

基于语言复杂度权衡假说的古代汉语词汇历时演变研究

上海外国语大学 语言科学研究院 雷 蕾

上海外国语大学 国际教育学院 韦瑶瑜

[摘 要] 近年来学界广泛关注语言复杂度的历时演变,尤其聚焦复杂度权衡假说的验证。现有研究多在词汇和句法层面探讨复杂度权衡,但缺乏具体维度的深入分析。本研究通过考察古代汉语两千余年的词汇复杂度变化,发现整体词汇多样性呈上升趋势,其中名词和形容词的多样性持续增长,而动词和副词多样性则显著下降。这一现象揭示了语言系统在词汇层面的动态调节机制:名词/形容词多样性的提升与动词/副词多样性的降低形成互补关系,反映了语言为平衡表达效率与经济性的自我调节。研究结果不仅深化了对古代汉语演变规律的认知,也为语言复杂度权衡假说提供了基于词汇层面的实证依据,拓展了相关研究的分析维度。

[关键词] 词汇复杂度 古代汉语 复杂度权衡假说 历时演变

引 言

语言复杂度权衡假说是历史语言学领域的重要理论框架,主张语言系统在演变过程中存在动态平衡机制,即当某一语言层面(如词汇)的复杂度增加时,其他层面(如句法)的复杂度可能相应降低,以维持整体表达效率与语言经济性的平衡(Coloma, 2017; Fenk-Oczlon & Fenk, 2014)。近年来,该假说在跨语言和历时研究中获得广泛验证。例如:Bentz等(2023)通过多指标分析了80种语言的复杂度,发现约三分之一的指标存在显著负相关,支持了语言内部的复杂度权衡效应。又如:Liu等(2025)通过考察拉丁语到现代罗曼语的演变,揭示了形态复杂度与语序自由度的此消彼长。

尽管已有研究为复杂度权衡假说提供了重要证据,但现有研究在维度与方法论层面仍存在明显局限。一方面,相关研究多集中于形态、词汇、句法等层面的交互,而对词汇复杂度内部的交互关系缺乏深入分析(Zhou *et al.*, 2023; Song & Lei, 2025)。另一方面,非印欧语系语言(如汉语)复杂度的演变研究较为薄弱,现有成果多基于横截面数据或单一指标,难以全面揭示语言长期演变的多维特征(Jenset & McGillivray, 2017)。本研究旨在基于团队构建的多维度词汇复杂度测量体系和工具(雷蕾、韦瑶瑜, 2025),系统分析古代汉语词汇内部不同参数间的复杂度权衡关系。具体来说,本研究将基于跨越两千余年的古代汉语文本数据,探索古代汉语词汇复杂度系统演变的动态调整机制,为深化复杂度权衡假说理论、拓展非印欧语系语言的历时研究提供新视角与方法论支持。

1 文献综述

1.1 语言复杂度权衡假说

历史语言学是研究语言随时间演变的研究领域,该领域通过研究语音、形态、词汇、句法、语义等语言维度的历时变化,以探索语言发展的规律(Campbell, 2013; McMahon, 2001)。国内外学者对不同语言的历时演变进行了系统研究(Hsiao, 2007; Labelle & Espinal, 2014; 王力, 2014),为我们理解不同语言层面的演变提供了重要参考。

语言复杂度的历时变化研究,近年来受到学界的广泛关注。围绕所有语言是否具有相同复杂度这一争论,研究者提出了复杂度权衡假说,即在语言演变的过程中,当语言某一个层面(如词汇)变得更加复杂时,则另一个层面(如句法)可能会趋于简化(Coloma, 2017; Fenk-Oczlon & Fenk, 2014)。复杂度权衡被视为语言整体复杂度相等的证据(Shosted, 2006; Sinnemäki, 2008),研究者从多种语言视角对此进行了大量验证。比如,Bentz *et al.* (2023)针对语言复杂度定义及其测量标准缺乏共识的问题,提出用多个指标构成的向量来描述语言的复杂度,并以此评估语言具有相等复杂性假说和复杂度权衡假说。该研究通过对80种语言的28个复杂度指标进行元分析,发现大约三分之一的指标之间存在负相关,部分支持了复杂度权衡假说,且较大程度证实了语言相等复杂度假说。该研究还发现,不同语言间尽管在形态和句法层面存在复杂度差异,但它们的整体复杂度向量却几乎没有区别。又如,Shcherbakova *et al.* (2023)考察了名词和动词在编码语法信息方面的关系,并验证了语言具有相等复杂度假说,即不同语法层面的编码具有相互权衡效应。该研究利用244种语言数据,发现名词和动词的特定语法特征之间存在协同演化的关系,但整体的语法编码量并未体现出平衡。该研究结果表明,在藏缅语系中名词和动词的编码量呈正相关,而在印欧语系中则呈负相关,说明语法信息的积累和损失具有族系特异性。再如,Nijs *et al.* (2024)以平行圣经译本为语料,运用信息论中的Kolmogorov复杂度,考察了三种西日耳曼语(英语、荷兰语、德语)的形态复杂度和词序复杂度。该研究结果表明,英语在形态复杂度方面持续低于荷兰语和德语,呈现出形态简化的趋势,并且形态复杂度和词序复杂度之间存在着显著的权衡关系。该研究结果基本支持了以往的相关发现,但并未观察到荷兰语和德语在复杂性上的差异。另外,Liu *et al.* (2025)通过量化分析拉丁语到现代罗曼语的演变过程,验证了复杂度权衡假说。该研究发现,形态丰富度和语序自由度之间存在着权衡关系,如拉丁语具有更高的形态复杂度和更自由的语序,而现代罗曼语则表现出更低的形态丰富度和更严格的语序。这一演变趋势表明某个层面语言特征的增加往往伴随着另一层面语言特征的减少。总之,已有研究大多证实了语言内部或跨语言的复杂度权衡假说,但不同语言具有相同整体复杂度这一假说却依然存在争议(Fenk-Oczlon & Fenk, 2014; Sinnemäki, 2014)。

