

《二语写作》第九辑

(2024年12月)

目录

卷首语

写作研究

跨学科视角下国际学术论文摘要的词汇和句法复杂度研究

.....董记华 张梦梦 邸湘红 (1)

中国高中英语学习者读后续写句法复杂度研究.....刘应亮 刘 慧 王建容 (17)

A Corpus-Based Study on Syntactic Complexity of English Abstracts in Chinese and English

Journal Articles.....Yuhan HU Qiuyan YAN (34)

研究述评

二语写作句法复杂性研究现状、趋势与前沿——基于知识图谱的计量分析对比研究

.....唐建华 朱慧敏 (52)

写作教学

对话式课堂展示对大学生写作反馈素养的影响研究.....张伟平 (63)

自由职业英语写作教师信念与实践的自我民族志研究.....王涌骅 (74)

中国大学生英语雅思写作中批判性思维的认知和习得研究.....张倩云 丁 鹏 (84)

二语学术写作中元认知体验和学习者能动性的关系探究

.....熊 慧 赵小亮 陈 萍 (98)

新作评介

诊断性评估助推二语写作能力发展——评《二语或外语写作诊断研究》

.....张海燕 苏惠琳 (110)

有声思维法的边界与潜能——兼评《有声思维法在二语写作中的反应性和真实性之

混合研究》.....刘晨妃 王俊菊 (115)

前沿动态..... (121)

英文摘要..... (128)

版权声明:

本刊已被《中国学术期刊网络出版总库》及CNKI系列数据库全文收录,为中国人文社会科学学术集刊AMI入库期刊,如作者不同意被收录,请在来稿时向本刊声明,本刊将作适当处理。文章版权所有,欢迎转载、评论、翻译、引用,但须注明。

Contents

A Cross-Disciplinary Investigation of Lexical and Syntactic Complexity in Abstracts of International Research Articles Jihua DONG, Mengmeng ZHANG, Xianghong DI	(1)
A Study on Syntactic Complexity in Continuation Writing by Chinese Senior High School EFL Learners Yingliang LIU, Hui LIU, Jianrong WANG	(17)
A Corpus-Based Study on Syntactic Complexity of English Abstracts in Chinese and English Journal Articles..... Yuhan HU, Qiuyan YAN	(34)
Current Status, Trends, and Frontiers in Syntactic Complexity Research of L2 Writing: A Comparative Bibliometric Analysis Based on Knowledge Mapping Jianhua TANG, Huimin ZHU	(52)
A Study on the Impact of Dialogic Presentation on College Students' Writing Feedback Literacy..... Weiping ZHANG	(63)
Freelance English Writing Instructor's Beliefs and Practices: An Autoethnography Yonghua WANG	(74)
A Study of the Cognition and Acquisition of Critical Thinking in Chinese University Students' IELTS Writing..... Qianyun ZHANG, Peng DING	(84)
A Study of the Relationship Between Learners' Meta-Cognitive Experiences and Agency in EAP Writing..... Hui XIONG, Xiaoliang ZHAO, Ping CHEN	(98)
Diagnostic Assessment as a Catalyst for the Development of Second Language Writing Ability: A Review of <i>The Diagonosis of Writing in a Second or Foreign Language</i> Haiyan ZHANG, Huilin SU	(110)
Boundaries and Potential of Think-Aloud Protocols: A Review of <i>Think-Aloud Protocols in Second Language Writing: A Mixed-Methods Study of Their Reactivity and Veridicality</i> Chenfei LIU, Junju WANG	(115)
Research Frontiers	(121)
English Abstracts	(128)

卷首语

创刊以来,《二语写作》集刊在学界同仁的鼎力支持下持续深耕、稳步前行,学术影响力不断扩大。集刊始终关注二语写作领域内的核心议题,以期引导学界的理论探索与实践创新。第九辑紧密追踪二语写作中的文本复杂度、反馈素养、批判性思维、学习者能动性等重点,力求通过跨学科视角、技术赋能路径与本土化案例,为读者呈现多维度的学术启迪。

“写作研究”栏目中,董记华、张梦梦、邱湘红基于跨学科语料库的词汇与句法复杂度对比研究,揭示了语言学与计算机科学摘要的学科特异性,为学术英语写作教学提供启发。刘应亮、刘慧、王建容针对高中英语读后续写任务开展句法复杂度分析,从发展性视角为中学生英语写作能力评价体系注入新思路。胡雨寒、闫秋燕对近二十年来中英文期刊文章摘要的句法复杂度进行了研究,发现中英文期刊文章摘要句法复杂度差异显著,摘要历时变化显著。

“研究述评”栏目中,唐建华、朱慧敏对比分析了近十年国际与国内期刊刊发的二语写作句法复杂性研究的演进脉络,指出国际与国内期刊中的相关研究整体呈前升后降趋势;二者均重点关注复杂性、准确性和流利性;但国际期刊论文中的语篇特征、认知因素、计算语言学等研究内容尚未引起国内期刊论文重视。

“写作教学”栏目中,张伟平基于自我调节学习理论,探讨对话式课堂展示对大学生写作反馈素养的干预效果,为动态反馈模式设计提供启示。王涌骅通过自我民族志,深度剖析自由职业写作教师的信念演变与角色重构,为教师专业发展研究提供新思路。张倩云、丁鹏通过雅思写作文本分析与半结构化访谈,提出“诊断性思维”概念,为批判性思维培养与写作技能提升搭建桥梁。熊慧、赵小亮、陈萍结合能动性理论框架,实证考察元认知体验与学习者能动性对学术写作成绩的交互影响,揭示了认知策略优化的重要机制。

“新作评介”栏目中,张海燕、苏惠琳对《二语或外语写作诊断研究》的评介,系统梳理了诊断性评估的理论框架与实践路径,为写作能力动态监测与教学干预提供方法论指导。刘晨妃、王俊菊从有声思维法的边界与潜能等视角切入,对《有声思维法在二语写作中的反应性和真

