

学术话语中的转述潜势生成机制研究

马晓龙

中国矿业大学(北京)文法学院, 北京

收稿日期: 2025年4月22日; 录用日期: 2025年6月16日; 发布日期: 2025年6月27日

摘要

转述结构作为学术话语多声性的核心载体, 其生成机制和功能意义尚未在不同学科中得到合理解释。本文基于系统功能语言学与构式语法, 构建跨学科学术语篇语料库, 提出“学术增强型语义-语用词嵌入模型(Academic Enhanced Semantic-Pragmatic Embedding)”, 量化分析转述动词的学科特异性语义分布规律; 并提出指数衰减模型(Exponential Decay Model), 揭示了不同转述动词的语义相似度衰减速率。研究发现: 1) 学科认识论驱动转述结构的句法分化, 人文社会科学倾向使用“主体-行为-内容”构式, 使用逻辑连接词以强化立场协商, 自然科学倾向使用“现象-结论”构式, 依赖名词化与简化主谓宾结构; 2) 转述潜势(Reporting Potential)是不同学科中句法规则与语法资源的概率化生成系统, 它通过及物性过程反映知识表征功能。本研究为学术写作规范与跨学科话语分析提供了理论框架与实证依据。

关键词

转述潜势, 学科认识论, 构式语法, 系统功能语言学

The Generative Mechanism of Reporting Potential in Academic Discourse

Xiaolong Ma

School of Law and Humanities, China University of Mining & Technology (Beijing), Beijing

Received: Apr. 22nd, 2025; accepted: Jun. 16th, 2025; published: Jun. 27th, 2025

Abstract

As a core carrier of heteroglossia in academic discourse, the generative mechanisms and functional significance of reporting structures remain underexplored across disciplinary contexts. This study develops a cross-disciplinary academic discourse corpus through the integration of Systemic Functional Linguistics and Construction Grammar. We propose an Academic-Enhanced

Semantic-Pragmatic Embedding model to quantify disciplinary-specific semantic distribution patterns of reporting verbs, complemented by an Exponential Decay Model that reveals semantic similarity attenuation rates across verb categories. Key findings demonstrate: 1) Disciplinary epistemologies drive syntactic differentiation of reporting structures, humanities and social sciences favor agent-process-content constructions with logical connectives for stance negotiation, while natural sciences prefer phenomenon-conclusion patterns through nominalization and simplified SVO structures; 2) Reporting potential constitutes a probabilistic generative system of syntactic conventions and grammatical resources, operationalized through transitivity processes to fulfill knowledge representation functions. This research establishes a theoretical framework and empirical foundation for academic writing standardization and interdisciplinary discourse analysis.

Keywords

Reporting Potential, Disciplinary Epistemology, Construction Grammar, Systemic Functional Linguistics

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

转述结构作为学术话语多声性的主要载体[1],既是知识建构的认知支架,也是一种特定的修辞策略[2][3]。现有关于转述行为的研究可归纳为三个主要方向:第一类聚焦引文类型的文本分析[4][5];第二类探讨立场标记的语用功能[6]-[9];第三类基于语料库开展引述词汇的量化研究[10]-[12]。

尽管这些研究初步揭示了转述行为的学科差异性,如不同学者群体的引用模式不一[13]。当前学术讨论仍存在三方面显著缺陷:首先,人工标注导致语料规模受限[14][15],致使跨学科结论的效度受到质疑[16][17];其次,转述结构的句法范式尚未形成系统化理论框架,缺乏可操作的定义标准与使用规范;最后,现有理论对转述功能存在概念简化倾向:主流研究多将转述行为简化为“作者-读者”二元互动模型[18],忽视了其在知识本体(ontology)表征[19][20]、人际立场协商[21]、信息层级组织[22]的多维功能属性。这种理论缺失尤其体现在对学科知识生产过程中功能语义机制的阐释不足上,未能揭示转述行为如何通过及物性系统实现概念意义的学科化重构。

从系统功能语言学(Systemic Functional Linguistics)理论范式考察,转述结构在学科话语中的功能实现存在三重理论困境。Martin[23]提出的介入系统(Engagement System)作为评价理论的核心机制,虽能通过调节态度资源分布实现学术立场的语篇建构[24],但其解释效力在转述实践中呈现显著局限性。现有研究多聚焦单模态书面文本的质性分析[25][26],未能揭示多模态学术语境中转述结构的语法实现过程。更为关键的是,当前理论框架尚未有效阐释学科实践对转述结构的重塑机制[27],特别是在知识生产层面,转述过程如何通过及物性系统实现概念功能的学科化重构仍属理论盲区。

值得注意的是,尽管已有研究尝试运用深度学习(Deep Learning)建模转述结构的句法规则[28],但方法论层面仍存在根本性矛盾:技术路径过度依赖表层形式特征抽取,未能建立学科认知范式与语法实现方式之间的参数化模型。这种理论与技术的双重局限致使核心命题持续悬置——转述结构的学科变异究竟是表层修辞策略的偶然分化,还是不同学科认识论在语言系统层的必然投射?该问题的解答亟待构建能同时解释知识建构机制与语法生成规则的双向分析模型。

计算语言学方法的突破为转述结构的层级化验证提供了三重实证路径。首先,基于学科类型的统计

建模验证了句法表征的显著差异[29][30],证实学科知识形态与语言编码策略存在系统性关联。其次,可视化分析技术成功追踪了转述词汇的历时语义演变轨迹,揭示其语义网络结构具有学科敏感性和领域适应性特征。最后,机器学习实验通过句法-功能耦合度分析[31]-[34],证明转述结构的功能实现受制于学科话语的深层规约机制。这些发现从方法论层面实证支持了“转述潜势”(Reporting Potential)的理论假设——即学术话语共同体在特定学科范式下策略性配置语法资源的系统性能力,具体表现为对词汇密度、句法复杂度及逻辑衔接模式的差异化选择。

然而,当前研究范式仍存在双重理论断裂:一方面,现有成果多局限于词汇层面的频率统计[35][36],未能结合系统功能语法框架解释转述潜在及物性系统、主位推进模式等语法层级上的生成机制[37];另一方面,计算语言学方法与功能语言学理论尚未形成有效对话,特别是缺乏对“学科知识建构→语法资源选择→功能意义实现”这一作用链的参数化建模。这种理论真空导致转述行为研究难以突破描述性分析的困境,无法揭示学科认识论差异如何实现知识表征的学科化重构。

为有效填补现有理论缺失与方法论局限,本研究以跨学科对比为核心,致力于推进转述结构的句法与功能分析工作,具体研究流程如下:第一阶段,构建跨学科语料库。以Goldberg[38][39]的构式形态学为理论基础,对不同学科语境中的转述结构展开系统梳理与总结。创新性地提出“学术增强型语义-语用词嵌入”(Academic Enhanced Semantic-Pragmatic Embedding)方法,在词向量空间内,精准建模并量化转述词与搭配词之间的语义关联程度。借此,深入挖掘转述动词在不同学科论文中的统计规律差异,进而全面且客观地考察学科语境对转述结构产生的影响。第二阶段,依据Halliday对及物性(Transitivity)过程的阐释,精准定义转述潜势,即在学科认识论的驱动下,将其视为写作者策略性调用句法模板与语义资源的概率化生成系统。随后,对本文核心假设进行严谨验证,即学科范式能够显著预测句法模板的选择,且转述功能对应遵循可形式化的句法规则。

