

大数据

观察下的宋词与元曲

杜肇昆 著



山西出版传媒集团



山西人民出版社

大数据

观察下的宋词与元曲

杜肇昆 著



图书在版编目 (C I P) 数据

大数据观察下的宋词元曲/杜肇昆著. — 太原 : 山西人民出版社, 2018. 1

ISBN 978-7-203-10083-6

I. ①大… II. ①杜… III. ①宋词—诗词研究②元曲—文学研究
IV. ①I207.23②I207.24

中国版本图书馆CIP数据核字(2017)第198786号

大数据观察下的宋词元曲

著 者: 杜肇昆
责任编辑: 魏 红
复 审: 刘小玲
终 审: 员荣亮
装帧设计: 刘彦杰

出 版 者: 山西出版传媒集团·山西人民出版社
地 址: 太原市建设南路 21 号
邮 编: 030012
发行营销: 0351—4922220 4955996 4956039 4922127 (传真)
天猫官网: <http://sxrmcbs.tmall.com> 电话: 0351—4922159
E-mail : sxskcb@163.com 发行部
sxskcb@126.com 总编室
网 址: www.sxskcb.com

经 销 者: 山西出版传媒集团·山西人民出版社
承 印 厂: 山西省教育学院印刷厂

开 本: 787mm×1092mm 1/16
印 张: 25.25
字 数: 202 千字
印 数: 1—1000 册
版 次: 2018 年 1 月 第 1 版
印 次: 2018 年 1 月 第 1 次印刷
书 号: ISBN 978-7-203-10083-6
定 价: 48.00 元

如有印装质量问题请与本社联系调换

用大数据思维揭示历史文化内涵（代序）

（一）

人类在认识客观世界的过程中，积累、总结出了科学实验、逻辑推理、仿真模拟、大数据四种认识方法及相应认识工具。

科学实验，是在创建特定条件下，通过干预和控制研究对象而观察和探索相关对象的一般规律的一种研究方法。科学实验是自然科学广泛应用的获得知识、检验知识、发现知识的一种认识方法。

逻辑推理，是运用抽象、演绎、综合的逻辑形式，借助逻辑和数学形成一系列概念、判断、推理表达出来的知识体系，形成特定理论，从而达到对客观事物本质、规律的认识。逻辑推理是人类认识自然和社会的重要思维方法。

仿真模拟，是在利用对真实系统的观察、测试数据信息基础上，建立数

学仿真模型，通过数据处理、模型仿真计算、相似转换等方法，实现对真实系统发展、演化规律性的描述，从而上升到反映真实系统运动规律的本质信息的认识。

大数据，是指用系统软件及其他工具进行捕捉、存储、管理和处理的巨量数据集合。大数据具有海量规模、快速流转、多样类型、价值密度低、在线运转五大特征，是世界物资资产外的另一种信息资产。大数据方法，是指不用传统随机分析（抽样调查）途径，而采用目标范围内全方位数据进行专业化分析处理，以揭示客观事物的状态、运动、趋势、规律，从而达到认识客体世界的目的。

计算机与网络技术的出现，为现代仿真技术和大数据方法提供了有效的手段和广阔的观察、认识空间。现代仿真软件广泛采用了可视化技术，通过图形、图表、甚至动画，生动逼真地演示出被仿真对象的各种状态，使模拟仿真的输出信息更加丰富、详尽，更加有利于对仿真结果的科学分析。现代大数据方法软件，能最大限度地、全面地利用信息系统，充分利用计算机数据处理能力，因而能形象地、逼真地从庞杂的数据背后，挖掘、分析出自然的、社会的客体行为的状态、演化趋势，形成对自然的、社会的一种新认识，为政治、军事、经济、商贸、文化等找到目标的新价值点。计算机与网络技术的融合才使大数据认识浮出水面并成为可能。

