确性方面,更促进了学习者对语言使用適切性的把握,同时培养了其自主发现和修正语言问题的能力,从而形成了良性循环的学习机制。在探究学习者与语料库的互动关系时,研究发现绝大多数学习对语料库持肯定态度,认为其提供的真实语境参照和丰富语言资源对写作提升和语言学习具有重要帮助。在充分发挥学习者-语料库动态交互方面,教师需要着力解决三个关键问题:通过精心设计的写作任务调动学习者的积极性,通过优化界面设计降低技术使用门槛,以及通过系统化的培训提升学生信息检索能力。这些措施对于建立高效的学习者与语料库互动机制至关重要。

作者简介:杨黥阳(1996-),女,汉族,重庆人,博士研究生,研究方向二语习得、心理语言学。

参考文献:

- [1]Jablonkai, R., & Csomay, E. The Routledge handbook of corpora of English language teaching and learning[M]. New York: Routledge,2022
- [2]陈秀娟,杜桂敏.基于语料库的英语写作教学模式建构及有效性研究[J].大学教育,2019,(11):105-107.
- [3]滕延江,刘世铸.基于语料库的英语写作教学新模式[J].外语电化教学,2006,(06):49-54.
- [4]Bitchener, J. A reflection on ‘the language learning potential’ of written CF[J]. Journal of Second Language Writing,2012,21(4):348–63.
- [5]Flowerdew, L. Using corpus-based research and online academic corpora to inform writing of the discussion section of a thesis[J]. Journal of English for Academic Purposes,2015,(20):58–68.
- [6]O’Keeffe, A., & McCarthy, M. The Routledge handbook of corpus linguistics[M]. London: Routledge,2010
- [7]Hyland, K. Teaching and researching writing (2nd ed.)[M]. London: Routledge,2013.
- [8]Chan TP, Liou HC. Effects of web-based concordancing instruction on EFL students’ learning of verb-noun collocations[J]. Computer Assisted Language Learning,2005,18(3):231–51
- [9]Ellis, R. A typology of written corrective feedback types[J]. ELT Journal. 2009;63(2):97–107
- [10]Boulton, A., & Landure, C. Using corpora in language teaching, learning and use[J]. Recherche et pratiques pédagogiques en langues de spécialité Cahiers de l’Aplut. 2016;35(2).
- [11]Boulton, A., & Vyatkina, N. Thirty years of data-driven learning: taking stock and charting new directions over time[J]. Lang Learn Tech,2021,25(3):66–89.

The application of data-driven learning in English writing revision

Yang Yiyang

School of Foreign Studies, University of Science and Technology Beijing, Beijing, China

100083

Abstract: This paper investigates the application of data-driven learning (DDL) in English as a Foreign Language (EFL) writing. By examining the numerical results alongside retrospective research methods such as questionnaires and stimulated-recall interviews, the study reveals the effectiveness of the corpus-based written corrective feedback. The findings demonstrate that the corpus-based error correction significantly improves learners’ use of vocabulary, collocations and phrases in EFL writing, promotes vocabulary acquisition, and thus enhances FL writing skills. The investigation contributes to research on learner–corpus interaction and offers valuable insights and practical guidance for educators on designing and implementing DDL tasks for EFL writing instruction.

Keywords: Data-driven learning (DDL); corpus; Written corrective feedback (WCF); EFL writing