

# 数字人文视域下文学研究的 现状、问题与趋势\*

郑 飞 逯 灵

---

**【摘 要】** 基于 Web of Science 及 CNKI 数据，运用文献计量方法，结合可视化分析工具，对国内外数字人文领域的文学研究进行统计分析和内容挖掘发现：数字文学核心要义与研究方法已渗入现当代文学研究领域，呈现客观、高效、可量化、多单元合作及可验证优势；而国内相关研究普遍存在体裁失衡、研究对象趋同、数据类型单一及数据解读不够深入等问题。鉴于此，从拓宽研究对象、创新研究路径、加深数据解读、转变学科思维等方面给出相关建议，旨在为我国数字化文学研究体系建设与发展提供参考。

**【关键词】** 数字人文 文学研究 计算机 语料库 跨学科

**【作者简介】** 郑飞，文学博士，北京航空航天大学外国语学院教授；逯灵，中国人民大学文学院 2022 级博士研究生。

**【中图分类号】** I0      **【文献标识码】** A

**【文章编号】** 2097 - 1125 (2023) 05 - 0005 - 13

---

在以计算机技术为基础的人工智能时代，海量信息引发的大规模数据加工业的蓬勃发展已成为大势所趋，数字化生产方式成为越来越多领域必须面临的挑战。数字技术强大的数据处理能力远胜过传统的材料处理方式，正以所向披靡之势迅速地改变世界，同时也为文学研究提出了富有挑战性的新课题。根据这一态势，本文试从宏观和微观两个层面总结分析当

---

\* 本文系国家社会科学基金重大项目“英国旅行文学史”（22&ZD288）、中央高校基本科研业务费项目“基于语料库的尤金·奥尼尔戏剧文体学研究”（YWF-17-SK-G-44）的阶段性成果。

下我国数字化文学研究的发展趋势和不足之处，探索数字技术在文学研究中的应用前景，以期为我国数字化文学研究体系的建设与发展提供参考。

## 一、数字化文学研究的现状与态势

在信息革命的激流冲击下，传统的学科壁垒在新型人机通信面前纷纷被打破。与自然科学的研究相比，传统的文学艺术乃至相关人文社会科学的资料处理和研究工作具有明显的模糊性和不确定性。随着当代各学科对综合性、精确性、科学性要求的不断提升，人们对文学研究的客观性和数字化日益充满期待，使得文学批评和研究活动日益呈现数字化的特征。

众所周知，工具的应用对实践的效果无不产生重大影响。计算机技术催生的电子档案、文本挖掘、3D建模、仿真模拟、图像识别以及数据可视化等数字化研究工具，使人文学科无论是在资料处理方面，还是在借助资料证实或证伪方面，都迈进了日新月异的变革进程。特别是在市场的推动下，各种文化产业应运而生，有力地促进了国家软实力的提升以及精神生产力的高速发展。

作为人文学科与数字技术相结合的产物，数字人文尽管尚无统一、准确的定义，但其显著特点却可大致归纳为：（1）在不断开发和更新中积极使用已经存在的数字语料库；（2）使用量化或建构模型的方式规划人文研究领域中的问题；（3）选择一个特定的学科去设计项目；（4）技能的广泛结合，诸如同时运用编程与批评分析等手段。<sup>①</sup>

以1946年意大利学者罗伯特·布沙（Robert Busa）用穿孔卡片分析托马斯·阿奎那（Thomas Aquinas）的作品为开端，计算机用于文本分析的研究范式初现雏形。<sup>②</sup>随着数字人文的日益成熟，计算机技术与文学研究的联系日益密切，现今已有数据语料库建设、文本挖掘、文学地图出版、社会网络分析、叙事方程分析、数字记忆建构、学术编辑等多种研究手段与形式问世。对于传统的文学研究而言，数字人文可帮助研究者通过编程、分析、可视化等手段刷新原有的结论、产生新观点、展示文学发展的宏观规律或趋势。同时，人文宗旨及其价值观的注入则会使数字技术的开发和应用更加符合人文学科的精神特质，甚至可以引导、拓展或改变研究者的工作方式。

在我国，2014年出版的《大数据大文化》一书对数字时代背景下文化

① 参见 Alan Liu, *The Meaning of the Digital Humanities*, *PMLA*, Vol. 128 (2), 2013, pp. 409–423.

② 参见 Chris Alen Sula and Heather V. Hill, *The Early History of Digital Humanities: An Analysis of Computers and the Humanities (1966–2004) and Literary and Linguistic Computing (1986–2004)*, *Digital Scholarship in the Humanities*, Vol. 34 (Supplement 1), 2019, pp. i190–i206.