（二）

信息化是当代的主要特征。1980年，美国著名的未来学家托夫勒在其所著的《第三次浪潮》中，将大数据称为“第三次浪潮的华彩乐章”。2008年，

末，“大数据”得到部分美国知名计算机科学研究机构的认可，业界组织计算社区联盟，发表了一份有影响力的白皮书《大数据计算：在商务、科学和社会领域创建革命性突破》，提出：大数据真正重要的是新用途和新见解，而非数据本身。此组织可以说是最早提出大数据概念的机构。2010年，肯尼斯·库克尔在《经济学人》上发表了长达14页的大数据专题报告《数据，无所不在的数据》。库克尔在报告中提到：世界上有着无法想象的巨量数字信息，并以极快的速度增长。从经济界到科学界，从政府部门到艺术领域，很多方面都已经感受到了这种巨量信息的影响。科学家和计算机工程师已经为这个现象创造了一个新词汇——大数据。库克尔也因此成为最早洞见大数据时代趋势的数据科学家之一。2011年，全球知名咨询公司麦肯锡全球研究院发布了一份报告——《大数据：创新、竞争和生产力的下一个新领域》。大数据开始备受关注，这也是专业机构第一次全方面地介绍和展望大数据。报告指出，大数据已经渗透到当今每一个行业和业务职能领域，成为重要的生产因素。2012年英国人维克托·迈尔·舍恩伯格《大数据时代》一书出版。《大数据时代》是国外大数据研究的先河之作，作者在书中前瞻性地指出，大数据带来的信息风暴正在变革我们的生活、工作和思维，大数据开启了一次重大的时代转型，并用三个部分讲述了大数据时代的思维变革、商业变革和管理变革。大数据的核心就是解析现状与预测。大数据将为人类的生活创造前所未有的可量化的维度。

相对于互联网的发展速度来说，大数据概念从提出到成为一个时代，不到30年，在这个短暂时间里，大数据被不同行业和领域所验证，但是结论几乎都只有一个中心点：大数据将成为社会变革的动力，将是时代进步的象征。

信息网络时代，一切信息都会以数据的形式存在、汇集、传播。频繁的

社交网络为大数据提供了信息汇集、分析的第一手资料。社交网络让人们越来越多地从数据中观察到人类社会的复杂行为模式及其利用价值。

大数据最早的应用领域是商业。商家从庞杂的数据背后挖掘、分析用户的行为习惯和喜好，找出更符合用户“口味”的产品和服务，并结合用户需求有针对性地调整和优化商业自身，从而谋取更大的商业利润。而社会学家则看到了社会行为大数据的社会价值，政治学家则看到了这些行为思想的政治价值，文化学者则看到了不同文化层次的文化动向需求，科学家则利用大数据洞察自然界的奥秘。

大数据已经成为一种战略资源。它在解读社会、现实等领域有极重要的意义和潜在前景。2012年，美国政府在白宫网站发布了《大数据研究和发展倡议》，并宣布用2亿美元投资大数据领域，把大数据技术从商业行为上升到国家科技战略行为。同时指出大数据技术领域的竞争，事关国家安全和未来。并表示，国家层面的竞争力将部分体现为一国拥有数据的规模、活性以及解释、运用的能力。国家数字主权体现对数据的占有和控制。数字主权将是继边防、海防、空防之后，另一个大国博弈的空间。2015年，我国国务院也印发了《促进大数据发展行动纲要》，明确提出将全面推进我国大数据发展和应用，加快建设数据强国，从战略高度发展大数据事业。

(三)

大数据应用在商务、科学和社会领域创建了革命性、突破性的诸多成果。运用大数据思维于文学作品的研究，近几十年来亦是波澜渐涌。

20世纪70年代中期，英国剑桥大学的两位师生，因运用计算机侦破伪

造莎士比亚作品的奇案而震动了西方文学界。他们分析一家出版商出版的莎士比亚的新作品时，用计算机对它的修辞和结构进行数据分析和比较，把新作品中的句型与公认的莎士比亚作品进行统计比对，发现莎士比亚一向不采用的修辞和用语频频出现，句型和结构也同惯用的方法不一样。他们将用计算机分析出来的大量证据向法院起诉，使伪造莎士比亚作品的出版商哑口无言，只得承认作伪。

