

机构 服务 会员 文学奖项 理论评论 文史 科幻 书汇 新作品 文艺报  
新闻 访谈 投稿 直播 艺术 世界文坛 民族文艺 网络文学 儿童文学 鲁迅文学院

中国作家协会 >> 理论评论 >> 理论热点

# 文学大数据与当代文学研究范式转型

来源：《当代作家评论》 | 杨丹丹 2022年02月23日08:17

分享到：

关键词：[文学大数据](#)

中国当代文学研究范式转型是一个常谈常新的话题。从当代文学发生之日起，学术界就面临如何阐释当代文学的问题，其中就涉及文学研究的方法和范式等内容。而且，随着文学社会语境、研究对象和诉求的变化，文学研究范式转型问题也越发突出和明显，通常显现在文学研究方法论争中，通过论争确立某一历史阶段通约有效的研究方法，并在长期实践中形成主流文学研究范式。从这一角度看，当代文学研究范式转型与当代文学演进和当代文学研究史变迁紧密关联。

当代文学与古代文学、近现代文学的本质差异之一是它的“当代”属性。“当代”的繁复多变和时刻在场，使当代文学研究需要直面最新的文学现场，成为“当前文学思潮、作品和现象最‘理想’的‘批评者’”，并运用“认同式”“讨论式”批评（1）契合当代文学的“当代性”。这预设了当代文学研究范式的不稳定，没有任何一种长久不变的研究范式能够有效应对不断新变的文学现场。这也使当代文学研究显现出强烈的危机感。当无法“对崭新的文学实践做出辨认、预判，疾言厉色或为之鼓与呼”时，（2）势必对已有研究范式做出修正和调整，甚至进行颠覆和重构，寻找新的研究范式替代。只有如此，才能保证文学研究的有效阐释力和公信力。因此，当代文学研究本身就包含对自身转型问题的持续探索。

从中国当代文学研究史来看，大体形成了文学政治研究、文学审美研究、文学文化研究和文学社会研究几种主流研究范式。它们与特定历史语境和时代诉求密切相关。大致对应20世纪50—70年代、80年代、90年代和2000年以来几个历史阶段。每种研究范式在相应的历史阶段都具有优先于其他研究范式的阐释权且处于流转状态。这说明任何一种文学研究范式都有其边界和限度，当新的社会语境、新的文学现象出现时，必然产生新的研究范式。或者说，文学研究范式转型问题是当代文学研究史自身生发的问题，一种新的文学研究范式的形成过程也是其他文学研究

## 大家都在看

- “新时代山乡巨变创作计划”引发热烈反…
- 中国当代文学信息库系统维护中
- 中国报告文学学会关于更新会员档案的通知
- 宅猪新作《择日飞升》：当网络作家关注…
- 网络公开课：文学作品改编为影视作品常见…
- 中国作家协会举办首次中国-拉美国国家文学…
- 中国作家协会党组理论学习中心组专题学…
- 2022年文学志愿服务示范性重点扶持项目…

## 编辑推荐

- 文学志愿服务示范性重点扶持项目公布
- 北京师范大学、鲁迅文学院联办研究生班…
- 古籍出版春风浩荡满目新
- 碎片化阅读的时代，你…
- 【独家】人物 | 石一枫：海胆，翻了个面
- 重读《狂人日记》：再… 补矣”
- 刘洋：科幻与科技的融合、超越与区别
- “首届扬子江网络文学最具IP潜力榜”发布



回顶部

范式的式微过程。可以说，“学术范式的转换和形成是一个相当复杂的过程，一方面出自社会文化转型的需要，另一方面有学术本身的发展特点”。（3）

因此，想辨识清楚当代文学研究范式转型问题，就需要阐明与此相关的文学新语境和新现象，以及原有文学研究范式的限度和新的文学研究范式的适用性问题。在此意义上，我们需要解释清楚文学研究在当下发生了怎样的新变化，文学研究方法和思维发生了怎样的新转变，出现了哪些新的文学研究实践路径，生产了哪些新知识等问题。

## 一、大数据时代与文学大数据

1980年，美国社会学家阿尔文·托夫勒在著作《第三次浪潮》中阐述了科技发展从农业社会到工业社会转型过程中起到的关键作用，并且预测电子科技进步将引发“第三次浪潮文明”：人类将由工业社会进入信息社会，“整个文明和制度、技术、文化都在变化的浪潮下日趋分裂”。（4）尤其是大数据技术在日常生活诸多方面的广泛应用，将推动大数据社会的形成。当下，随着互联网、物联网和社交网络的快速发展，阿尔文·托夫勒的预测开始变为现实，“根据国际权威机构Statista的统计和预测，2020年全球数据产生量预计达到47ZB，而到2035年，这一数字将达到2142ZB，全球数据量即将迎来更大规模的爆发”。（5）

