

October 2022

Media Formation and Paradigm Production: Evolution of the Method of “Distant Reading” and Its Prospects

Nan Zheng

Follow this and additional works at: <https://tsla.researchcommons.org/journal>

Recommended Citation

Zheng, Nan. 2022. "Media Formation and Paradigm Production: Evolution of the Method of “Distant Reading” and Its Prospects." *Theoretical Studies in Literature and Art* 42, (4): pp.104-112.
<https://tsla.researchcommons.org/journal/vol42/iss4/11>

This Research Article is brought to you for free and open access by Theoretical Studies in Literature and Art. It has been accepted for inclusion by an authorized editor of Theoretical Studies in Literature and Art.

媒介构造与范式生产：“远读”方法的演变及其前景

郑楠

摘要：弗朗科·莫莱蒂所提出的“远读”对西方学界产生了极大的影响。学者多将之归结为数字人文研究的一部分，而鲜有人指出远读单独作为研究方法的重要意义。本文提出以“阅读媒介”的概念来理解远读方法的演变过程。阅读媒介是一种信息反馈装置，具有改变读者文本认知的功能。远读方法的演变实则是不同的阅读媒介对研究者认知能力的改变；他人研究将文本转变为专家意见，量化分析将文本转变为数据，建构模型则在量化的基础上将数据转变为文学模型。一系列认知变化带来了研究方法的变化，远读方法在演变过程中融合了多学科思维方式，具备了成为新文科研究范式的潜力。

关键词：远读；阅读媒介；量化分析；建模；新文科

作者简介：郑楠，暨南大学文学院比较文学与世界文学专业在读博士生，主要从事数字人文与科幻小说研究。通讯地址：广东省广州市天河区黄埔大道西601号暨南大学文学院，510632。电子邮箱：zeng_nang@163.com。

Title: Media Formation and Paradigm Production: Evolution of the Method of “Distant Reading” and Its Prospects

Abstract: Franco Moretti’s “distant reading” has had a significant impact on Western scholarship. While scholars mainly categorize it as part of the research in digital humanities, few have noted the importance of distant reading as a research method in itself. This article proposes the concept of “reading media” as an approach to understanding the evolution of distant reading as a method. “Reading media” is a kind of information feedback apparatus capable of changing readers’ perception of a text. The evolution of the distant reading method indicates that different reading media have transformed the researchers’ cognitive ability of texts; that is, other people’s research transforms the texts into expert opinions, quantitative analysis transforms texts into data, and modeling transforms data into literary models based on quantification. A series of cognitive changes have brought about changes in research methods. The distant reading approach in particular has evolved to integrate multidisciplinary thinking and, therefore, has the potential to become a new paradigm for liberal arts research.

Keywords: distant reading; reading media; quantitative analysis; modeling; new liberal arts

Author: Zheng Nan is a Ph. D. candidate in the College of Liberal Arts, Jinan University. His areas of academic specialty cover digital humanities and science fiction studies. Address: College of Liberal Arts, Jinan University, 601 West Huangpu Avenue, Guangzhou 510632, Guangdong Province, China. Email: zeng_nang@163.com.

在传统的文学研究中，文学作品常被视为一种浑然天成的艺术奇观，是作家伟大心灵的文字反映，因此研究者需要依赖“细读”（close reading）来理解文学作品，即“对文本的语言、结构、象征、修辞、音韵、文体等因素进行仔细解读”（赵一凡 630）。因为能够让研究者对文本产生精细独到的解释，所以细读在很长时间内都是文学研究的范式。但由此也带来了一个显而易见的问题，正如雷蒙德·威廉斯（Raymond Williams）所言：“没有人真的

懂19世纪的小说，因为没有人读完了或是能够读完从单行本到报纸连载小说在内的全部作品。”（58）确实如此。假如说细读几十乃至几百本小说尚具有可行性，但当研究对象的数量扩大到成千上万本时，细读便显得捉襟见肘。

细读的局限性正是本文研究“远读”（distant reading）方法的一个切入点。本文指出，造成这一局限性的原因并非在于细读方法本身，而在于研究者对于研究对象的

认知。对文本的认知方式决定了文学的研究方法,因此想要解决运用细读方法时所面临的困境,研究者需要转变自身对于文本的认知。而基于细读在面对大量文本时不可避免的局限性,有学者提出应该尝试由细读改为远读,远读的文学研究方法应运而生。远读在何种意义上改变了研究者对于文本的认知?它能弥补细读的缺陷吗?远读本身又有怎样的特点和问题?它有可能成为未来文学研究的新范式吗?本文将尝试对这些问题作出学理化的探讨。

一、远读 1.0: 阅读媒介

阐释远读,必须从斯坦福大学教授弗朗科·莫莱蒂(Franco Moretti)谈起。2000年,莫莱蒂在《新左翼评论》(*New Left Review*)上发表了题为《对世界文学的猜想》(“Conjectures on World Literature”,下简称《猜想》)的文章,并在其中明确提出了远读的概念:“远离阅读。我要重申的是:距离,是认知的条件。有了距离,你就可以关注那些比文本小得多或大得多的单位:手法、主题、修辞——或是体裁和系统。”(莫莱蒂 12)此后,远读概念以及这篇文章受到越来越多的关注。截至2021年7月21日,在谷歌学术(Google Scholar)上,《猜想》一文的被引量已经高达2614次。另外,以在Web of Science的核心合集中以“distant reading”为主题而检索到的247篇论文为例,经由VOSviewer软件做作者共被引分析(author co-citation analysis),发现莫莱蒂已经成了理解远读的关键人物。因此,本文试图将莫莱蒂的《猜想》作为考察当代远读的起点,以此探析莫莱蒂的远读究竟为何会引起学界的高度关注。

在《猜想》中,莫莱蒂用自己的方式承接了威廉斯当初所提出的问题:“为人所知的19世纪英国小说就有3万部,而还有4万、5万、6万部不为人知,更未曾读过,将来也没人会去读。”(9)也就是说,第一,对于任何试图进行文学总体研究的学者而言,研究对象的体量问题是显而易见的,研究者无法穷尽所有的研究对象,他们所细读过的文本必然远远少于其未细读过的文本。第二,在海量的文献中选择可进行细读的研究对象,研究者的筛选标准必须是“为人所知”,即研究对象要与文学的经典化(canonization)密切相关。正所谓“除非你觉得极少数文本确实重要,否则你不会花费那么多气力去研读”(11)。但问题在于,即使研究者只选择各国的经典文本去展开文学讨论,研究对象也未必能被研究者尽数细读,更何况关于“何为经典”的讨论又涉及另一个巨大的、边界含糊的学术难题。莫莱蒂认为,如果仅靠传统的细读研究法来研究文学,“世界文学”就很难有真正来临的那一天。为此,他提出了一种探索“世界文学”的可能性,即远读。