该研究的实践意义超越理论语言学范畴。在学术撤稿和不当引用的背景下,转述话语规范研究成为维护学术诚信的关键。国际期刊日益强调结构化摘要格式[40][41],本研究为学科特异性写作指南提供实证基础。最终,本研究阐明转述结构的语法深层结构如何同时反应学科认识论与构建认知权威,这一双重角色在当前话语分析中尚未得到充分理论化[42]。

2. 转述结构的研究进展

早期转述结构研究聚焦引语类型学分析,通过人工标注揭示学科差异特征,人文科学倾向采用直接引语以保持原始权威性,而自然科学更偏好间接引语实现客观化表述。此类研究虽初步确立学科差异假设,但其方法论存在双重制约:语料规模局限与分类标准表层化,未能深入句法形式化分析。

近年来,文本分析范式逐渐转向数字驱动。例如,Huang[43]定义并探讨了文献之间直接引用关系(DCCPs)。研究者每篇论文构建了一个以该论文为中心的引文网络,包含所有引用该论文的文献及其相互之间的引用关系。通过使用微软学术图谱(MAG)中的计算机科学领域数据集(MAG-CS)分析发现,DCCPs有助于区分具有深远影响或依赖性影响的研究,丰富了对共引和文献耦合关系的理解。这一发现验证了早期研究的学科差异假设,但进一步揭示了转述结构的复杂性。Maton[44]探讨了如何通过语言表达的复杂性来研究知识实践的复杂性,特别是通过语义密度(Semantic Density)来分析话语中的意义如何被浓缩和相互关联。简而言之,直接引语在人文科学中常嵌套于多层从句,而自然科学中的间接引语则更倾向于简化句式。当前研究虽通过句法复杂度指标(Syntactic Complexity Index)实现大规模量化验证[45][46],但其解释框架仍存在根本缺陷:过度依赖词汇频率统计(如直接/间接引语分布比例),缺乏对转述机制的系统性理论建模[47][48]。

语料库研究[49]通过对转述动词与评价副词的研究,揭示了学术立场建构策略。Hyland[6]的跨学科

比较表明,人文科学倾向使用高介入性标记,而自然科学偏好隐性评价策略。Varttala [50]进一步发现,动词选择反映学科认识论取向——人文科学多采用认知型动词,自然科学偏好实证型动词。此类研究虽深化对立场表达机制的理解,但其方法论存在显著局限:局限于词汇层面分析,未能解释句法结构对学术修辞的制约作用。

系统功能语言学为突破此局限提供理论工具。Martin [51]的语篇语义框架揭示,转述结构通过及物性系统(Transitivity system)实现双重功能。名词化结构压缩信息流强化概念功能,而评价性转述动词激活人际功能。后续研究[52]证明,学科差异本质上是语法资源的选择差异——人文科学利用过程类型扩展实现立场对比,而自然科学通过主谓结构简化提升信息密度。这些发现暗示存在系统性语法资源调配机制,本文将其理论化为转述潜势系统(Reporting Potential System),即学科共同体基于认识论需求对句法资源进行的概率化选择模式。

此外,转述潜势的概念虽未明确定义,但其机制在后续计算语言学领域得到初步验证。Zhang [34]探索了任务感知专业化和指令调整两种技术,通过 Mixture-of-Experts Transformer 架构和任务特定指令的输入,实现了任务感知的输出,此研究尝试论证学术写作中的转述结构是否遵循概率化句法规则。人文科学论文中,转述结构较多强调对转述内容和作者主体观点的重视,自然科学论文则注重对客观事实的呈现与因果关系的推导。这种转述结构的差异折射出两种认识论立场的深层分野:解释主义(Interpretivism)与实证主义(Positivism)。解释主义强调对文本、思想与学术立场的解读,注重主体性在知识建构中的作用;而实证主义则以科学现象的客观描述与实验结论的逻辑推导为核心,追求知识的普遍性和证伪性。因此,不同学科的转述结构不仅是句法选择的结果,更是其认识论立场在语言表达中的体现。

然而当前研究存在理论断裂:词汇频率统计无法解释及物性系统与主位推进等语法层级的生成机制。缺乏对“学科认识论 → 语法选择 → 功能实现”双向作用链的参数化建模[37]。

综上所述,尽管转述潜势的概念具有理论吸引力,其形式化机制与功能意义仍面临挑战。有鉴于此,本文拟以构式形态学理论和系统功能语言学的元功能理论为基础,整合计算语言学方法,构建跨学科研究框架。本文旨在形式化转述结构的句法规则,提出转述潜势系统的定义,讨论这一系统的功能意义并揭示其与学科认识论的深层关联。

3. 研究设计

3.1. 研究问题

本研究的核心假设为:其一,转述结构的句法规则与学科认识论范式显著相关;其二,转述潜势的实现遵循形式化的句法规则,并映射对应的及物性过程。结合研究假设,本文着重讨论以下三个问题:

- 1) 不同学科论文的转述结构在句法形式上是否存在显著差异?
- 2) 学科语境是否显著影响转述动词的语义规律?
- 3) 转述潜势的实现是否遵循形式化的句法规则,其所反映的语言功能在不同学科中是否呈现系统性差异?

3.2. 语料来源

参考国内外期刊近年的影响因子和研究范围,本文以 Web of Science 平台作为学术论文语料的数据来源,自建跨学科学术语篇语料库。Web of Science 平台收录了来自全球的高质量学术期刊、会议录、图书、专利和网络资源等多种类型的文献,其核心合集包括 Science Citation Index Expanded (SCIE)、Social Sciences Citation Index (SSCI)、Arts & Humanities Citation Index (A & HCI)等多个子库,论文数据质量具有保障。在本研究中,为平衡学科分布,语料选取涵盖自然科学(Natural Science)、人文社会科学(Humanities

& Social Science)的学术论文语篇。考虑到研究成果的时效性,时间跨度为2022年至2024年,检索论文数量为638篇。自然科学论文检索关键词为Natural Science,人文社会科学论文检索关键词为Social Science和Humanities,所选论文均为高被引论文,保证学术水平符合要求。最终建立的语料库中,token数5,832,831,type数162,754。

3.3. 研究步骤

首先,从构式形态学视角出发,讨论转述结构的语法规则并概括转述结构,在自建对比语料库中检索对应的代表性转述词。采用词嵌入方法将上述代表性转述词从高维离散的符号空间映射到低维连续向量空间,计算转述词与其搭配词之间的语义关联度。词嵌入是一种展现语义分布规律的可视化方法,近年来已成为研究词汇语义变化的常用方法[53],其优势在于通过捕捉词与词之间的语义和语法关系,将离散的词符号转换为连续的向量表示,当所有的词向量聚合在一起即形成一个词向量空间[54]。在词向量空间中,每一个词向量可以看作作为一个坐标点,各点之间的空间距离可用于判断词语之间的语义相似度,此“空间距离”即余弦相似度,其基础计算方法如(1)所示:

$$\text{cosine_similarity}(A, B) = \frac{A \cdot B}{\|A\| \cdot \|B\|} \quad (1)$$

假设两个向量分别为 A 和 B , $A \cdot B$ 是向量 A 和 B 的点积, $\|A\|$ 和 $\|B\|$ 分别是向量 A 和 B 的模长。余弦相似度的取值范围在 $[-1, 1]$ 之间。值越接近1,表示两个向量越相似;值越接近-1,表示两个向量越不相似;值为0时,表示两个向量正交(在高维空间中可能不相关)。本文考察转述词与搭配词之间的相关性,若相关度高,则转述词与搭配词会相互聚集,将聚集结果放置在同一张图中则可以呈现不同学科中转述结构的搭配词差异。考察转述词与搭配词的语义关联度后,提出“指数衰减模型”(Exponential Decay Model, EDM)计算转述动词的语义衰减规律。另外,“语义向量空间”(Semantic Vector Space, SVC)能够有效捕捉和量化构式的语义特征,这也为研究转述结构的句法规则提供了新的思考方式[55]。语言功能层级方面,引用系统功能语言学的元功能理论考察不同学科中转述结构所映射的概念功能,并提出转述结构的高级范畴概念,即转述潜势。