事实上，系统认识和理解大数据最先从学术界开始。《自然》《科学》《ERCIM News》等业界顶尖期刊都曾开设“大数据专辑”，对数据的维护、再利用，数据的可解释性和数据处理主体等问题进行探讨。虽然，学术界的讨论主要集中在学理层面，但都表明一个事实：大数据可能成为引领未来社会发展的支配性力量。而且，随着大数据理论不断成熟，大数据讨论开始溢出学理层面，转而在社会实践应用维度发力，进而上升为国家发展战略。中国、美国和欧盟等国家/地区纷纷制定了大数据发展政策。中国早在2014年就将“大数据”写入政府工作报告，明确实施国家大数据战略。更为重要的是，“一场由大数据带来的技术革命”和“智能革命”（6）开始重塑现实世界，渗入日常生活的各个角落，“从商业科技到医疗、政府、教育、经济、人文以及社会的其他各个领域”（7）都有大数据的魅影。人们可以直接感受到自己生活在一个被数字编织的数据网中。那么，在大数据时代，数据技术是如何改变文学的？产生的文学大数据呈现出哪些新的特征？文学研究方法出现哪些变化？

回答上述问题的前提是首先明确“文学大数据”的概念。目前为止，学术界对“大数据”的概念仍存在较大分歧，麦肯锡、维克托·迈尔·舍恩伯格和肯尼思·库克耶等学者都曾界定过大数据，但都没有形成共识。（8）所以，笔者只能从现有概念中提取大数据的一些基本特征：数据规模大，数据来源、种类多，数据增长、处理速度快，数据价值密度低等。（9）这些特征决定了大数据的“大”不仅是数据数量的大，更是指数据的复杂多变，以及通过数据获取新知识的能力。以此为标准，可以为“文学大数据”暂时确立一个相对稳固的概念：文学大数据与传统意义上的文学数据不同，不是指文学作品和文学文献电子化后产生的电子文本数据，而是指文学作品在互联网、物联网和社交网传播过程中产生的与文学相关的数据，包括日志、音频、视频、图片、弹幕、点赞等结构化、半结构化和非结构化多种类型



总第4875期

## 专题

更多



庆祝中国共产党成立100周年



网络文学名家谈写作

## 访谈

更多



新乡土、创意写作与80后的文学症候

“我生而为人，自然对人好奇，对人所安身立命的这个世界也很好奇。文学是我表达“好奇”的方式。”

## 文史

更多



重读《狂人日记》：  
再引 侯补



白话正文分属不同语言世界，但一紧密相

回顶部

大众文化与晚清《鲁滨孙》 中国化

朱自清的“诗”与“思”：从五四到后五四

穆涛：子不语

数据。这些数据数量巨大，来源、种类繁多，成碎片化分布，且处于实时更新状态，有价值 and 可利用的数据密度较低，需要利用计算技术分析这些数据。例如，余华的小说《文城》出版后，在微信、微博等社交媒体和豆瓣读书、京东图书等网络购书平台，以及抖音、快手等短视频媒体中迅速传播，产生了数量庞大的文学数据，包括读者截取的小说原文段落、长篇专业评论、简短的即时感受文字、表达阅读情绪的图片、讲述阅读体会的视频、购书推荐评语、销售数字表格、小说影响力排行表和小说的知识图谱等。从这些数据中可以分析出读者对小说的总体评价、专业批评家与普通读者的接受差异、读者的地理信息、小说在不同地域的文化影响力、当下社会情绪的走势、读者的审美变化和小说的未来发展趋势等内容。而且，不同类型数据之间的重组又可以产生新数据。可以说，当《文城》进入互联网和社交网被数据化之后，就不再是一部单纯的文学作品，转而成为生产文学数据及其相关新知识的载体。

文学大数据既是纸媒文学在互联网和社交网传播的结果，又是依托网络、利用数字技术生产数字化文学的产物。这些“数字文学”（10）在利用文学大数据的同时，又再次产生了文学大数据。人工智能文学、超文本文学、视觉文学和界面文学等“数字文学”的诞生不仅是新旧文学形态的更迭，而且是文学写作生态的全方位变革。它的“非线性叙事、非稳定结构、读者参与性以及多媒体特征被视为对传统文学理论以及文学审美的挑战”，同时，其特有的“编程功能、数字代码、计算机算法等数字文学内部运行机制被视为新美学基础和美学特征”。（11）可以说，这些新的文学特性都与文学大数据紧密关联。例如，超文本文学就超越了传统语义文本的限制，把词语、地理位置和个体信息等众多主题元素按照多种组合方式重组，甚至“任何可以看见、可以听见的都可以构成文本的纹理”。（12）而且，写作主体既可以是人，也可以是智能机器。进一步而言，超文本形成的基础是文学大数据，通过分析和处理文学大数据，把文学大数据存储在各个文字节点上。然后，在这些文字节点之间确立相关关系，再以链接的方式把各个文字节点串联起来，形成一种网状结构。同时，超文本在互联网和社交网传播过程又生产了新的文学数据，并无限循环往复。林森的小说《白毛女在1971》就是一部典型的超文本小说，作家在文本中设置了多个链接符号，这些链接符号与故事情节叙述密切相关，读者只有阅读完这些链接网页的内容后，才能继续阅读小说剩余的内容。（13）