已有相关研究主要集中在探索形态、语序、句法等语言层面的历时变化(Nijs *et al.*, 2024; Shcherbakova *et al.*, 2023),对词汇复杂度历时变化的研究十分有限,仅有少量工作探讨了特定语体(如口语和学术写作)中的词汇复杂度演变(Zhou *et al.*, 2023; Zhu & Lei, 2018),更有研究考察语言演变过程中词汇复杂度与句法等其它层面特征的相互作用或权衡演进。比如,Zhou *et al.* (2023)聚焦学术写作中词汇复杂度的历时演变。该研究者通过自建学术文本百年语料库,运用词汇密度、复杂度与多样性等九项指标,系统考察了学术写作在词汇层面的百年变迁。该研究发现,学术写作的词汇复杂度呈现显著的压缩化和程式化趋势,这种

演变与学术写作整体向信息密度提升和表达常规化发展的方向相吻合。该成果不仅揭示了科学共同体写作实践的历史轨迹,其关于词汇复杂度与文本可读性关系的发现,也为学术写作教学提供了实证依据。又如,Zhu & Lei(2018)探讨了现代英语中屈折变化的历时演变特征。该研究针对传统个例观察研究的局限性,通过屈折熵指标对20世纪英语语料进行分析,发现现代英语呈现显著的屈折衰减趋势。这一发现为最小努力原则(Principle of Least Effort)、熵最小化原则(Principle of Entropy Minimization)等语言学理论提供了实证支持,揭示了语言演化过程中形式简化与认知经济性的内在关联。Song & Lei(2025)的研究是目前极少数考察语言词汇和句法复杂度相关权衡演进的。该研究以古汉语为研究对象,通过对熵和依存距离等指标的分析,探讨了词汇复杂度和句法复杂度之间的权衡关系,发现古汉语在两千多年的发展过程中,词汇复杂度显著增加,而句法复杂度则呈下降趋势。这一结果支持了在单一语言内部存在复杂度权衡的观点。

1.2 古代汉语词汇复杂度测量工具及研究

AlphaLexClassicalChinese(ALCC)是首个针对古代汉语文本词汇复杂度的系统性评估工具(雷蕾、韦瑶瑜,2025)。该工具以现代汉语词汇复杂度研究框架(Zhang & Lei, 2025)为基础,依据古代汉语的语言特征进行有针对性的优化重构,旨在通过量化分析文本的词汇特征,为古代汉语语言研究、文本难度评估、教学应用提供实证依据。

ALCC的词汇复杂度评估体系包含三个核心维度:词汇密度、词汇成熟度与词汇多样性,共51个指标。词汇密度通过实词占比反映文本信息密度,涵盖词汇总数、实词词数、词汇密度值及移动平均词汇密度等4个指标。词汇成熟度以词频等级划分来测量词汇的高阶性,依托自建的阿尔法古代汉语语料库,将词汇划分为1K至5K及其他六个等级,构建了共13项指标,包括各等级词汇占比、移动平均频次占比等指标。词汇多样性通过34项指标全面刻画词汇丰富性,包含基础统计量(如前50词不同词数)、类符-形符比及其转换形式、词类专项变异度(实词、动词、修饰语等)、熵值测量等维度,以凸显语言本体特征。

相较于现代汉语词汇复杂度测量工具(Zhang & Lei, 2025),ALCC在技术层面实现了四项关键优化(雷蕾、韦瑶瑜,2025)。首先,底层文本处理采用专为古代汉语设计的roberta-classical-chinese-large-upos模型(Yasuoka, 2023),该模型在分词、词性标注与依存句法分析三个方面的准确率均达90%以上,确保了古代汉语词汇复杂度等语言特征测量的精度。其次,针对汉语形容词分类的学术争议(Pulleyblank, 2014;王力,2013),ALCC通过“动词+形态标记”标注转换,解决了汉语形容词标注问题,保留汉语语法特点。第三,重构词汇成熟度指标体系。ALCC以阿尔法古代汉语语料库为基础建立词频分级标准,替代原有现代汉语分级词表。最后,修正了词汇多样性计算中的默认因子值(McCarthy & Jarvis, 2010)。经阿尔法古代汉语语料库验证发现,古代汉语MTLD计算需采用0.890阈值(现代汉语为0.720),该调整使复杂度测量更符合古代汉语实际特征。

雷蕾、韦瑶瑜(2025)运用ALCC对部编版语文小学、初中、高中课本中的文言文课文文本进行词汇复杂度分析,并采用随机森林算法进行上述古代汉语文本难度预测。该研究发现,模型对上述小学、初中、高中三级文本难度预测的总体准确率达71%。该实验也验证了ALCC可用于古代汉语文本词汇复杂度研究。