这项研究成果引起了国内外学者的兴趣与重视，他们利用计算机数据化来分析研判古典文学作品，导致计算机闯进“红学”研究园地。

1980年6月，在美国梦斗湖畔的威斯康星大学召开的首届国际《红楼梦》研讨会上，威斯康星华裔学者陈炳藻先生宣读了一篇《从词汇上的统计论〈红楼梦〉的作者问题》的文章，引起了国际红学界的注意和兴趣。1986年，陈炳藻教授又发表了《电脑在文学上的应用：〈红楼梦〉与〈儿女英雄传〉两书作者用词的比较》一文，之后又出版了《电脑红学：论〈红楼梦〉作者》的专著。陈先生利用计算机对《红楼梦》前八十回和后四十回的用字进行了测定，并从数理统计学的观点出发，探讨《红楼梦》前后用字的相关程度，并同《儿女英雄传》进行对比研究，运用数理语言学进行编排、统计、比较和处理，进而找出各组相关度。结果发现《红楼梦》前八十回与后四十回所用的词汇正相关度达78.57%，而《红楼梦》与《儿女英雄传》所用词的正相关度是32.14%。由此推断得出前八十回与后四十回的作者均为曹雪芹一人的结论。尽管学界还存有争议，但这种用计算机数据思维进行探索的创造性研究方法，给人们留下了深刻的印象。

在中国电脑红学研究也有创造性的成果。

1983年，江苏学者彭昆仑等用计算机探索《红楼梦》时间进程和人物年

龄问题，取得了突破性进展，解开了《红楼梦》中的一些年龄之谜。对于林黛玉入京都这一年究竟是几岁，红学家们历来就有多种说法，有13岁、11岁、9岁、8岁、6岁等不同结论。彭先生选定贾宝玉、林黛玉、薛宝钗等72名人物为信息数据存储对象，按小说各章节所描写的人物年龄和时令特征，参照有关红学家的研究成果和各种说法，将信息编码录入计算机，运用数理逻辑对年龄、年序进行分析、筛选，结果认定林黛玉入京都“9岁论”说比较符合生活和文学逻辑。中央电视台拍摄的电视连续剧《红楼梦》中，有关故事纪年、季节和人物年龄的处理意见，均采用了这项研究成果。

深圳大学曾研制《红楼梦》多功能计算机自动检索系统，从语言、文学和中国古代文化等方面对《红楼梦》进行了系统研究。研究人员将中国艺术研究院《红楼梦》研究所校注、人民文学出版社1982年出版的《红楼梦》作为基础原文，全文输入计算机。人们利用这个检索系统第一次得到了《红楼梦》中一系列重要的统计数据。如《红楼梦》全书的精确字数是731017字；书中使用不同的汉字为4462个；使用最多的是“了”字，达21176次；全书有1623个不同的四字格成语，用得最多的是“不在话下”；书中采用了24种修辞手法，用得最多的是“比喻”，共408条。该研究还根据电脑检索系统提供的资料，对红学研究中的一些问题提出了新的看法。例如，在对《红楼梦》前八十回与后四十回的一些语言风格要素与风格手段，即某些用字、用词及结尾处理的差异做了统计比较研究后，得出了前八十回与后四十回语言风格存在明显差异的结论，为两者出现不同作者之手提供了有力的证据。这与从词汇数量统计的结果相比较，似乎更具有说服力。1987年10月在慕尼黑举办的联邦德国第一届中文信息电脑处理研讨会上，《红楼梦》电脑检索系统成为一个最引人注目的项目，受到了与会者的高度关注和赞扬。