在《猜想》中,莫莱蒂展示了自己依据众多民族文学

专家的研究材料而进行远读的过程。他首先借用了“世界体系理论”(The Theory of World System)的经济假设,即世界资本主义体系“同为一体,但不平等”(10)。由此,或许可以认为,在相互关联的世界文学体系内,强势的中心文学(西欧文学)会入侵并改变弱势的边缘文学(非西欧文学)。为了验证这一猜想,莫莱蒂以“西方小说形式如何影响本土文学内容”为切入点,不直接阅读文学文本,而通过阅读二十多种文学批评专著,最终从众多民族文学研究者的研究中得到了一个共同点,即大量“边缘地区”(指非西欧国家)的文学,在特定时期都受到过“中心地区”(西欧国家)小说形式的影响,而外来的“西方形式”在遭遇“本土内容”时,双方会相互妥协。妥协的方式与程度虽各有不同,但这一普遍规律却足以证明:世界文学是一个不平等的整体。

可以看到,一开始的远读方法并没有涉及任何量化分析(quantitative analysis)。莫莱蒂所做的,不过是“把别人的研究成果拼凑起来”(11),并从中分析出小范围的细读所难以反映的文学规律。这其中的远读更像是对于他人研究的“综合”(synthesis)。大概也正是因此,在《猜想》一文发表后,学界的讨论更多聚焦在莫莱蒂对“世界文学体系”的描述上(Moretti, “More Conjectures” 73)。有学者认为,所谓的远读方法并没有像莫莱蒂宣称的那般具有革命性(Spivak 107-109)。因为远读方法似乎只是一种变形的文献综述,综述在传统文学研究中并不罕见,而莫莱蒂所采用的民族文学研究也有对其他研究的借鉴与总结。伊利诺伊大学厄巴纳-香槟分校的学者泰德·安德伍德(Ted Underwood)就认为,远读“是一种将历史知识探究描述为实验的实践”(33)。还有的学者据此把远读归为“数字人文”(digital humanities)研究的一种。

本文认为,上述说法虽有一定真知灼见,但其并没有深入探讨远读的方法论意义,甚至部分混淆了远读在不同阶段的内涵。更准确地说,远读是一种借鉴科学实验思维的文学研究方法,与纯粹的文学研究不同,它正是通过“阅读媒介”实现了“假设-验证”的实验闭环,它具有鲜明的实验性。

所谓阅读媒介,指的是改变研究者认知文本方式的信息反馈装置。我们知道,远读方法的全过程是,先由研究者对文学文本进行理论上的假设,再由阅读媒介根据研究假设对文学文本进行特定的处理,然后反馈相应的文本信息,最后由研究者通过分析该信息以验证假设是否成立。在这一过程中,研究者对文本的认知逐渐从原始的文字材料转变为反馈的信息类型。为便于理解,本文以阅读媒介为中心,构建了《猜想》中远读方法的分层模型,如图1所示。

我们可以从整个远读的分层模型入手来进一步理解该流程。《猜想》中的远读方法实际上出现了金字塔型的三个层次。底层为原始材料,即整个远读研究所涉及的

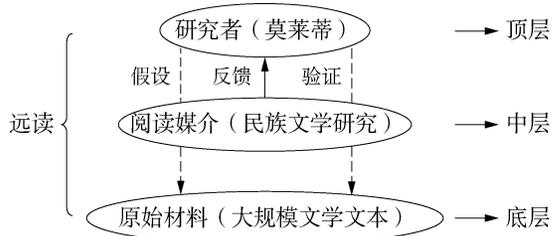


图1

大规模文学文本,中层则是阅读媒介,即莫莱蒂利用的二十多本民族文学研究专著,顶层则是作为研究者的莫莱蒂。从这个模型来看远读的具体运作流程,我们或许更容易理解远读的实验特性。所谓实验,即“在受控条件下进行的检验,以证明已知的真理或检验假设的有效性”(Muijs 13)。“假设-验证”是实验方法区别于其他研究方法的关键所在。结合波普尔的证伪理论,科学研究需要坚持对研究假设进行检测。只有当研究假设可被检验时,研究者才不会坚持不可靠的研究前提,否则研究者就有可能为了符合假设而歪曲事实或经验材料。而之所以称远读为“文学实验”,关键就在于阅读媒介,正是这一反馈装置使得远读方法中“假设-验证”的流程得以完成。具体到莫莱蒂的远读方法,这个过程逻辑就是:“世界文学体系”是莫莱蒂的研究问题,“现代小说是西方小说形式与本土内容相互妥协的成果”则是莫莱蒂的研究假设。莫莱蒂不需要在具体的文本语境中证明这一假设,而是通过收集不同的民族文学研究组成阅读媒介,这一阅读媒介则根据莫莱蒂的假设,反馈了“西方小说形式同本土内容存在结合问题”的信息。莫莱蒂根据这一反馈的信息,验证了研究假设最终成立,从而得出世界文学同一而不平等的结论。

尽管传统文学研究中有研究设想,但却鲜有形成基于信息反馈的验证机制,因而研究者很难清晰地知道自己的研究设想是否可靠,或存在哪些问题。特别是依赖细读方法的研究,其文本分析更多聚焦在对研究设想的阐释,虽然这一做法丰富了文学研究的进路与结论,但在缺乏明确验证机制的情况下,很容易出现“强制阐释”,即用理论框定文本,致使文学研究频频陷入寻求自圆其说的僵局,而这一做法显然无益于文学知识的生产。莫莱蒂在《猜想》中所展现出的“假设-验证”机制,无疑证明了文学研究能够引入一种更为科学的实验方法。

