本研究以语体分析为核心，围绕"剽窃的界定与态势""跨国视域与国内案例"以及"剽窃认定的路径"三个方面展开探讨。本文以 2003 年《法律语言学研究》中"语体分析"章节为基础进行重修，旨在为相关研究提供参考。

剽窃认定的语体分析

刘蔚铭

2026/9/9

摘　要：随着学术传播与数字媒介的变革，剽窃行为愈发隐蔽。本文从法律语言学视角出发，分析剽窃认定中的司法困境与认定难点。在此基础上，从多个关键要素对文本进行综合考量，提出“四层分析维度”语体分析体系，将主观的“抄袭感”转化为客观的语言证据，为剽窃认定提供科学、可操作的方法论路径。

关键词：剽窃认定；法律语言学；语体分析；四层分析维度

一、剽窃的界定与态势

剽窃与作者鉴别虽同属可疑文本的作者身份问题，但二者性质迥异。作者鉴别旨在通过语体分析，提取稳定的语言特征，以判定可疑文本的作者归属；而剽窃则涉及对他人作品的非法侵占，兼具法律违规与学术不端的双重属性。根据英国法律语言学研究所（Forensic Linguistics Institute）的界定：Plagiarism is the deliberate theft of the work of another without attempt to acknowledge their text as a source——剽窃是指故意窃取他人作品，且未试图注明出处的行为。基于此，剽窃的本质即是一种蓄意的“侵占”行为——无论手段是原封不动照搬，还是改头换面据为己有，只要未标明出处，即构成剽窃。法律语言学中的语体分析，正是针对此类可疑文本展开系统鉴别，穿透文本表层伪装，揭示隐蔽的侵占行为，从而为剽窃认定提供客观的语言证据。

剽窃是一种古今中外都有的普遍社会现象，其蔓延已渗透至社会各阶层与多领域。在学术

与科研领域，论文抄袭、专利侵权及数据窃取事件屡禁不止，涉及从博导教授到基层研究人员的广泛群体；在文化与创意产业，影视剧本、音乐作品及文学创作中的权属争议频发，折射出原创保护的困境；在科技与医药领域，核心算法、临床试验数据乃至传统中药秘方被窃取或流失海外的案例，更凸显了知识产权保护的严峻性；而在数字时代，IT 行业的代码复制、网络内容抓取等新型剽窃行为层出不穷。值得注意的是，此类行为甚至已下沉至基础教育阶段，中学生作文抄袭现象亦非个案。这种全领域的泛化趋势，使得剽窃的精准认定成为亟待解决的现实课题。

剽窃行为的泛化并非我国独有，国际学界亦面临严峻挑战。例如，澳大利亚莫纳什大学（Monash University）前副校长因早年学术生涯中的剽窃丑闻被迫辞职；即便是诺贝尔文学奖得主亦未能豁免，1989 年获奖者卡米洛·何塞·塞拉（Camilo José Cela）晚年便深陷抄袭争议——其小说《圣安德烈斯的十字架》（The Cross of Saint Andrew）被法院裁定抄袭女作家卡门·福莫索（Carmen Formoso）的作品《卡门、卡梅拉、卡米尼亚》（Carmen, Carmela, Carmiña）。上述案例表明，剽窃行为具有跨越国界与阶层的普遍性，且往往具有高度的隐蔽性。这进一步凸显了借助语体分析等技术手段，对可疑文本进行客观、精准鉴别的迫切性与必要性。

二、跨国视域与国内案例

剽窃认定的复杂性，不仅源于其法律与道德属性的交织，更在于其形态随学术传播与技术变革不断演化。为此，本章通过梳理具有代表性的跨国与国内案例，呈现不同语境下剽窃行为的表现形式与认定困境，进而揭示：单纯依赖法律条文或量化比例，往往难以应对日益隐蔽的文本挪用，而语体分析正是破解此类难题的关键路径。