1987年，复旦大学数学系副教授李贤平在美国威斯康星大学运用计算机技术中的模式识别法和统计学家使用的数据分析法，对《红楼梦》进行统计分析、风格分析。他把《红楼梦》一百二十回本作为一个整体，以47个虚字为识别特征，对它们在书中各回的出现频率进行统计分析，输入计算机后绘成图表，提出了又一次震惊红学界的《红楼梦》成书过程新观点，证明了《红楼梦》各回写作风格具有不同的类别，各部分实际上是由作者在不同时期里完成的。李贤平认为：“《红楼梦》前八十回是曹雪芹据《石头记》增删而成，其中插入他早年著的《金瓶梅》式小说《风月宝鉴》，并增写了具有深刻内涵的许多内容。《红楼梦》后四十回是曹家亲友在曹雪芹全书尚未完成就突然去世之后，搜集整理原稿并加工补写而成。程伟元将全稿以活字版印刷刊行。高鹗校勘异文补遗订讹。”他的这一看法否定了被红学界一直视为曹雪芹作前八十回，高鹗续后四十回的定论。这一挑战传统的观点，引起国内外红学界广泛的关注。

利用大数据思维以计算机为手段的现代科学技术来研究《红楼梦》这样伟大的作品，确实别有一番新天地，拓展了视野，增添了新认识，产生了传统研究方法无法得出的认识成果。

文学作品是语言文字的组合，语言是人类用来表达和交流思想的工具，文字是记录语言的符号。一部文学作品就是一部符号系统，任何符号系统都能数据化，都能同计算机对接处理。反之，任何文学作品所表示出的数字系统，都隐藏着反映作品特性、特点、规律的东西，甚至反映那个时代背景的有价值东西，这在于我们去挖掘。当现代科学技术与古典文学相结合时，往往会产生出意想不到的认识结果和价值，让我们感受到这种研究方法的艺术魅力。

(四)

数据统计思维在文学的多个领域都能够得到应用。

20 世纪 90 年代，笔者根据多年的资料积累，拟写一部关于宋词创作方面的著作《填词探要》，为喜欢填词者提供一部便捷而系统的类工具书。全书拟分“词谱”“词韵”“词格”“词风”“词论”五个部分。在“词谱”部分的构思中出现了一个问题：宋词有一千多个词牌，应选多少个词牌，既有代表性又简洁明了？其依据又是什么？查清朝舒梦兰的《白香词谱》，列词牌 100 例，按字数排列，从少到多；龙榆生撰的《唐宋词格律》，列不同韵格谱例 153 例，亦按字数排列；王力的《汉语诗律学》在第三章“词谱举例”中，共列谱例 206 例，亦是按字数排列。这几部词谱有很多优点，但其共同的短处是，在宋词词牌里常见且作品量较大的若干词牌并未收入，连王力书谱例 206 例中亦未收录“醉落魄”“醉蓬莱”等词牌。为了在词谱中有依有据地体现宋词常见词牌的客观性，笔者着手运用数理统计工具，对《宋词全集》进行全方位的统计处理，完成了《宋词常用词牌的统计分析》一文，结果发现：宋词流行度最高的“常用词牌”有 40 个，其中作品量最大的词牌有 28 个，“多见词牌”63 个，并根据出现频率排队列出图表，予以展示。文章还分析了词牌的结构特征。对《宋词全集》信息的数据处理结果，在茫茫宋词海洋里揭示、还原、提纯了宋词流行度的现状，使宋词的精华数量化、简约化、清晰地浮出水面。文章在《中国文学网》发表后，引起了广泛关注，访问量很大。百度文库、百度学术、豆丁网、语文网、精英家教网等以及个人博客均有转载。中国人民大学附中、北京朝阳区、延庆五中、佛山高中等学

校高考模拟冲刺卷，关于词的部分内容均标明取自此文。以该文的统计结果为基础，在 63 个多见词牌的基础上，又向外扩展到 100 个词牌，选 200 个例词完成了“词谱”一章。为既简约涵盖又不遗漏宋词的精华，找到了客观的依据。《填词探要》一书 2001 年由北岳文艺出版社出版，并被香港大学图书馆收藏。