实性之混合研究》进行了评介，该书探讨了有声思维法在二语写作中的反应性和真实性，对深化写作过程研究、完善数据采集范式具有重要参考价值。

作为本刊特色，“前沿动态”板块整理了2024年国内二语写作领域的重要期刊文献，帮助学者和广大师生第一时间了解学术前沿。

《二语写作》第九辑通过对上述议题的深耕，既回应了学界对热点问题的持续性追问，又以方法论创新与理论框架拓展为突破口，推动二语写作研究从“现象描述”向“机制阐释”的纵深化转型。本辑成果不仅为教学实践提供了可借鉴的解决方案，更为二语写作研究贡献了中国智慧与声音。今后，《二语写作》将继续扎根中国语境，拓展国际对话，以严谨学风与开放姿态，见证并推动二语写作教学与研究的高质量发展。

跨学科视角下国际学术论文摘要的词汇和句法复杂度研究

董记华¹ 张梦梦¹ 邱湘红²

¹ 山东大学
² 潍坊新华中学

摘要: 本文旨在从跨学科视角探究语言学和计算机科学的国际学术论文摘要中词汇和句法的使用情况。研究利用自然语言处理软件 (TAALED、TAALES、TAASSC) 对自建跨学科语料库进行词汇和句法复杂度自动标注测量。曼-惠特尼 U 检验的结果显示: 1) 在词汇复杂度方面, 计算机科学摘要更倾向使用具有学科特色、难度高的词汇, 而语言学摘要使用的词汇更多样, 且两个学科的摘要在词汇密度方面没有显著差异; 2) 在句法复杂度方面, 语言学学术论文摘要使用了更长的语言单位, 包括更多的从属结构、并列子句和短语, 以及复杂名词和动词短语结构。本研究不仅为学术英语写作教学提供了实证参考, 而且有助于促进语言学与计算机科学间的跨学科交流与合作。

关键词: 跨学科研究; 词汇复杂度; 句法复杂度; 学术论文摘要

1. 引言

学术论文摘要是对论文内容的高度凝练 (Li et al. 2023), 兼具信息传递与学术成果推广的重要功能 (Ansarifar et al. 2018; Zhao et al. 2023)。目前围绕学术论文摘要的研究尤其关注多维度的语言结构, 如词汇 (Tankó 2017)、句法 (Li et al. 2023)、修辞结构 (Jiang & Hyland 2017) 等。这些研究不仅为学术论文摘要提供了语言使用的参考, 促使其能够有效发挥激发读者兴趣、引导读者深入探索全文内容的体裁功能 (Zhao et al. 2023), 也有助于为学术英语写作教学提供文本特征层面的理论支撑与实践指导 (Ansarifar et al. 2018; Li et al. 2023)。

在诸多文本特征中, 词汇复杂度与句法复杂度作为书面语质量的两大关键语言指标 (郑咏滢、冯予力 2017), 尤其受到了学术写作领域的关注 (Liu & Li 2024; 鞠玉梅 2020)。研究或选取学生学术写作 (如 Staples et al. 2016), 或聚焦于期刊论文 (如吴雪 2017) 及其部分体裁 (part-genre), 如摘要 (Tankó 2017)、引言 (Zhou, W. et al. 2023)、讨论 (Ziaecian & Golparvar 2022) 等。这些研究为学者如何运用语言资源、提升学术写作质量提供了有价值的参考。

学术语篇的语言特征反映了特定学科知识构建的方式 (Ken Hyland 等 2023)。学科通常指具有特定理论、方法和对象的研究领域 (刘辉 2015)。不同学科在话语协商和知识构建策略等方面存在差异, 如软学科较硬学科更大程度上依赖作者个人

鲜明的话语说服力及说服力所承载的权威性 (Hyland 2005; 姜峰 2023), 这尤其体现在学术论文部分体裁的词汇使用与句法结构中 (Jin et al. 2021; 赵永青等 2019)。然而, 对于学术论文摘要这一部分体裁, 尽管前人研究已探讨了特定学科的词汇或句法复杂度特征, 如科学文本 (Jin et al. 2021)、海洋科学文本 (Liu & Li 2024) 等, 但对摘要中词汇和句法复杂度的学科差异及其反映的学科特性关注不足。对学术论文摘要的词汇与句法结构开展跨学科对比, 有助于揭示不同学科背景下的学术作者如何在摘要这一“自足的独立系统” (Li et al. 2023: 322) 中调用语言资源有效传递关键信息, 从而反映其特定语篇实践中的学科范式。

鉴于此, 本研究基于自建语言学与计算机科学国际学术期论文摘要语料库, 对比分析语言学与计算机科学论文摘要的词汇复杂度与句法复杂度差异。语言学与计算机科学均致力于信息的生成与解读 (Dong et al. 2024), 但研究对象和写作风格不同。语言学侧重对语言现象的研究, 探讨语言结构、规律及其与思维、文化和社会的关系, 涵盖丰富的人文和社会要素 (Saussure 2011; Yule 2022)。语言学家对语言非常敏感 (Li 2021), 能够更细腻地传达作者身份 (Dong et al. 2024)。而计算机科学基于数学与逻辑基础, 强调算法优化与技术创新, 关注计算机系统的描述与验证 (Knuth 1974)。该领域的作者倾向于客观的写作风格, 以提升其内容的客观性, 弱化自身声音, 直接呈现信息 (Hyland 2001)。两者的交叉融合推动了计算语言学的发展, 促进了自然语言处理、机器学习及大语言模型的进步 (Mitkov 2022)。对比分析这两个学科的学术论文摘要, 可以直观展现学科间的不同思维方式和学术传统, 加深对各自学科特性以及相互关系的认识, 为学科的自我完善和发展提供新视角; 同时, 有助于打破学科壁垒, 促进语言学与计算机科学的交流与合作, 推动新的研究方法, 为解决复杂学术问题提供新思路。