（一）跨国语境下的剽窃认定

剽窃认定在国际学术语境中同样面临复杂挑战。剽窃行为并不局限于当代语境，亦跨越国界与时间维度，甚至侵入经典与权威文本领域。据人民网 2003 年 9 月 3 日报道，我国儿童文学作家邱勋在梳理美国儿童文学丛书《少男少女丛书》（The Bookshelf for Boys & Girls）的源流时发现，该书系中若干文本与蒲松龄《聊斋志异》相应篇目存在高度同构。该书由美国大学学会出版公司（The University Society, Inc.）出版，畅销近百年；其中，该丛书第三卷《童话故事卷》（Folk and Fairy Tales）中的《奇妙的梨树》（The Wonderful Pear Tree）无论从故事框架、人物设定、主要情节还是细节描写，与蒲松龄《种梨》几乎如出一辙——唯一的改动是将原作者蒲松龄的名字替换为 Frances Carpenter。另一篇《镜中少女》（The Maid in the Mirror）则更为隐蔽：它并未整篇照抄，而是截取《凤仙》的关键情节独立成篇，且连人物名字也未作更改——男名保留“Lu（刘）”，女名直书“Feng Hsien（凤仙）”，直接暴露出其与文言小说人物谱系之间的源流关系。

上述案例说明：跨文化转写、署名替换与篇目拆并，虽可制造“异文化新作”的表象，但叙事母题、情节序列、人名语音及文体痕迹仍会留存源文本指纹；这也正是语体分析在跨国剽窃认定中发挥作用的前提。

无独有偶，剽窃行文亦出现在西方学术传记写作中。苏格兰历史学家、传记作家詹姆斯·麦凯（James Mackay）以罗伯特·彭斯（Robert Burns）研究闻名，但其 1997 年出版的《亚历山大·格拉汉姆·贝尔》（Alexander Graham Bell）却使其陷入严重的学术诚信危机。指控显示，麦凯不仅大段挪用他人研究成果，甚至连篇首作者献辞亦原样抄袭。更有甚者，麦凯的多部著作亦被证实存在大范围文本重用。波士顿大学荣誉教授、贝尔研究领域的同行学者罗伯特·布鲁斯（Robert Bruce）在系统比对后提出严厉指控，指出麦凯在撰写贝尔传记时，剽窃了其本人及多位前人的研究成果与文献材料。

针对这一指控，法律语言学专家马尔科姆·库尔萨德（Malcolm Coulthard）介入调查，对相关作品段落进行了缜密的文本分析与检验，最终以确凿的语言证据证实了剽窃事实。这一

案例再次印证，剽窃者手段如何隐蔽，语体分析终将让谎言现形。随着剽窃事实的确认，麦凯的学术声誉遭受重创，曾经备受尊敬的历史学家，声誉从此一落千丈，其作家生涯实质上也由此画上了句号。

该案例表明：剽窃并不因作者“专业”而自动消失；正相反，专业作者的文本更具稳定的语体习惯，一旦发生跨文本搬运，便会在献辞、称谓、引证方式、句法节奏与材料组织等层面留下可识别痕迹。

（二）国内典型案例分析

反观国内，学界的相关案例，则进一步揭示了司法实践中的认定困境，围绕文本权属的学术争议亦屡见不鲜，其中部分案件因情节恶劣、影响广泛，已成为学术诚信治理中的标志性案例。这些案件不仅折射出学术诚信危机，也凸显了借助专业手段甄别剽窃、捍卫学术规范的艰难与紧迫。下文仅以语言学界两起具有代表性的案件为例，便足以窥见剽窃现象的普遍性与严重性。

1. 方言辞书著作权纠纷案

据《经济参考报》1999年6月12日报道，一起发生于1993年5月的语言学著作纠纷案，曾引发广泛的社会关注。案件双方均为方言研究者：原告郭建荣与被告温端政之间存在师承关系，温端政为郭建荣的业师。

20世纪80年代，温端政出版了方言研究专著《太原方言词汇》；此后，郭建荣发表了《太原方言民俗词语》一文。部分学者在《语文研究》上撰文指出，郭文与温著高度重合，涉嫌文本抄袭，由此双方陷入长期的学术与法律争议。1993年5月，郭建荣向太原市中级人民法院提起诉讼，指控温端政及《语文研究》编辑部侵犯其著作权与名誉权；温端政与编辑部则提起反诉，主张郭文抄袭温著，侵犯了温端政的著作权，同时指称郭建荣对外散发的相关材料侵害其名誉权。该案由此形成双向指控、互为原被告的复杂诉讼格局。