对当代文学研究而言，面对文学大数据及新的文学样态，需要认真思考如何调整和重塑文学研究方法，与此新变相契合，进而建构新的研究范式。因此，“如何在设计算法、挖掘工具、可视化工具及存档技术时凸显价值、解释和意义变得更加重要”。（14）

## 二、文学数据研究中“小数据”与“大数据”

为了应对文学的数字化和数据化，文学研究引入了“计算”方法，把信息技术和计算机技术应用到文学研究。文学计算的两大核心是：文学数据和计算技术。尤其是计算技术至关重要，它决定了文学数据研究的上限。因为，文学数据研究“具有高度技术化的内涵与本质，它必须依托数字系统建设、程序代码编写或者其他手段来实现研究过程或呈现研究结果”。（15）现阶段，这些计算技术包括可视化与

一名特殊党员的代表证

盛唐诗歌的庄严气象

### 原创推荐

更多



散文 | 读莲

读莲，仿佛看到漫天飞升的莲花向我飘来，向我们飘来。莲花的清香，莲花的精灵，弥漫在人间。莲花天生丽质，自香透骨，让躁动的魂灵沾染莲开的香韵，让我们从沉迷中醒来，安静的行走。



散文 | 月光之夜

在月光的世界里，良田曲线被月光一路点缀，让良田变得十分美丽。因为随着月光的变幻，游离在月光世界里的曲线美，却把良田带进了幸福的世界。这种曲线美不断延伸的结果，却代表了良



散文 | 中国作家网散文参赛作品：…

家军是个身负使命的作家，其使命就在于他不断地精神还乡。生斯养斯的那片神奇的土地常使家军眼中饱含泪水。而对家乡的深情回望也是他作品一贯的题材。家军对乡土总是怀着浓烈的感情



散文 | 中国作家网散文参赛作品…

祖国，地大物博，山川秀美；阿拉善，多姿多彩，锦上添花。



诗歌 | 雪山草地间（组诗）

一组关于大渡河、泸定桥及藏地风情的诗歌



回顶部

数据设计技术，文本分析、聚合与数据挖掘技术，地理信息系统技术和仿真技术等，这些技术的大规模开发和普遍应用推进了文学数据研究的深入发展。这也可以解释为何在文学数据研究初期，大部分研究只能停留在单体文本词语抽取、词频统计和语法分析上。因为文学数据量不够，计算技术不先进，文学数据研究指涉的对象只能是小数据样本和简单计算分析。20世纪80年代，钱锋、（16）陈大康、

（17）丁宁、（18）、林兴宅（19）等人曾提出文学与计算机科学、系统科学相结合的理论设想，利用数学建模和计算机自动计算的方法，对文学作品的文体特征、审美风格和作者身份辨识进行研究。但在具体实践中，受制于文学数据库和计算技术，大部分研究仍然集中在作品修辞风格的计量分析上，“通过建立语料库，对文本语言的字符、段落、词长、词类、词性以及高频词分布等特征项进行统计，运用计算公式来计算平均段落长度、词长变化、文本聚类等”。（20）例如，武汉大学语言自动处理研究组通过统计老舍的《骆驼祥子》的词频，来分析小说的审美特征；（21）刘宾运用数学建模技术分析《福乐智慧》，经过电脑检索、排序，编制成索引词典的框架；（22）钱锋、陈光磊通过比较巴金的《我们会见了彭德怀司令员》和倪海曙的《三轮摩托车》中的辞格、音律和章法，确认巴金和倪海曙的小说语言风格；（23）与此类似，王景丹通过统计曹禺等8位剧作家文本的句频，确认各自文体风格；（24）吴礼权通过分析词语、句式和修辞运用，总结“简约”与“繁丰”审美风格的基本特征。（25）从中不难看出，初始阶段的文学数据研究以简单统计文本修辞为主，很少使用复杂的计算模型和计算技术。