2. 文献综述

2.1 词汇复杂度及其学术语篇中的跨学科对比

词汇复杂度指文本中词汇的重复程度和复杂程度 (Lu 2012), 涵盖词汇难度 (lexical sophistication)、词汇密度 (lexical density) 和词汇多样性 (lexical diversity/variation) 三个维度 (Bui 2021; Tabari et al. 2023)。其中, 词汇难度指文本中高级词汇或低频词汇的使用情况 (Read 2000)。词汇密度指文本中实词的占比 (Ure 1971), 词汇密度越高表明文本的信息密度越大 (Bui 2021)。词汇多样性则测量文本词汇的变化或重复程度 (Carroll 1938; Tabari et al. 2023)。在传统词汇多样性指标类符形符比 (type-token ratio, 简称 TTR) (Templin 1957) 的基础之上, 学者提出了更为丰富的词汇多样性指标如 MLTD、voc-d 和 HD-D 等, 以消除文本长度对结果的影响 (McCarthy & Jarvis 2010)。在学术写作中, 适当复杂的词汇资源能够帮助作者更为严谨、深入地阐释学术观点, 从而增强学术写作论证的说服力 (Zhou, X. et al. 2023)。

近年来, 学术写作中词汇复杂度的跨学科对比受到了广泛关注 (如 Huynh Le &

Tan Ha 2023; Khany & Kafshgar 2016; Müngen 2023; 刘辉 2015)。例如, 刘辉 (2015) 对比了化学、生物学和语言学学术论文中研究方法部分的词汇复杂度, 发现语言学使用了更多的副词和人称代词, 但这三类学科在词汇多样性方面不存在显著差异。Khany和Kafshgar (2016) 研究了人文科学、生命科学和物理学讨论部分的词汇多样性和词汇密度, 发现人文科学比生命科学和物理学中的词汇更多样, 而这两门自然科学在词汇密度方面则未见显著差异。此外, Huynh Le和Tan Ha (2023) 探究了英国学术书面语语料库 (British Academic Written English, 简称BAWE) 语料库中不同学科学术论文的词汇在英国国家语料库 (British National Corpus, 简称BNC)、美国当代英语语料库 (Corpus of Contemporary American English, 简称COCA)、学术词汇表 (Academic Word List, 简称AWL) 中的分布特征。其研究结果显示, 各学科领域的词汇在分布上呈现出显著差异, 其中社会科学学术论文更倾向使用常用词汇。上述研究为揭示不同学科间词汇复杂度的差异提供了实证依据, 但多聚焦于词汇多样性和词汇密度, 对词汇难度测量指标的关注不足, 且尚未对摘要这一特定部分体裁进行较为详细的探讨。

2.2 句法复杂度及其学术语篇中的跨学科对比

句法复杂度指语言产出中句法结构的多样性、复杂性及拓展性 (Kim & Lu 2024; Lu 2017; Norris & Ortega 2009)。句法复杂度也被视为一个多维度概念 (Kyle 2016; Lu 2017), 且每个维度都可以通过多种指标进行较为全面的测量 (陆小飞、许琪 2016)。作者可以基于不同文本的句法复杂度特征, 在具体的写作语境中 (如学术英语写作) 选择更符合其规范的句法资源, 从而在句法层面上提升学术写作质量, 增强学术交流能力 (Casal & Lu 2021)。

前期研究围绕学术写作中句法复杂度的学科差异开展了系列探讨 (Dong et al. 2023; Shen et al. 2023; Staples et al. 2016)。例如, Dong et al. (2023) 基于BAWE语料库探究了四个学科群组及其下具体学科的句法特征, 分析发现两个软学科群组在从句和短语层面的复杂度高于另外两个硬学科群组。不同学科也呈现出具体的学科差异, 比如计算机科学使用更多的动词短语结构。此外, 近些年研究者尤其关注了学术语篇中特定部分体裁的句法复杂度差异 (Lu et al. 2021; Ziaecian & Golparvar 2022)。例如, Ziaecian和Golparvar (2022) 基于细颗粒度的从句和短语复杂度指标对比了应用语言学、经济学和化学学术论文的讨论部分。研究结果显示, 软学科——即应用语言学和经济学——在从句层面展现出更高的复杂度, 而化学则使用了更复杂的短语结构。再如, Zhou, W. et al. (2023) 对比了基础硬学科 (化学、数学、物理学) 和应用硬学科 (化学工程、计算机科学、机械工程) 学术论文引言部分句法结构的使用情况, 发现二者未呈现显著差异。

上述文献虽然揭示了不同学科间句法复杂度的差异特征, 但在研究对象和测量指标上仍有待拓展。首先, 目前已经开展的研究多关注学术论文的整体文本或特定部分体裁, 如研究方法、讨论等, 而摘要的词汇和句法复杂度特征尚未得到较为全

面和深入的探讨。其次，现有研究虽然对比分析了不同学科的词汇复杂度，但主要集中在词汇多样性和词汇密度等方面，对于词汇难度的测量和探讨不够充分。因此，本研究旨在探究语言学和计算机科学学术论文摘要中词汇复杂度和句法复杂度的差异及其所体现的学科范式。研究发现有助于完善学术论文摘要在不同学科间的语言特征研究体系，揭示不同学科摘要在词汇和句法使用上的差异；帮助学生在学术写作中更加精准地选择和运用语言资源，从而提高学术论文摘要的质量；同时，也可作为学术英语写作（EAP）教学中的文本特征教学提供实证依据，为培养学生的跨学科写作能力提供实践指导。具体而言，本研究拟回答以下两个研究问题：