4. 研究结果及讨论

4.1. 学科范式驱动的转述构式

构式语法视角下,语法是“形式-意义-功能”相互配对的复杂结构[56],其由固定成分与空位组成,具有定位性、能产性和黏着语义特征。实际上,单个词或者句法结构也都是构式,区别仅在于内部复杂性不同。构式中的各个要素通过承继层级(inheritance hierarchies)相互联系。例如派生词缀构式X-ment, -ment所构成的词由动词义转为名词义,意为行为、结果或状态,其构词型式如下所示:

$[[X]_V\text{-ment}]_N \rightarrow \text{“Actions, outcomes, or states.”}$

X-ment的构词型式通过承继层级上接更抽象的图示,下含作为其示例的具体词汇,从而形成层级关系。换言之,这种构词型式是对体现形式语义系统性配对的复杂词组概括而成的一种抽象图示[57]。人们在接触大量的同一类型词语时,会从中推断出某种抽象图示,并将其扩展用于构造其他相关的同类型词语[55],例如X-er派生出的词义为“动作的执行者”。由此可见,构式结构是抽象性概括,具有图示性,且包含下一层级的词汇或句法结构示例。

基于以上论述,本研究将转述话语视为学术语篇中的一种句法构式,其本质为“固定句法结构+可变语言内容”的配对体。转述构式的核心特征符合构式语法,解释如下:定位性上,转述构式中存在固

定成分,例如转述动词、引用标记词等占据固定的句法位置,形成如主句谓语或从句引导词。能产性上,转述构式替换空位内容生成大量实例,比如替换被引作者,自然科学论文中则需要考虑对数据、公式等的引述规范。黏着语义性上,固定成分的语义需要依赖构式整体激活,即不可以脱离构式考察转述词,例如“argue”在人文科学论文中可能隐含的语义是对学术观点的辩论。综上,转述结构可抽象概括为三类核心句法构式,且具有图式性,具体分析如下:

1) “主体-行为-内容”构式,常见于人文社会科学论文,固定成分包括作者主体、转述动词、引述标记,可变空位包括被引用内容。句法形式上表现为“作者主体+立场性转述动词(如 argue, claim)+that 引导的引述内容”,其通过显性作者介入与立场标记传递学术观点,且转述动词一般隐含作者立场[24],例如“Foucault argues that discourse shapes power relations.”。核心依存关系为主句动词(V)支配从句引导词(COMP),形成主从结构。人文社会科学论文表述中,主体为作者(NP),行为是立场性转述动词(如 argue, claim),内容为 that 引导的从句(CP)。因此,人文社会科学论文中的转述结构可概括为:[主体(NP)]+[行为(V): argue/claim...]+[被引内容(CP)]。

2) “实证现象/推理结果-结论”构式,常见于自然科学论文,固定成分包括现象描述、转述动词、引述标记,可变空位包括研究结论、数据对比等(注意,空位的句法结构一般为名词化短语或被动结构)。其句法形式为“现象主体(如 data, results)+实证性转述动词(如 show, demonstrate)+that 引导的结论内容”,通过非人称化表达,强调客观现象与结论的直接映射,例如“The data demonstrate that the hypothesis is valid.”。例示的上层抽象图示为主谓宾框架,固定句法角色为“现象+结论+被引内容”,核心依存关系表现为主句动词(V)支配名词化短语,体现非人称化特征,实证动词则需依赖构式激活可验证性语义。在自然科学论文表述中,现象为数据或结果(NP),结论为实证性动词(如 show, demonstrate),内容为名词化短语或被动结构(CP)。因此,自然科学论文中的转述结构可以概括为:[实验现象/结果(NP): data...]+[结论(V): show/demonstrate...]+[被引内容(CP)]。

3) 元话语构式,跨学科通用,固定成分包括引述标记、作者信息,可变空位包括引述内容。句法形式为“引述标记(According to...)+作者信息+引述内容(直接或间接引语)”,通过中立引述标记传递信息,平衡客观性与立场表达[33],例如“According to Smith (2022), the results suggest a correlation.”。结构上表现为介词短语+主谓结构(PP+NP+CP),固定句法角色为“标记+来源+内容”,元话语构式表现出的学科特征为跨学科通用,标记为引述介词(如 According to.../Based on...),来源为作者信息(NP),内容为直接或间接引语(CP)。核心依存关系中,介词(P)支配来源(NP),形成状语结构,引述标记需依赖构式激活中立性语义。因此,元话语转述结构可以概括为:[引述标记(NP)]+[转述来源(NP)]+[引用内容(CP)]。

上述句法结构的形成包含三个主要模块。其一,输入端,包含自由句法结构;其二,定位端,部分词汇或句式的高频使用导致构式中的某一成分定位化;其三,构式图示化,空位允许不同变量的填充,语义黏着性增加。

综上所述,本研究在自建跨学科语料库中提取三类转述构式,并对高频转述词进行频率统计。为确保数据准确,提取时合并同一动词的不同屈折形式。通过对 583 万形符的语料库分析,提取出 10 个使用频率最高的转述词,用于后续计算转述动词相关参数,结果见表 1。

Table 1. The statistical results of high-frequency reporting words
表 1. 高频转述词统计结果

转述词	总计频次	自然科学	人文社会科学
show	5413	2675	2738

续表

state	4732	1776	2956
indicate	3366	1340	2026
suggest	2800	886	1914
note	2207	633	1574
propose	2160	999	1161
demonstrate	1414	683	731
confirm	956	467	489
argue	681	125	556
claim	484	138	346

4.2. 转述动词的语义分布规律

为深入探究转述结构的句法规则与学科认识论范式的关联,依据前文对转述结构中高频转述词的统计,本研究选取三种转述结构中出现的高频转述词“show”,“argue”,“suggest”,通过词嵌入方法,提出适用于计算学术语篇词义相似度的词向量公式。由此,分析不同学科转述结构中转述词与其搭配词之间的语义关联程度,探讨不同学科语境中转述动词的语义分布规律。

传统词嵌入模型,如 Word2Vec、GloVe 等主要依赖上下文窗口内的词汇共现信息。然而,从语言学视角来看,词不仅仅是机器层面的字符串,除去计算机意义外,还具有语言意义与价值。早在 1950 年, M. Joos 指出:“一个语素的语言学意义……可以定义为该语素与上下文中的所有其他的语素出现的条件概率的集合”[58]。换言之,词的价值不是由标志它的客观对象的实体来确定的,而是由它对其他词的关系及其在该语言中的地位来决定的[59]。这说明词向量在语言学上具有一定“可解释性”(explainable),反映了语言中的统计学规律,确实可以作为描述自然语言数学面貌的可行研究方法,且可以通过语言规律分析语言现象。