21世纪以来，随着数字人文理论在国内的传播，以及众多“关系型数据库的建设、语言信息处理技术的发展和会计量方法向人文领域的持续渗透”，（26）文学数据研究取得长足进步。例如，罗男运用基于改进的向量空间模型特征权重计算方法（TF-IDF）和潜在语义分析技术（LSI），分析15部金庸小说中的字、词汇、句子和段落4个语法单位，以此探究金庸小说的语言风格，并辨识《卧龙记》的真实作者；（27）张小宇在建立鲁迅杂文语料库、鲁迅小说语料库和鲁迅同时期作家杂文语料库基础上，使用语料库检索工具AntConc，计算鲁迅杂文的词长、句长、单音节词和双音节词等17个语言特征出现比例，并与同时代作家的杂文语言特征进行比较，以此总结鲁迅杂文的语言风格；（28）余韵为了证明巴金前后期小说语言风格的一致性，把巴金小说划分为6个阶段，并建立相应的语料库，在此基础上，计算不同阶段巴金小说的词长分布，词汇丰富度，共现词、独有词、实词和虚词使用频率；（29）与此类似，金迪对格非和余华小说语言风格的研究，（30）时季对阎连科和刘震云小说语言风格的研究，（31）以及王少康、（32）年洪东、（33）李慧（34）对现代作家身份识别和作品辨伪的研究都属于此种研究类型。但上述文学数据研究仍没有突破文体学、修辞学和语言学研究框架。

但我们也不能据此掩盖一些学者在文学数据研究方面做出的开创性探索，尤其是在小说人物功能和社会网络分析方面取得的突破。例如，吴蕙羽利用优化后的Python技术、Pyecharts数据库和Echarts数据可视化JS库，分析陈忠实的《白鹿原》中的人物关系和社会网络，并涉及人物情感分析。（35）但吴蕙羽的研究重心在计算技术，而非小说叙事，因而没有得出与小说叙事相关的结论。相对于此，赵薇对李劫人的“《大波》三部曲”的社会网络和人物功能分析更为深入和透彻。赵



↑  
回顶部

运用社会网络分析(SNA)方法,利用数据挖掘、中心性计算和可视化技术,分析《死水微澜》《暴风雨前》《大波》前后两个版本5部小说中的人物关系和社会网络,“对加权网络中最高中介中心性节点所提示的关键人物的叙事功能进行了深入探讨”,(36)认为“小人物”具有重要叙事功能,承担了讲述重大历史事件的责任。除此以外,通过分析文学研究数据库,发现某一阶段文学研究主题、(37)热点、(38)发展趋势和历史轨迹,(39)以及学者影响力、(40)学科关联(41)和学术刊物影响力(42)等内容,成为文学数据研究的重要趋向。例如,丁帆、赵普光以当代文学研究中的论题词词频为考察中心,通过分析论题词的频率、集中程度和论题词的显隐等数据,客观呈现当代文学研究70年发展轨迹。但文章没有引入知识图谱分析技术,仍属于传统意义上的计量统计范畴。而谢靖、章鑫鑫在对CSSCI数据库数据进行清洗和格式转化,以及设定频次、中介中心度和聚类指标等参数基础上,利用CtiespaceⅢ软件计算2000—2011年期间CSSCI收录的中国文学研究期刊发文情况,形成文学研究期刊关键词共现图谱,立体呈现这一期间中国文学研究热点、发展趋势和学科知识图谱。(43)而且,通过统计中国文学期刊引文数据,分析中国文学图书学术影响力。(44)王贺则通过比较《解放日报》数据库的电子目录、索引与纸质版的差异,探讨“数字人文”与传统学术的关系。(45)

虽然文学数据研究在广度和深度上都取得了一定程度的进步,但还不是真正意义上的文学大数据研究:一方面,当代文学人物关系和社会网络研究、当代文学关联数据库研究、当代作家档案数据研究、当代文学学术地图和知识谱系研究、基于共现分析的文本知识挖掘研究、当代文学研究者的隐性合作关系研究等诸多方面仍然存在进一步深挖的空间;另一方面,互联网、物联网和社交网络生产的文学大数据还没得到充分利用,基本处于搁置状态。例如,豆瓣读书、微信、微博生产的文学数据及其关涉的人物情感和社会心态、文学的网络主题模型、游戏文学的虚拟现实、界面文学的知识挖掘、新媒体文学的知识图谱等众多研究话题还没有充分展开。之所以出现这种态势,是因为现阶段文学大数据研究存在一些困境和症结,为文学大数据研究从理论构想进入现实实践从而产生真正价值设置了壁垒。这也是当代文学研究范式转型亟须解决的问题。

### 三、文学大数据研究的困境和症结

当代文学大数据研究仍处于起步阶段,难以避免出现诸多困境,我们需要直面这些症结,寻找解决问题的方法和路径,在此基础上推进当代文学研究的发展。现阶段,当代文学大数据研究面临专题数据库少、跨学科合作度低、研究方法认同度低、研究范围窄、研究学者少、专业学术期刊少、学术评价体系缺失、学科专业建设缓慢、课堂教学滞后和网络基础设施建设薄弱等问题。