- 1) 语言学和计算机科学国际学术论文摘要在词汇复杂度上有何差异？
- 2) 语言学和计算机科学国际学术论文摘要在句法复杂度上有何差异？

3. 研究方法

3.1 语料库

本研究构建了涵盖语言学与计算机科学国际期刊学术论文摘要的语料库。该语料库源自 Web of Science 数据库，具体收集过程如下。首先，为确保所选期刊的权威性，本研究征询了语言学与计算机科学领域专家的意见，并选取 Journal Citation Reports 2022 中排名靠前的期刊，两个学科分别选取 10 本期刊，具体期刊名称详见附录。其次，考虑到数据库中摘要的可获取性和时效性，本研究设定语料检索的时间范围为 1991 年至 2024 年，每本期刊每年随机选取五篇摘要。最终，语言学语料库包含 1700 篇摘要，299,591 字符；计算机科学语料库包含 1700 篇摘要，271,814 字符。

3.2 词汇和句法复杂度指标

本研究使用词汇难度自动分析工具（Tool for the Automatic Analysis of Lexical Sophistication，简称 TAALES，Kyle et al. 2018）、词汇多样性自动分析工具（Tool for the Automatic Analysis of Lexical Diversity，简称 TAALED，Kyle et al. 2021）、句法复杂度自动分析工具（Tool for the Automatic Analysis of Syntactic Sophistication and Complexity，简称 TAASSC，Kyle 2016）自动计算词汇复杂度和句法复杂度，这些软件在词汇与句法复杂度研究中已被广泛接受与应用（如 Tabari et al. 2023；Zhang & Lu 2024）。

为全面测量词汇复杂度的学科差异，本研究参照 Zhou, X. et al. (2023) 选取了词汇难度、词汇密度、词汇多样性三个维度。表 1 呈现了所使用词汇复杂度指标的信息。在词汇难度方面，LC1, LC2, LC5, LC6 分别以布朗语料库（Brown Corpus）和美国当代英语语料库（COCA）的学术子语料库为参照语料库，考察了词汇在参照语料库中出现的频率和范围。词汇出现的频率越低，范围越小，则表明词汇难度越高（Tabari et al. 2023）。同时，LC3 基于 MRC 心理语言学数据库测量文本中词汇的平均熟悉度得分（Coltheart 1981），得分越低，说明目标文本中的词汇对于英语母语者更陌生，即词汇难度更高。LC4 则基于预先构建的心理词典数据库，

根据母语者习得词汇的年龄来计算词汇复杂度 (Kuperman et al. 2012), 数值越高, 表明词汇习得时间越晚, 词汇越难。在词汇密度方面, 选用LC7测量实词类符数除以总单词类符数。在词汇多样性方面, 选用LC8和LC9两种类符符比变体指标, 以消除文本长度的影响 (Kyle et al. 2021)。

表1 词汇复杂度指标 (Zhou, X. et al. 2023)

维度	缩写	指标	测量工具
词汇难度	LC1	KF_Freq_AW_Log	词汇难度自动分析工具 (Kyle et al. 2018)
	LC2	KF_Nsamp_AW	
	LC3	MRC_Familiarity_AW	
	LC4	Kuperman_AoA_AW	
	LC5	COCA_Academic_Range_Log_AW	
	LC6	COCA_Academic_Frequency_Log_AW	
词汇密度	LC7	lexical_density_types	
词汇多样性	LC8	mtld_original_aw	词汇多样性自动分析工具 (Kyle et al. 2021)
	LC9	mtld_ma_bi_aw	

在句法复杂度测量中, Lu (2010) 根据 Wolfe-Quintero et al. (1998) 和 Ortega (2003) 提出二语写作句法复杂度测量指标并总结了包含单位长度、从属子句、并列结构、特定短语结构以及句子复杂度五个维度共 14 项句法复杂度指标。表 2 展示了指标的具体信息。

表2 句法复杂度指标 (陆小飞、许琪2016)

维度	缩写	指标	测量工具
单位长度	MLS	平均句子长度	句法复杂度自动分析工具 (Kyle 2016)
	MLT	平均 T 单位长度	
	MLC	平均子句长度	
从属子句使用量	C/T	每个 T 单位中的子句数量	
	DC/C	从属子句比率	
	DC/T	每个 T 单位中从属子句数量	
	CT/T	复杂 T 单位比率	
并列结构使用量	T/S	并列句比率	
	CP/T	每个 T 单位中的并列短语数量	
	CP/C	每个子句中的并列短语数量	
特定短语结构	CN/T	每个 T 单位中的复杂名词短语数量	
	CN/C	每个子句中的复杂名词短语数量	
	VP/T	每个 T 单位中的动词短语数量	
句子复杂度	C/S	每个句子中的子句数量	

3.3 数据分析

本研究的数据分析步骤如下：首先，采用上述三个分析工具分别计算语言学和计算机科学学术论文摘要的词汇和句法复杂度指标数值。之后，对词汇复杂度和句法复杂度结果进行描述性分析。由于所分析数据不符合正态分布且未通过方差齐性检验，故使用曼-惠特尼U检验（Mann-Whitney U test）来探究语言学和计算机科学在词汇复杂度和句法复杂度上是否存在显著差异。

4. 结果与讨论

4.1 词汇复杂度的学科差异

表3展示了语言学和计算机科学论文摘要中词汇复杂度的平均值、标准差以及曼-惠特尼U检验的结果。研究表明，计算机科学摘要中的词汇难度显著更高，语言学摘要中的词汇更加多样，而两个学科在词汇密度上不存在显著差异。