关于远读的理论设想是非常美好的,但在实际运用中却存在不少困难。最重要的一个困难,就是缺乏对他人研究成果的选择标准。因为即使是针对同一个研究对象,不同研究者的研究思路和结论都有可能差别很大,而在不进行细读的情况下,似乎很难作出客观公正的筛选。莫莱蒂的《猜想》也没有给出解决方案。不仅如此,莫莱蒂在文中所使用的民族文学研究大多为英文著作,因而在关乎方法与主题等内容性的选择标准之外,还存在着

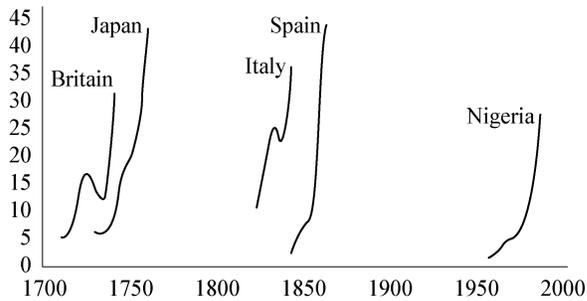
语言这一层筛选机制。是否有其他尚未进入英语世界的民族文学研究,其与莫莱蒂所引用的著作相比,彼此的阐释恰好相悖?对于因语言而产生的研究障碍,这更是一个无解的难题。第二个困难就是对他人的研究难以验证。正如有学者指出,《猜想》中的远读其实是一个基于信任的方案(Parla 123)。莫莱蒂所利用的民族文学研究,归根结底是他人细读文本的产物。由于莫莱蒂并没有直接面对底层的文本,因此他必然难以亲自判断他人的细读在何种程度上是可靠的。因此,莫莱蒂对他人研究的采用只能建基于信任之上:信任文学专家的文本收集与文本分析,信任科学共同体对相关成果的评价,唯独缺失属于研究者基于文本信息的自身判断。

不难发现,上述两个弊端均集中于作为阅读媒介的他人研究。这也是本文强调阅读媒介是一个反馈装置的重要原因。民族文学专家的细读,其反馈的不仅是文本包含的信息,更是专家的价值判断。研究者需要依靠解读这些反馈,才能完成对于假设的验证。假如研究者既无法解释对人为价值判断的选择标准,又无法验证反馈信息的真实性,那就如同在远读方法中引入了一个信息处理的“黑箱”(black box):一个具有信息反馈功能,却不清楚内部运作结构的筛选系统。这便是《猜想》中远读所遗留下来的挑战——其在实践中仍存在着“信息黑箱”。修复阅读媒介的弊端,进一步优化其性能,成为改进远读方法的核心任务。

二、远读 2.0: 从量化分析到建构模型

发表《猜想》之后,莫莱蒂于2003年又在《新左翼批评》上发表了《图表、地图、树型图(一)》(“Graphs, Maps, Trees-I”),并于次年展示了第二、三部分的成果,最终在2005年集集成同名书籍(下文简称《图表》)。《图表》是莫莱蒂的另一部代表作,展示了《猜想》之后莫莱蒂对于阅读媒介的改进,即将他人的细读替换为量化分析。具体而言,为了研究文学史中“小说的崛起”(the rise of novel),莫莱蒂将目标聚焦于书籍史(book history)——对书籍文本的印刷、发行与流通等的研究——通过量化书籍史研究中的数据,统计不同国家小说出版的时间、数量以及地点等信息,并将之形成图表、地图以及树形图。如图2所示,在统计了英国、日本、意大利、西班牙以及尼日利亚分别在1700年至2000年间小说的出版数量后,莫莱蒂将数据绘制成图表。借助图表,莫莱蒂发现,上述五个国家的小说发行都经历过一个快速上升的阶段。

用量化分析来取代《猜想》中他人的细读,解决了《猜想》中阅读媒介难以被检验的问题。正如莫莱蒂所言:“量化数据之所以有用,是因为它们独立于解释。”(Graphs, Maps, Trees 30)数据的统计方法是清晰而明确的,能够单独地被其他研究者分析和理解,而数据本身则

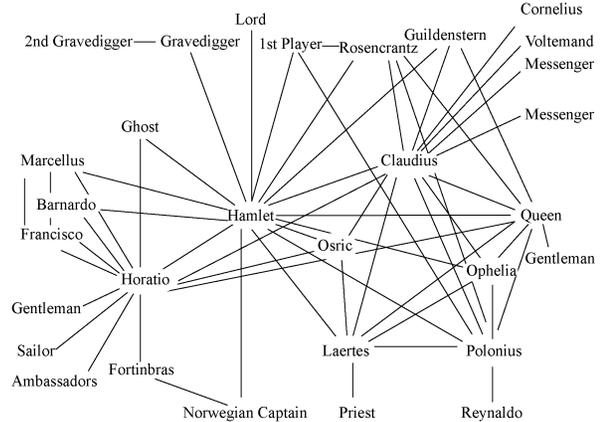
图2 (Moretti, *Graphs, Maps, Trees* 6)

是客观存在的,不因不同研究者的阅读而发生改变。因此,不同于细读所隐含的价值判断,量化分析这一阅读媒介所反馈的,是关于文本的客观说明。

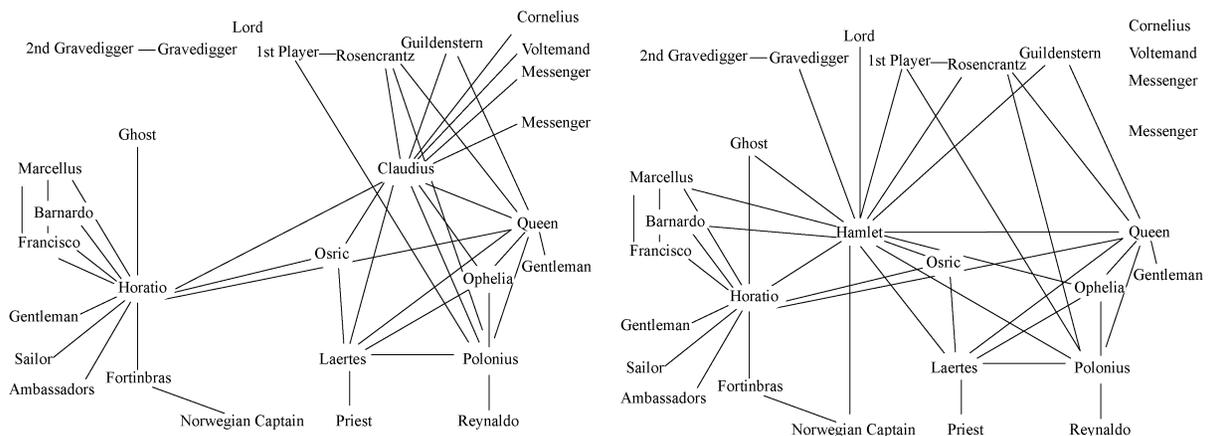
文学研究的量化分析也成为莫莱蒂及其团队后来的工作方向。2011年,莫莱蒂出版了论文合集《远读》(*Distant Reading*),其中收录的《风格公司:对1740至1850年间七千本英国小说标题的反思》(“Style, Inc: Reflections 7000 Titles [British Novels, 1740-1850]”)一文就引入了考察小说标题字数的平均数和中位数的统计方法,通过计算分析1740年到1850年随着出版市场的扩大,小说标题字数呈现出下滑的趋势,并综合当时书籍购买者的喜好与标题长度所带来的广告效益等方面的思考,研究者得出了市场对小说呈现方式(the presentation of novels)存在影响的结论。(*Distant Reading* 179-210)