在《宋词常用词牌的统计分析》一文的基础上，笔者又对宋词的音韵倾向进行了统计分析。关于宋词的用韵特点，龙榆生先生在《唐宋词格律》一书中，将宋词的用韵分为五种类型：平韵格、仄韵格、平仄韵转换格、平仄韵通叶格、平仄韵错叶格五种类型。统计的结果是：仄韵格的作品量最高，占作品总量的 43.47%；其次是平韵格，占作品总量的 37.57%。在宋词中，单韵格（仄韵格、平韵格）占绝大部分，两项总计占到 81.04%；复韵格（平仄韵转换格、平仄韵错叶格、平仄韵通叶格）三项合起来只占 18.97%。文章还分析了词牌音乐属性与文学属性共体的宋词特性，是形成宋词音韵方式的基础因素。这一揭示对研究宋朝时期的音乐流行度和音乐特质提供了客观依据，对研究宋朝音乐史是有价值的。

21 世纪初，随着传统文化的复兴，社会诗词曲创作十分活跃。2004 年，中华诗词学会发布了《中华新韵（十四韵）》。一时创作界关于坚持旧韵还是应用新韵的争论硝烟四起。但争论的核心都没有触及为什么在中国历代会有不同的韵书，音韵与语言语音变迁有何关系的实质问题。为了澄清这个问题，笔者写了《十六韵及其他》一文。文章从历朝历代有影响的韵书产生的背景、特点入手，指出每代韵书的产生都是对前一代韵书的修正与改革，这种改革的动因是因为语言语音的变迁和不断精简、简化韵部的实用需要而至。文章还是从统计数据入手，将历朝历代有影响的韵书韵部数量的演变作了统计，

并作出变化曲线图。从图上一目了然地观察到韵目不断递减的规律，从而指出改革韵书、提倡新韵是历史的必然。文章进一步提出新时代诗词曲创作应提倡“三韵一统”的观点。2008年，文章在《中国文学网》发表后，引起了强烈反响，两派争吵不休。仅中国文学网和中华诗词论坛两大网站的访问量就超过1.1万人次，还不算其他网站和个人博客的转载。

为了更深层次地阐述这种观点，笔者从汉字起源、汉字的特点出发，特别提出汉字的时空特性，是产生诗词格律的基础的观点。但汉字的语音不是一成不变的，在中国漫长的历史演进中，经过了上古语音、中古语音、近古语音、现代语音的缓慢变化，因而产生了相应的音韵系统。从隋朝陆法言《切韵》始，到清已成型的《十三辙》，历朝历代都有自己的韵书。在1400多年里，音韵从193韵、206韵一直演变、递降到13韵，这是一个不断适应语音变化的改革历程。为了加深这种认识，还从历史典籍中用大量例证和数据予以说明，写成了《诗词曲韵案》一书，2009年由华艺出版社出版。网上公布后，引起学界的关注，网站纷纷转载。美国哈佛大学、芝加哥大学图书馆通过国内机构来函购书馆藏。

在《诗词曲韵案》出版之后，因学术活动结识了不少散曲研究、创作界的朋友，并参加了一些全国散曲研究学术会议。在接触一些网络散曲作品后，发现太缺乏有关散曲创作方面的著作，遂萌生写一部有关从创作角度系统认识元曲的书，拟名为《散曲探要》，使其成为《填词探要》的姐妹篇。随即开始研究、整理散曲有关资料。当完成一般性的叙述如元朝进驻中原的历史、元朝的政策特点、社会矛盾的深化、元戏曲与散曲的形成与发展背景等30万字的上卷，开始围绕散曲创作必须深化认识的诸问题的下卷时，新的问题出现了。

创作绕不开的第一个问题就是“谱”的问题。周德清在《中原音韵》中列举北曲有 335 个曲牌，历代谱书又有增加。哪些曲牌在元朝是引领散曲创作的中心、具有最大的流行度，学习创作的人必须掌握和深透了解的最基本的曲牌是那些，依据又是什么，检索以往资料找不到任何可借鉴的答案。