本文对转述词与其搭配词之间的语义关联度作重点考察,采用的计算方法为词向量夹角的余弦(cosine)值。余弦值的数学逻辑是线性代数中的“点积”(dot product),其定义如公式(2)所示:

$$v \cdot u = \sum_{i=1}^N v_i u_i \quad (2)$$

传统词向量长度计算公式为:

$$|v| = \sqrt{\sum_{i=1}^N v_i^2} \quad (3)$$

然而,如果词向量越长,那么点积就会越大,导致的问题是高频单词可能更加倾向于和更多的单词共现,由此导致计算难度变大。因此改进方法在于将词向量长度归一化,即用两个词向量的长度来除以点积的值,得到“归一化点积”(normalized dot product),也就是两个词向量夹角的余弦值 $\cos \theta$,其公式定义如(4)、(5)所示:

$$a \cdot b = |a| |b| \cos \theta \quad (4)$$

$$\frac{a \cdot b}{|a| |b|} = \cos \theta \quad (5)$$

由此，两个词向量之间的余弦值可以使用公式(6)：

$$\cos(\mathbf{v}, \mathbf{w}) = \frac{\mathbf{v} \cdot \mathbf{w}}{\|\mathbf{v}\| \|\mathbf{w}\|} = \frac{\sum_{i=1}^N v_i w_i}{\sqrt{\sum_{i=1}^N v_i^2} \sqrt{\sum_{i=1}^N w_i^2}} \quad (6)$$

本节所讨论的转述词与其搭配词之间语义关联使用的即为以上基本计算逻辑。然而，学术语篇中转述词的功能高度依赖学科语境，因此本研究在原有词向量公式的基础上提出修改，提出学术语篇转述词专用词嵌入模型：“学术增强型语义-语用词嵌入公式”(Academic Enhanced Semantic-Pragmatic Embedding, AESP Embedding)，如(7)所示：

$$\mathbf{w}_i = \lambda \cdot \mathbf{w}_b + \mu \cdot \left(\sum_{d \in D} \omega_d \cdot \mathbf{w}_d \right) + \nu \cdot \left(\sum_{s \in S} \rho_s \cdot \mathbf{w}_s \right) \quad (7)$$

参数解释如下：

其一，基础向量 $\mathbf{w}_b$ 基于传统 Skip-gram 模型计算，捕捉通用语义特征，其训练语料来自于大规模学术文本，上下文窗口扩展至 10 个词以覆盖长句结构。计算方法如(8)所示：

$$\mathbf{w}_b = \frac{1}{N} \sum_{c \in \text{Context}(w_i)} \mathbf{c} \quad (8)$$

其中， N 表示上下文词数， $\mathbf{c}$ 表示上下文词向量。

其二，关于学科加权向量 $\left(\sum_{d \in D} \omega_d \cdot \mathbf{w}_d \right)$ 的计算。学科权重 ω_d 通过学科标签动态调整，计算方式如(9)所示：

$$\omega_d = \frac{\text{TF-IDF}(w_i, d)}{\sum_{d' \in D} \text{TF-IDF}(w_i, d')} \quad (9)$$

其中， $\text{TF-IDF}(w_i, d)$ 表示词 w_i 在学科 d 中的权重值，用于衡量学科特异性。学科向量 ω_d 基于学科子语料库训练的独立词向量，反映学科内转述词的共现模式。

其三，句法依存向量 $\left(\sum_{s \in S} \rho_s \cdot \mathbf{w}_s \right)$ 。依存关系权重 ρ_s 根据句法角色 s (如主谓、动宾)的重要性赋值，计算方式如(10)所示：

$$\rho_s = \frac{\text{Freq}(s)}{\max_{s' \in S} \text{Freq}(s')} \quad (10)$$

其中， $\text{Freq}(s)$ 表示句法角色 s 在学术语篇中出现的概率。句法向量 $\mathbf{w}_s$ 通过依存句法树提取目标词的句法角色。

另外，在此优化公式中，调和系数 λ 、 μ 、 ν 用于平衡通用性、学科特异性间关系，避免计算数据过拟合。

综上，本文将采用(7)在学术语篇中计算特定词的词向量。当考察转述词与搭配词的语义关联度时，传统的余弦相似度计算方式可能导致某些高频词出现次数过多，影响结果，因此本文提出对余弦相似度进行改进：

$$\text{AESP-Sim}(w_i, w_j) = \frac{w_i \cdot w_j}{\|w_i\| \cdot \|w_j\|} \times \log(1 + \text{Co-Occur}(w_i, w_j)) \quad (11)$$

其中，改进点在于引入共现频次的对数项 $\log(1 + \text{Co-Occur}(w_i, w_j))$ ，增强了高频搭配词的关联权重。

根据实验结果，“argue”，“show”，“suggest”与其语义最相关的前 8 个词汇间的计算结果分别见表 2~4。

Table 2. The top 8 words most semantically related to “argue”
表 2. 与 “argue” 语义最相关的前 8 个单词

lead	may	due	expect	thus	however	find	also
0.495	0.488	0.474	0.469	0.458	0.451	0.449	0.447

Table 3. The top 8 words most semantically related to “show”
表 3. 与 “show” 语义最相关的前 8 个单词

result	fig	compare	indicate	two	significantly	also	respectively
0.671	0.668	0.625	0.611	0.593	0.587	0.578	0.574

Table 4. The top 8 words most semantically related to “suggest”
表 4. 与 “suggest” 语义最相关的前 8 个单词

finding	find	also	however	may	significantly	indicate	result
0.724	0.702	0.694	0.685	0.667	0.655	0.643	0.641

根据以上计算结果，本文发现学科语境显著影响转述结构的句法表达。人文社会科学论文呈现高语义复杂度特征，其转述结构偏好使用嵌套从句以承载学术立场协商。以高频转述动词“argue”（频次 556）为例，其与逻辑连接词“however”（0.451）、“thus”（0.458）及认知情态词“may”（0.488）的强关联性，反映出该学科通过句法复杂化实现多声对话的典型策略。这种句法选择与解释主义范式的认识论需求相契合，即通过逻辑连接词与评价性副词（如“critically”共现率 0.086）的协同作用，表现不同学术观点的辩证性。

相比之下，自然科学论文则遵循低语义复杂度原则，其转述结构呈现显著的信息压缩特征。高频转述动词“show”与实证标记词“result”（0.671）、“fig”（0.668）及“significantly”（0.587）形成稳定搭配模式，表明该学科倾向于采用主句动词直接支配名词化短语的句法构式。此类结构通过抑制修饰成分与多重从句，实现现象与结论的直接对应，符合实证主义范式对客观权威性的语法化要求。这种句法简化策略不仅优化了信息传递效率，更通过关系过程主导的及物性网络，强化了科学话语中概念功能的单向传递机制。

为更好描述转述动词与其相关词汇关系，本文提出指数衰减模型拟合每个动词的相似度随排名变化的趋势：

$$y = a \cdot e^{-b \cdot x} \tag{12}$$

其中， x 为相关词汇排名， y 为相似度得分， a ， b ， c 为每个动词的独立拟合参数，根据公式(11)计算结果，通过最小二乘法得到各转述动词参数，如表 5 所示。图 1 根据表 5 计算结果呈现了公式(12)的函数拟合图像，可以更好地观察转述动词与其语义高相关词汇间关系。

Table 5. The calculation parameters of three types of reporting verbs
表 5. 三类转述动词计算参数

转述动词	a	b	模型公式
argue	0.495	0.015	$y = 0.671 \cdot e^{-0.012x}$
show	0.671	0.012	$y = 0.724 \cdot e^{-0.018x}$
suggest	0.724	0.018	$y = 0.495 \cdot e^{-0.015x}$