表3 语言学和计算机科学的词汇复杂度差异

指标	语言学	计算机科学	Z	p	Cohen's <i>d</i>
	M (SD)	M (SD)			
LC1 ¹	2.82 (0.13)	2.78 (0.15)	-8.20	**0.001	0.17
LC2	254.76 (20.84)	251.03 (23.43)	-4.80	**0.001	0.10
LC3	585.04 (5.23)	584.52 (6.44)	-1.78	*0.032	0.04
LC4	6.52 (0.34)	6.78 (0.39)	-19.52	**0.001	0.40
LC5	-0.55 (0.09)	-0.60 (0.11)	-13.26	**0.001	0.26
LC6	2.78 (0.14)	2.70 (0.16)	-15.34	**0.001	0.31
LC7	0.71 (0.05)	0.71 (0.05)	-1.70	0.069	0.03
LC8	67.68 (20.01)	64.41 (20.41)	-5.04	**0.001	0.10
LC9	59.60 (20.12)	53.39 (20.95)	-8.80	**0.001	0.17

注：* $p < 0.05$ ，** $p < 0.001$

在词汇难度方面，语言学LC1、LC2、LC5和LC6四个指标的均值（ $M=2.82$ ， $SD=0.13$ ； $M=254.76$ ， $SD=20.84$ ； $M=-0.55$ ， $SD=0.09$ ； $M=2.78$ ， $SD=0.14$ ）显著高于计算机科学（ $M=2.78$ ， $SD=0.15$ ； $M=251.03$ ， $SD=23.43$ ； $M=-0.60$ ， $SD=0.11$ ； $M=0.27$ ， $SD=0.16$ ），具有小到中等效应量（所有 $p < 0.05$ ；Cohen's $d=0.17$ 、 0.10 、 0.26 、 0.31 ）。这说明语言学学术论文摘要中的词汇在COCA和Brown两个参考语料库中出现的频率更高且范围更广，词汇难度更低。此外，语言学（ $M=585.04$ ； $SD=5.23$ ）在LC3指标上显著高于计算机科学（ $M=584.52$ ， $SD=6.44$ ； $p < 0.05$ ，Cohen's $d=0.04$ ），说明英语母语者对语言学中使用的词汇更熟悉。语言学（ $M=6.52$ ， $SD=0.34$ ）在

1 具体指标详见表1

LC4指标上的得分显著低于计算机科学 ($M=6.78$, $SD=0.39$) 且具有中等效应量 ($p<0.05$, $Cohen's\ d=0.40$), 这表明语言学摘要中的词汇是英语母语者在早期习得的词汇, 而这些词汇多为生活必须用词 (Maouene et al. 2008), 词汇难度较低。与其相比, 计算机科学摘要使用的词汇在参照语料库中出现的频率较低, 范围较小, 英语母语者习得时间较晚且对其熟悉程度较低, 说明计算机科学摘要使用的词汇难度更高。这与先前研究的结果一致 (Huynh Le & Tan Ha 2023), 即硬学科倾向于使用更复杂的词汇。

这种差异可能是语言学和计算机科学研究对象和探讨现象的不同所导致的。语言学主要涵盖社会和人文事件, 其使用的词汇往往对读者而言更加熟悉和易于理解 (Huynh Le & Tan Ha 2023), 如例句1在汇报结果发现时使用的皆为常用词汇。而计算机科学中可能会出现很多术语的缩写, 比如例句2中的QFD即为quality function deployment术语的缩写, 这些专业名词及其缩写在计算机科学中较为常见, 而在其他学科中使用较少, 因此计算机科学的词汇在参照语料库中出现的频率较低, 范围较小。

例句 1: Results indicate that enhancement did not affect the request strategy or amount of internal modification. (语言学)

例句 2: This report suggests a new agility development method for dealing with the interface and alignment issues among the agility drivers, capabilities and providers using the QFD relationship matrix and fuzzy logic. (计算机科学)

在词汇密度方面 (LC7), 语言学 ($M=0.71$, $SD=0.05$) 和计算机科学 ($M=0.71$, $SD=0.05$) 未呈现显著差异 ($p>0.05$; $Cohens\ d=0.03$), 说明两个学科摘要中的实词类符数占总单词类符数的比例无显著差异。语言学 and 计算机科学都注重通过较高的实词比例, 以凝练的语言在有限篇幅的摘要部分实现信息的高效传递 (Liu & Dou 2023; Zhou, W. et al. 2023)。如例句1和2的词汇密度分别为0.60和0.72, 二者均通过较高密度的实词完成对研究发现的总结与概括。

在词汇多样性方面, 语言学摘要 ($M=67.68$, $SD=20.01$; $M=59.60$, $SD=20.12$) 在LC8和LC9上显著高于计算机科学 ($M=64.41$, $SD=20.41$; $M=53.39$, $SD=20.95$) 且具有小效应量 (二者 $p<0.05$; $Cohen's\ d=0.10$, 0.17), 说明语言学摘要中使用的词汇更多样。这一发现与先前研究的结果相似 (Khany & Kafshgar 2016), 即软学科词汇多样性高于硬学科。相较于计算机科学, 语言学的研究对象涵盖了丰富的人文与社会因素, 其领域的词汇展现出更高的可替代性, 从而在摘要中呈现出更为多样的词汇。如例句3的前半句engagement, 在后半句中则描述为a construct beyond any one single paradigm, 增添新信息的同时也提升了词汇多样性 (见粗体部分)。

例句4中计算机科学学科术语 GADMS、GA 和 DMS 的重复使用则有可能会降低摘要中词汇的多样性。

例句 3: Numerous theoretical frameworks have been tied to **engagement**, indicating its importance as **a construct beyond any one single paradigm**. (语言学)

例句 4: **GADMS** is a hybrid method composed of a genetic algorithm (**GA**) and a local search strategy named direct multisearch (**DMS**). In **GADMS**, we combine binary encoding with real value encoding to utilize the advantages of **GAs** and **DMS**. (计算机科学)

4.2 句法复杂度的学科差异

表4展示了语言学和计算机科学摘要中句法复杂度的描述性统计分析以及曼-惠特尼U检验的结果。研究表明,相较于计算机科学,语言学的摘要使用了更长的句子、T单位、子句,以及更多的从属结构、并列结构、复杂名词短语和动词短语。