上文提到,安德伍德更倾向于将远读视为一种文学史研究,因为文学史研究更多集中于对文学规律的研究,相比对文学作品的美学研究而言,这种观点认为远读更追求“普遍客观”,这也使得在文学史研究中运用量化分析显得更为合适。这个看法固然有其合理性,但实际上只关注到远读探索文学史规律的案例事实,而忽略了远读中阅读媒介在改变文学认知模式上的潜力。在《远读》的最后一篇论文《网络理论,情节分析》(“Network Theory, Plot Analysis”)中,莫莱蒂将远读运用在对《哈姆雷特》的情节分析上。通过对人物对话的量化计算,莫莱

蒂设计出《哈姆雷特》的人物网络图,如图3所示:

图3 (Moretti, *Distant Reading* 213)

在上述过程中,作为阅读媒介的量化分析先将文本转化为数字,而后再通过可视化呈现将数字转化为网络图谱,这便是阅读媒介实现认知改变的方法——研究者眼中的小说从线性文字变成了空间网络。莫莱蒂介绍了网络图谱的四个优势:一、将历时性的文字阅读变成共时性的网络结构;二、情节可视化;三、构建文学模型;四、通过文学模型进行“文学实验”。(*Distant Reading* 215-220)更为关键的是,可视化网络能够直接呈现量化数据难以呈现的事实。如图3所示,尽管哈姆雷特(Hamlet)和克劳狄斯(Claudius)在网络中心性的数值上相差不远(中心性等于与其他网络节点最短距离的平均数,相邻两个节点的距离为1):哈姆雷特为1.62,克劳狄斯则是1.45,似乎二者同样是故事的中心人物,但分别将二者抽离出网络后,可以发现,哈姆雷特对网络的影响力要比克劳狄斯大,这证明了哈姆雷特才是故事的主角。对数据可视化的使用使得远读不只限于大规模的文学文本研究,也使得作品批评得以展开。利用网络图谱,远读可以直观地呈现数据统计所隐藏的内部联系,如图4所示:

图4 (Moretti, *Distant Reading* 221)

尽管网络图谱对原先的问题有所改进,不过以上研究仍然存在问题。虽然量化分析能够确保阅读媒介反馈信息的客观性,但对文本的量化往往只是为了用数据来验证已有研究的正确与否。而事实上,不需要关系网络,读者也能够轻易判断出哈姆雷特才是文本的主角,这样的远读方法所能带来的研究意义极为有限。同时,目前的阅读媒介大多是简单的量化分析方法,研究者难以利用这些数据产生新的文学知识。为何选择此种数据的变化来证明研究假设,而非另外一种?这需要研究者进行说明与解释,但选择依据依旧很少出现在远读的研究中。美国罗格斯大学学者安德鲁·戈德斯通(Andrew Goldstone)就指出,实践远读的研究者未能系统地说明分析特定数据的理由何在(“Distant Reading: More Work to be Done”)。上述莫莱蒂的《风格公司》便是一例,加拿大麦吉尔大学学者苏真(Richard Jean So)认为,《风格公司》在数据的处理上仍有缺陷,理由在于莫莱蒂对数据的处理更像是由直觉所驱动的(“All Models Are Wrong” 671)。利用平均数和中位数确实能展示出某种数据的变化,但却没办法告诉研究者这一变化趋势的强度,有可能上述两个指标所展示出来的变化只是某种暂时性的噪音(noise),因而莫莱蒂在这一文章中的分析方法显得不够系统化(“All Models Are Wrong” 670-671)。综上所述,量化分析需要运用多种数理知识,但文学研究者往往缺乏这一方面的知识,从而导致采用的量化分析方式过于简单与唯一,并且缺乏选择依据。这是造成在文学研究中运用量化分析难以产生高价值数据的重要原因。因此,深化对量化分析的认识,是避免远读变成简单数字统计的重中之重。

如何深化对于文本的量化分析?苏真给出的建议是建模(modeling)。针对《风格公司》的直觉驱动问题,苏真提出了一个关于方法上的建议,即在计算机上利用回归模型(regression model)对数据进行再处理,分析小说标题随年份递增后字数上升或下降趋势的强弱。通过建立回归模型,苏真为《风格公司》引入了拟合(fitting)的概念,研究者可以根据模型拟合数据的精确度作出模型的调整,由此分析统计数据所呈现出的真正趋势。(“All Models Are Wrong” 671)

如果说量化是将文本转化为客观数据的集合,那么建模则是在数据集合中搭建结构关系。不同的量化方式会使文本产生不同的数据,在这个意义上,文本能够产出的数据集合是多维的(multi-dimensional)。但假如研究者只关注单一维度的数据,只因该数据的变动幅度非常大,就宣称该数据的变化代表了整个数据集合的变化,这就很容易产生错误的研究结论。因为有可能该数据本身的权重很小,因此用它来进行整体评价是不恰当的。只有当研究者利用多维数据建构出模型,并通过模型挖掘出多维数据的结构关系,才能从整体上去把握某个数据的变化是不是对于整体具有显著意义,也才能判定数据与