ALCC为古代汉语词汇复杂度的历时变化研究提供了多维度的测量工具,并为深入探讨词汇内部不同参数间的复杂度权衡关系及其在语言演变中的动态调整机制提供了方法论支持,从而有助于弥补现有研究在非印欧语系语言尤其是古代汉语历时演变大规模定量研究的不足。

1.3 现有研究局限性

传统历史语言学研究在方法论层面面临着若干挑战(Jenset & McGillivray, 2017)。一方面,历史语言学长期依赖例证分析方法,未能充分采用更具方法透明度的定量语料库研究范式。另一方面,研究方法也常缺乏统计严谨性,研究者往往过度依赖原始词频统计和事后解释性分析,而非采用假设检验框架和效应量评估等科学验证手段。近年来,随着语料库数据和定量方法的广泛应用,历史语言学研究范式也逐渐从基于实例的观察转向更系统化和大规模的数据分析(Jenset & McGillivray, 2017),这也标志着该领域研究范式的重大转型。

然而,从语言历时演变测量的指标来看,目前研究主要使用的复杂度权衡指标较为有限,主要包括频次、熵值(Shannon, 1948)以及依存距离(Liu *et al.*, 2017)等少数几类指标。这些指标在一定程度上能够反映语言系统的词汇和句法特征,但其局限性也较为明显。首先,频次分析仅能体现词语或结构的使用频次,但无法深入揭示更复杂的语言规律。其次,熵值通过衡量语言特征的非均匀分布和信息含量,被广泛应用于形态、词汇和句法层面的复杂度研究(Zhu & Lei, 2018; 雷蕾等, 2024),但其解释范围仍然有限。最后,依存距离则聚焦于句法结构中词语间的线性关系,更多被视为工作记忆负担的指标(Lei & Wen, 2020; Liu *et al.*, 2017)。由于现有测量手段主要集中在上述指标,现有研究难以全面捕捉语言历时变化中的多维度特征及其复杂度权衡关系。

除了上述方法论相关的不足以外,已有研究在探索语言复杂度历时演变方面仍存在一定局限性。首先,已有研究对语言内部不同层面(如词汇和句法)之间的复杂度权衡已有一定探讨(Song & Lei, 2025),但尚未深入到某一具体层面上的更细致分析。例如,在某一特定领域内是否也存在多个参数间的复杂度权衡,目前尚缺乏系统性研究。其次,现有研究成果主要集中在欧洲语言(Liu *et al.*, 2025; Nijs *et al.*, 2024),而对汉语等非印欧语系语言的关注明显不足,尤其是对古代语言的复杂度演变研究非常少。最后,已有研究大多基于非连续时段、横截面抽样语料进行研究(Bentz *et al.*, 2014; Labelle & Espinal, 2014),这可能不足以全面捕捉语言复杂度的长期演变规律。

2 研究设计

2.1 研究目的与研究问题

本研究旨在探索古代汉语两千余年词汇复杂度的演变趋势及词汇复杂度权衡,以为语言复杂度权衡假设提供新证据。本研究将回答如下研究问题。

研究问题1:古代汉语词汇复杂度各指标的演变趋势是什么?

研究问题2:古代汉语词汇复杂度各指标间是否存在此消彼长的权衡趋势?

2.2 研究数据

本研究所用语料为《阿尔法中国古代历史史籍语料库》(Song & Lei, 2025)。该语料库涵盖官修二十四史、《清史稿》以及部分先秦经典史籍文献,如《尚书》《国语》《春秋左传》《春秋公羊传》《春秋穀梁传》等,共30本中国古代历史史籍著作。值得注意的是,除了官修二十四史,我们还收录了《清史稿》。《清史稿》由赵尔巽主编,其虽未定稿,但由于其主编和参编人员多为清末民初人士,其语言代表了清末民初古代汉语文本特

征,因此具有重要的语言学研究价值。

语料库文本时间跨度超过2000年,从先秦到民初,共分为11个历史阶段,包括先秦、汉、魏晋、南北朝、唐、五代十国、宋、元、明、清和民国等。上述历史时期的划分能够清晰地追踪古代汉语语言演变的轨迹。另外,语料库文本在体裁上具有较高一致性,有助于控制不同文体对语言特征的影响,从而更准确地分析古代汉语词汇复杂度在2000年历史长河中的历时变化。语料库文本共包括两千余万字,其详细统计信息见Song & Lei(2025)。

2.3 数据处理

数据处理包括如下三个主要步骤。首先,我们运用古代汉语文本词汇复杂度分析工具ALCC对《阿尔法中国古代历史史籍语料库》所有文本进行词汇复杂度分析。上述11个历史阶段中,若某个历史阶段包含多本史籍文本,则对51个指标分别取平均值。

我们对所有51个指标跨越11个历史阶段的演变是否有显著上升或下降趋势做统计检验。统计检验分两步进行。第一步,我们基于Python语言statsmodels工具包中的普通最小二乘回归(Ordinary Least Squares Regression)方法,对所有词汇复杂度指标进行二次趋势检验(The Quadratic Trend Test),以验证每个复杂度指标是否具有显著非线性演变趋势。第二步,对没有显著非线性演变趋势的指标,采用Python语言py-mannkendall工具包中的Mann-Kendall趋势检验方法,做线性演变趋势检验。

