

中国社会科学网

WWW.CSSN.CN 中国社会科学院主办
中国社会科学杂志社承办

[首页](#) >> [学科](#) >> [新兴学科](#)

构建量化数据库 算出更精确的历史——访香港科技大学人文社会科学学

院院长李中清教授

2014年06月30日 07:53 来源：《中国社会科学报》2014年6月30日第614期 作者：记者 雷家琼 字号

[打印](#) [纠错](#) [分享](#) [推荐](#) [浏览量 147](#)





■1979年,李中清(右一)与父亲在新疆。



■李中清(右一)、何炳棣、康文林合影

构建量化历史数据库、采用量化分析研究的模式并不过分强调量化分析方法本身的高明与否。大量零散但又有一定体系的史料扩展了史学研究的材料范围,大规模的数据化历史资料平台,也为整个学术界提供了更丰富、灵活而有效的研究资源。这种方法兼及史料整理和问题研究,既是史学研究社会科学化的重要途径,也是各相关学科发展的重要推力。

随着计算机技术的发展,将量化数据库方法运用于学术研究受到越来越多的关注。量化数据库与传统学术研究方法相较,具有什么特点?未来会呈现什么样的发展态势?

对中国的学术研究有什么意义和价值？日前，记者就这些问题采访了运用量化数据库研究历史的先行者李中清教授。

治学之路逃不开父亲的影响

《中国社会科学报》：李先生，您好。感谢您能接受采访。令尊李政道先生在物理学界的成就广为人知。他对您要求严格吗？对子女有什么期待？

李中清：和中国传统的父母一样，父亲对子女的教育要求非常严格。小时候，每次我和弟弟考试成绩不理想，父亲会很不高兴地批评我们，还亲自辅导我的数学，不过只要他坐在我身边，我的大脑就会一片空白，后来他只能放弃（笑）。但他对子女未来志向和生活的选择却很宽容，从不干涉，任由我们选择喜欢的领域。父亲常说，人和动物之所以不一样，不是因为人拥有多少物质或财富，而是因为人可以用自己的大脑去思考和写作，能创造知识与文明，并将其代代相传。因此，一个人做哪行并不重要，重要的是能脚踏实地，真正做一些于创造知识和传承文明有益的工作。父亲的这种思想一直影响着我和弟弟。

《中国社会科学报》：您受您父亲的影响很深，但为什么最终选择了历史专业？

李中清：受家庭环境的影响，我和弟弟从小就立志要从事学术研究。年少时，我与弟弟讨论最多的便是将来会成为哪个学科的教授（笑）。进入大学前，我曾将经济学、城市规划和历史作为三个可选的方向。但因为经济学太依赖数学，而数学又是父亲的强项，我不想受父亲影响太多，就放弃了；城市规划太过实用；历史是基础学科，更为厚重，所以我选择了历史。

正当我考虑去哪所大学读历史时，突然接到个电话，电话那头的人说：“我是芝加哥大学的何炳棣，你是李中清吗？”我回答：“是的。”他接着说：“听说你想学中国史，你会中文吗？”我说：“我不会中文。”他说：“那你本科不要来芝加哥念，芝加哥中文教得不好。美国好大学中耶鲁和哥伦比亚的中文教得最好，你可以选择其一。待你本科毕业时再与我联系。”这是我第一次跟何先生说话，我当时甚至不知道他是谁，但我还是听取了他的意见，去耶鲁大学读本科。

进入耶鲁后，我主要跟随张光直先生学习。张先生的专长是中国考古，便劝我集中精力学习、研究中国古代史。大学二年级，我学了高年级本科生的中国史课程，三年级上了研究生的中国史课程。本科毕业时，我的成绩在全班名列前茅，哈佛、普林斯顿、剑桥和耶鲁都愿意录取我，并为我提供研究生奖学金。那时的我已经读了何先生的全部著作，对他的学术很是仰慕，加上之前的约定，最后我还是决定去芝加哥大学，成为何先生的研究生。何先生以严厉著称，很多人私下称他为“老虎”，一位朋友知道我的决定后，特意跑来跟我说：“你决定去‘老虎’的宝库了，真的很勇敢。”

