

公共话语中的社会分层关注度*

——基于书籍大数据的实证分析(1949-2008)

柳建坤 陈云松

提要:本文利用谷歌图书的汉语书籍语料库和其他宏观经济社会数据,考察并分析了1949-2008年中国公共话语中有关社会结构的话语定义的变迁模式及其影响机制。研究发现,改革开放以来,由官方意识形态建构的“阶级”话语在社会分层的话语体系中逐渐消退,而以公众为面向的“阶层”话语在公共话语中的重要性不断提升。本文利用主成分分析法生成阶层关注度指数,并与宏观经济和政治指标等时间序列进行格兰杰因果关系检验,结果表明:改革以来,公众对阶层问题的关注度受到来自国民经济的总体走势、收入差距及政治参与程度的影响;收入不平等因素比宏观经济指标更能够影响民众对阶层议题的关注;官方舆论导向并不对民众的阶层关注度构成影响,但反过来受到民众阶层关注度的影响。本文为基于个体样本的传统主观阶层研究提供了大历史跨度和宏观层次的验证。

关键词:大数据 阶层意识 阶级意识 社会分层

一、引言

改革开放后,社会结构快速变迁带来的种种新变化和新问题推动了中国社会分层研究的发展。早期的社会分层研究集中对客观阶层的构成方式、结构特征以及流动机制进行了描述分析(李强,1993;李培林,1995;孙立平,1996)。从20世纪90年代后期开始,阶层意识开始成为这一研究领域的重要议题,许多学者通过个体或群体对自身社会经济地位的主观性评价来考察宏观社会结构变迁在微观层面的影响(卢汉龙,1996;刘欣,2001;李培林、张翼,2008;范晓光、陈云松,

* 本研究获得2016年国家社会科学基金一般项目“中国民众主观阶层的结构特征和动力机制”(项目号:16BSH011)的支持。本文初稿曾在2017年“社会分层与流动”冬季论坛上宣读。感谢刘欣、秦广强、朱斌等师友的批评指正,还要特别感谢匿名评审人的宝贵意见和修改建议。文责自负。

2015)。上述研究对改革开放前后中国社会结构的变迁轨迹、利益关系演化及其背后的结构性逻辑进行了非常充分的探讨,并且开辟了从个体认知视角剖析社会结构变迁的研究维度。

不过,目前国内学界在主观维度的社会分层研究尚有待完善。首先,囿于数据资料的时空局限,现有文献关于阶层意识的讨论缺乏对大历史跨度中社会阶层意识的全景描述,仅仅集中分析了近年来市场转型中个体的自我阶层定位。其次,对阶层意识形成和变动机制的解释主要集中于客观地位、相对地位及地位变动等个体层面因素,尽管近年来有研究开始探讨主观阶层地位与收入不平等的关联(陈云松、范晓光,2016),但对宏观因素的检视仍然不够全面。再次,以往的阶层意识研究试图探究社会成员如何理解自身或他人的社会经济地位状态,但对塑造这种认识的根源以及形成过程,即关于社会结构的话语定义问题尚未进行审视和讨论。事实上,改革开放前后,中国社会结构的属性在公共话语体系中经历了一场从“阶级”到“阶层”的剧烈转型。这一变化不仅与中国政治经济体制的调整存在着密切联系,更凸显出在制度转型背景下国家意志与公众态度在社会结构形态的话语定义权上的力量变化。虽然已有不少政治学者对此问题进行了详细阐述,但其在社会分层研究中的意义尚未完全彰显,而且也没有获得实证意义上的验证。

本文将从上述不足之处出发,进一步拓展以往的阶层意识研究,并由此提出两个基本问题:第一,以大历史跨度的思路来看,1949年以来公共话语中关于社会分层的定义是否经历了从阶级型到阶层型的转变?国家与公众在这一话语变迁中究竟发挥了怎样的作用?第二,体制改革所产生的宏观经济政治效应与社会结构的话语转变有何内在关联?要对此加以学术回应,应当首先保证分析数据的规模性、代表性和时空广覆性,并且需结合中国独特的社会转型背景来完善解释框架。在最新的国际研究中,陈云松和严飞(Chen & Yan, 2016)利用书籍文化大数据对美国社会在整个20世纪一百年间的阶层话语与通货膨胀、就业率、基尼系数等指标进行了时间序列分析,发现宏观经济指标与社会公众的阶层关注度之间存在着紧密的统计关联。借鉴这种大数据方法和分析逻辑,我们认为可以针对中国情境进行经验层面的回应和解释,从而为社会分层文献提供宏观层次的中国案例。

二、文献回顾

在阶层意识研究的早期阶段,研究者主要试图通过了解个体成员对自身阶层地位的认识来描述阶层结构的整体特征。无论是针对欧美发达国家还是东欧以及东亚国家,大量的实证研究均显示,绝大多数人对“阶层”这一概念有着较为清晰的认识(Jackman & Jackman 1983; Evans et al., 2004; Shirahase, 2010)。并且在考虑了社会经济背景可能产生的影响后,大部分人倾向于将自己视为中间阶层的一员(Evans et al., 1992)。但在针对中国民众的阶层意识的研究中,国内学者发现中国人的阶层自我定位明显低于同期的欧美国家居民(刘欣, 2001; 李培林等, 2005; 陈云松、范晓光, 2016)。此外,无论在城市还是农村,人们的客观社会经济地位与对于自身的主观阶层地位认知之间均存在着不小的偏差(仇立平、韩钰, 2015; 范晓光、陈云松, 2015)。