3 研究结果

3.1 显著非线性演变词汇复杂度指标

二次趋势检验结果表明,下列10个词汇复杂度指标具有显著非线性演变趋势。统计结果见表1,各指标演变趋势如下页图1所示。上述10个指标均为词汇多样性维度的指标,它们都经历了从先秦时期到汉及魏晋时期的上升趋势,然后各指标沿着不同路径非线性演变。具体来说,MSTTR、MATTR、MSEAW、MAEAW、MAELW等指标呈现非线性整体上升趋势,而MTLD、MALV、MAVV1、MAModV、MSELW等指标呈现非线性整体下降趋势。

表1 二次趋势检验结果

指标	指标代码	t ²	t ² 系数	p值	趋势
均值分段类符-形符比	Mean Segmental TTR (MSTTR)	0.69	0.00	0.00	上升
移动平均类符-形符比	Moving Average TTR (MATTR)	0.69	0.00	0.00	上升
文本词汇多样性	Measure of Textual Lexical Diversity (MTLD)	0.64	-0.55	0.01	下降
移动平均实词变异度	Moving Average Lexical Variation (MALV)	0.75	0.00	0.00	下降
移动平均动词变异度1	Moving Average Verb Variation 1 (MAVV1)	0.72	0.00	0.00	下降
移动平均修饰语变异度	Moving Average Modifier Variation (MAModV)	0.65	0.00	0.01	下降
全词熵均值分段	Mean Segmental EAW (MSEAW)	0.71	0.00	0.00	上升
实词熵均值分段	Mean Segmental ELW (MSELW)	0.75	0.00	0.00	下降
移动平均所有词熵值	Moving Average EAW (MAEAW)	0.71	0.00	0.00	上升
移动平均实词熵值	Moving Average ELW (MAELW)	0.71	-0.01	0.00	上升