发展战略的调整提出了若干设想。<sup>①</sup> 2020年出版的《数字人文研究》总结了近十几年间数字人文在音乐、建筑、医疗、媒体等领域的发展状况。<sup>②</sup> 目前，国内已涌现出较多数字人文研究的成果，相关的科研立项也如雨后春笋般遍布各地。只是由于数字技术尚处于初兴阶段，故而中央和地方的科研项目大多仍集中在图书馆服务及档案资源管理方面，仅有少量项目将数字化研究应用于中国古籍经典分析<sup>③</sup>或中英平行语料库的创建等，<sup>④</sup> 至于数字人文用于文学研究的立项占比就更低了。这一状况显然与国外以团队形式将数字化技术应用于人文社会学科研究，并带动相关领域迅速改观的总体形势不相协同。

相比之下，在数字化文学研究方面的个人著述却数量可观，而且大多属综述类与应用类研究，前者介绍数字人文领域的文学研究优势、方法范式及发展现状，后者多利用各种技术手段对文本进行数据化处理与解读。这一研究态势表明，数字化文学研究正在崛起，势在必行，只要把握住机遇、方向与规律，必将得到世界各国的普遍认可和广泛利用。

为了调查国内外数字化文学研究的现状与热点，本文以“数字人文”“语料库”“文学研究”为检索关键词，搜集国内以CNKI（中国学术期刊网络版）为来源的相关文献220篇（2002—2019年）及国外以Web of Science（核心期刊）为来源的189篇相关文献（1985—2019年）作为研究样本，数据分析依托美国华裔学者陈超美开发的CiteSpace软件，通过关键词共现和聚类分析等方法对数据统计结果进行可视化分析。

图1和图2是利用CiteSpace的“关键词共现”和“聚类分析”功能法得出的结果，从中可见国内外数字化文学研究领域的前沿与典型话题。其中，符号“#”代表该领域某一主题的文献聚类，节点大小代表某主题的频度，直线代表某主题的历年文献，曲线代表文献被引频次。图1中“literature”和“linguistics”两个节点半径较大、直线延伸较长、曲线数量较多，表明二者是自2002年至2019年国外数字人文的高频话题。图2中的“语料库”节点半径最大、直线延伸最长、曲线数量最多，表明“语料库”

① 参见崔成泉、王晓芳编著：《大数据大文化》，云南大学出版社2014年版。

② 参见孟建主编：《数字人文研究》，复旦大学出版社2020年版。

③ 如国家社会科学基金重大项目“基于大数据技术的古代文学经典文本分析与研究”（18ZDA238）、国家社会科学基金一般项目“图书馆古籍文献的数字人文开发与应用模式研究”（17XTQ003）、国家社会科学基金重点项目“中古流贬文学文献整理与研究”（18AZW014）、国家社会科学基金重大项目“汉魏六朝文学编年地图平台建设”（19ZDA253）等。

④ 如国家社会科学基金重大项目“大规模英汉平行语料库的建立与加工”（10ZD&127）等。



式主义以及法国的年鉴学派。<sup>①</sup> 19 世纪末，美国学者托马斯·C. 曼德霍尔(Thomas C. Mendenhall) 曾统计某作家作品文本词汇的平均长度，并与其他作家作品的词汇长度加以对比，以确定该作家的写作习惯，这一类似科学实验的方法后来发展为计量文体学。在我国，20 世纪 20 年代梁启超曾提出“历史统计学”的概念，用统计学原理分析历史数据，以认识历史发展的宏观规律。当时，俄国形式主义理论家什克洛夫斯基和民间文学专家弗拉基米尔·普洛普也开始探索更为科学化的文学研究方法。1946 年，布沙创建的文本编码转为可量化数据技术开启了后来的计算机检索，大幅提高了数字化研究效率。由此可见，早期数字人文的研究方法虽然简单，却为后来的研究开辟了新路，也使数学思维和统计方法在文学研究中得到了大力推广。这一技术不仅加速了文学研究从“天马行空”到“脚踏实地”的转变，更使世界范围内的文学比较研究成为可能。

当下，数字人文的发展如火如荼。但值得注意的是，在数字人文以其庞大的存储空间、精确的统计能力以及直观的结果展示为文学研究注入活力的同时，其在文学研究中的应用也不可避免地出现单一化、模式化、集中化等问题。需要肯定的是，计算机技术在文学研究中的应用在很大程度上影响甚至改变了文学研究者的研究内容、研究方法乃至研究思维。具体而言，数字化文学研究的优势体现在三个方面。