还是借用大数据统计来解开这个谜。隋树森的《全元散曲》是记载元散曲最权威的本子，具有大量的信息，以全部作品作为统计基础，以曲牌出现频率高低与作品数量多寡为目标值，根据边界条件进行统计分析，结果发现：在周德清《中原音韵》提到的 335 个曲牌中，流行度最高的“常用曲牌”只有 12 个，占曲牌总数的 5.53%；产生大量作品群的曲牌有 17 支；超过曲牌作品平均数的“多见曲牌”有 45 个。对套曲及宫调亦作了统计分析，并全部排队列表展示。2009 年写成了《元散曲常用曲牌的统计分析》一文，发表在中华诗词网上。对散曲常用曲牌的统计界定后，掌握了这些曲牌及其作品，就等于掌握了元散曲的本质精华，对散曲就有了基本的认识基础，而不是凭着感觉来确定。在此基础上撰写、核定了《元散曲四十五个常用曲牌格律》的理论谱与实用谱两篇文稿，提供散曲创作者参考，发表之后一时网上纷纷转载。

谱的问题解决之后，还有一个作家群的问题。最具影响力和主导元朝文坛的作家群落是哪些，弄清这个问题才会对元朝文坛的主流及特点有个清晰的认识。《全元散曲》提供了元朝 213 位作家及相应作品量，根据这些数据进行全员统计，统计结果显示：有影响的主体作家共 37 位，占作家总数的 17.37%，但其作品总和却占到作品总量的 74.48%，接近 3/4。这些作家在元散曲中占有绝对的主导地位，代表全元散曲的主体性。数量分析还表明，37 位主体作家中，胥吏作家与才会作家的作品量占绝大多数，表明全元散曲的

文学大厦，主要是由胥吏作家与才会作家两根巨柱支撑着。进一步分析这两个群落的特点，就自然与元朝社会的政治背景联系起来，对那个社会的文化风貌有了更深的认识。2012年写出了《元散曲主体作家群及其结构的统计分析》一文，发表在中国文学网上，用数据揭示了散曲、作家、社会与元历史的内在联系。

在研究有关散曲的多种著述后，发现关于散曲的认识尚有三个问题须待清晰与界定：一个是元散曲的历史分期问题，一个是元散曲的艺术特色问题，一个是元散曲风格流派的表述问题。

关于元散曲的历史分期，学界一直有“两分期”“三分期”“四分期”之说。究竟哪一种更客观反映元朝的历史现状，笔者还是以《全元散曲》提供的数据为依据，参考诸多文献关于元曲作家的生平，将作家与作品与活动轨迹，统计出数据表，描述到分层设计的元朝历史的时空图上。分析时空图，明显地看出具有相对集中的两个作家散点群落，分布在元朝前后两个时间段，其分界线在1290~1300年间。由此可以验证“两分期”说具有更坚实的数据基础。分析作家结构时，数据证明，官宦作家作品总量大于平民作家作品总量，由此得出元散曲的支撑力量是官宦作家这一与以往文献完全不同的结论。2012年写成了《元散曲历史分期的统计研究》一文，发表在《文学遗产》（网络版）2012年6期上。中国传媒大学古代文学专业一研究生发表的一篇论文《元散曲分期问题初论》就引用了这篇文章的观点与资料。

关于元散曲的艺术特色，涉及评估这一代文学价值尺量问题，涉及与唐诗、宋词艺术特色比较问题，也涉及当代散曲创作借鉴与继承问题，所以需要有一个系统而全面的剖析与论述。所谓“全面”需要建立在对《全元散曲》作品总体统计、归类的基础上；所谓“系统”需要建立在对散曲的音乐属性