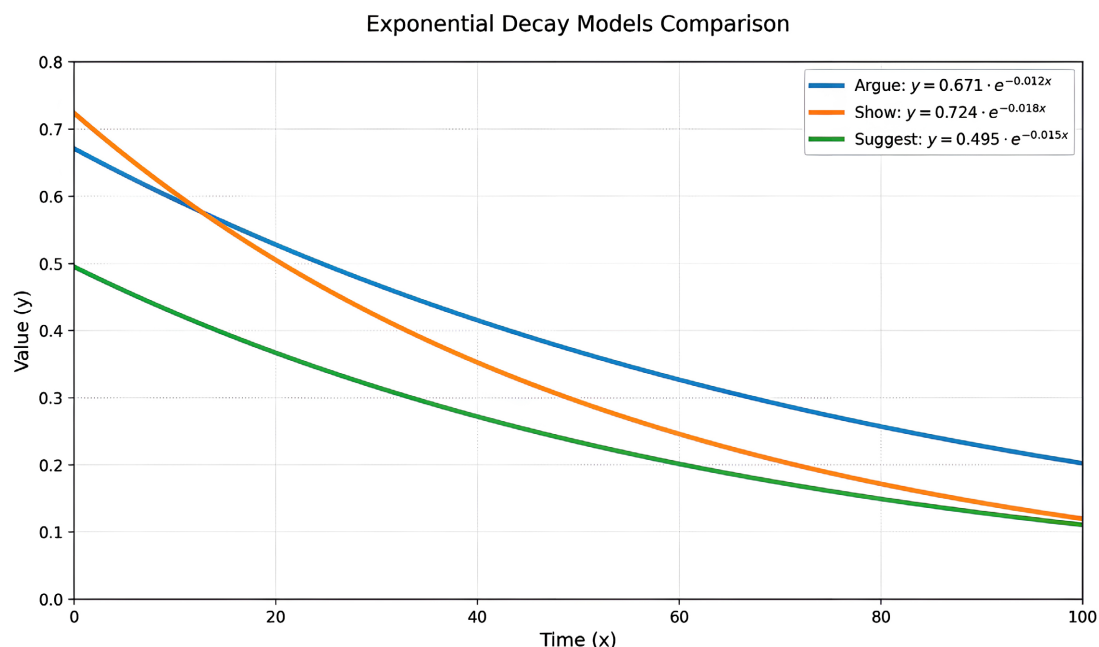


Figure 1. Graph of function fitting for the exponential decay model of reporting verbs

图 1. 转述动词指数衰减模型的函数拟合图像

结合(11)计算结果,通过公式(12)计算得出 3 个转述动词与其前 8 个词汇的语义相似度变化结果,由此可以观察转述动词的语义变化规律,具体结果见表 6,其中 1~8 的整数为指数衰减模型中 x 取值,即与转述动词语义相关度最高的前 8 个单词。图 2 则根据公式(12)呈现了三种转述动词与其语义相关度最高的 8 个单词之间的相关性变化趋势。

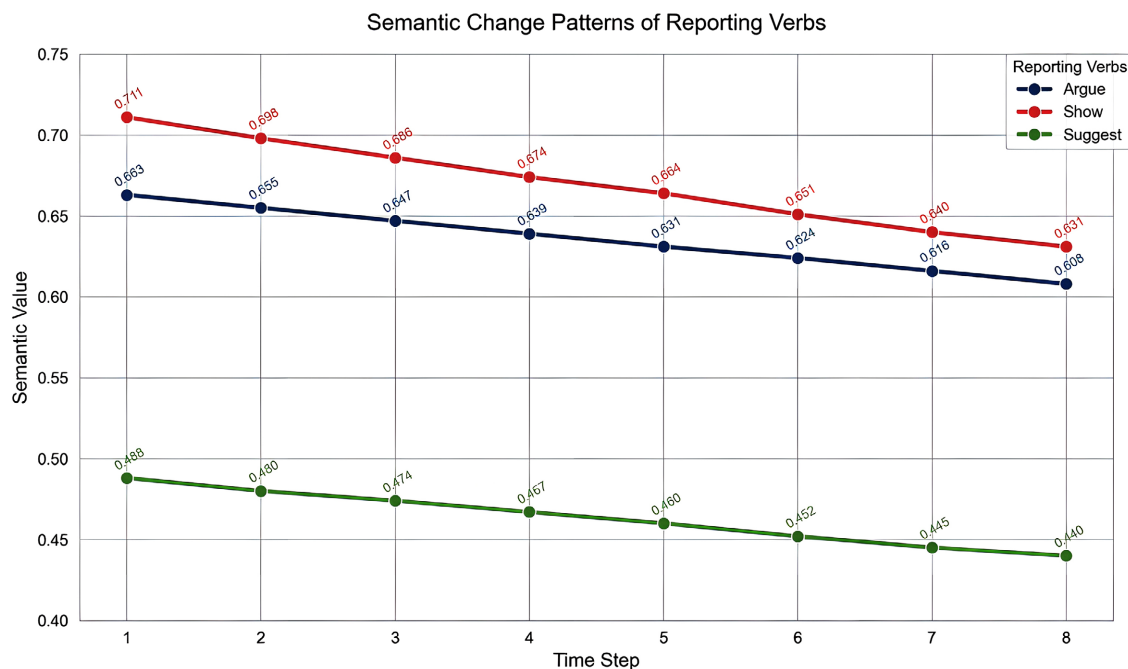


Figure 2. The semantic change patterns of reporting verbs

图 2. 转述动词的语义变化规律

Table 6. The calculation results of semantic changes in reporting verbs
表 6. 转述动词语义变化结果

转述动词	1	2	3	4	5	6	7	8
argue	0.663	0.655	0.647	0.639	0.631	0.624	0.616	0.608
show	0.711	0.698	0.686	0.674	0.664	0.651	0.640	0.631
suggest	0.488	0.480	0.474	0.467	0.460	0.452	0.445	0.440

根据上述结果,不同转述动词的语义相似度衰减速率(b 值)与所属构式的句法复杂度呈负相关。“show”($b = 0.012$)因依赖固定实证模板(SVO 结构),其搭配词语义集中,衰减速率最缓(表 6 中第 8 位相似度仍达 0.631);“suggest”($b = 0.018$)因常介入假设性内容,例如“suggest that...may...”,搭配词多样性更高,衰减速率最快(第 8 位相似度降至 0.440);“argue”($b = 0.015$)介于两者之间,反映其嵌套从句结构需兼容立场标记与逻辑连接词,语义稳定性中等。

综上所述,“argue”等社科类转述动词在社会科学中的语义向量受高介入性立场标记影响,其词向量空间分布离散;“show”等实证类转述动词在自然科学中则因句法角色单一,词向量聚类显著。这表明转述结构的实现本质为概率化系统。换言之,不同学科范式通过调节 λ 、 μ 、 ν 参数(公式 7),策略性分配句法模板与语义资源的调用概率。

4.3. 转述潜势的功能系统

Halliday 指出,及物性系统根据过程类型的选择反映语言使用者对经验世界的认知框架。在转述学术概念时,不同学科的句法构式选择直接体现其认识论范式对知识表征的偏好。人文社会科学论文中,“主体-行为-内容”构式以动作过程(Material Process)为核心,通过显性作者介入和嵌套从句或递归性(Recursiveness)从句构建多声对话网络。例如,“argue”隐含“主张-反驳”的语义链,通过及物性网络扩展概念关系,服务于解释主义对学术争议的人际协商过程,表明写作者对转述内容中暗含的争论内容的反驳或辩证。人文社会科学论文中的转述潜势系统通过高度复杂的句法结构(如多层从句)容纳立场标记与逻辑连接词(例如表 2 中的 however、thus),反映社会科学对辩证推理的依赖。如例(1),动词“found”属于典型的动作过程,用于转述已有研究的结论,并通过被动语态强调研究行为的客观性,同时引出研究结果。例(2)中,动词“presented”和“found”分别表示研究者的行为与发现,构成动作过程的双重嵌套。本句结构采用“研究者(Wang et al.) + 动作动词(presented/found) + 从句内容”的模板,符合人文科学转述结构,强化了学术话语的权威性与说服力。

例(1)Previous research has also **found** that when they lack explicit information about an internet-based source’s accuracy or inaccuracy, young children may be unsure about the source’s capacity to provide accurate information.