文学大数据研究的基础是数据,包括数字文学文本和文献史料及其数据化后形成的各种文学专题数据库,也包括文学文本在互联网、物联网和社交网络传播产生的文学大数据。但大数据的海量、动态和价值密度低等特性使数据库建设存在较大难度。比较便捷和高效的方式是利用数据挖掘、分析和显示技术针对特定大数据进行研究。因此,文学专题数据库和文学研究专题数据库建设尤为重要。但现实情况



↑  
回顶部

是，已建成的中国现当代文学及其研究专题数据库仍然很少，主要集中在“民国图书数据库”（46）“晚清和民国全文期刊数据库”（47）“《申报》全文数据库”（48）“《大公报》（1902—1949）数据库”（49）“近代上海方志资料数据库”（50）“民国文献大全数据库”（51）“解放前报刊题录数据库”（52）“民国图片资源库”（53）“民国时期文献数据库”（54）“大成故纸堆”（55）“典海民国图书资源平台”（56）“民国图书数字化资源库”（57）等民国文学数据库上。而当代作家作品及其研究文献数据库、当代文学期刊数据库和当代文学史数据库等与当代文学研究直接相关的专题数据库非常稀少。这一问题已得到学术界的重视，陆续启动了“网络文学文献数据库建设”“汉译文学编年考录及数据库建设”“中国现代文学报刊作品系年及数据库建设”“中国新诗传播接受文献集成、研究及数据库建设”“抗战大后方文学史料数据库建设”“‘学衡派’年谱长编及文献数据库建设”“延安时期未刊文献资料收集、整理与数据库建设”（58）等项目。只有完成基础性的专题数据建设，才能为文学大数据研究提供支持。

文学大数据研究作为新兴的研究范式，从大规模的理论引入研究实践不过十几年的时间，学术界对这种研究范式始终保持足够的距离和警惕，这种研究范式还没有形成共识。这种认同情境来源于三个方面：第一，文学大数据研究运用的数据计量分析、实证性方法与文学的审美性、精神性特征相冲突，削减了文学的人文价值。人们担心对“计算”方法的推崇，会使文学研究滑向技术中心主义，形成算法霸权，成为技术的注脚，丧失自身的独立性。因而，这种研究范式“对于文学研究而言，却只能是一种手段和工具，研究者的才情、趣味、智性决定了研究的深度和走向”。（59）第二，文学大数据研究的根基是数据库和计算技术，只有不断优化和更新相关计算技术，才能推进文学大数据研究的深入发展。一旦计算技术更新缓慢或停滞不前，这种研究范式也将处于悬停状态。因而人们对这种研究范式的长效性持怀疑态度，“是否能够从历史的长线来对文学史现象和作家作品做出相对超越性的评价，是存有疑问和不无难度的”。（60）第三，文学大数据研究需要研究者熟练掌握和运用计算技术，而计算技术体系和理论结构较为复杂，涉及数字化技术、数据管理技术和数据分析技术等多个层面，仅文本挖掘技术就涉及信息抽取、文本分类、文本聚类、关联规则和模式发现等众多方面，相关的计算机软件更是种类繁多。对人文学者而言，想要短期内掌握计算技术，且能够熟练使用计算机软件，存在较大难度。而能够独立“运用计算方法解答、探讨或是思索人文问题的研究者”（61）更是寥寥无几。这就不难理解为何大部分文学大数据研究都集中在理论反思和论争层面，很少拓展到计算技术分析层面。

文学大数据研究是一种跨学科研究方法，它与统计学、信息管理学和计算机科学等其他学科有直接关联。同时，文学也以其独特的人文特性反思文学大数据本身的价值和意义，实现文学与数据和计算之间的多重互动。在此意义上，文学大数据研究具有明显的交叉学科研究特征，它“促使各人文学科突破学科边界，加强不同学科理论与方法的交流与合作”。（62）这就决定了研究者需要进行跨学科、跨专业的协同创新研究。但当前文学大数据研究还没有真正实现跨学科研究，文学研究者的关注点集中在文学大数据研究的人文反思上，其他学科的研究者聚焦在计算技