就“抄袭”争议，法院先后两次委托国家版权局进行专业鉴定。鉴定结论明确：温端政专著出版在先，郭建荣论文发表在后，且二者在词条内容与编排结构上存在系统性重合。量化比对结果显示，郭建荣论文收录的词条中，有84.09%在文字表述或结构安排上与温著完全相同或高度相似。基于上述客观事实，一审法院采纳鉴定意见，认定郭建荣构成著作权侵权，判令其公开赔礼道歉，并酌情赔偿损失10000元。

然而，案件并未就此尘埃落定。双方均不服一审判决，案件进入山西省高级人民法院二审程序。终审裁判出现了戏剧性的反转：法院以郭建荣论文的抄袭比例“未达到法定10%的界限”为由，撤销了一审判决，改判郭建荣不构成抄袭。

二审的改判依据，源于特定的司法实践惯例。尽管当时的《著作权法》中并无“抄袭量”的具体规定，但审判实践常参照20世纪80年代颁布的《图书期刊保护试行条例实施细则》中“以10%作为抄袭界限”的标准。由于郭建荣的文本未超过该比例阈值，终审判决不仅推翻了一审对其侵权的认定，还认定温端政与《语文研究》编辑部的行为侵害了郭建荣名誉权，判令二者公开赔礼道歉，并在《语文研究》上刊登声明，为郭建荣恢复名誉。

为确证事实真相，山西省方言学会委托六所高校的八位语言学博士生导师，对郭建荣的论文是否构成抄袭进行了客观的复核鉴定。鉴定结论明确指出：郭建荣论文的抄袭比例超过80%，已构成实质性剽窃。尤为关键的是，温著中的若干明显讹误，在郭文中亦被原样复制。这一事实从根本上排除了“独立创作”的可能性，为剽窃认定提供了具有高度专业说服力的文本证据。

这一专家鉴定结论，不仅厘清了案件的事实争议，也凸显出此类纠纷背后的制度性隐忧。它表明，剽窃并非孤立的道德失范行为，而是会触发学术评价、法律适用与社会信任等多重连锁反应。此类行为不仅侵蚀学术公平，也暴露出法律在认定标准上的模糊地带，其负面影响具有长期性与扩散性。因此，学术剽窃的认定，不应止步于机械的比例计算，而必须建立在严谨的语体分析与学术逻辑之上，以维护学术诚信的根本底线。

2. 电子辞书侵权纠纷案

另一类更具隐蔽性与复杂性的侵权纠纷，则频发于外语辞书与数字化语言工具的交叉领域。据《北京晨报》等诸多媒体报道，北京外国语大学教师编纂的《现代法汉汉法词典》于 1994 年 8 月首次出版。其后，编者发现某电脑公司推出的“文曲星”品牌电子辞典，其法汉与汉法模块的内容与《现代法汉汉法词典》存在高度重合，涉嫌著作权侵权。为维护自身合法权益，编者遂以侵犯著作权为由，将生产商诉至法院，请求判令停止侵权、消除影响并赔偿损失。

本案庭审的核心争议，在于如何界定辞书作品的“独创性”边界。被告辩称，其电子辞书并非对《现代法汉汉法词典》的简单复制，而是在原有基础上新增六万余条词目，并集成检索、发音、分类等多项功能，体现了足以构成新作品的独创性表达，应依法享有独立的著作权。

然而，本案最终未进入实体判决阶段。2002 年底，双方经协商达成著作权许可使用协议，原告撤回起诉，被告公开致歉并支付许可使用费。这一结果虽未通过司法裁判明确侵权边界，却以一种务实的方式化解了纠纷，既保障了原创者的合法权益，也为数字时代辞书类作品的知识产权纠纷提供了更具弹性的解决路径。

上述典型案例表明，随着学术研究边界的跨学科化、国际化以及传播媒介的数字化变革，剽窃行为的内涵与表现形式日趋多元、隐蔽与复杂。面对不断演化的抄袭行为，司法裁判自不可缺席；然而，如何在法律的形式界定与学术的内在逻辑之间寻求平衡，实现对各类隐蔽侵权行为的精准甄别与有效规制，已成为维护学术诚信与知识产权秩序的关键所在。