例(2)Wang et al. (2019) **presented** 5- to 8-year-old Chinese children with an unspecified internet source and a teacher. They **found** that when confronted with unfamiliar scientific questions, 5- and 6-year-olds did not trust the internet source’s answers over those provided by a teacher or a peer, and even 7- and 8-year-olds trusted a teacher over an internet source.

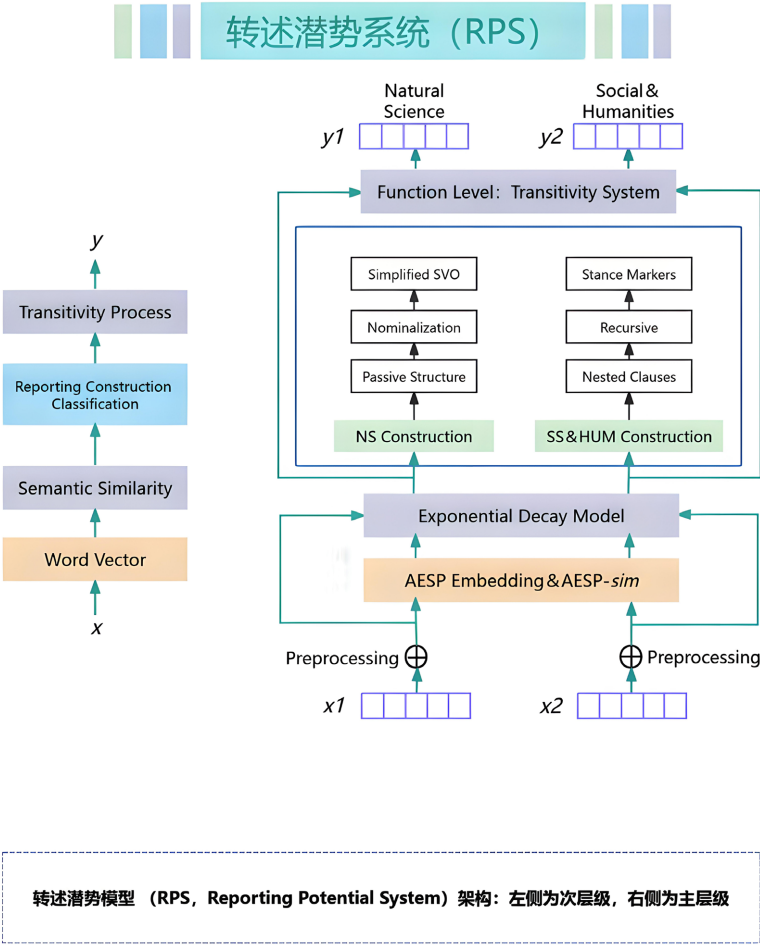
自然科学论文中,“实证现象/推理结果-结论”构式则依赖关系过程(Relational Process),通过非人称化主语和实证性动词直接映射现象与结论。自然科学论文写作风格因此趋向于简化及物性网络,压缩信息密度,较多选择名词化短语、被动结构,或直接转述公式、图片等,强调实证主义对客观性与可验证性的追求。在自然科学论文的转述潜势系统中,句法角色单一、语法复杂度低与语义集中性符合实验

科学线性因果逻辑的表达需求。例(3)中,“demonstrate”属于关系过程,“The data”完全去主体化,突出客观性与可验证性,且本句中出现的被动结构(is associated)、名词化短语(a 27% increased risk)均符合自然科学论文压缩信息,追求高密度语义、低复杂语法的文本特征。

例(3) The data **demonstrate** that the inability to afford medical services is associated with a 27% increased risk of mortality among adults with liver disease.

本节的上述内容研究了转述潜势如何采用及物性系统完成不同学科的知识表征,而前文通过 AESP Embedding 模型与指数衰减模型揭示了转述潜势的概率化生成机制。其一,学科语境显著影响转述潜势系统的句法规则,社会科学中多重嵌套从句的概率显著高于自然科学,表明学科范式通过调节 λ 、 μ 、 ν 参数(公式 7)动态分配语法资源。其二,转述动词的语义稳定性与其句法复杂程度负相关,当转述结构趋向于复杂,则转述动词的语义复杂性会降低,减少学术论文中的表达歧义。其三,转述潜势通过概念功能构建知识网络,不同的学科语境会根据所需引用内容采取不同的及物性过程,目的为更好地完成知识表征。

综上所述,本文提出转述潜势系统(RPS, Reporting Potential System)(见图 3)。系统分为主层级与次层级,主层级为右侧内容,核心逻辑是学科驱动的转述潜势生成流程。首先在输入端选择不同的学科类型,分为“Natural Science”与“Social & Humanities”,分别对应学科认识论范式,即实证主义与解释主义。



搜集语料并划分为 x_1 与 x_2 , 随后进入预处理阶段, 包括清洗(去除噪音)、分词(提取转述动词及搭配词)、语义标注(划定句法角色)。依照本文提出的 AESP 公式, 捕捉转述词的通用语义(调节 λ 参数), 衡量学科特异性(调节 μ 参数)以及引入句法依存向量(调节 ν 参数), 输出对不同学科语境中敏感程度高的转述动词。根据本文提出的指数衰减模型(Exponential Decay Model)量化转述动词的语义衰减速率, 揭示学科差异性。根据以上计算结果, 对语料中的转述结构进行总结, 输入至不同的转述构式当中, 利用其中的通用句法规则撰写论文, 自然科学论文偏好使用简化主谓宾(Simplified SVO), 被动结构(Passive Structure)和名词化(Nominalization), 人文社会科学论文偏好使用嵌套从句(Nested Clauses), 递归性(Recursive)和立场标记词(Stance Markers), 句法规则之间互相联通并循环使用, 以达到组成完整学术表达的效果。最后, 根据句法规则完成的语篇进入功能层级并讨论及物性, 随后输出结果 y_1 和 y_2 , 也就是符合标准的学科性学术论文。

而在次层级中, 显示了通用语料 x 的处理过程。首先使用 AESP 输出转述词的增强语义向量。其次, 利用改进的余弦相似度(AESP-Sim)加权共现频次, 识别高频搭配词。随后, x 进入转述构式归类, 根据计算结果套用对应的转述结构, 输出适配学科规范的句法结构。最后, 参考及物性系统进行功能维度的考察。

综上, RPS 模型的核心逻辑合计三点。其一, 学科范式驱动认识论, “Natural Science”与“Social & Humanities”的范式差异贯穿整个流程。其二, 概率化生成机制, 在考察学术话语时, 通过调节 λ 、 μ 、 ν 参数动态平衡通用性、学科特异性与句法特征, 实现转述潜势的灵活调用。其三, 形式化可验证性, EDM, AESP Embedding, AESP-Sim 提供量化依据, 确保句法规则与功能映射的可解释性。