回顶部

术上，二者处于不同的研究向度，无法真正交融在一起。类似陈大康、钱锋、刘石、赵薇等学者能够实现二者之间的融合，是因为他们本身具有跨学科背景和扎实的数学功底，能够熟知计算原理，独立构建数学模型和使用计算工具。但大部分研究者仍在技术外围摸索和徘徊。这种局面的形成与国内数字人文专业设置和相关教学环节薄弱密切相关。目前，只有中国人民大学信息资源管理学院在图书情报与档案管理一级学科下，设立了数字人文二级学科。而根据国家数字人文组织联盟统计，在2007—2014年期间，国外设置数字人文专业的高校多达180余所，<sup>（63）</sup>并形成了从本科到研究生完备的培养体系。<sup>（64）</sup>而且，国内外高校在相关课程设置方面也有较大差距。截止到2020年，开设数字人文相关课程的美国高校就有41所，<sup>（65）</sup>包括斯坦福大学、耶鲁大学和伊利诺伊大学香槟分校等世界知名高校。而国内只有清华大学、北京大学、武汉大学、西安交通大学、汕头大学、南京大学、上海师范大学等少数高校开设了数字人文课程。同时，在课程的数量、教学体系建设和实践应用方面，国内外也存在较大差距。国内数字人文研究中心和人文计算实验室建设推进缓慢，只有武汉大学、清华大学和上海师范大学等为数不多的高校成立了数字人文研究中心。而国外的数字人文研究中心和实验室多达百余家，并且不断输出学术成果。<sup>（66）</sup>尤其是斯坦福文学实验室、伦敦国王学院数字人文实验室、耶鲁大学数字人文实验室和普林斯顿数字人文中心在世界范围内都产生了重要影响力。

虽然文学大数据研究还存在诸多困境，但不能就此否认它的价值。除了为当代文学研究提供新的研究技术和方法，也为文学研究共同体的形成提供了可能性和实践路径。谢刚、江震龙在《现代中国民族文学观与共同体诗学建构》一文中深入阐释了民族文学共同体意识和共同体诗学，认为其实质“是一与多、己与群、局部与整体、差异性与共同性的辩证统一”，<sup>（67）</sup>并体现为民族审美共同体意识、辩证的诗学特征、包容共生的文学伦理观和以人民为中心的民族文艺本质。谢刚、江震龙对共同体意识和共同体诗学的理论建构深刻而透彻，既没有完全模仿西方的共同体理论，也没有拘泥于中国传统的共同体观念，而是在中西理论之间寻找到恰切的平衡点。或者说，是站在中国文学的立场上构建世界文学的共同体意识和诗学，因此具有很强的理论价值和实践参考意义。按照谢刚、江震龙的逻辑推论，文学研究共同体的实质也是“多”与“一”的辩证关系。而当代文学研究没能处理好这种辩证关系，走向了共同体的反面。当下反全球化和逆全球化思潮在不同领域设置了构建共同体的壁垒。或者说，共同体和反共同体在发生机制、运作模式和实践路径等方面遵循相同的原则和规律，二者之间的界限既泾渭分明又暧昧模糊。这种共同体危机意识和反共同体情态也在当代文学研究中存在，以文学研究范式的断裂、颠覆和重建为表征，具体体现为文学政治研究、文学社会研究、文学历史研究、文学审美研究和文学技术研究几种研究范式之间的矛盾及研究范式的内部冲突。基于此，当代文学研究史可以看作文学研究共同体与反共同体的关系史。但关键问题不是几种研究范式是如何断裂的，而是如何继承和转换的。一种研究范式的兴起如何从另一种消隐的研究范式中汲取有价值的资源。也就是说，文学大数据研究的兴起并不全然意味着原有研究范式的失效，而是研究范式之间的整合和重组。文学大数据研究的实质是从数据角度挖掘和呈现文学新变，但新变不是颠覆文学的审美性和精神



↑  
回顶部

性特质，而是以数据进一步证实和丰富文学的审美性和精神性，并发现其中的新内容、新样态和新趋势。文学研究应在数据技术加持下变得更客观、理性，更符合当下对文学研究提出的要求。文学大数据研究建构的技术化研究秩序不是划出文学研究的固定边界，而是提倡文学研究的敞开性，为人与科技共同体搭建一条顺畅通道。我们提倡的文学研究共同体意识，针对的不仅是反共同体观念，而是以共同体意识为根基，接续和重建不同研究范式之间的内在关联，“借以窥探某一历史时刻的集体性是什么，它的需求是什么，它的矛盾性何在，它的问题何在等问题”，

(68) 寻找解决原有文学研究范式症结的路径，从而建立不同文学研究范式之间的结构关系，因为“它们既有联系又有差异，共同推进文学创作和文学研究的发展”。(69) 而且，一种新的文学研究范式的价值不仅体现在方法论层面，更体现在价值论层面。在不同研究范式之间寻求价值观的最大公约数，在此基础上构建当今社会具有引领作用的价值观念，这是推进当代文学研究发展的重要前提条件之一。在此意义上，当代文学研究范式转型绝不是线性时间链条上的替代问题，而是不同研究范式在对话中的相互补充和再造。它既是当下的又是历史的也是未来的，但绝不是唯一的、霸权的和僵化的。