数据之间呈现出何种联系。如此一来,研究者对于数据的选择与分析会更加系统化,更加有据可依,而不是仅凭自身直觉来决定研究方向。

除了对系统性的强调外,建模的另外一个意义在于迭代性。模型,本质上是对研究对象的一种数学关系表达,是一种理想化的产物,因此在实践中,模型并不一定能准确地描述文本,只能尽量地去接近文本。知名统计学家乔治·博克斯(George Box)的名言“所有的模型都是错的”(All models are wrong)便是对这一事实的高度总结。但这并非意味着模型就是无效的,恰恰相反,模型的有效之处就在于其“无效”。换言之,建构模型的一大意义便在于寻找模型所反馈的错误。模型中出现的错误一般分为两种:在代入数据后,要么模型中出现了研究假设认为不应该存在的现象,要么没有出现假设认为应该存在的现象。在上述两种情况中,或者是建构的模型出现了问题,或者是研究假设出现了问题,不管是何种情况,都迫使研究者不得不重新调整自身建构的模型或假设,使之更加符合数据的情况。这种调整并非一次性的,而是永无止境的,因为模型只是对象的一种表征,它虽不是真实的对象,但却可以无限地接近对象,而正是建模这种不断在假设、模型、数据之间来回往返的迭代性,才使得模型得以揭示人类凭借直觉所感知不到的事实。

苏真本人使用具体的实践展示如何理解建构模型的意义。他和芝加哥大学的霍伊特·朗(Hoyt Long)联手,在《文学模式识别:细读与机器学习之间的现代主义》(“Literary Pattern Recognition: Modernism between Close Reading and Machine Learning”)中进行了一次以俳句(haiku)为对象的建模研究。具体而言,苏真和朗首先发现,通过传统的文本细读法进行俳句的本体研究,无法同通过历史主义批评的俳句传播研究很好地结合起来,更无法揭示出俳句确立为俳句的本质性和特殊性,究竟在它从日本传播到西方的过程之中发生了什么变化。两位研究者试图找出一种能够展示文体形式层面与历史文化层面之间相互影响的研究方法。为此,苏真和朗构建了一个在传统文本细读中被确定为俳句的英语俳句语料库(由翻译作品和改编作品组成)与一个非英语俳句语料库,他们先为俳句做了一个通用的文本表示模型(词包模型),并给出一个关于区分俳句与非俳句的假设:俳句可以通过共有的措辞和音节数模式来区别于非俳句。在这个假设的基础上,他们运用机器学习(machine learning)让电脑程序在两个语料库自动识别分类依据。苏真和朗发现,模型对俳句翻译作品的判断率较高,改编作品却较低。他们一开始怀疑原因是假设中对模型的限定条件过多,于是删除了音节数的特征。但再进行测试后却发现,在考察条件变得更加宽松后,机器的判断率却下降了。这种分类错误反而揭示了俳句和非俳句语料库具有重叠之处的可能。苏真与朗通过对这些错误分类的重新审

视,发现了在美国诗歌领域中可能存在更为广泛的“东方主义共鸣”。这种作为统计学意义上的俳句模式为研究者重新描绘了俳句的定义与传播路径,并提示了一种在流传过程中所保持的“俳句性”。

在上述例子中,苏真展示了不断调整假设与模型,最终完成文学批评的全过程。这种批评并非一次性完成的,而是要通过不断的迭代才能达到。部分学者提出以“尺度阅读”(scalable reading)来形容这种研究,即研究者“形成一些松散的猜想→大规模的文献挖掘(所谓‘远读’)→找出离群值→通过回归细读来精准定位→提出随意阅读难以发现的问题——又或者,继续在细读、远读间回返,不断精确模型,无限逼近想要再现的问题”(赵薇 46)。麦吉尔大学学者安德鲁·派博(Andrew Piper)就在其《枚举:数据与文学研究》(*Enumerations: Data and Literary Study*, 2018)中用示意图展示了此种远读与细读的结合,如图 5 所示:

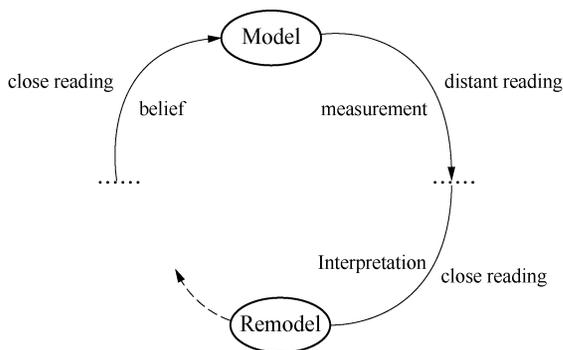


图 5 (Piper 10)

本文认为,上述说法有值得商榷之处。基于本文在图 1 提出的远读分层模型,这种循环往复的研究实际上一直发生在远读方法的内部,而非如图 5 所示是远读、建模与细读三个独立部分的组合。正是因为目前关于远读的研究并没有清晰地提出阅读媒介这一概念,才使得学界对远读方法产生众多的误解。远读从来都不是与细读相互割裂的研究方法——从远读 1.0 到远读 2.0 的演变中,远读一直能够包含细读,而这一包含方式,正是通过阅读媒介实现的。

首先,在远读 1.0 中,莫莱蒂的《猜想》其实有着两个层次的细读,第一重是作为阅读媒介的民族文学专家对底层文本的细读,第二重是莫莱蒂对于作为阅读媒介的民族文学专家研究的细读。不难发现,《猜想》中的第一重细读是为第二重细读服务的,换言之,远读 1.0 本身可以视为经由阅读媒介这一概念所连接起来的双重细读。

其次,在远读 2.0 中,当阅读媒介被置换为更加“客观”的数理方法之后,细读的方式和意义均发生了改变。例如,通过阅读媒介对文本的转换,细读可以无须读“字”,而是读“数据”和“图”。如上文所示,对文学文本从“字”到“数据”与“图”的认知变化无疑打开了一个更

为广阔的文学思考空间。同时,远读 2.0 中的细读意义并不在于探讨个别文本的神圣之处。与之相反,此番细读实际上是在探究一个整体结构中的异常之处。细读异常能够给予我们关于整体更深刻的洞见。正如学者吉云飞提出的,数字时代下文学研究一大张力在于“对局部的理解总是离不开对整体的理解”(181),这已然揭示了远读与细读的辩证法——如果说基于计算的远读 2.0 是对研究对象的系统性把握,那么细读就是对这一系统把握的不断调整。而这种调整机制,正是通过引入跨学科的阅读媒介实现的。作为当代远读方法的持续探索者,苏真的看法呼应了上述观点:“文本细读的结果为我们反思远读方式的局限性提供了关键的机会。”(“Race and Distant Reading” 72)所以,在远读 2.0 中,局部细读是系统远读的调整基础,后者始终包含着前者。