表4 语言学和计算机科学的句法复杂度差异

指标	语言学	计算机科学	Z	p	Cohen's d
MLS ¹	27.45 (8.38)	21.21 (6.70)	-27.14	**0.001	0.56
MLT	26.12 (8.18)	21.26 (6.41)	-22.10	**0.001	0.46
MLC	16.31 (4.36)	15.55 (3.99)	-5.07	**0.001	0.10
C/T	1.66 (0.66)	1.41 (0.46)	-17.21	**0.001	0.34
DC/C	0.37 (0.15)	0.29 (0.15)	-15.16	**0.001	0.30
DC/T	0.68 (0.59)	0.46 (0.38)	-16.14	**0.001	0.32
CT/T	0.47 (0.24)	0.36 (0.22)	-14.35	**0.001	0.29
T/S	1.07 (0.19)	1.00 (0.16)	-12.66	**0.001	0.24
CP/T	0.90 (0.52)	0.61 (0.41)	-18.15	**0.001	0.37
CP/C	0.58 (0.35)	0.45 (0.30)	-11.41	**0.001	0.23
CN/T	4.12 (1.50)	3.18 (1.11)	-23.23	**0.001	0.48
CN/C	2.57 (0.81)	2.32 (0.72)	-9.91	**0.001	0.20
VP/T	2.42 (0.90)	2.10 (0.63)	-13.71	**0.001	0.29
C/S	1.76 (0.69)	1.42 (0.51)	-20.65	**0.001	0.42

注: * $p < 0.05$; ** $p < 0.001$

在单位长度方面,语言学摘要的句子、T单位以及子句的平均长度(M=27.45, SD=8.38; M=26.12, SD=8.18; M=16.31, SD=4.36)显著高于计算机科学摘要

1 具体指标详见表2

($M=21.21$, $SD=6.70$; $M=21.26$, $SD=6.41$; $M=15.55$, $SD=3.99$), 且具有大到小效应量 (所有 $p<0.05$; Cohen's $d=0.56$ 、 0.46 、 0.10)。这一发现与以往研究的结果相似 (Dong et al. 2023; Lu et al. 2021), 即软学科学术论文比硬学科学术论文使用的语言单位平均长度更长。相较于短句, 长句能够更为有效地将丰富的信息整合至一个句子中, 从而提升了单个句子的信息密度 (Wu et al. 2020)。如例句5, 尽管该句并未包含特别复杂或生僻的词汇, 但凭借短语及子句的运用, 显著增加了句子的长度, 使得语言表达更加贴合学术论文正式严谨的特点。而计算机科学的句子较短, 这表明该学科摘要的写作风格可能更加简洁。这可能是由于包含较多专业术语的复杂长句容易给读者带来理解上的挑战, 进而降低其摘要的可读性 (Lankamp 1988)。因此, 计算机科学中短句的使用能够更有效地凸显学科内的专业术语, 以提升读者的阅读效率 (见例句4)。

例句 5: Sociolinguistic and sociocultural frameworks are blended with use of a multidimensional psycholinguistic frame, CAF, to analyze speech samples produced by 10 adult learners of English across different developmental levels as they all produced narratives containing constructed dialogue. (语言学)

在从属子句方面, 相较于计算机科学 ($M=1.41$, $SD=0.46$; $M=0.29$, $SD=0.15$; $M=0.46$, $SD=0.38$; $M=0.36$, $SD=0.22$), 语言学论文摘要 ($M=1.66$, $SD=0.66$; $M=0.37$, $SD=0.15$; $M=0.68$, $SD=0.59$; $M=0.47$, $SD=0.24$) 明显包含了更多的子句、复杂T单位以及从属子句, 具有中等到小的效应量 (所有 $p<0.05$; Cohen's $d=0.34$ 、 0.30 、 0.32 、 0.29)。这一结果表明语言学在信息传递过程中更倾向于使用从属结构。此发现与既有研究结论相契合 (Ziaeeian & Golparvar 2022), 即软学科倾向于使用更多的从句结构。从句结构通过显性化句子成分间的逻辑关联, 促进了内容的流畅表达 (Wu et al. 2020), 有助于读者的阅读。相较于计算机科学, 语言学领域内名词普遍具有更宽泛的含义 (Huynh Le & Tan Ha 2023), 往往需借助更为详尽的修饰成分如从句, 以精确传达特定语境下的含义, 从而实现信息的有效传递。以例句6为例, 作者通过从句结构 (见画线部分) 对核心名词 linguistic features 进行修饰限定。而在例句7中, 高度凝练的专业术语 AND-logic 则不需要多余的成分来修饰限定就可以实现信息的精准传达。

例句 6: Therefore, the current research intends to identify the linguistic features that distinguish students' writing samples across CEFR levels by adopting a machine-learning method, decision tree, which provides the direct visualization of decisions made in each step of the classification procedure. (语言学)

例句 7: Those algorithms mainly focus on twig pattern query with the AND-logic. (计算机科学)

在并列结构中, 语言学论文摘要的并列子句和并列短语 ($M=1.07$, $SD=0.19$; $M=0.90$, $SD=0.52$; $M=0.58$, $SD=0.35$) 显著多于计算机科学论文摘要 ($M=1.00$, $SD=0.16$; $M=0.61$, $SD=0.41$; $M=0.45$, $SD=0.30$), 且达到了中等到小的效应量 (所有 $p<0.05$; Cohen's $d=0.37$ 、 0.24 、 0.23)。这说明语言学摘要在信息传递中更注重使用并列结构来达到逻辑连贯的目的。这印证了前期研究关于软学科使用更多并列结构的观点 (Dong et al. 2023)。如例句 8 通过并列词 *and* 将两个动词短语 (见画线部分) 连接起来实现内容表达上的连贯。值得注意的是, 本研究只关注了并列子句与并列短语的使用情况, 并不能据此断定计算机科学摘要忽视内容的连贯性。计算机科学可能采用其他语言策略, 如例句 9 所示, 通过相邻句子间词汇 (如 *AR games*) 的重叠, 来实现连贯内容的构建。