首先，对海量文本的处理效率提高。21 世纪初，意大利学者弗朗科·莫莱蒂(Franco Moretti) 在《关于世界文学的猜想》一文中提出“世界文学体系”和“远读”概念，引发数字化文学研究热潮。<sup>②</sup> 这使得探索世界文学发展规律、世界文学创作模式等一系列总体性问题成为可能。例如，麦高恩(Joseph P. McGowan) 对英国拉丁语文学的整体性研究、<sup>③</sup> 菲舍尔-斯塔克(Bettina Fischer-Starcke) 对简·奥斯汀和其同时代作家的对比研究<sup>④</sup> 以及博尔托拉托(Claudia Bortolato) 对男女作家用词差异的研究等。<sup>⑤</sup> 此类宏

① 参见 [美] 戴安德、姜文涛：《数字人文作为一种方法：西方研究现状及展望》，赵薇译，《山东社会科学》2016 年第 11 期，第 26～33 页。

② 参见 Franco Moretti, *Conjectures on World Literature*, *New Left Review*, Vol. 1, 2000, pp. 54-68.

③ 参见 Joseph P. McGowan, *An Introduction to the Corpus of Anglo-Latin literature*, in Phillip Pulsiano and Elaine Treharne, eds., *A Companion to Anglo-Saxon Literature*, Oxford: Blackwell Publishers Ltd., 2017, pp. 11-49.

④ 参见 Bettina Fischer-Starcke, *Corpus Linguistics in Literary Analysis: Jane Austen and Her Contemporaries*, London: Bloomsbury Publishing, 2010.

⑤ 参见 Claudia Bortolato, *Intertextual Distance of Function Words as a Tool to Detect an Author's Gender: A Corpus-Based Study on Contemporary Italian Literature*, *Glottometrics*, Vol. 34, 2016, pp. 28-43.

观研究需要对大量的文本进行加工处理，这对单纯依赖人力的传统文学研究来说，其难度如同庄子所说的“吾生也有涯，而知也无涯。以有涯随无涯，殆已”，<sup>①</sup>但计算机技术的介入使得在有限的时空内处理无限的文本成为可能，从而有效弥补了传统文学研究在宏观层面的不足。

其次，文学研究的精确度日益提升。除了在宏观层面的广泛应用，数字人文在文学研究的微观层面也大有可为，文学作品的词汇、句型、人物、情节、主题等要素分析纷纷进入了数据运算体系。在传统的文学研究中，由于文学研究材料庞杂，加之研究者精力有限，常常难以对文学作品中的细枝末节进行全面、准确的锁定与分析。因此，研究者很有可能在研究过程中有所遗漏，从而影响研究结论的准确性。现今数字人文领域的文学研究仍旧以分析高频词（词簇）及关键词为主，如马尔伯格曾使用词簇或重复的单词序列研究狄更斯的文体风格，对狄更斯的人物刻画技巧以及作品中肢体语言和言语的表现方式得出了新结论；<sup>②</sup>王梦君运用语料库分析《圣经·启示录》的词频、字节、形符、类符等，得出《圣经·启示录》用词简练、文笔精湛的结论；<sup>③</sup>陈心怡等基于关键词和高频词组分析《道连·格雷的画像》中的道德倾向及语言特色；<sup>④</sup>袁文晗利用 AntConc 中的关键词表功能发现，《追风筝的人》中的地点主要为阿富汗、喀布尔、美国，而描述感情的关键词依次是微笑、希望、喜欢、痛、爱，由此推断主人公的精神救赎之路；<sup>⑤</sup>在陈建生等对《尤利西斯》的意识流技巧的研究中，他们通过分析标点符号、主题词及词簇挖掘乔伊斯的语言特征。<sup>⑥</sup>以上研究提及的诸如关键词、词频、标点符号等内容，在传统的文学研究中，其搜集和整理既会耗费大量的时间和精力，统计和分析的准确性也有待商榷，但在计算机技术的辅助下，文学研究中的数据收集及分析体现出精准、高效等优势，在节省研究者时间、精力的同时充分提高了其工作效率。

最后，文学研究结论的可验证性增强。传统的文学研究主观意味浓厚，

① 陈鼓应注释：《庄子今注今译》，商务印书馆2007年版，第113页。

② 参见 Michaela Mahlberg, *Corpus Stylistics and Dickens's Fiction*, New York: Routledge, 2015, pp. 51-60。