更进一步来看,将远读同细读对立起来的态度,在一定程度上也反映了对学科范式转变过程的误解。一提起范式转变(paradigm shift),人们往往会联想到托马斯·库恩(Thomas Kuhn)所提出的“科学革命”,仿佛范式转变的过程就是一场范式之间的权力斗争,例如有数字人文学者表示,数字人文研究(远读)需要融入主流人文学界,展示其在人文研究领域的领导力(Liu 495)。这种说法其实是变相地表达了数字人文研究(远读)与传统人文研究(细读)存在竞争关系的观点。问题在于,范式转变并非逼迫研究者作出非此即彼的选择。理解范式转变的意义,要根植于库恩如下的判断:“范式改变的确使科学家对他们研究所及的世界的看法变了。”(94)新范式的出现意味着我们关于对象的认识方式切实发生了改变,但并不代表过去的研究方法全然失效了。范式转变的意义更多在于充盈我们对研究对象的理解,而不在于掩埋我们到达今日这一认识现状的道路。

回到远读与细读的问题上,本文强调这并不是两种截然独立的研究范式。归根结底,远读是一种多学科研究方法的融合,传统细读自然是其中一种。而理解阅读媒介所扮演的角色,则是理解这一方法融合的关键所在。在如今的数字时代,与其说媒介是人的延伸,不如说人分布于媒介之中。在认知科学领域,已有众多学者提出以“分布式认知”(distributed cognition)来解释人类与技术的交互方式。所谓分布式认知,指的是主体对客体的认知发生不仅局限于个体之内,而是分布于媒介、文化以及社会等多种外部因素中(Michael Engestrom 42-43)。认知能力的分布并非简单地将研究者的认知任务委托给搜索引擎、数据库以及文献分析软件等种种现代媒介,而是意识到“认知任务分布于内部表征和外部表征之中,分布式认知活动源于二者的交互作用”(周国梅 傅小兰 150)。而细读融入远读的方式,正是研究者通过阅读媒介从而与文本进行多次交互所实现的:研究者可以利用远读从更大范围去验证细读的猜想,而通过细读,研究者反过来

不断地调整远读的方向。因此,远读对细读的涵盖亦可转换为如下表达——远读是在阅读媒介加持下的交互式细读。远读不是对细读的取代,而是对细读在特定使用条件下所存缺陷的弥补。远读的出现宣告了新时代中细读的变化,而非细读的消亡。而基于上述论证,本文也确认了文学研究者之主体性在远读实践中的重要地位。文学研究者作为远读方法的顶层存在,其主观能动性在远读方法中仍然非常重要。没有研究者的细读,就没有不断精进完善的远读。如果失去身处顶层研究者的主体构思,仅凭阅读媒介不可能产出有价值的文学研究。

综上所述,阅读媒介是远读方法中具备改变认知功能的信息反馈装置。他人研究、量化分析和构建模型是阅读媒介三种不同的呈现方式,这三者都使得读者意识到文学文本并非必须依赖亲自细读才能体会个中真意的“神圣文字”,文本可以从多个角度进行理解——他人研究将文本转化为专家意见,量化分析将文本转化为数据,建构模型则是将数据转化为系统。前两者因为缺乏可验证性、系统性和迭代性,因而产出的是一种静态的认知改变。构建模型由于兼具这三个特性,因而是一种动态的认知改变。也就是说,研究者通过中层阅读媒介所反馈的信息,去查看底层“原始材料”中为何会出现离群值,或调整阅读媒介,或调整假设,最终产生有价值的文学批评。显而易见,这种文学批评是通过改变研究者对文本的认知方式所达到的,离开阅读媒介这个概念,很容易将远读仅仅理解为一种数字运算的过程。

三、远读的未来:新文科的研究范式

2019年4月29日,教育部、中央政法委以及科技部等多部委在天津联合召开“六卓越一拔尖”计划2.0启动大会,建设新文科成为热点话题。结合目前学界对新文科的讨论,新文科之“新”,在于其融合新兴产业领域、转变传统研究范式以及适应社会发展需求(王铭玉 张涛)。而作为融合了文学、统计学以及计算机科学等多学科研究方法的远读,正有成为新文科研究范式的潜力。

当然,想要让远读符合建设新文科的要求,我们仍有不少问题需要解决。首先,最关键的因素,即文学研究者该如何提出合适的问题。所谓合适的问题,指的是需要利用以及能够利用远读去解决的问题。例如鉴赏陶渊明《归园田居》中的诗人心态,由于其文本量过小,加之心态这一目标难以量化,可能就不太适合使用远读。而大规模文本的问题就非常需要远读的方法,就像文学史研究一样。本文甚至认为,大部分文学史研究,都有可能在运用远读后被重新改写。

而在提出合适的问题这一方面,目前国内较为常见的一种做法是,组成一个跨学科的团队,由文学学者提出问题,再由统计学者将之转化为统计模型,然后由计算机

学者进行编程和计算,最后再由文学学者进行阐释(项蕾等 175)。团队合作的方式能够较好地弥补文学研究者缺乏数理知识背景的缺陷,但这也涉及一个难点,就是从文学学者所提出的文学问题,到统计学者(或者计算机学者)的实践操作之间,究竟该如何转换。为了填补这一学科之间的鸿沟,文学研究者是否需要了解一定的统计与编程的概念及方法?