4.4. 转述潜势的生成机制

转述潜势的生成机制可解构为三个交互层级: 学科规范驱动的语法资源调用、知识表征的功能适配以及概率化句法模板的动态选择, 三者共同作用于学术话语的建构过程。

在学科规范层面, 语法资源的调用机制体现出显著的认识论分野。人文社会科学基于解释主义范式, 其知识生产依赖于作者主体间的立场协商, 因而倾向于采用多声部逻辑连接词、立场性转述动词与嵌套从句的协同使用, 构建对话性论证网络。这种语法选择策略不仅实现了不同学术观点的互动, 更强化了研究过程的辩证维度。相较而言, 自然科学受实证主义范式主导, 其语法资源选择呈现去主体化特征, 具体表现为高密度名词化结构、简化主谓宾框架以及关系过程的优先使用。此类句法策略通过压制作者主体性介入, 将知识表征锚定于可验证的经验事实。

在功能适配层面, 转述潜势的实现机制体现为及物性系统的选择性调用。人文社会科学论文通过动作过程的高频使用激活及物性网络, 转述动词与逻辑连接词的共现不仅传递学术立场, 更重构了学术话语的互文性特征。而自然科学论文则通过关系过程的主导性使用, 压缩及物性网络层级, 将现象与结论直接对应。这种功能适配差异实质上是学科认识论在语言维度上的投射。

在句法实现层面, 概率化模板的选择受制于学科权重参数与句法依存参数的动态调节。学术写作者基于学科认识论框架与修辞目标, 策略性分配语法资源: 人文社会科学侧重句法结构的复杂化以实现多声性, 自然科学则优先采用信息压缩结构以提升论证效度。这种选择机制本质上构成了语法资源的再语境化过程, 即将抽象的学术范式转化为具体的语言实践。

综上, 转述潜势的生成机制本质上是学科认识论通过语法资源再语境化实现的符号实践。这一过程通过句法模板的概率化调用与及物性过程选用, 最终建构起学科话语的认知权威与多声性特质。各层级的交互作用表明, 学术话语的建构既是语法选择的结果, 更是学科规约制约下的策略性实践。

5. 总结

本研究基于跨学科理论框架,系统阐释学术话语中转述潜势(Reporting Potential)的生成机制,揭示学科认识论如何通过句法规则与语义功能的双向映射塑造转述结构。研究提出,转述潜势本质上是由学科认识论驱动的概率化生成系统,其运作机制体现为句法形式化规则与语义资源的协同作用,动态调节通用语言规范、学科语境约束及句法依存关系三者间的互动平衡。通过构建“转述潜势系统”(Reporting Potential System, RPS)理论模型,研究揭示了句法规则与学科范式之间的动态适配规律,为学术话语分析提供了可操作化的分析框架。

需要指出的是,本研究存在以下局限:语料采集范围限于英语学术期刊论文,未涵盖非英语学术语篇及专著、会议论文等非期刊文本类型。这种语料局限性可能导致研究结论在跨语言与跨体裁情境中的适用性受限。后续研究需通过扩展多语种、多模态学术语料库,进一步验证转述潜势系统的理论解释力与跨文化适应性。

参考文献

- [1] Bazerman, C., et al. (2005) Reference Guide to Writing across the Curriculum. Palor Press.
- [2] Hyland, K. (1999) Academic Attribution: Citation and the Construction of Disciplinary Knowledge. *Applied Linguistics*, 20, 341-367. <https://doi.org/10.1093/applin/20.3.341>
- [3] 桂诗春, 冯志伟. 杨惠中, 等. 语料库语言学与中国外语教学[J]. 现代外语, 2010, 33(4): 419-426.
- [4] Chang, Y.Y. and Swales, J.M. (1999) Informal Elements in English Academic Writing: Threats or Opportunities for Advanced Non-Native Speakers. In: Candlin, C.N. and Hyland, K., Eds., *Writing: Texts, Processes and Practices*, Longman, 145-167.
- [5] 唐玮, 林正军. 学术英语写作中引语研究的现状与展望[J]. 外语教学理论与实践, 2022(2): 94-101.
- [6] Hyland, K. (2005) Stance and Engagement: A Model of Interaction in Academic Discourse. *Discourse Studies*, 7, 173-192. <https://doi.org/10.1177/1461445605050365>
- [7] 杨拓. 学术语篇中作者立场表达研究[J]. 黑龙江教育(理论与实践), 2016(3): 17-18.
- [8] 张英, 雷蕾. 学术语篇中“名词 + 补足语”结构立场表达——基于语料库的跨学科研究[J]. 解放军外国语学院学报, 2018, 41(1): 39-47+159-160.
- [9] 李玖, 王建华. 学术写作中的立场表达研究范式述评[J]. 外语教育, 2019(00): 105-114.
- [10] Charles, M. (2006) Phraseological Patterns in Reporting Clauses Used in Citation: A Corpus-Based Study of Theses in Two Disciplines. *English for Specific Purposes*, 25, 310-331. <https://doi.org/10.1016/j.esp.2005.05.003>
- [11] 任明珍. 中外英语语言学类期刊中转述动词对比研究[J]. 考试与评价(大学英语教研版), 2017(4): 115-119.
- [12] 刘永厚, 司显柱. 中外学者学术评价能力对比研究——以国际发表中的英语转述动词为例[J]. 中国外语, 2022, 19(2): 69-77.
- [13] Boyack, K.W., van Eck, N.J., Colavizza, G. and Waltman, L. (2018) Characterizing in-Text Citations in Scientific Articles: A Large-Scale Analysis. *Journal of Informetrics*, 12, 59-73. <https://doi.org/10.1016/j.joi.2017.11.005>
- [14] Casal, J.E. and Lee, J.J. (2019) Syntactic Complexity and Writing Quality in Assessed First-Year L2 Writing. *Journal of Second Language Writing*, 44, 51-62. <https://doi.org/10.1016/j.jslw.2019.03.005>
- [15] 刘锐, 黄启庆, 王珊. 汉语学术语篇转述标记的形式、功能与分布[J]. 当代修辞学, 2021(6): 60-72.
- [16] Biber, D. (1995) Dimensions of Register Variation. Cambridge University Press. <https://doi.org/10.1017/cbo9780511519871>
- [17] Biber, D. (2014) Using Multi-Dimensional Analysis to Explore Cross-Linguistic Universals of Register Variation. *Languages in Contrast*, 14, 7-34. <https://doi.org/10.1075/lic.14.1.02bib>
- [18] Hood, S. (2010) Appraising Research: Evaluation in Academic Writing. Springer.
- [19] Halliday, M.A.K. and Matthiessen, C.M.I.M. (2004) An Introduction to Functional Grammar. 3rd Edition, Hodder Arnold.
- [20] Halliday, M.A.K. and Matthiessen, C.M.I.M. (2014) An Introduction to Functional Grammar. Routledge.