③ 参见王梦君：《基于语料库的〈圣经·启示录〉语言特征检索分析》，《广西教育学院学报》2015年第4期，第85~87页。

④ 参见陈心怡、张荔：《〈道连·格雷的画像〉中的道德倾向及语言特色——基于关键词和高频词组的语料库研究》，《外语与翻译》2016年第2期，第76~85页。

⑤ 参见袁文晗：《基于语料库的卡勒德·胡赛尼作品〈追风筝的人〉多维度分析》，《文教资料》2017年第21期，第27~29页。

⑥ 参见陈建生、张珊：《基于语料库文体学的〈尤利西斯〉意识流研究》，《外国语言文学》2018年第3期，第304~317页。

难以实现客观、科学的证实或证伪，而计算机技术的应用则为验证结论提供了有效支撑。例如，莫莱蒂曾利用 1 个包含 44 个单词和标点符号的特征集来检验文本风格并达到了预期效果，<sup>①</sup> 这些词后来被称为 MFV（常见词），在语料库文体学研究中得到了广泛应用；张婷婷利用语料库生成的词表佐证《简·爱》中年轻女性追求爱情的主题；<sup>②</sup> 杨元媛运用语料库检索发现《哈克贝利·费恩历险记》中的高频词词簇多为主动语态且多用于描述数量、地点，从而论证了该作品富含口语化特征。<sup>③</sup> 与传统文学研究的结论验证体现的主观性、随意性相比，客观数据的支撑使文学研究结论的验证更富有可靠性、说服力，从而加强了文学研究的科学性。

近年来，数字化文学研究虽取得了一定的成果和进展，但由于国内文学研究者对数字人文的认识不足、文学领域的数字人文基础设施薄弱、国际及地区之间的交流合作系统有待完善，故而数字化文学研究也呈现一系列亟待解决的问题。

首先，研究中的文学体裁占比严重失衡。统计发现，在数字化文学研究成果中（如图 3 所示，综述类、点评类除外），多数研究针对的文学体裁为小说，而诗歌、戏剧等其他体裁的占比却微乎其微。

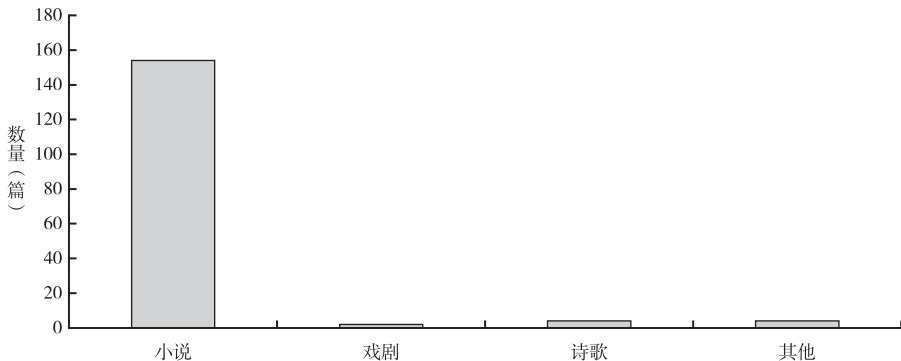


图 3 数字化文学研究针对各类文学体裁的成果数量

不可否认的是，由于数字人文极强的数据存储及分析能力，小说（尤其是长篇小说）因庞大的篇幅成为其天生的研究对象。国内外文学研究者运用数字化手段对《圣经·启示录》《尤利西斯》《哈克贝利·费恩历险记》