与文学属性与结构属性的综合剖析与表述的基础上。根据这个前提,笔者以隋树森先生《全元散曲》所收录全部作品为依据,列举了元散曲在“体例艺术”“语意艺术”“声韵艺术”“音律艺术”四个方面的40种表现形式,并统计筛选254首例曲予以说明佐证。文章在与唐诗、宋词的比较研究中阐述了元散曲鲜明的时代个性和艺术个性,并分析了元散曲艺术特色形成的历史文化及社会政策因素。《元散曲艺术特色及其表现形式》于2013年发表在《中国当代诗词网》上,2014年应邀参加天津第十三届中国散曲学术研讨会并在会上交流。该文网上访问量目前已超过九千人次。

关于元散曲风格流派的表述问题,在曲界有两种倾向:一种认为散曲风格主要存在“豪放、清丽”二类,一种认为戏谑俗白是元散曲风格主要特征。两种观点都有其片面性。笔者仍然通过《全元散曲》全部作品为依据,通过梳理和数量统计,对传统文献对散曲风格主要存在“豪放、清丽”的二分表述提出了质疑,并比较宋词的“豪放”与散曲的“豪放”的本质差别,指出这一词界定元散曲的不准确性。在对《全元散曲》全部作品分类的基础上,提出了元散曲客观上存在着三种相互关联又相互有别的风格体系,主张用“俗俏、旷达、清丽”来表述与界定,并用关汉卿、马致远、张可久为首的作家群若干作品实例予以佐证。统计分析认为,“豪放、清丽”类作品约占《全元散曲》的90%左右,俗俏戏谑类作品只占10%左右。因而得出结论,“豪放、清丽”类作品是元散曲主流,但“俗俏、戏谑”类作品抢眼、亮眼,是与元散曲“豪放、清丽”类平等的一个风格特点。三分表述对元散曲风格评价应属比较客观。2015年写成《元散曲风格流派二分界定辩疑——兼论俗俏、旷达、清丽三分界定表达》的文章,发表在湘潭大学与中国韵文学会合办的《中国韵文学刊》2015年第4期上。现已被《中国知网》《万方数据》

《文史哲数据图书馆》《中文科技期刊》《维普网》等知名网站收录数据库中。

散曲经元以后，到明到清仍在延续着。但一些文献认为散曲元以后便处于衰落状态。元明清散曲演变到底是个什么状态，不能仅凭思辨感觉来论述，还是让大数据来回答这个问题。

《全元散曲》《全明散曲》《全清散曲》三部文集记载着元明清三朝散曲作者与作品的全部状况，具有丰富的信息量。笔者即利用这三部文集提供的数据为依据，运用统计分析方法，比较了元明清三朝六百多年散曲发展变化的脉络。分析数据显示，无论散曲在总量、分量、均量上，明朝均远远高于元朝，明朝具有远高于元朝的散曲生产密度，说明明朝是散曲的进一步扩张期。到清朝散曲总量、分量、均量，则全低于元朝，说明到清朝才是散曲的衰落期。纵观从元到清的677年间，元明清三朝散曲运动脉络，经历了成熟发展——更高发展——衰落的抛物线形态过程。文章的结论证明，以往诸多文献中认为元朝以后散曲便衰落了的观点是不正确的。根据分析结果，2015年写成了《元、明、清散曲跨代变量态势统计比较分析》一文，阐述这些观点。

唐诗宋词元曲成为中国大一统朝代唐宋元的主流文化形态，它不仅塑造了各自朝代的文化形象，也深深影响了历代国人文化素养的偏好。唐宋元三朝由于历史背景的不同、社会政策的不同，其文化的社会氛围也大不相同。为了弄清唐诗、宋词、元曲，不同韵体在不同朝代的发展状况，以《全唐诗》《全宋词》《全元散曲》三部著作提供的数据，运用数据统计方法，比较分析了唐、宋、元三朝韵文文学发展态势的异同，分析唐诗宋词元曲时代的差异性和延续的相似性；指出唐诗宋词元曲在唐宋元三朝虽然均量态势、产量密度不同，但都形成了当朝极具特色的主流文化现象，同时又都是诗体文学自