注释：

(1) 程光炜：《当代文学学科的“历史化”》，《文艺研究》2008年第4期。

(2) 陈培浩：《“现代汉诗”与中国诗学“当代性”的生成》，《当代作家评论》2021年第3期。

(3) 张弘：《学术范式转型与批判意识》，《学术月刊》2003年第7期。

(4) [美]阿尔文·托夫勒：《第三次浪潮》，第117页，黄明坚译，北京，中信出版集团，2018。

(5) 中国信息通信研究院：《大数据白皮书（2020年）》，引自  
<http://www.ideadata.com.cn/temp/article/file/20210115/1610676847871064775.html>。

(6) 吴军：《智能时代：大数据与智能革命重新定义未来》，第40页，北京，中信出版社，2016。

(7) [英]维克托·迈尔·舍恩伯格、[英]肯尼思·库克耶：《大数据时代：生活、工作与思维的大变革》，第15页，周涛译，杭州，浙江人民出版社，2012。

(8) 见孟小峰、慈祥：《大数据管理：概念、技术与挑战》，《计算机研究与发展》2013年第1期。

(9) 马建光、姜巍：《大数据的概念、特征及其应用》，《国防科技》2013年第2期。



回顶部

- (10) 单小曦：《数字文学的命名及其生产类型》，《中州学刊》2011年第6期。
- (11) 李斌：《数字技术、虚拟现实与网络功能——数字文学的审美新变化与发展新趋势》，《当代外国文学》2021年第1期。
- (12) 熊澄宇编：《新媒介与创新思维》，第291页，北京，清华大学出版社，2001。
- (13) 范玉刚：《网络文学：生成于文学与技术之间》，《文学评论》2008年第2期。
- (14) [美]安妮·伯迪克等：《数字人文：改变知识创新与分享的游戏规划》，第43页，马林青、韩若画译，北京，中国人民大学出版社，2018。
- (15) [英]大卫·M.贝里、[挪威]安德斯·费格约德：《数字人文：数字时代的知识与批判》，第129页，王晓光等译，大连，东北财经大学出版社，2019。
- (16) 钱锋、陈光磊：《文学·数学·计算机……》，《自然杂志》1981年第6期。
- (17) 陈大康：《文学、数学与电子计算机》，《自然杂志》1988年第12期。
- (18) 丁宁：《系统研究：文艺理论跃迁的契机》，《文艺理论研究》1985年第3期。
- (19) 林兴宅：《论系统科学方法论在文艺研究中的运用》，《文学评论》1986年第1期。
- (20) 杨建军：《定量分析法在中国现当代文学研究中的运用》，《厦门大学学报》（哲学社会科学版）2016年第4期。
- (21) 陈静：《当下中国“数字人文”研究状况及意义》，《山东社会科学》2018年第7期。
- (22) 刘宾：《电子计算机与文学研究——兼谈〈福乐智慧〉（维文版）的微机处理》，《新疆社会科学》1987年第6期。
- (23) 曾毅平、朱晓文：《计算方法在汉语风格学研究中的应用》，《福建师范大学学报》（哲学社会科学版）2006年第1期。
- (24) 王景丹：《从句频分析看八位剧作家的风格异同》，《修辞学习》2003年第4期。
- (25) 吴礼权：《从统计分析看“简约”与“繁丰”的修辞特征及其风格建构的原则》，《修辞学习》2003年第2期。