在本文看来,这一回答是肯定的。远读既然要发展为新文科,自然应当是由人文学者主导研究,而在这种情况下,主导者的跨学科思维就显得非常重要。完全依靠团队中理工科出身的学者来完成具体实践是不够的,从苏真对俳句的研究中就可以看出,光是模型的迭代就同时需要文学、统计学以及计算机科学等多种学科知识的有机结合,这并不是一段段能够独立切割开来的任务流程,并不是说人文学者提出问题后就可以高枕无忧,等着统计学家与计算机学家生产出数据后再进行分析。跨学科知识的结合是全过程的,而参与这一过程必然要求人文学者理解和掌握更多其他学科的知识。

虽然这并不意味着文学研究者就必须成为统计专家或者计算机专家,但跨学科的知识掌握究竟需要多少才算足够?这倒确实是一个亟待解决的问题。以编程这个问题为例,目前学界对此就有明显的争议。例如,有学者认为团队合作就足够了,人文学者并不需要掌握编程技术(Ore “To Code or Not To Code”)。而朗则建议人文学者需要了解一定的编程知识,同时也要具备实际操作的能力(朗 苏真 林懿 52)。本文认同朗的观点,而这也是本文提出阅读媒介的基础,即不同的认知工具会带来不一样的认知结果。莫莱蒂本人曾表示过编程对改变认知方式的重要性:

我在年轻的研究生和同事身上看到,编码赋予了他们一种我所不具备的,而且将来也不可能获得的智慧和直觉。这种智慧体现在脚本编写上。但是,在编写脚本的过程中,某个概念也会逐渐成形。虽然这个概念往往不会以概念的形式呈现,但是你能够看到它就隐藏在编写的脚本里[……]因此,我认为开展数字人文研究项目的高校,不管这些项目是大是小,都应该确保每个人都有机会获得这种智慧。(丁斯曼 莫蕾蒂 36)

这可以证明,编程或建模不仅仅是一种分析方法,更是一种认知工具。正是基于对认识方式的改变,研究者对诸如风格或流派划分等问题才可能有全新的认识,因为类似的问题完全有可能转化为统计学意义上的词语特征问题。斯坦福文学实验室(Stanford Literary Lab)所发布的《小册子4》(Pamphlets 4)就佐证了这一见解。在这一名为《2958部19世纪英国小说的定量文学史:语义群集法》(“A Quantitative Literary History of 2,958 Nineteenth

Century British Novels: The Semantic Cohort Method”)的文章中,莫莱蒂的两位研究生利用算法提取并分析出近3000本小说中的词义群变化,判断出19世纪英国小说叙事模式经历了从“讲述”到“展示”(from telling to showing)的变化,这一结果亦符合当时不断扩大的社会空间的模糊性与流动性特征。因此,本文认为人文学者至少要了解其他学科的核心研究范式,从而获得一种改变自身认知文学的能力,才有可能在学科融合的基础上提出运用远读才能解决的重要问题。

亟须解决的第二个问题是文学文本数据的匮乏。尽管量化或建模的工作可以利用人工完成(莫莱蒂的《哈姆雷特》关系网络就是手工绘制的),但目前越来越多运用远读方法的研究已经需要依靠计算机来完成量化与建模的工作。而运用远读的前提是计算机必须能够“读到”文本,这就涉及文本的电子化。从目前来说,获取电子化文本有两种方式。第一种是利用现成的文本数据库,例如谷歌图书(Google Books)、青空文库(Aozora Bunko)、古登堡计划(Project Gutenberg),等等。但现有的数据库未必包含研究者想要研究的文本,因此第二种方式则是研究者将想要研究的纸质文本转化为电子文本。而为了实现纸质文本的电子化,研究者需要先扫描纸质文本,然后将获得的扫描图像进行分析处理,最终将其转化为计算机能够识别的字符,这一过程便是光学字符识别(optical character recognition,简称OCR)。

然而当研究对象多达成千上万本,甚至需要构建一个数据库的时候,OCR就非常耗时耗力,这种工作往往需要以年为时间单位进行计算,这也是目前文本数据匮乏的重要原因之一。同时,OCR的识别率很大程度上依赖于文字检测算法。就目前而言,OCR对现代印刷品的识别率较高,但对较为久远的印刷品以及纸稿的识别率就比较低。除非文字检测算法能够取得更加重大的进步,否则当面对无法进行OCR的文本时就不得不考虑手动输入,而这又加大了文本电子化的难度。以上两种情况均可以考虑用外包(outsourcing)或者众包(crowdsourcing)的方式来处理。外包,就是将文字识别、转录、矫正等工作交由专业的文字识别机构来完成,而众包则是将该工作以自由自愿的形式交给公众志愿者来完成,例如上海市图书馆开展的盛宣怀档案抄录项目,选取盛宣怀档案中与辛亥革命相关的日记、文稿以及信札等纸质材料,供广大民众在线认领转录。

最后便是研究成果的评价问题。因为远读方法本身是一种多学科方法的融合,因此对采用了远读的文学研究的评价工作就不能由单个领域的评审专家来做。例如,传统文学论文期刊评审文学论文的过程基本上要先由编辑初审,再交给对应领域的文学专家多轮盲审,最终决定刊发与否。然而作为一个多学科合作的文学研究成果,单凭文学专家又如何能够评判该成果所包含的统计学

运算?同时,运用远读的文学研究很多时候会涉及计算机代码的运算,而这些代码往往是研究成功的关键所在,那么期刊的文学专家又该如何基于这些代码去复现研究过程是否可靠?因此,面对新兴的远读方法,文学期刊或许需要扩大盲审专家的学科领域。同时,学界还需要创建更多有能力评价人文学科与理工学科交叉成果的期刊,例如清华大学主办的《数字人文》与中国人民大学主办的《数字人文研究》,等等。学术期刊之所以重要,是因为其是学术研究成果最为核心的展示平台。缺乏更多合适的平台,远读便难以进入学界的视野中心,其中蕴含的多学科研究方法便难以得到推广与理解,学科融合更将举步维艰。

不难发现,远读方法极大地改变了文学领域的研究思路。本文试图指出阅读媒介是内嵌于远读方法中具备改变认知功能的信息反馈装置,以此理解远读方法的出发点。研究者通过阅读媒介所反馈的信息,改变自身认知文学文本的方式,进而改变了文学研究的方法。从他人研究、量化分析到建构模型,阅读媒介的进化是远读方法演变的核心。在这一演变过程中,远读方法展示出对跨学科研究方法的融合,并揭示了这一融合的价值,最终指明了传统文科向新文科转变的可能路径。文学研究者只有深入地去了解远读方法的核心,而不是止步于量化、数据与建模等概念,才能够理解实践远读方法的关键,以获得远读方法对文学研究范式转变的启示。