① 参见 Franco Moretti, *Distant Reading*, London: Verso, 2013, p. 181。

② 参见张婷婷：《语料库文体学视角下〈简·爱〉特色的研究》，《海外英语》2019年第5期，第227~228页。

③ 参见杨元媛：《语料库驱动下的〈哈克贝利·费恩历险记〉词簇分析》，《商丘师范学院学报》2019年第10期，第36~39页。

《简·爱》《兔子四部曲》《德伯家的苔丝》《大卫·科波菲尔》《远大前程》等作品的研究，充分说明了长篇小说在数字化文学研究中的受欢迎程度，而戏剧和诗歌等其他文学体裁在数字化文学研究中的表现却略有逊色。近年来，数字化文学研究中诗歌和戏剧未受重视的现象已引起了部分学者的注意，学界逐渐涌现出一批利用数字化手段研究诗歌、戏剧、演讲等文学体裁的成果。国外如克莱仕（Mohammed Rameez Qureshi）等根据词汇特征判断不同文学体裁的特征；<sup>①</sup>凯勒（Elisa Cajiao Cuéllar）利用语料库探究狄金森的诗歌《我的生命——一把上膛的枪》及《我品尝未酿之酒》的艺术内容；<sup>②</sup>玛格多那（Lehmann Magdolna）探索并创建了爱尔兰戏剧家库迪（Mária Kurdi）的戏剧词汇语料库，分析若干戏剧的学术价值、母题及潜在受众等信息；<sup>③</sup>科菲（Neil Coffee）等对维吉尔的《埃涅阿斯纪》和卢坎的《内战：法沙利亚》两首长诗进行修辞、意象、感情等方面的研究，以论证文本之间的互文性。<sup>④</sup>国内如方颖利用沃斯（Paul Werth）提出的文本世界假设<sup>⑤</sup>分析阿瑟·米勒两部时空剧中的人物形象，并运用语料库的关键语义域剖析主人公的生活经历；<sup>⑥</sup>郑飞通过语料库分析奥尼尔《悲悼》中的海岛意象及“海岛”一词的搭配、寓意，证明了海岛与爱的隐喻关联；<sup>⑦</sup>詹宏伟等利用数据探究《罗密欧与朱丽叶》中男女主人公表达爱意的不同方式。<sup>⑧</sup>

① 参见 Mohammed Rameez Qureshi, Sidharth Ranjan and Rajakrishnan Rajkumar et al., A Simple Approach to Classify Fictional and Non-Fictional Genres, in Francis Ferraro, Ting-Hao Kenneth Huang and Stephanie M. Lukin et al., eds., *Proceedings of the Second Workshop on Storytelling*, Florence: Association for Computational Linguistics, 2019, pp. 81–89.

② 参见 Elisa Cajiao Cuéllar, Disentangling Emily Dickinson's Riddles and Encoded Voices in "My Life Had Stood-A Loaded Gun" and "I Taste a Liquor Never Brewed", *Journal of the Spanish Association of Anglo-American Studies*, Vol. 37 (2), 2015, pp. 27–43.

③ 参见 Lehmann Magdolna, Corpora in Literature: A Corpus-Based Study of Literary Texts on Irish Drama, in Zsuzsa Csikai and Andrew C. Rouse, eds., *Critical Essays in Honour of Mária Kurdi: Tanulmányok Kurdi Maria Tiszteletére*, Martonfa: SPECHEL e-editions, 2017.

④ 参见 Neil Coffee, Jean-Pierre Koenig and Shakthi Poornima et al., Intertextuality in the Digital Age, *Transactions of the American Philological Association*, Vol. 142 (2), 2012, pp. 383–422.

⑤ 参见 Paul Werth, *Text Worlds: Representing Conceptual Space in Discourse*, London: Longman, 1999.

⑥ 参见方颖：《文本世界与戏剧人物塑造——以阿瑟·米勒的两部时空剧为例》，《外国语文》2016年第6期，第26~31页。

⑦ 参见郑飞：《试用语料库分析〈悲悼〉中海岛的意象》，谢群、陈立华主编：《当代美国戏剧研究：第14届全国美国戏剧研讨会论文集》，北京理工大学出版社2010年版，第41~46页。

⑧ 参见詹宏伟、黄四宏：《大数据时代的文学经典解读——〈罗密欧与朱丽叶〉计量文体分析》，《外语与翻译》2017年第2期，第63~68页。



诗歌及戏剧在数字化研究中占比的日益提升，表明小说研究在数字人文领域“一家独大”的地位逐渐被消解，更预示着数字化文学研究的多元化发展趋向，也在一定程度上启发了国内数字化文学研究的新方向。

其次，研究对象趋同。笔者依据国内外相关研究文献的中心论点，将其分为以下几类：文体特征、多维分析、综述文章、语言特征、人物形象、主题分析、叙事特征及其他，国外学者多将重心集中于对理论的探究和总体性批评，涉猎面比国内更加广泛，时常结合历史、政治、文化等话题的讨论。而国内学者显然对作品的微观分析更为专注，给人以零敲碎打之感。