例句 8: After the course, the students had increased tolerance for English diversity and understood the ways that other English varieties are realistically hybridized or dehegemonized across countries. (语言学)

例句 9: In this article, we present a game-theoretic framework, called **advance reservation (AR) games**, to analyze this strategic behavior. We use **AR games** to analyze the impact of pricing, charging, and information sharing policies on the economic equilibria of the system and on its dynamic behavior. (计算机科学)

对于短语结构, 与计算机科学相比 ($M=3.18$, $SD=1.11$; $M=2.32$, $SD=0.72$; $M=2.10$, $SD=0.63$), 语言学论文摘要 ($M=4.12$, $SD=1.50$; $M=2.57$, $SD=0.81$; $M=2.42$, $SD=0.90$) 使用了更多的复杂名词短语和动词短语, 且达到了显著水平, 具有中到小效应量 (所有 $p<0.05$; Cohen's $d=0.48$ 、 0.20 、 0.29)。这一结果表明, 语言学摘要可能倾向于采用更为复杂的短语结构。一方面, 学术研究急剧扩展, 并伴随着众多高度专业化的子学科不断涌现, 这些子学科在主题范畴与受众定位上趋向精细化 (Biber & Clark 2002)。这种“信息爆炸”现象要求信息以高效、简洁的方式进行呈现, 这在语言学摘要中表现为短语结构的较多使用。在例句 6 中, 语言学摘要通过将信息凝练并整合至以 *decision* 为中心词的名词短语结构内, 实现了信息高效且简洁的表达。另一方面, 这些短语结构有助于提升语言的精确性和丰富性。对于计算机科学而言 (可参考例句 7), 这些表达精确性的功能主要是通过更多精细化的学科术语来实现的, 从而减少了复杂短语结构的使用。

在句子复杂度方面, 语言学摘要句子中的子句数量 ($M=1.76$, $SD=0.69$) 显

著高于计算机科学 ($M=1.42$, $SD=0.51$), 且具有中等效应量 ($p<0.05$; Cohen's $d=0.42$), 表明语言学比计算机科学更常使用子句结构。这一发现与Dong et al. (2023) 结果相似, 即在软学科中作者更倾向于使用子句来增强解释力和文本连贯性, 分析文本中的论点并构建立场。软学科的知识建构更依赖于作者个人鲜明的话语说服力 (Hyland 2005), 而子句的使用有助于作者立场的建构 (Hyland & Tse 2005)。如例句 10, 语言学作者在摘要中通过子句表达了跨学科合作对语言评估具有重要意义这一观点。

例句 10: Finally, I suggest that interdisciplinary alliances with cognitive and computational neuroscience and artificial intelligence (AI) fields should be forged to meet the demands of language assessment in the 21st century. (语言学)

由此可见, 语言学和计算机科学摘要在词汇和句法特征上具有显著的学科差异。语言学倾向于使用高频基础词汇, 呈现出更高的词汇多样性和更复杂的句法结构; 而计算机科学则多依赖专业术语等高难度词汇, 但句法结构较为简单。这些差异本质上是语言学和计算机科学在知识建构方式、研究对象等方面的不同所导致的。

5. 结论

本研究考察了语言学和计算机科学国际学术论文摘要中词汇和句法的使用情况, 并结合学科特点对这些差异进行了分析。本研究发现, 在词汇复杂度方面, 计算机科学摘要更倾向于使用具有学科特色、难度高的词汇, 而语言学摘要使用的词汇更多样, 且两个学科的摘要在词汇密度方面没有显著差异。在句法复杂度方面, 相较于计算机科学, 语言学学术论文摘要使用了更长的语言单位, 以及更多的从属结构、并列结构、复杂名词短语和动词短语。

本研究通过定性分析印证了特定的句法结构往往实现了学科特定的修辞功能这一观点 (Dong et al. 2023)。本研究通过探究语言学和计算机科学两个学科学术论文摘要在词汇和句法上的使用特点, 不仅拓宽了词汇复杂度和句法复杂度研究的视野, 更为语言学和计算机科学的跨学科学术交流提供了参考, 同时也为未来探究不同学科学术文本中词汇与句法的互动模式提供了启示。此外, 本研究对学术英语写作具有指导意义: 对于教师而言, 本研究揭示的词汇句法特征可作为教学素材, 为不同学科背景的学生提供更加精准有效的词汇与句法教学策略; 对于学生而言, 理解不同学科运用特定的词汇和句法资源来实现其修辞功能的规律, 有助于提升他们的学术写作能力与跨学科交流能力。

尽管本研究在词汇与句法指标的选取上力求多样, 但仍存在一定的局限性。具

体而言, 本研究仅聚焦语言学与计算机科学, 未来研究可考虑纳入更为丰富的学科视角。其次, 本研究的语料仅限于学术论文的摘要, 未来研究可对比分析不同部分体裁的词汇与句法特征, 以揭示其更全面而深入的语言规范。此外, 本研究亦揭示了不同学科论文摘要对词汇与句法资源使用上的侧重, 但对于词汇与句法复杂度在不同学科学术写作中的互动机制及其如何共同服务于学术信息的表达, 仍需继续深入探讨。