三、语体分析：剽窃认定的路径

面对上述复杂且争议频发的剽窃认定难题，法律语言学的核心任务，在于对可疑文本展开系统的语体分析（stylistic analysis）。该方法并非仅限于字数与词汇重合率的表层统计，而

是立足于语言使用的深层规律，从词汇偏好、句法结构、修辞模式及语域特征等多维视角，系统评估文本间的实质性相似度。无论是隐蔽的词汇替换，还是系统性的语体重构，科学客观的剽窃鉴定都必须建立在可量化、可复现的判定维度之上。基于此，本部分将通过文本预处理与综合判定的协同运作，从四层维度构建系统与可复现的司法鉴定框架，为司法实践提供可靠的语言学证据。

（一）文本获取与预处理

语体分析的首要环节，在于为后续比对建立可比性基础，这类似于精密实验前的“备料”阶段。此阶段需系统获取两类文本：可疑文本（即待鉴定文本）与比对文本（即已知来源文本）。为确保分析结果的科学性与可比性，须对两类文本进行标准化预处理，使其在文本编码、段落划分及标点使用等方面保持严格一致，以避免因格式差异导致的误判，确保鉴定过程的证据效力。

在此基础上，需对文本进行精细化预处理，以消除非语体性干扰因素。该环节主要包括：剔除标点符号等表层形式要素、统一字母大小写、准确切分段落边界。针对多语言、方言或特殊字符文本，还需实施严格的字符标准化，确保文本素材的格式一致性。唯有确保预处理过程的系统性与规范性，方能保障后续语体特征分析的准确性。

（二）四层分析维度

1. 词汇维度

词汇分析是语体识别的基石，其核心任务在于揭示文本在微观语言选择上是否呈现高度一致性。具体而言，需比对可疑文本与比对文本在词目选择、词频分布、专业术语密度及惯用搭配等维度是否存在系统性重合。在剽窃鉴定中，若可疑文本与对比文本的词汇重合度异常之高，且词条搭配模式高度一致，即便剽窃者进行了少量同义词替换，亦难以掩盖抄袭本质，构成剽窃认定的关键语言证据。反之，若可疑文本构建了与原文截然不同的词汇体系，且对既有文本的吸收比例极低、功能明确，则更符合合理引用的范畴。

2. 句法维度

句法分析主要考察文本在句法结构、句式复杂度、语序模式及语态选择等方面的系统性特征，是揭示深层文本依赖、穿透表层改写的关键维度。若可疑文本不仅在词汇层面存在借用，更在深层结构上系统性保留了比对文本的句法框架与嵌套逻辑，即便辅以同义替换、语序调整或局部删减，其高度一致的句法组织模式仍可作为认定实质性相似乃至剽窃的重要依据。

3. 语篇维度

语篇分析主要考察文本的宏观系统性特征，包括段落架构、论证逻辑及信息推进模式，旨在超越表层句法结构，揭示深层文本依赖关系。具体可从以下三个维度展开：首先考察篇章结构，通过剖析段落间的递进、转折及衔接关系，判断其宏观组织模式是否与比对文本存在系统性重合；其次聚焦论证链条，重点检验论点与论据的排列次序、推导逻辑及论证强度，以识别是否存在论证结构的整体性复制；最后分析信息结构，通过对比主位（Theme）与述位（Rheme）的分布特征及信息焦点的定位，判断微观信息组织模式是否雷同。上述从宏观篇章到微观信息的系统性比对，能够有效穿透表层改写，揭示具有高度隐蔽性的深层剽窃行为，为司法认定提供坚实的语言学依据。

4. 风格维度

风格分析的核心，在于识别特定作者或文本来源所独有的“风格指纹”，即那些具有高度个人化特征的语言习惯。这种个体性语言标记，具体体现于修辞手法、情态表达、语域选择及标点偏好等可观测的微观特征之中。可疑文本与比对文本的系统性重合度，构成了区分剽窃与合理使用的核心判据，尤其在应对高度隐蔽的改写型剽窃时，风格分析往往具有一锤定音的作用。在实务中，可通过语料库检索、词频统计、可读性公式及机器学习分类器等技术手段，对风格指纹进行量化提取与比对，为鉴定结论提供可复现的数据支撑。