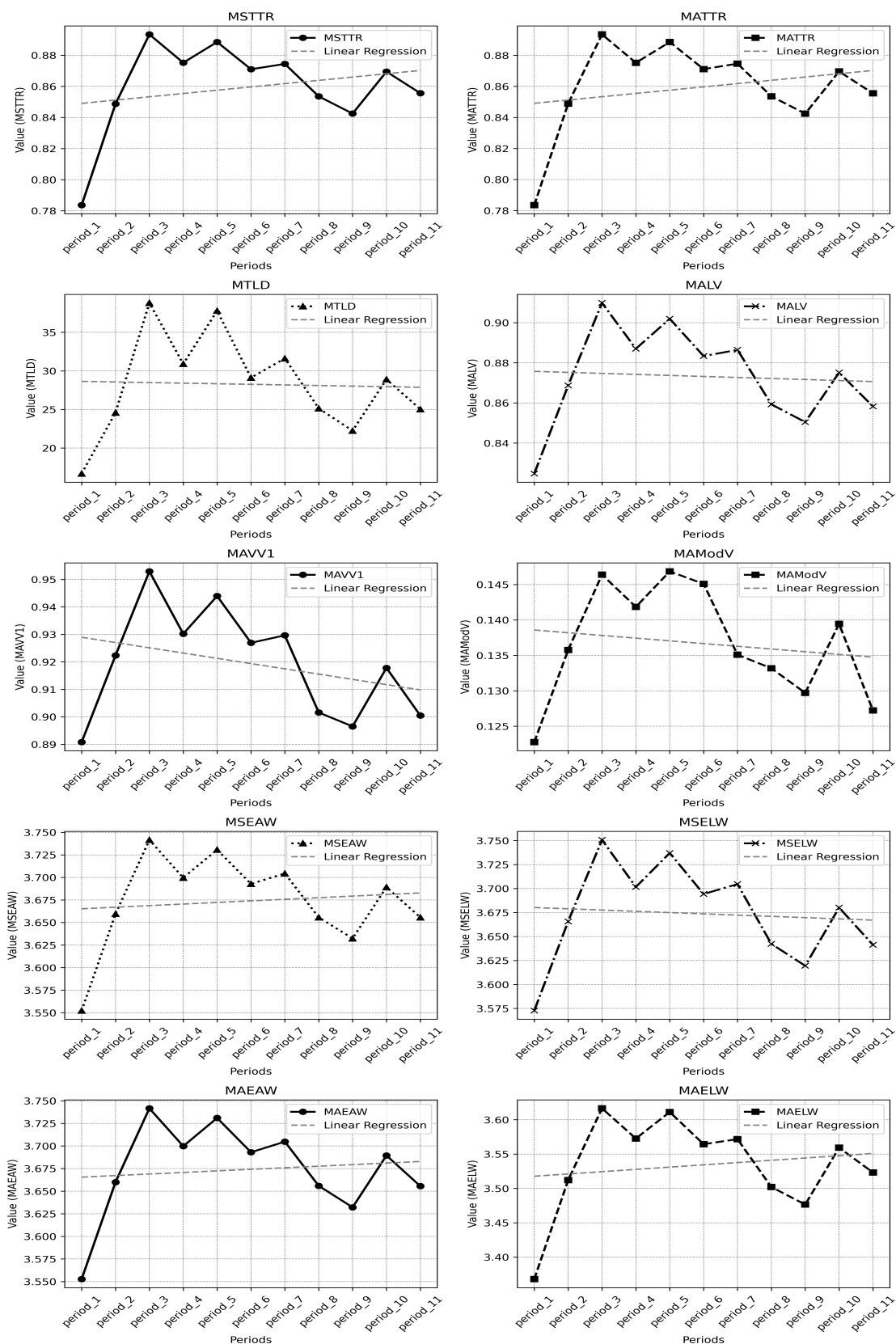


图1 非线性演变词汇复杂度指标趋势图

3.2 显著线性演变词汇复杂度指标

Mann - Kendall 趋势检验结果表明,下列4个词汇复杂度指标具有显著线性演变趋势。统计结果见表2,各指标演变趋势如图2所示。其中,MAFR2K为词汇成熟度维度指标,MAVV2、MANV、MAAdjV为词汇多样性维度的指标。具体来说,MANV、MAAdjV等指标呈现整体线性上升趋势,而MAFR2K、MAVV2等指标呈现整体线性下降趋势。

表2 Mann - Kendall趋势检验结果

指标	指标代码	Z 值	Tau 值	p 值	趋势
2K 移动平均频次占比	Moving Average Frequency Ratio 2K (MAFR2K)	-2.02	-0.49	0.04	下降
移动平均动词变异度 2	Moving Average Verb Variation 2 (MAVV2)	-2.18	-0.53	0.03	下降
移动平均名词变异度	Moving Average Noun Variation (MANV)	2.65	0.64	0.01	上升
移动平均形容词变异度	Moving Average Adjective Variation (MAAdjV)	2.18	0.53	0.03	上升

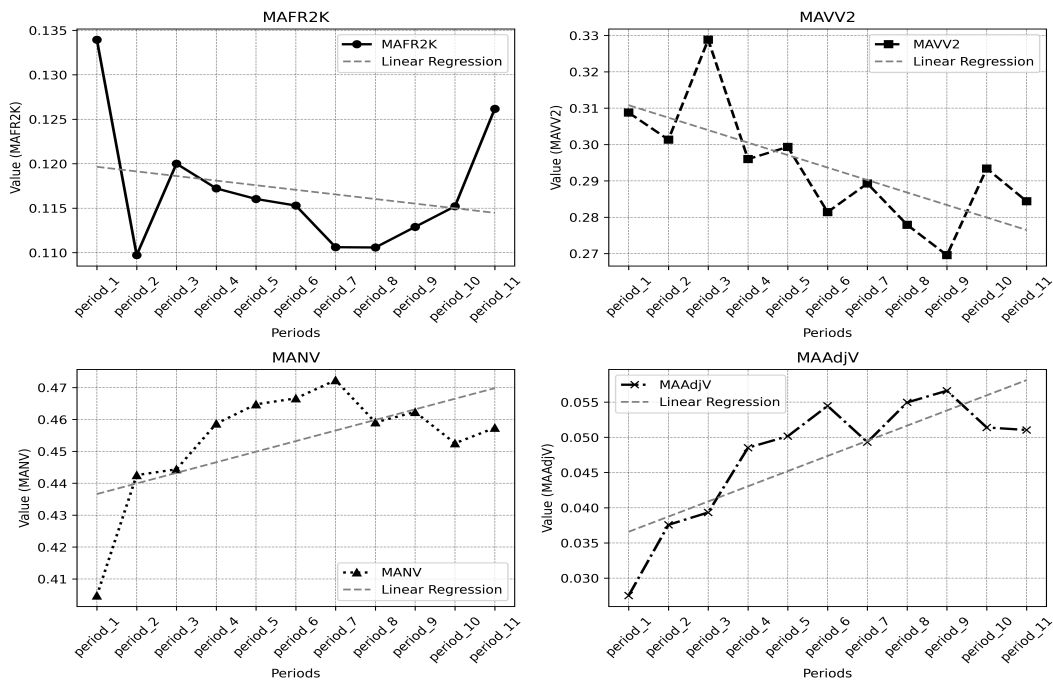


图2 线性演变词汇复杂度指标趋势图

本研究主要发现归纳如下。其一,绝大多数呈现显著演变趋势的词汇复杂度指标,整体虽有起伏变化,但在先秦至汉魏晋时期均表现出明显的上升趋势。其二,类符/形符比和熵值等词汇整体多样性维度指标整体呈现上升趋势,特别是名词和形容词等多样性指标基本呈持续上升趋势。然而,实词及其熵值、动词和副词等多样性维度指标却呈现出下降趋势。其三,古代汉语词汇复杂度在不同维度和类别上表现出差异化的演变特征,并且不同类别之间的词汇复杂度指标之间存在权衡关系,即词汇整体多样性、名词和形容词多样性指标的提升可能伴随着实词、动词、副词多样性指标的下降。