参考文献

- Ansarifar, A., H. Shahriari & R. Pishghadam. 2018. Phrasal complexity in academic writing: A comparison of abstracts written by graduate students and expert writers in applied linguistics [J]. *Journal of English for Academic Purposes* 31: 58-71.
- Biber, D. & V. Clark. 2002. Historical shifts in modification patterns with complex noun phrase structures: How long can you go without a verb? [A]. In T. Fanego, M. J. López-Couso & J. Pérez-Guerra (eds.). *English Historical Syntax and Morphology* [C]. Amsterdam: John Benjamins. 43-66.
- Bui, G. 2021. Influence of learners' prior knowledge, L2 proficiency and pre-task planning on L2 lexical complexity [J]. *International Review of Applied Linguistics in Language Teaching* 59(4): 543-567.
- Carroll, J. B. 1938. Diversity of vocabulary and the harmonic series law of word-frequency distribution [J]. *The Psychological Record* 2: 379-386.
- Casal, J. E. & X. Lu. 2021. "Maybe complicated is a better word": Second language English graduate student responses to syntactic complexity in a genre-based academic writing course [J]. *International Journal of English for Academic Purposes: Research and Practice* 1(1): 95-114.
- Coltheart, M. 1981. The MRC psycholinguistic database [J]. *The Quarterly Journal of Experimental Psychology: Section A* 33(4): 497-505.
- Dong, J., H. Wang & L. Buckingham. 2023. Mapping out the disciplinary variation of syntactic complexity in student academic writing [J]. *System* 113. doi: 10.1016/j.system.2022.102974.
- Dong, J., K. Du & L. Buckingham. 2024. Author self-reference: A cross-linguistic/cultural and cross-disciplinary analysis [J]. *Journal of English for Academic Purposes* 68. doi: 10.1016/j.jeap.2024.101352.
- Huynh Le, N. & H. Tan Ha. 2023. Lexical demands of academic written English: From students' assignments to scholarly publications [J]. *SAGE Open* 13(4). doi:10.1177/21582440231216292.
- Hyland, K. 2001. Humble servants of the discipline? Self-Reference in research articles [J]. *English for Specific Purposes*, 20(3): 207-226.
- Hyland, K. 2005. *Metadiscourse: Exploring Interaction in Writing* [M]. London: Continuum.
- Hyland, K. & P. Tse. 2005. Hooking the reader: A corpus study of evaluative that in abstracts [J]. *English for Specific Purposes* 24(2): 123-139.
- Jiang, F. K. & K. Hyland. 2017. Metadiscursive nouns: Interaction and cohesion in abstract moves [J]. *English for Specific Purposes* 46. doi:10.1016/j.esp.2016.11.001.
- Jin, T., H. Duan, X. Lu, J. Ni & K. Guo. 2021. Do research articles with more readable abstracts receive higher online attention? Evidence from science [J]. *Scientometrics* 126: 8471-8490.

- Kim, M. & X. Lu. 2024. L2 English speaking syntactic complexity: Data preprocessing issues, reliability of automated analysis, and the effects of proficiency, L1 background, and topic [J]. *The Modern Language Journal* 108(1): 270-296.
- Khany, R. & N. B. Kafshgar. 2016. Analysing texts through their linguistic properties: A cross-disciplinary study [J]. *Journal of Quantitative Linguistics* 23(3): 278-294.
- Knuth, D. E. 1974. Computer Science and its relation to mathematics [J]. *The American Mathematical Monthly* 81(4): 323-343.
- Kuperman, V., H. Stadthagen-Gonzalez & M. Brysbaert. 2012. Age-of-acquisition ratings for 30,000 English words [J]. *Behavior Research Methods* 44(4): 978-990.
- Kyle, K. 2016. Measuring syntactic development in L2 writing: Fine grained indices of syntactic complexity and usage-based indices of syntactic sophistication [D]. Ph.D. dissertation. Atlanta: Georgia State University.
- Kyle, K., S. A. Crossley & S. Jarvis. 2021. Assessing the validity of lexical diversity indices using direct judgements [J]. *Language Assessment Quarterly* 18(2): 154-170.
- Kyle, K., S. Crossley & C. Berger. 2018. The tool for the automatic analysis of lexical sophistication (TAALES): Version 2.0 [J]. *Behavior Research Methods* 50: 1030-1046.
- Lankamp, R. E. 1988. *A Study on the Effect of Terminology on L2 Reading Comprehension: Should Specialist Terms in Medical Texts be Avoided?* [M]. Netherlands: Rodopi.
- Li, Y., Y. Gao & X. Lu. 2023. Effects of word limit on sentence length and clause length in academic journal article abstracts: A synergetic linguistic perspective [J]. *Journal of Quantitative Linguistics* 30(3-4): 322-342.
- Li, Z. 2021. Authorial presence in research article abstracts: A diachronic investigation of the use of first person pronouns [J]. *Journal of English for Academic Purposes* 51. doi: 10.1016/j.jeap.2021.100977.
- Liu, Y. & T. Li. 2024. Comparing the syntactic complexity of plain language summaries and abstracts: A case study of marine science academic writing [J]. *Journal of English for Academic Purposes* 68. doi: 10.1016/j.jeap.2024.101350.
- Liu, Z. & J. Dou. 2023. Lexical density, lexical diversity, and lexical sophistication in simultaneously interpreted texts: A cognitive perspective [J]. *Frontiers in Psychology* 14. doi: 10.3389/fpsyg.2023.1276705.
- Lu, X. 2010. Automatic analysis of syntactic complexity in second language writing [J]. *International Journal of Corpus Linguistics* 15(4): 474-496.
- Lu, X. 2012. The relationship of lexical richness to the quality of ESL learners' oral narratives [J]. *The Modern Language Journal* 96(2): 190-208.
- Lu, X. 2017. Automated measurement of syntactic complexity in corpus-based L2 writing research and implications for writing assessment [J]. *Language Testing* 34(4): 493-511.
- Lu, X., J. E. Casal, Y. Liu, O. Kisselev & J. Yoon. 2021. The relationship between syntactic complexity and rhetorical move-steps in research article introductions: Variation among four social science and engineering disciplines [J]. *Journal of English for Academic Purposes* 52. doi: 10.1016/j.jeap.2021.101006.
- Maouene, J., S. Hidaka & L. B. Smith. 2008. Body parts and early-learned verbs [J]. *Cognitive Science* 32(7): 1200-1216.