接受多位一流史学家指导

《中国社会科学报》：在学术道路上，您曾受教于多位著名史学家，他们对您有什么影响？

李中清：在芝加哥大学期间，何先生曾推荐唐长孺等人的著作给我们。他要求我必须广博而扎实的历史基础，学习和掌握从黄土文明的起源到中华民国的各方面历史。研一时，何先生要我写作《中国移民史》一文，这对一个一年级的研究生来说，压力非常大。我花了一年时间，整理了从东周到清代的历代移民状况，发现中国移民和国外有一个非常重要的差别，即中国历代移民往往是政府组织的，直到明代以后才出现“逃人”等社会移民，且数量也不多。当时是1975年，国内对移民问题的研究还没起步。

我最初致力研究上古史。何先生要求我写“汉武帝传”。经过一年的努力，我发现这是一个21岁的年轻人无法完成的任务。于是何先生便要求我做中国边疆史，最后逐渐聚焦西南社会经济史。为了完成这一题目，我用了10年时间，遍访美国、欧洲和亚洲各地所藏的中国西南地方史料。在日本的尊经阁文库，我找到崇祯时期的《贵州通志》，这是全世界独一无二的孤本。此后，我又用了15年完成写作，到正式刊行《中国西南社会经济史》，前后共25年。

1973年，经周恩来总理特批，我第一次来到中国大陆。除了在上海柴油机厂和周边农村参与劳动，感受革命外，何先生还联系复旦大学的谭其骧先生，单独为我讲授魏晋南北朝史。1978年底，我第二次来到中国大陆前，在日本和中国香港访学了9个月。在日本，斯波义信先生给了我很多指导。我也去东洋文库、内阁文库、尊经阁等图书馆阅读了各种中国古籍善本。到中国后，我先在上海图书馆看古籍，之后到北京大学进修，跟随考古系的邹衡先生学习金文。1979年，我还去听谢国桢先生的课，学习明清史。这些都是何先生要求的。谢老告诉我，学习明清史，必须掌握两把“钥匙”：一把叫傅衣凌，另一把叫郭松义。郭先生对明清官方文献和文集非常熟悉，傅先生则对方志和民间材料了如指掌。由于傅先生远在厦门，我只能经常向郭松义先生求教，他也非常热心地指导我。后来，我的博士论文决定研究西南社会经济史，中国政府又向我推荐了云南大学的方国瑜先生，学习西南少数民族历史。这几位先生的教诲让我获益匪浅。

博士毕业之前，我曾被推荐为密歇根大学中国研究中心的Junior Fellow（类似于现在的博士后），在一年的自主研究时期里，我开始研究东周和西汉的“亭”制度，通过它来研究当时中国地方社会与国家的关系。恰在这一年，杨向奎先生来美访学，就在密大中国研究中心，与我共用一个办公室。杨先生当时已年过七旬，独自在密大且不擅烹饪，因此我几乎每天中午都请他吃饭，既为改善他的饮食，也为陪他说话解闷。利用这一机会，杨先生每天教我一两个钟头的中国思想史和古文字。

1972—1982年的10年时间里，我接受了多位一流中国史学家的指导，大大开阔了我的学术视野，也让我深深明白中国史研究的艰辛，要达到这些一流学者的水平还非常困难。这促使我认真反思今后的治学方向和方法。

量化分析让史学从“繁”至“简”