4 讨 论

本小节我们对本研究主要发现进行讨论。首先,本研究发现,绝大多数具有显著演变趋势的词汇复杂度指标,无论整体演变趋势如何,其从先秦至汉、魏晋都有明显的上升趋势,即词汇使用变得越来越复杂。Song & Lei(2025)通过单词熵值来研究古代汉语复杂度的历时变化,也发现了类似的现象。这表明词汇复杂度在这一历史阶段呈现增长趋势。词汇复杂度在这一时期的提升可以归因于多种可能因素,其中之一是语言接触带来的词汇增长。语言接触是指不同语言或方言使用者之间的频繁互动(Lavidas & Bergs, 2020; Nijs *et al.*, 2024)。汉代时期,丝绸之路的开辟促进了中国与其他国家和地区之间的交流。在这一历史背景下,新的文化、科技、生活等相关概念不断引入,丰富了汉语词汇系统,使其能够更好地反映这些文化创新(储泰松, 2025)。这与我们研究中发现的从先秦到汉代词汇复杂度的快速提升趋势似乎是一致的。另一方面,从先秦到汉代,随着国力增强,社会经济发展和文化的繁荣,汉语词汇系统也经历了系统性重构。中央集权制度的确立推动了行政管理需求的精细化,促使大量官僚体系术语的规范化和复杂化(刘晓满, 2020)。同时,汉赋等文学体裁的兴盛要求语言表达承载更丰富的修辞功能,推动了比喻性词汇和复合词的大量创造。此外,儒学经学体系的完善以及佛教思想的传入,也促使哲学、宗教领域产生大量专业术语,并进入了文学创新和日常生活中(兰宇冬, 2012)。这些社会文化变迁可能在一定程度上扩大了词汇的使用场景,并通过语义分化和词义扩展机制,促使原有词汇产生了新的语义维度,从而在某种程度上提升了词汇系统的复杂度。这种由内生文化需求驱动的词汇演变,与外源性语言接触共同推动了词汇复杂度的提升。

其次,本研究发现,在古代汉语两千余年演变发展历程中,类符/形符比和熵值等词汇整体多样性维度指标整体呈现上升趋势,特别是名词和形容词等多样性指标基本呈持续上升趋势。名词多样性的持续增长可从以下三方面解释。其一,社会文化与技术进步可能催生新概念与新事物的命名需求。随着农业工具革新、天文历法发展以及哲学思想深化,汉语中不断出现与具体事物或抽象理念相关的专有名词。这种词汇扩展在一定程度上与前文所述语言接触驱动词汇复杂度上升相呼应,即外部文化输入和内部认知深化共同推动了词汇系统的更新。其二,社会复杂化带来的专业化趋势可能推动术语系统扩展。例如,医学领域随着中医药研究的系统化而可能产生大量专业名词(臧文华等, 2018),随着法律体系的完善而形成专有术语(刘晓林, 2023),这些领域的术语积累在一定程度上提升了名词多样性。其三,知识体系的深化可能促使词汇系统细化。以《水经注》为例,其对地理特征的精准描述或可反映对自然现象认知的细化(马伊明、艾冲, 2025),而《齐民要术》中对农作物品种的命名则可能体现农业技术进步对词汇体系的推动(熊帝兵、惠富平, 2021)。这些因素共同构成了名词多样性持续上升的深层动力。

在此基础上,形容词多样性的上升趋势可能由以下机制驱动。作为名词的修饰性语法成分,形容词的扩展与名词系统的丰富性存在关联性。随着各领域名词数量的增加,与之对应的修饰性形容词亦呈同步扩展态势。我们分析了本研究11个历史阶段数据中形容词修饰名词结构在依存结构中的占比变化,其比例呈现阶段性增长特征:从先秦时期的1.97%逐步提升至汉代的3.20%和魏晋时期的3.31%,最终在明代达到峰值4.89%。这也印证了名词和形容词的协同演进关系。

再次,本研究发现,古代汉语的实词及其熵值、动词和副词等多样性维度指标却呈现出下降趋势。这一现象可归因于语言系统内部的动态平衡机制,具体表现为两个核心维度的相互作用。首先,语法结构的变

化导致词类功能转移。随着汉语语法体系的成熟,原本由动词承担的表达功能被其他语言手段部分替代。例如,动词通过语法化减少了独立动词的使用频次,从而压缩了动词多样性(庄卉洁, 2018)。这种语法资源的再分配,本质上是语言系统为提升表达效率而进行的结构性调整。另外,语言经济性原则的内在驱动深刻影响了词汇系统的演化方向。词汇系统的熵值变化遵循“最小努力原则”,即通过减少冗余、优化资源分配来降低认知与使用成本。具体而言,高频动词通过语义泛化覆盖更多场景,减少了新动词的创造需求。同时,动词通过构词法生成新词,这种策略比单纯增加独立动词更符合语言经济性原则。而作为动词修饰语的副词,其使用多样性也随着动词多样性的下降而下降。上述机制共同作用,导致动词和副词的多样性指标下降,而语言系统的整体表达效能得以维持。

先前关于古代汉语历时演变的研究发现,古代汉语在其两千年演变发展过程中呈现了词汇复杂度和句法复杂度的权衡演进的特点,即其词汇复杂度呈上升趋势,而其句法复杂度呈下降趋势(Song & Lei, 2025)。本研究聚焦古代汉语两千年的词汇复杂度演变,首次发现古代汉语在词汇维度内部的演变过程中亦呈现了权衡演进的现象,即在词汇复杂度多个指标整体呈现上升趋势的同时,名词和形容词多样性呈现上升趋势,而动词和副词多样性呈现下降趋势。