- [21] Halliday, M.A.K. and Matthiessen, C.M.I.M. (2013) *Halliday's Introduction to Functional Grammar*. Routledge. <https://doi.org/10.4324/9780203431269>
- [22] 杨炳钧, 覃朝宪. 系统功能语言学中的元功能思想[J]. 中山大学学报(社会科学版), 2001(1): 47-56.
- [23] Martin, J.R. and White, P.R. (2005) *Language of Evaluation: Appraisal in English*. Palgrave Macmillan.
- [24] Martin, J.R. and White, P.R. (2003) *The Language of Evaluation* (Vol. 2). Palgrave Macmillan.
- [25] Dontcheva-Navratilova, O. (2023) Persuasion in Multimodal Digital Genres: Building Credibility in Video Abstracts. *ESP Today*, **11**, 213-236. <https://doi.org/10.18485/esptoday.2023.11.2.2>
- [26] 梁红艳, 曾蕾. 功能语言学视阈下的投射语言翻译研究[J]. 北京科技大学学报(社会科学版), 2016, 32(4): 12-16+22.
- [27] 邓泳笙. 多维度融合的引文推荐算法[D]: [硕士学位论文]. 广州: 华南理工大学, 2019.
- [28] Kitaev, N. and Klein, D. (2018) Constituency Parsing with a Self-Attentive Encoder. *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Long Papers)*, Melbourne, 15-20 July 2018, 2676-2686. <https://doi.org/10.18653/v1/p18-1249>
- [29] Marín-Arrese, J.I. (2021) Stance, Emotion and Persuasion: Terrorism and the Press. *Journal of Pragmatics*, **177**, 135-148. <https://doi.org/10.1016/j.pragma.2021.01.022>
- [30] 刘旭. 基于 Python 自然语言处理工具包在语料库研究中的运用[J]. 昆明冶金高等专科学校学报, 2015, 31(5): 65-69+93.
- [31] Levy, O. and Goldberg, Y. (2014) Dependency-Based Word Embeddings. *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Short Papers)*, Baltimore, 23-25 June 2014, 302-308. <https://doi.org/10.3115/v1/p14-2050>
- [32] Levy, O. and Goldberg, Y. (2014) Linguistic Regularities in Sparse and Explicit Word Representations. *Proceedings of the Eighteenth Conference on Computational Natural Language Learning*, Ann Arbor, 26-27 June 2014, 171-180. <https://doi.org/10.3115/v1/w14-1618>
- [33] Levy, O. and Goldberg, Y. (2014) Neural Word Embedding as Implicit Matrix Factorization. *Proceedings of the 28th International Conference on Neural Information Processing Systems*, Montreal, 8-13 December 2014, 2177-2185.
- [34] Zhang, Y., Cheng, H., Shen, Z., Liu, X., Wang, Y. and Gao, J. (2023) Pre-Training Multi-Task Contrastive Learning Models for Scientific Literature Understanding. *Findings of the Association for Computational Linguistics: EMNLP 2023*, Singapore, 6-10 December 2023, 12259-12275. <https://doi.org/10.18653/v1/2023.findings-emnlp.820>
- [35] Anthony, L. and Lashkia, G.V. (2003) Mover: A Machine Learning Tool to Assist in the Reading and Writing of Technical Papers. *IEEE Transactions on Professional Communication*, **46**, 185-193. <https://doi.org/10.1109/tpc.2003.816789>
- [36] 饶元, 吴连伟, 王一鸣, 等. 基于语义分析的情感计算技术研究进展[J]. 软件学报, 2018, 29(8): 2397-2426.
- [37] Jurafsky, D. (2000) *Speech and Language Processing*. Prentice Hall.
- [38] Goldberg, A.E. (1995) *Constructions: A Construction Grammar Approach to Argument Structure*. University of Chicago Press.
- [39] Goldberg, A.E. (2006) *Constructions at Work: The Nature of Generalization in Language*. OUP. <https://doi.org/10.1093/acprof:oso/9780199268511.001.0001>
- [40] Huth, J., Brogan, M., Dancik, B., et al. (1994) *Scientific Format and Style: The CBE Manual for Authors, Editors*. Cambridge University Press.
- [41] 熊淑慧. 研究生新手写作者与专家的文献综述写作对比研究[J]. 上海理工大学学报(社会科学版), 2021, 43(3): 228-235.
- [42] 于晖, 宋金戈, 王乐. 大语言模型辅助的功能语篇分析: 理论、方法与实践[J]. 外语电化教学, 2024(5): 43-51+111.
- [43] Huang, Y., Bu, Y., Ding, Y. and Lu, W. (2020) Exploring Direct Citations between Citing Publications. *Journal of Information Science*, **47**, 615-626. <https://doi.org/10.1177/0165551520917654>
- [44] Maton, K. and Doran, Y.J. (2017) Semantic Density: A Translation Device for Revealing Complexity of Knowledge Practices in Discourse, Part 1-Wording. *Onomázein Revista de lingüística, Filología y Traducción*, 46-76.
- [45] Liu, Q. (2022) Text Complexity Analysis of Chinese and Foreign Academic English Writing via Mobile Devices Based on Neural Network and Deep Learning. *Library Hi Tech*, **41**, 1317-1332. <https://doi.org/10.1108/lht-11-2021-0383>
- [46] Shen, C., Guo, J., Shi, P., Qu, S. and Tian, J. (2023) A Corpus-Based Comparison of Syntactic Complexity in Academic Writing of L1 and L2 English Students across Years and Disciplines. *PLOS ONE*, **18**, e0292688. <https://doi.org/10.1371/journal.pone.0292688>
- [47] Uba, S.Y. (2019) Semantic Categories of Reporting Verbs across Four Disciplines in Research Articles. *English*

- Language Teaching*, **13**, 89-98. <https://doi.org/10.5539/elt.v13n1p89>
- [48] Yang, L. (2013) Evaluative Functions of Reporting Evidentials in English Research Articles of Applied Linguistics. *Open Journal of Modern Linguistics*, **3**, 119-126. <https://doi.org/10.4236/ojml.2013.32016>
- [49] Flowerdew, L. and Petrić, B. (2024) Corpus, Concordance, Collocation: How Corpus Linguistics Can Inform Academic Writing Pedagogy. *English for Specific Purposes*.
- [50] Varttala, T. (2002) Hedging in Scientific Research Articles: A Cross-Disciplinary Study. *Domain-Specific English: Textual Practices Across Communities and Classrooms*, Peter Lang, 141-174.
- [51] Martin, J.R. (1992) *English Text*. John Benjamins Publishing Company. <https://doi.org/10.1075/z.59>
- [52] Martin, J.R. and Rose, D. (2008) *Genre Relations: Mapping Culture*. University of Toronto Press, 181-198.
- [53] Garg, N., Schiebinger, L., Jurafsky, D. and Zou, J. (2018) Word Embeddings Quantify 100 Years of Gender and Ethnic Stereotypes. *Proceedings of the National Academy of Sciences*, **115**, E3635-E3644. <https://doi.org/10.1073/pnas.1720347115>
- [54] Mikolov, T., Sutskever, I., Chen, K., et al. (2013) Distributed Representations of Words and Phrases and Their Compositionality. *Advances in Neural Information Processing Systems*, **26**, 3111-3119.
- [55] 吴侠, 邵斌, 王贵. 构式化视角下英语临界复杂介词的固化研究[J]. *现代外语*, 2024, 47(3): 316-330.
- [56] Michaelis, L.A. and Lambrecht, K. (1996) Toward a Construction-Based Theory of Language Function: The Case of Nominal Extraposition. *Language*, **72**, 215-247. <https://doi.org/10.2307/416650>
- [57] Booij, G. (2009) Compounding and Construction Morphology. In: Lieber, R. and Stekauer, P., Eds., *The Oxford Handbook of Compounding*, Oxford University Press, 201-216.
- [58] Joos, M. (1950) Description of Language Design. *The Journal of the Acoustical Society of America*, **22**, 701-707. <https://doi.org/10.1121/1.1906674>
- [59] 冯志伟. 计算语言学方法研究[M]. 上海: 上海外语教育出版社, 2023: 701.