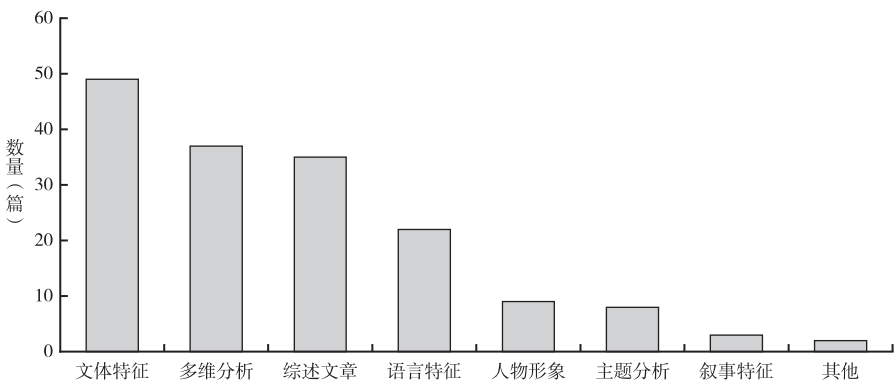


图 4 研究对象数量统计

目前，我国数字化文学研究的热点主要集中于对文学作品的文体特征、语言特征、人物形象以及意象蕴含进行分析（见图 4），而对文学作品的修辞手法、叙事特征、艺术规律、美学原理的研究甚少，研究内容的高度集中意味着我国当下数字化文学研究的趋同化程度较高，缺乏开拓与创新的品格。时至今日，语料库技术、语言学、文体学以及计算机技术固然促成了文学研究范式的革新，但就国内外已有的研究成果而言，大多数研究仍集中于对已有结论的证实或证伪，因此研究的创新性仅体现在研究手法上，尚未推动文学研究的实质性飞跃。以分析文学作品的“人物形象”为例，国外学者罗德里格斯-萨拉斯（Gerardo Rodríguez-Salas）通过分析凯瑟琳·曼斯菲尔德的传记和作品剖析作家的情感经历及她与弟弟的情感联系；<sup>①</sup> 莫莱蒂以量化方式研究《哈姆雷特》中的人物关系网络。<sup>②</sup> 国内学者也大多通过词汇

① 参见 Gerardo Rodríguez-Salas, “I Am Just as Much Dead as He Is”: Community, Finitude and Sibling Intimacy in Katherine Mansfield, *Atlantis*, Vol. 38 (2), 2016, pp. 63–82。

② 参见 Franco Moretti, “Operationalizing”: or, the Function of Measurement in Modern Literary Theory, *The Journal of English Language and Literature*, Vol. 60 (1), 2014, pp. 3–19。

和句法论证人物的心理状态和性格特征，如耿林璐运用 AntConc 软件分类整理《别让我走》中描写克隆人感情的词汇，从而发现该作品中的克隆人的确有与人类相似的情感；<sup>①</sup> 闫长红利用语料库研究《红字》中的主要人物形象；<sup>②</sup> 谢婷婷利用 WordSmith 分析《伊坦·弗洛美》中的人物塑造技巧等。<sup>③</sup> 其他微观层面的研究与人物形象研究的情形也大抵相同，即通过引进数据来描述现象，却缺乏对数据背后深层问题的挖掘，导致研究结论难以升华。这表明研究者亟须改变“数据本位”的思想，将数据看作达成目标的一种手段，而不是研究内容，更不是研究目标本身。

再次，数据类型单一。数字人文诞生伊始，其技术目标便指向对文本进行数据化整理、统计、分析和表达。从19世纪末到21世纪初，大量数字人文领域的文学研究成果都依赖于对文本中高频词（词簇）和关键词的探究，以此解读人物形象特征、作品主题、意象内涵、语言特征及文体特征等内容。但高频词或关键词多为实义词，即名词、实义动词、形容词及副词，因而导致对功能词即代词、数词、冠词、助动词、情态动词等词类关注过少。究其原因，主要有两点：第一，功能词数量多、语境复杂、指示模糊，较难得出确切可靠的结论；第二，功能词研究需要扎实的语言学知识，这方面较为薄弱的研究者必定要在功能词研究中耗费大量工夫。由于这一缘故，研究者对功能词语句结构等形式要素的关注远远不够，在运用语言学相关知识开辟诗歌的韵律、语调或语音研究，以及戏剧的句法、方言、非标准词研究等方面，仍大有可为。

最后，数据解读浅尝辄止。与先前较为模糊、主观的定性分析相比，数字人文的确使文学研究向科学化方向迈进了一大步，但对于数据的深入解读仍属缺失状态。由于研究者将重心放在利用计算机技术处理文本的词汇、句法等微观单元上，运用语言学知识进行分析便自然而然地成为研究方法之首，统计结果也证实我国的相关文献往往以文体学理论为基础，占比高达19%，除此之外还有两种情况占比较高，即没有理论支撑或只有极少数其他学科理论的支撑。反观国外研究，人类学、社会学、哲学、文化学、美学、后现代主义等均与语言学相配合，新视角层出不穷。例如，金（Evgeny Kim）等在语料库基础上运用心理学理论探究同人小说的人物感情交流方式