《中国社会科学报》：您什么时候开始将量化分析方法应用于史学研究呢？

李中清：受何先生影响，读研期间我就思考中国人口史的研究。1983年，我在中国第一历史档案馆参加学术会议，偶然听闻辽宁省档案馆有一大批历史资料在向一档馆申请销毁。我知道后立即要求一档馆帮我联系，去辽宁省档案馆查阅这批档案。后来我带了5大本乾隆时期的人口登记资料回美国。如何分析这种系统的人口史料，当时包括美国的中国史学界都没有经验，我也不知道如何研究。后来得知南加州一所大学有位学者分析过类似的日本人口史料，便前去请教。在他的帮助下，我找来当时还很专业的IBM电脑，摸索着学习构建数据库，并采用量化的分析方法来研究这批资料。此后的20多年里，我始终致力于收集中国东北各地的历史人口资料，扩充人口数据库并开展定量研究。这种研究方法使得我的研究常有不一样的发现。

《中国社会科学报》：量化分析方法与传统的研究方法相比较，区别在哪里？有何优点？

李中清：量化分析有一些明显的优势，比如它擅长多变量分析，不仅能深入理解各种因素的变迁与彼此的关系，还能兼顾所有个体的影响与权重，避免选材时的疏漏与偏废。它在研究大规模群体型、连续性材料上非常有效，为克服史料繁芜的困难提供了重要思路，有利于推进史学的整体研究。

由于习俗、文化等差异，很多历史现象蕴含着复杂的信息，直接进行跨地域、跨文化比较是很困难的，而量化分析往往是从复杂历史现象中抽取出相对简单、独立的指标，这些指标大多可以直接进行比较，如财富分配、出生/死亡率、家庭规模与结构、代际流动与社会分层等。我曾经主持过由多国社会科学家参与的“欧亚人口和家庭史项目”，虽然我不会瑞典语，无法阅读瑞典材料，但却可以依靠数据库化后的量化分析，比较北欧、西欧和中国明清时期的人口出生率、死亡率，这一比较研究也得到了国际人口学界的广泛认同。总体上，传统人文学研究大多是在做解释工作，而我们是在做新的计算，发现新的史实。

定量研究通常需要处理大规模的历史材料和多科学家的参与，因此需要大量人力和学术的合作，这也与传统研究的“单打独斗”不太一样。作为一种研究方法，量化分析可能在对部分系统性史料的深入信息挖掘上更为有效，但与传统方法一样，尊重史料、依靠可靠材料依然是量化分析的前提。因此，我研究的重点并不在方法运用上，而是关注史料背后能提供给我的客观信息，当然，研究的终极关怀与传统学者也是一致的。

《中国社会科学报》：放眼当前的国际计量史学动态，您认为计量史学未来的发展趋势是怎样的？

李中清：从1990年代中期开始，一些重大历史量化数据库对学界开放，计算机辅助下的量化分析成为国际社会科学学术研究中一股新风潮，许多大型量化数据库得以

构建并为学界服务。随着计算技术的进步和跨学科交流的频繁，这种研究模式可能会成为今后定量研究的主要发展趋向。

构建量化历史数据库、采用量化分析研究的模式与早期的计量史学不太一样，它重视史料整理和数据化平台的建设，但不过分强调量化分析方法本身的高明与否。对历史学来说，大量零散但又有一定体系的史料得以进入研究视野，扩展了史学研究的材料范围。大规模的数据化历史资料平台，也为整个学术界提供了更丰富、灵活而有效的研究资源。它们不仅是历史研究的重要基础，也越来越被其他相关学科所看重。因此，这种方法兼及史料整理和问题研究，既是史学研究社会科学化的重要途径，也是各相关学科发展的重要推力。

量化数据库助史学“科学化、国际化”

《中国社会科学报》：您和您的团队正从事量化历史数据库方面的建设，请介绍下相关情况。

李中清：从1983年获得辽宁人口史料起，我便致力于从人口史切入，通过构建大规模、可量化的历史数据库，采用定量计算的方法发现不一样的中国历史。在此过程中，我逐步组建研究团队，也逐步从人口史领域向其他领域扩展。例如，从20世纪80年代起，我们前后花费了20多年，建立了基于八旗户口册的中国多代人口数据库系列。该系列数据库包含辽宁、双城和皇族三个子数据库，前两个已经对全球学界免费公开。辽宁数据库涵盖了1749—1909年间辽东地区26万人的150余万条记录。双城数据库涵盖1866—1913年间双城县10万人的130余万条记录，并尝试与不同时段的家户地亩资料相连接。这些大规模、长时段的微观历史数据包含了丰富的人口和社会经济信息，具有时间上的深度和空间上的广度，对人口统计学、家与家族、亲属关系、社会分层与流动、卫生健康等多个研究领域有重要价值。