从四层维度的内在关系来看，风格分析最具综合性与统摄力。当剽窃者试图通过局部改写规避痕迹时，往往会在词汇和句法层面进行同义替换或句式重组，但其风格特征仍与比对文本保持高度一致。对此，需特别防范这种“改头换面”式的隐蔽剽窃行为。正因深层风格具有难以伪装的高度稳定性，风格层面的比对不仅是语体分析中最具综合性的维度，更是判定隐蔽抄袭、揭示文本真实来源的关键锚点。至此，词汇、句法、语篇与风格四层维度相互支撑，共同

构成了一个从表层到深层、从可伪装到不可伪装的完整鉴定体系。

（三）综合判定与报告撰写

语体分析的科学性，核心在于四层维度的协同运作与深度互证。该框架不仅为法律语言学鉴定提供了一套严密且具可操作性的方法论工具，也为综合判定提供了坚实的逻辑支撑。应当明确，任何单一维度的审视均可能导致以偏概全的误判；只有将词汇、句法、语篇与风格四个层面进行综合考量与交叉印证，才能得出经得起质证与检验的鉴定结论。在此基础上，鉴定人需将四层维度的分析结果系统整合，形成结构清晰、论据扎实的司法鉴定报告，为法庭提供可直接采信的语言学证据。

在此基础上形成的语体分析鉴定报告，必须具备高度的专业性与透明度，并确保分析过程的可追溯性与结论的可复现性。最终的鉴定报告不仅应系统呈现从文本预处理到四层维度分析的全流程，还需辅以量化的统计数据与可视化的对比图表，最终形成基于证据链的明确综合结论，为司法审判与学术裁决提供可直接采信的专业依据。这也正是语体分析从理论建构走向实践应用的核心价值所在。

在司法实践中，法律语言学家通常作为专家证人出庭，依托四层维度的系统性分析结论出具专业意见。其核心职责在于将高度专业的语体分析转化为清晰、可被采信的鉴定结论，帮助法庭厘清文本间的渊源与复制关系，从而为公正裁决提供核心科学证据。这一过程对分析者提出了跨学科的双重素养要求：既需具备系统的语言学理论素养与语体分析实践能力，又必须深刻理解司法程序对证据可采性与证明力的严苛标准。从本质上看，语体分析正是通过严谨的科学路径，将主观层面的“抄袭感”彻底转化为客观化、可量化且可复现的司法认定依据，为构建公平、透明的学术与知识产权秩序提供坚实的语言学支撑。

四、结语

基于上述分析，剽窃已不再仅仅是单纯的道德失范，而是演变为渗透于生活方方面面、挑战法律边界的复杂社会现象，同时也是一种跨领域的学术不端行为。面对隐蔽抄袭，机械的量化标准往往力有不逮，而语体分析则为司法实践提供了科学、可操作的方法论工具。随着自然

语言处理等技术的不断进步，语体分析在可疑文本鉴别领域将发挥更为精准、高效且可复现的作用。剽窃的认定，从来不只是数字的“游戏”——它关乎语言使用的深层规律、司法认定的标准，以及学术诚信的底线。

本文提出的“四层维度”语体分析体系，通过词汇、句法、语篇与风格的协同分析与交叉比对，将主观的“抄袭感”转化为客观、可量化、可复现的司法判定标准。这不仅为司法实践提供了坚实的语言证据，也为维护学术公正与知识产权保护奠定了科学基础。维护原创尊严，不仅依赖于司法裁判的公正，更需依靠科学严谨的语体分析；唯有二者相辅相成，方能真正净化学术生态，捍卫学术与法律的尊严。

文献参照：

刘蔚铭：《法律语言学研究》，中国经济出版社，2003 年 10 月，162-167.

文献资源：

学术网站：法律语言学研究 (http://www.flrchina.com)

微信公众号：法律语学 (forensicling)

微信视频号：微明檐影