① 参见耿林璐：《基于语料库分析〈别让我走〉中克隆人的情感》，《比较文学与跨文化研究》2018年第2期，第73~81页。

② 参见闫长红：《基于语料库的〈红字〉研究》，《黑龙江工业学院学报》（综合版）2017年第6期，第120~124页。

③ 参见谢婷婷：《〈伊坦·弗洛美〉的语料库检索分析》，《中国西部科技》2014年第11期，第118~120页。

与特征；<sup>①</sup> Lerena 通过建立小型语料库反驳感伤理论不适用于叙事体裁的观点等。<sup>②</sup> 由于广义的文体学是一门跨学科、多角度的综合学科，在运用过程中常需要相关学科的配合，因此运用文体学对文本进行分析时不宜止步于狭义的文体学理论。国内大部分运用文体学分析文本的著述仅以创作风格或文体特征等为研究对象，并未结合其他学科对其核心内容或深层意义予以阐释，以致对数据的解读不够全面、充分。目前来看，如何正确地解读数据依旧是数字化文学研究的一大挑战，研究者只有广泛涉猎社会人文乃至自然科学等各学科领域，才能充分阐发数据背后的丰富内涵。

任何新事物在诞生之初，总是不完善的、弱小的，目前国内数字化文学研究存在的问题恰是数字人文在我国波浪式前进、螺旋式上升的有力说明，因而对于数字化文学研究在发展过程中偶尔出现的波折，我们应当客观、理性地加以对待，既要防止守旧主义对新事物的盲目歪曲，也要警惕虚无主义对旧事物的彻底否定。

### 三、数字化文学研究的趋势与前景

数字化文学研究发展前景如何？这是学界普遍关心的问题。很显然，数字化的优势和意义决定这一前景的明暗。2021年10月16日—17日在北京召开的“第十届文学伦理学批评国际学术研讨会”的主题为“人工智能时代的文学伦理学批评”，来自世界100余所高校的近500位专家探讨了人工智能时代面临的重大伦理问题，这次会议就是数字化与人文研究全方面融合的直接证据。

在数字技术迅猛发展的今天，具有跨学科、实证性特征的数字人文顺应信息时代符号化和人机对话的趋势，正在成为学术研究的新领域。数字人文的客观性、集约性、高效性、可靠性、运算性和可控性，以空前的规模展示文学活动的真相和全景，正在从整体上逐步改变和影响个体，并逐渐推进科学技术与人文精神的相互融合。这对人类未来更加“诗意地栖居”于世界显然大有助益。科学在过去的时代主要用于征服自然，在生态主义旗帜下，科学将逐渐转向征服自我，而文学恰好是人类最显明的自我。因此，数字人

① 参见 Evgeny Kim and Roman Klinger, *An Analysis of Emotion Communication Channels in Fan Fiction: Towards Emotional Storytelling*, in Francis Ferraro, Ting-Hao Kenneth Huang and Stephanie M. Lukin et al., eds., *Proceedings of the Second Workshop on Storytelling*, Florence: Association for Computational Linguistics, 2019, pp. 56 - 64.

② 参见 María Jesús Hernández Lerena, *Pain and Narrative Shape: Beyond the Indocility of Trauma in Three Newfoundland Novels*, *Journal of the Spanish Association of Anglo-American Studies*, Vol. 41 (2), 2019, pp. 143 - 160.

文不仅为人文事业增添了科学的品格，而且将刷新人文研究的前景。

除了数字化技术令传统人文研究无法望其项背，数字人文的意义也同样不可小觑。工具和手段总是服务于目标和意义，数字人文在具体研究领域的应用也总是受到哲学等人文科学的观念制约。亦即数字化既带来文学艺术活动本身的本体论变革，也将影响到文学艺术中的主体性因素，而本体论和主体论在更高层次的结合与统一，正是文学艺术本身作为自然和人类更加和谐地共同发展的表现。事实上，早在20世纪初柏格森的理论体系中业已显露出综合并超越传统的唯理论和经验论，向全新的本体论和主体论统一方向发展的端倪。其后哲学的主流发展，按照日本学者金森修的说法，便是“意识哲学”和“概念哲学”分别在20世纪中期呈现存在主义和结构主义双峰并峙的格局。<sup>①</sup>