回顶部

(26) 赵薇：《网络分析与人物理论》，《文艺理论与批评》2020年第2期。

(27) 罗男：《基于统计学的金庸小说个人语言风格及疑似作品的证伪研究》，华中师范大学硕士学位论文，2020年5月。

(28) 张小宇：《基于语料库的鲁迅杂文语言风格研究》，南京师范大学硕士学位论文，2016年5月。

(29) 余韵：《巴金前后期小说的计量风格学研究》，华中师范大学硕士学位论文，2017年5月。

(30) 金迪：《基于语料库的格非、余华小说计量风格学研究》，南京师范大学硕士学位论文，2018年5月。

(31) 时季：《基于语料库的阎连科、刘震云小说计量风格学比较研究》，南京师范大学硕士学位论文，2019年5月。

(32) 王少康、董科军、阎保平：《基于语句节奏特征的作者身份识别研究》，《计算机工程》2011年第9期。

(33) 年洪东、陈小荷、王东波：《现当代文学作品的作者身份识别研究》，《计算机工程与应用》2010年第4期。

(34) 李惠、刘颖：《基于语言模型和特征分类的抄袭判定》，《计算机工程》2013年第5期。

(35) 吴蕙羽：《基于Python技术分析小说人物关系和社交网络》，《电脑编程技巧与维护》2020年第6期。

(36) 赵薇：《社会网络分析与“〈大波〉三部曲”的人物功能》，《山东社会科学》2018年第9期。

(37) 赵宪章、苏新宁：《基于CSSCI的中国文学研究主题词分析（二〇〇〇—二〇〇四）》，《当代作家评论》2006年第6期。

(38) 谢靖、章鑫鑫：《基于CSSCI（2000—2011年）的中国文学研究热点知识图谱分析》，《西南民族大学学报》（人文社会科学版）2014年第5期。

(39) 丁帆、赵普光：《历史的轨迹：中国现当代文学研究七十年的实证分析——以论题词频的统计为中心》，《文艺研究》2019年第9期。

(40) 丁楠、潘有能、叶鹰：《基于CSSCI的文科学者h指数实证研究》，《大学图书馆学报》2009年第2期。

(41) 王昊、苏新宁：《基于CSSCI本体的学科关联分析》，《现代图书情报技术》2010年第10期。



↑  
回顶部

(42) 钱玲飞:《“985”高校文科学报h指数分析——基于CSSCI(2004—2007年度)的数据》,《西南民族大学学报》(人文社科版)2009年第10期。

(43) 谢靖、章鑫鑫:《基于CSSCI(2000—2011)的中国文学学科知识图谱研究》,《图书与情报》(人文社会科学版)2014年第2期。

(44) 谢靖:《中国文学图书学术影响力分析(国内学术著作)——基于CSSCI(2000—2007)》,《东岳论丛》2009年第10期。

(45) 王贺:《“数字人文”与传统学术——以〈解放日报〉目录、索引及数据库为中心》,《文艺争鸣》2020年第10期。

(46) 见<http://172.20.32.35/library/publish/default/IndexBook.jsp>。

(47) 见<http://www.cnbkcsy.cn>。

(48) 见<http://shenbao.lib.shnu.edu.cn/WEB/INDEX.html>。

(49) 见<http://tk.cepiec.com.cn/tknewsc/tknewskm?@@@0.26952861971221864>。

(50) 见<http://10.10.184.11:81>。

(51) 见<http://www.cadal.zju.edu.cn/index>。

(52) 见<http://10.10.184.11:8001>。

(53) 见<http://www.minguotupian.com>。

(54) 见<http://hollis.harvard.edu/primolibweb>。

(55) 见<http://www.dachengdata.com>。

(56) 见[www.dian-hai.com](http://www.dian-hai.com)。

(57) 见<http://mylib.nlc.cn/web/guest/minguotushu>。

(58) 王贺:《“数字人文”如何与现代文学研究结合》,《现代中文学刊》2019年第1期。

(59) 孙桂荣:《文学研究的当代性与大数据时代的实证学风》,《湘潭大学学报》(哲学社会科学版)2018年第2期。

(60) 张福贵:《文学史学研究中技术分析法的有效性》,《探索与争鸣》2020年第5期。

(61) [美]梅丽莎·丁斯曼、[美]苏真:《人文研究中的数字:苏真访谈录》,向俊译,《山东社会科学》2018年第9期。



↑  
回顶部

(62) 甘琳、钱烨：《数字人文学科交叉特征探究——基于作者关键词共现网络的实证研究》，刘石、孙茂松、顾青主编：《数字人文》第3期，第123页，北京，中华书局，2020。

(63) Alliance of Digital Humanities Organizations. Introduction of Center Net [EB/OL] [2014-01-26]，引自<http://digitalhumanities.org>。

(64) 吴加琪、董梅香、赵子菲：《国外数字人文专业研究生教育调查》，《图书馆论坛》2018年第6期。

(65) 加小双、冯慧玲：《“SCP2”数字人文教育综合体系的构建与应用》，《图书馆论坛》2020年4期。

(66) 赵生辉、朱学芳：《我国高校数字人文中心建设初探》，《图书情报工作》2014年第6期。

(67) 谢刚、江震龙：《现代中国民族文学观与共同体诗学建构》，《中国社会科学》2021年第10期。

(68) 张旭东：《读书报专访杰姆逊：“理论已死”？理论何为？》，《中华读书报》2012年12月5日。

(69) 张福贵：《当代中国文学研究话语体系的建构》，《中国社会科学》2019年第10期。

中国作家出版集团版权所有 | 京ICP备16044554号 | 京公网安备110402440007号 | 互联网新闻信息服务许可证(10120180023)

地址：北京市朝阳区农展馆南里10号15层 联系电话：010-65389115 邮箱：cnwriter@126.com

本网站有部分内容来自互联网，如媒体、公司、企业或个人对该部分主张知识产权，请来电或致函告之，本网站将采取适当措施，否则，与之有关的知识产权纠纷  
本网站不承担任何责任。



↑  
回顶部