本研究的发现揭示了语言系统在词汇复杂度内部层面的动态调节机制。这种名词/形容词多样性上升与动词/副词多样性下降的权衡现象,本质上是语言系统遵循的经济性与表达力相互平衡的自我调节过程。一方面,名词和形容词的扩展通过命名新事物、修饰新概念,为语言系统提供了更精细的表达工具;另一方面,动词和副词的精简则通过语法化和构词法优化了语言表达的效率。这种调整机制既避免了词汇系统的过度复杂,又维持了语言表达的核心功能。值得注意的是,词汇复杂度的内部权衡并非简单的此消彼长,而是通过功能转移和资源再分配实现的协同演进,即当动词功能被构词法和语法结构部分替代时,形容词作为名词修饰成分的扩展恰好填补了表达需求的空白。本研究发现不仅深化了我们对古代汉语演变规律的理解,更为语言复杂度理论提供了新的分析维度,即在语言演变过程中,不同词类的动态平衡可能构成语言系统稳定性的微观基础,这种平衡既受制于社会文化变迁的宏观驱动,也遵循语言系统内部的自组织规律。

5 结 语

本研究探索了古代汉语词汇复杂度从先秦至明清的历时演变规律。研究发现,古代汉语词汇复杂度在词汇整体多样性、名词及形容词多样性等维度呈现上升趋势,而实词、动词及副词多样性等维度则持续下降。研究揭示了不同词类复杂度指标间的动态权衡关系。这一演变机制主要由三重因素共同作用。其一,语言接触与社会文化变迁引入的新概念新术语及知识体系的深化,提升了名词与形容词的多样性;其二,语法结构成熟化与语言经济性原则可能促使动词功能通过语法化等向其他语言手段转移,导致动词及副词多样性下降;其三,语言系统内部可能通过资源再分配,以实现表达效率与认知成本的平衡,如名词/形容词扩展填补了动词精简带来的表达需求空间,从而在词汇维度内部形成协同演进的自适应调节机制。

本研究发现对未来相关研究和语言教学具有启示意义。首先,本研究揭示的汉语词汇复杂度动态平衡机制,为汉语和日语等东亚语言比较研究提供了新思路。比如,通过对比汉语与日语词汇复杂度演变的异同,可深化对语言系统经济性调节机制的认知,同时为复杂度权衡假说的跨语言验证提供实证基础,以推动构建更具解释力的普适性语言理论模型。其次,今后研究可构建包含文言文、白话文及口语化文本(如明清

小说、戏曲对话、民间文书等)等多体裁、多语体语料库,通过跨体裁、跨语体的对比分析,探讨书面语与口语在词汇复杂度演变中的互动关系。另外,今后研究可引入社会、经济、文化等相关变量,通过量化分析或质性研究探讨其与语言复杂度动态调节的中介作用机制,从而更全面地揭示社会、经济、文化等因素对语言演变的潜在影响。今后研究还可结合古代汉语音系简化、构词模式演变等变量,考察语音-词汇联动对语言复杂度调节机制的深层影响,从而更全面揭示古代汉语演变的多维动因。最后,汉语词汇复杂度的历时演变规律为日语等语言的词汇教学设计提供了参考。在分级教学中,可借鉴动态调节理念,针对日语等语言的动词/副词等易混淆词类,优化词汇难度梯度与表达效率的平衡策略,通过科学配置词汇复杂度,提升学习者语言输出的准确性与表达多样性。

[本文为2025年国家社会科学基金年度一般项目“古代汉语文本难度智能评测系统开发与研究”(批准号:25BYY100)阶段研究成果。项目负责人:雷蕾]

参考文献

- 马伊明,艾冲.东汉时期护乌桓校尉府驻地新探——以《水经注》记载为基础[J].中国历史地理论丛,2025,40(2).
- 王力.中国语言学史[M].北京:中华书局,2013.
- 王力.汉语语法史[M].北京:中华书局,2014.
- 兰宇冬.佛经翻译对中古诗歌创新之影响——以“色”字为核心的考察[J].文史哲,2012(3).
- 庄卉洁.汉语动词“代”与“替”的历时演变研究[J].语文研究,2018(3).
- 刘晓林.唐律中的“亦如之”:立法语言的形成、特征与价值之一例[J].社会科学战线,2023(4).
- 刘晓满.汉代的皇帝责问[J].历史研究,2020(1).
- 萧秦英.辞汇竞争、不对称与语法演变:蒙古语否定结构与时间系统的历时演变[J].语言暨语言学,2007,8(2).
- 储泰松.丝绸之路东段早期语言景观与语言接触[J].中国语文,2025(1).
- 雷蕾,韦瑶瑜,刘康龙.AlphaReadabilityChinese:汉语文本可读性工具开发与应用[J].外语与外语教学,2024(1).
- 雷蕾,韦瑶瑜.古代汉语词汇复杂度:指标体系与测量工具[J].外语电化教学,2025(6).
- 臧文华,白红霞,蔡永敏.“引经”术语源流考[J].中华中医药杂志,2018,33(10).
- 熊帝兵,惠富平.中国古代踏粪技术传承与变迁[J].自然科学史研究,2021,40(2).
- Bentz C, Gutierrez-Vasques X, Sozinova O, et al. Complexity trade-offs and equi-complexity in natural languages: A meta-analysis[J]. Linguistics Vanguard, 2023(9).
- Bentz C, Kiela D, Hill F, et al. Zipf's law and the grammar of languages: A quantitative study of Old and Modern English parallel texts[J]. Corpus Linguistics and Linguistic Theory, 2014(2).
- Campbell L. Historical linguistics an introduction[C]//Campbell L. Historical Linguistics: An Introduction. 3rd ed. Edinburgh: Edinburgh University Press, 2013.
- Coloma G. The existence of negative correlation between linguistic measures across languages[J]. Corpus Linguistics and Linguistic Theory, 2017, 13(1).
- Fenk-Oczlon G, Fenk A. Complexity trade-offs do not prove the equal complexity hypothesis[J]. Poznan Studies in Contemporary Linguistics, 2014, 50(2).
- Jenset G B, McGillivray B. Quantitative Historical Linguistics: A Corpus Framework[M]. Oxford: Oxford University Press, 2017.
- Labelle M, Espinal M T. Diachronic changes in negative expressions: The case of French[J]. Lingua, 2014, 145.
- Lavidas N, Bergs A. On historical language contact in English and its types: State of the art and new directions[J]. Linguistics Vanguard, 2020(6).
- Lei L, Wen J. Is dependency distance experiencing a process of minimization? A diachronic study based on the State of the Union addresses[J]. Lingua, 2020, 239.
- Liu H, Xu C, Liang J. Dependency distance: A new perspective on syntactic patterns in natural languages[J]. Physics of Life Reviews, 2017(21).
- Liu S, Yan J, Liu H. The complexity trade-off between morphological richness and word order freedom in Romance languages: A quantitative perspective[J]. Zeitschrift Für Romanische Philologie, 2025, 141(2).