从现代文学的角度看，具有本体论性质的理性主义的兴起与西方近代的观念革命、德国古典哲学和科学哲学的发展密不可分。但是在斯达尔夫人、泰纳等那里，文学研究还带有浓重的社会学色彩，注重的是文学作品产生的时代、种族、环境、习俗等宏观背景，由19世纪孔德的实证主义进到20世纪的形式主义、结构主义，关注的则是文本，这两个方向从文学创作的外部条件和静态结果方面阐释文学，却忽略了文学创作主体，以致造成只见背景不见心灵、只见作品不见作家的倾向，抽掉了生动的创作过程，使文学现象成了无源之水、无本之木。

在另一个方向上，伴随着现代主义思潮的全面内转，轻视创作主体的倾向在20世纪各种强调主体的创作和研究中得到了修正，象征主义、印象主义、意识流、精神分析、表现主义和存在主义等均对此做出了有力反拨，刷新了文学对人的哲学概括和对主体的历史评价。这些努力和探索反映了在新的世界格局，特别是在西方发达国家社会危机日重的形势下产生的新人本主义要求。

然而，历史发展到今天，已经对本体论和主体论的统一提出了明确要求。究竟是歌德创作了《浮士德》，还是浮士德创造了歌德；是塞万提斯创作了《唐·吉珂德》，还是唐·吉珂德创造了塞万提斯。先有鸡还是先有蛋？实际上唯有两种视角的统一才能得出更为合理的解释。如果说这是一种作家本位主义和历史本位主义的联合，那么以当今方兴未艾的生态哲学观看来，对于作家与作品的关系、生活与文学的关系的观点理当发生新的跨越、换上新的视角，即从历史的和作家的本位主义，转向自然本位主义。如此说来，不是浮士德和唐·吉珂德创造了歌德和塞万提斯，或是歌德创作了《浮

<sup>①</sup> 参见[日]金森修：《巴什拉：科学与诗》，武青艳、包国光译，河北教育出版社2001年版，第20~24页。

士德》和塞万提斯创作了《唐·吉珂德》，而是浮士德和唐·吉珂德所从属的自然史创造了歌德和塞万提斯以及他们各自的杰作，因为社会历史只是宇宙运动总过程的一个组成部分。事实上，类似法国哲学家兼艺术家巴什拉的哲学思想和瑞典诺贝尔文学奖获得者托马斯·特朗斯特罗姆的“自然·人文统一观”的诗歌创作都显示出自然与技术共存的新世界观倾向。毫无疑问，从全球性的观念角度看，这一倾向对于克服人与自然的对峙，克服唯生产力论乃至唯物质、科技生产力论，都具有革命性的意义。

总之，从文学的主体论批评出发，研究作家和读者如何共同创作了作品，固然顺理成章；从历史主义出发，研究社会生活基础如何塑造并支配了作家和读者，也自言之成理。但从更广阔视野和更根本的层次看，文学现象犹如自然的参天大树上绽放的审美之花，灌注其营养、支撑其枝干的大地，才是真正深厚博大的基础。因此，就有必要采用更科学、更准确、更高效、更有涵盖力的研究方法和技术手段，在采集和处理更全面的信息数据的基础上刷新文学研究的方向、规模和深度。数字化方法只有在整合起原有的各自为政的科学与人文学研究的力量，联合起技术和诗的宗旨与精神之后，才可望实现理性与情感、逻辑与想象、现实与愿望、实用与艺术、物质与精神的新的统一，进而让从语言学、心理学的转向发展到哲学转向的一切积极成果，诸如本体论与主体论的统一以及新自然哲学的创立，都及时地为文学研究所用。

数字人文，虽然仍名之以人文，却有别于传统人文，因为科学方法的运用将以其无可辩驳的事实极大地改变人文研究的面目。只是这种改变非但不因其科学性而消弭人文精神，反而会在自觉的人文精神支配下使文学的创作与研究增添更多客观性因素和理性因素，使先前已不足以满足现实要求的盲目性和主观性得到有效的克服。当然，莫莱蒂倡导的数字化世界文学图景也必将克服文学沙文主义，有力地扩展世界文学的疆域。

我们并不想夸大数字人文的改造作用，更对方法论崇拜保持着戒心，但毋庸讳言，手段与方法的改变必然对研究的方向和目标产生重大影响。如果我们回顾计算机的应用带给世界以何种改变，以及 ChatGPT 在今天的轰动性效应，就不难理解这一点。因此，我们的结论是：唯有走向这一新的统一，人类开拓的新世界才能是繁荣和炫美的、奋发和公正的、自然的生息永存和人的全面发展的；唯有走向这一新的统一，人类才能从原有的本位主义摆脱出来，超越柏格森的生命哲学、直觉主义乃至创化论，达到以宇宙为本位同时也是更根本的人本主义的历史高度。

（责任编辑：陈华积 权 达）