同时,我们也需要意识到,远读对整个文学学科的建制提出了更新的要求。建设新文科是文学学科发展不可逆转的潮流,这是文学在面对当代科技冲击时所作出的积极反应。不过新文科的建设并非是一朝一夕便能够完成的,从远读方法的演变过程中可以发现,学科间不同方法的借鉴与融合是一个需要反复探索的工作。为此,文学研究者既要更加重视跨学科合作,也要促进自身的思维转变,理解其他学科的研究范式,做到取长补短。同时,学界也要为新文科的成果建立完善的评价体系,创造出更多的展示机会。唯有打通人文学科与其他学科之间的知识壁垒,才能为新文科的发展奠定坚实基础,使文学研究在新时代焕发更强大的活力。

引用作品[Works Cited]

梅丽莎·丁斯曼 弗兰科·莫雷蒂:《人文研究中的数字:弗兰科·莫雷蒂访谈》,向俊译,《山东社会科学》9(2017):32—37。

[Dinsman, Melissa, and Franco Moretti. “The Digital in the Humanities: An Interview with Franco Moretti.” Trans. Xiang Jun. *Shandong Social Sciences* 9 (2017): 32—37.]

Goldstone, Andrew. “Distant Reading: More Work to Be Done”, 8 August 2015. 3 May 2021 < <https://andrewgoldstone.com/blog/2015/08/08/distant/> >

- Jean So, Richard. "All Models Are Wrong". *PMLA* 132.3 (2017): 669-673.
- Jean So, Richard and Edwin Roland. "Race and Distant Reading." *PMLA* 135.1(2020): 59-73.
- 吉云飞:《作为“计算批评”的“远读”——以网络小说“升级文”中的节奏与情绪为例》,《中国现代文学研究丛刊》8(2020): 177—187。
- [Ji, Yunfei. "Distant Reading as 'Computational Criticism' — A Case Study of the Rhythm and Emotion in the 'Upgrade Game' Genre of Online Fiction". *Modern Chinese Literature Studies* 8(2020): 177-187.]
- 托马斯·库恩:《科学革命的结构》,金吾伦、胡新和译。北京:北京大学出版社,2003年。
- [Kuhn, Thomas. *The Structure of Science Revolution*. Trans. Jin Wulun and Hu Xinhe. Beijing: Peking University Press, 2003.]
- Liu, Alan. "Where Is Cultural Criticism in the Digital Humanities?" *Debates in the Digital Humanities*. Ed. Matthew K Gold. Minneapolis: University of Minnesota Press, 2012. 490-509.
- 霍伊特·朗 苏真:《文学模式识别:文本细读与机器学习之间的现代主义》,林懿译,《山东社会科学》11(2016): 34—53。
- [Long, Hoyt, and Richard Jean So. "Literary Pattern Recognition: Modernism between Close Reading and Machine Learning." Trans. Lin Yi. *Shandong Social Sciences* 11(2016): 34-53.]
- Michael, Cole, and Yrjö Engeström. "A Cultural-Historical Approach to Distributed Cognition." *Distributed Cognitions: Psychological and Educational Considerations*. Ed. Gavriel Salomon. New York: Cambridge University Press, 1993. 1-46.
- 弗兰科·莫莱蒂:《对世界文学的猜想》,诗怡译,《中国比较文学》2(2010): 9—20。
- [Moretti, Franco. "Conjecture on World Literature." Tran. Shi Yi. *Comparative Literature in China* 2(2010): 9-20.]
- Moretti, Franco. *Distant Reading*. London and New York: Verso Books, 2013.
- . *Graphs, Maps, Trees: Abstract Models for a Literary History*. London and New York: Verso, 2005.
- . "More Conjectures." *New Left Review* 2(2003): 73-81.
- Muijs, Daniel. *Doing Quantitative Research in Education with SPSS*. London: Sage, 2010.
- Ore, Gabriela. "To Code or Not to Code." *Digital Humanities*, 4 December 2017. 3 May 2021 <https://www.vanderbilt.edu/digitalhumanities/gabriela-ore-to-code-or-not-to-code/>
- Parla, Jale. "The Object of Comparison." *Comparative Literature Studies* 41.1(2004): 116-125.
- Piper, Andrew. *Enumerations: Data and Literary Study*. Chicago: University of Chicago Press, 2018.
- Spivak, Gayatri Chakravorty. *Death of a Discipline*. New York: Columbia University Press, 2003.
- 泰德·安德伍德:《远读的系谱》,伊豆原英悟译,《数字人文》2(2020): 31—48。
- [Underwood, Ted. "A Genealogy of Distant Reading." Trans. Eigo Izuhara. *Digital Humanities* 2(2020): 31-48.]
- 王铭玉 张涛:《高校“新文科”建设:概念与行动》,《中国社会科学报》2019年3月21日第4版。
- [Wang, Mingyu, and Zhang Tao. "Constructing 'New Liberal Arts' in Universities: Concepts and Actions." *Chinese Social Sciences Today*. 21 March 2019.]
- 雷蒙德·威廉斯:《漫长的革命》,倪伟译。上海:上海人民出版社,2013年。
- [Williams, Raymond. *The Long Revolution*. Trans. Ni Wei. Shanghai: Shanghai People's Publishing House, 2013.]
- 项蕾等:《数字与文学的对话——“数字人文规范对传统文学研究方法的挑战”研讨会纪要》。《中国现代文学研究丛刊》8(2020): 158—176。
- [Xiang, Lei, et al. "Conversation between Digits and Literature: Minutes of Conference on 'The Challenge of Digital Humanities to Traditional Literary Research Methods'." *Modern Chinese Literature Studies* 8(2020): 158-176.]
- 赵薇:《数字时代的“世界文学”研究:从概念模型到计算批评》,《外国文学动态研究》3(2020): 35—48。
- [Zhao, Wei. "Study of 'World Literature' in a Digital Age: From Conceptual Models to Computational Criticism." *New Perspectives on World Literature* 3(2020): 35-48.]
- 赵一凡:《西方文论关键词》。北京:外语教学与研究出版社,2006年。
- [Zhao, Yifan. *Keywords in Western Literary Theories*. Beijing: Foreign Language Teaching and Research Press, 2006.]
- 周国梅 傅小兰:《分布式认知——一种新的认知观点》,《心理科学进展》2(2002): 147—153。
- [Zhou, Guomei, and Fu Xiaolan. "Distributed Cognition: A New Cognitive Perspective." *Journal of Developments in Psychology* 2(2002): 147-153.]

(责任编辑:王嘉军)