- McCarthy P M, Jarvis S. MTL D, vocd-D, and HD-D: a validation study of sophisticated approaches to lexical diversity assessment[J]. Behavior Research Methods, 2010, 42(2).
- McMahon A. Historical linguistics: Overview[C]//Wright J D, eds. International Encyclopedia of the Social & Behavioral Sciences. Amsterdam: Elsevier, 2001.
- Nijs J, Van De Velde F, Cuyckens H. An information-theoretic approach to morphosyntactic complexity in English, Dutch and German[J]. Journal of Quantitative Linguistics, 2024, 31(4).
- Pulleyblank E G. Outline of Classical Chinese Grammar[M]. Vancouver: University of British Columbia Press, 2014.
- Shannon C E. A Mathematical Theory of Communication[J]. Bell System Technical Journal, 1948, 27(3).
- Shcherbakova O, Gast V, Blasi D E, et al. A quantitative global test of the complexity trade-off hypothesis: The case of nominal and verbal grammatical marking[J]. Linguistics Vanguard, 2023(9).
- Shosted R K. Correlating complexity: A typological approach[J]. Linguistic Typology, 2006, 10(1).
- Sinnemäki K. Complexity trade-offs in core argument marking[C]//Miestamo M, Sinnemäki K, Karlsson F, eds. Studies in Language Companion Series. Amsterdam: John Benjamins Publishing Company, 2008.
- Sinnemäki K. Global optimization and complexity trade-offs[J]. Poznan Studies in Contemporary Linguistics, 2014, 50(2).
- Song J, Lei L. Lexical bloom, syntactic retreat: Examining complexity trade-offs within Classical Chinese evolution across two millennia[J]. Corpus Linguistics and Linguistic Theory, 2025.
- Yasuoka K. Sequence-labeling RoBERTa model for dependency-parsing in Classical Chinese and its application to Vietnamese and Thai[C]//2023 8th International Conference on Business and Industrial Research(ICBIR), 2023.
- Zhang H, Lei L. AlphaLexChinese: Measuring lexical complexity in Chinese texts and its predictive validity for L2 writing scores[J]. System, 2025, 134.
- Zhou X, Gao Y, Lu X. Lexical complexity changes in 100 years' academic writing: Evidence from Nature Biology Letters[J]. Journal of English for Academic Purposes, 2023, 64.
- Zhu H, Lei L. Is Modern English becoming less inflectionally diversified? Evidence from entropy-based algorithm[J]. Lingua, 2018, 216.

作者简介:雷蕾 男 汉族 上海外国语大学语言科学研究院教授 研究方向:语言数据科学、语料库语言学、语言数字人文

联系方式:E-mail:leileien@126.com

作者简介:韦瑶瑜 女 汉族 上海外国语大学国际教育学院副教授 研究方向:语料库语言学

联系方式:E-mail:yaoyu_wei@outlook.com

The Historical Evolution of Lexical Complexity in Classical Chinese: A Trade-off Perspective

LEI Lei & WEI Yaoyu

(Shanghai International Studies University)

Abstract: The diachronic evolution of linguistic complexity has attracted much attention, particularly in relation to the Complexity Trade-off Hypothesis. While existing research often examines complexity trade-off at the lexical and syntactic levels, detailed analysis of specific dimensions remains limited. Our study investigates lexical complexity in Classical Chinese over a period of two millennia, revealing an overall increase in lexical diversity. Specifically, diversity metrics for nouns and adjectives demonstrate consistent growth, while those for verbs and adverbs exhibit a significant decline. The pattern suggests a dynamic regulatory mechanism within the linguistic system, where increasing diversity in nouns and adjectives is complemented by decreasing diversity in verbs and adverbs, reflecting a self-regulating process to optimize communication and economy. The findings contribute to an in-depth understanding of the evolutionary patterns of Classical Chinese and provide empirical evidence from a lexical perspective to support the hypothesis of linguistic complexity trade-off, broadening the scope of relevant research.

Keywords: lexical complexity; classical Chinese; Complexity Trade-off Hypothesis; diachronic evolution

(责任编辑:王晨萌)