从2002年起，我的研究团队先后与北京大学、苏州大学合作，将新中国成立后两所大学15万余学生的学籍材料输入电脑，构建起中国大学生数据库。目前，我们又先后与华中师范大学、山西大学和上海交通大学达成协议，一方面将这些学校的学生材料也输入电脑，扩展大学生数据库，另一方面与它们的校友会等机构合作，追踪、调查这些学生毕业后的工作与生活状况，为今后更全面地分析新中国成立以来高等教育的功能和社会阶层的流动提供基础。2009年以来，在梁晨的建议和组织下，我们还将全国各省市档案馆所藏民国时期大学生的个人信息材料输入电脑，构建民国大学生数据库，目前已完成近10万学生的信息录入，希望将新中国成立前后的大学生数据库逐步对接，分析100年来近代高等教育在中国的引入、生长及其与社会的互动等问题。

这些史料格式统一、数量庞大，与“大人口”（big population）的历史关系密切。通过分析大学生数据库，重新认识了中国高考制度对社会流动的作用，这种认识对高教

改革也有一定借鉴意义。如果不借助史料数据库化和量化分析，这种信息背后的“大人口”历史规律与经验很难被准确解读和发现。

《中国社会科学报》：您认为量化数据库对中国史学有何影响？对于中国史学未来的发展，您有何期许与建议？

李中清：构建大规模历史量化数据库，进行定量研究，可以让史学研究更“科学化”和“国际化”，中国史学将更有条件系统展现人类社会在某些方面的长时段与大转变，中国史学者的工作也必然会更引起世界的关注。同时，构建量化数据库还可为国内培养史学专业学生提供新素材和新方向，国内史学界应该给予必要的关注和重视。

从国际层面来说，一国学术研究的影响力与其国家实力和发展状况密切相关。中国的崛起应该是本世纪最重要的国际变革，因此，中国史学界应该有自信和觉悟，在充实自身研究的同时，多参与国际合作，向全世界展示更多的中国成就和经验。这种经验不一定是几十年前常谈到的“唐宋文明”或“明清盛世”，而可能是 20 世纪以来的历史，比如革命历程与古老中国现代化的转变等。正如我们在分析高等教育材料时发现的那样，百年来被认为落后的中国，在很多制度设计和社会效果上所取得的成绩可能是西方社会无法比拟的。转换角度、扎实研究，中国史学对国际学术的贡献会大不一样。

（本采访得到梁晨先生的大力协助，特此致谢！）

李中清（James Lee）：美籍华裔历史学家、社会学家，诺贝尔物理学奖得主李政道之子，美国芝加哥大学历史学博士。曾在美国加州理工学院人文社科系、密歇根大学历史及社会学系任教，曾任密歇根大学中国研究中心主任，2009 年至今任香港科技大学人文社会科学学院院长。他是用社会科学方法治史的倡导者，亦是将量化数据库应用于历史研究的先行者。他和研究团队将档案资料、社会调查、家谱、碑铭及口述资料联系起来，创立从 18 世纪至今的大规模中国历史数据资料库。

分享到：转载请注明来源：[中国社会科学网](#)（责编：李秀伟）

相关文章

- [分布式学术搜索引擎研制及其大数据应用](#)
- [“中国文化海外传播动态数据库”平台系统发布](#)
- [“中国文化海外传播动态数据库”发布](#)
- [中国口头文学遗产建数据库](#)
- [中国最大文学数据库诞生 共有 8.878 亿字](#)
- [叶舒宪：数据库是口头文学遗产原生态的土壤](#)
- [刘晔原：建设口头文学遗产数据库走在时代前沿](#)