关于主观阶层意识的形成机制,研究者主要从三个维度提供了实证解释。首先,个体实际占有社会经济资源的多寡会对其阶层地位的认知形成产生决定性影响,这在现实世界中具体表现为人们在教育、收入和职业等客观地位指标上的差异(Hodge & Treiman, 1968)。其次,个体对自身阶层的定位也会受到主观因素的影响。例如,针对中国城市地区的研究表明,除了党员身份、教育年限、收入、住房产权等客观的社会经济因素外,公平感、生存焦虑和社会流动感知等同样也会影响个体的阶层认同(翁定军, 2010; 陈光金, 2013; 范晓光、陈云松, 2015)。最后,某些宏观因素,如收入不平等也被证实会对个体的阶层认同产生显著的负向影响(陈云松、范晓光, 2016)。

以上研究发现较为完整地描述了国内外民众阶层地位认同的基本结构特征,对于个体的阶层自我定位的理论解释也提供了诸多洞见,极具启发意义。不过,此领域的研究仍存在一些不足。首先,以往阶层意识研究的分析对象集中在个体层面,即使其经验结果来自于全国性调查,但抽样调查本身的缺陷也使其缺乏针对公众整体的解释意义。在时间维度上,国内学者在该领域的研究主要聚焦于改革后的情况,并且数据的历史跨度多为1-10年,时间较短,因而无法展现改革前民众个体或整体的阶层意识的历史图景。其次,在解释阶层意识变迁这一问题上,国内研究者大多采取的是社会学的微观理论范式,强调个体自身

的社会经济地位、主观心理态度以及与他人之间的社会比较的作用,但对宏观结构因素之于阶层自我定位的重要影响没有给予足够的重视(陈云松、范晓光,2016)。但近年来国外的实证研究表明,人们的阶层意识形成过程有着深刻的社会经济根源,一些宏观经济指标(如国民生产总值和失业率)、社会不平等程度以及舆论导向因素都会对个体的阶层自我定位产生显著影响(Andersen & Curtis, 2012; Curtis, 2015)。

除了在研究对象层次、分析跨度以及解释机制等方面存在的不足外,另一个关键的问题长期以来未得到相关研究者的重视,也即以往关于阶层意识的讨论往往是在“事实”层面展开的,所探究的仅是社会成员如何评判自身以及他人和社会结构中的位置状态,并未考虑到人们如何形成关于特定社会结构的认知。而这涉及社会结构的看似客观、实则由外部力量所塑造的话语实践属性。社会结构的话语属性凸显在中国因体制转型而形成的各个差异巨大的历史阶段中。随着1978年后中国国家发展的方向逐渐脱离高度紧张的政治化特征,与社会结构有关的公共话语体系也发生着“去阶级化”的变革(张乐天,2001)。在这一背景下,词性更加中立的“阶层”话语构成了官方以及学术界用于定义转型时期中国社会结构的话语系统(陆学艺,2002)。有关中国社会结构的话语表述在改革前后所呈现的巨大差异,显示了制度转型在改变公共话语体系的主导力量上所产生的重要影响,也进一步凸显了拓展分析对象层次和延伸分析时段的必要性。

基于对阶层意识研究脉络的审视和评论,本研究试图从话语建构的角度重新开展主观维度的社会分层研究。具体而言,我们将以大历史观来概览新中国建立以来公共话语中关于社会结构的定义在不同历史阶段的变迁历程,并着重从制度转型的时代背景中探寻国家意志和公众态度在建构社会分层话语中所发挥的重要作用,从而展示改革开放对中国国家发展的独特意义。

三、理论背景

在改革前相当长的时间内,受到指令性经济体制、强制性政治动员以及革命性舆论导向体系的影响,阶级话语以及由此激发的各种政治

运动交织于绝大多数中国民众的日常生活之中(郭正林,2003)。而自1978年肇始的制度改革进程则将中国推向了政治经济体系剧烈变动的新时代,社会结构也由此开始快速分化。与此同时,公众也对迅速变化的客观社会形势积极回应,不断革新对阶级与阶层这两种社会结构概念的理解。具体而言,中国制度转型对公众的社会结构认知变迁的影响主要体现在两个方面:经济领域的市场化改革和政治领域的参与式民主发展以及舆论导向的动态性调整。

首先,市场化转型带来的最显著的成效是使中国经济保持了三十多年的高速稳定增长。正是在宏观经济利好和经济结构日益多元化的激励下,强调个人自致属性的职业机制逐渐成为新的社会分化机制,社会经济资源的获取途径日益多样化,从而使阶层结构在改革之后进入到快速分化的阶段(陆学艺,2003)。这一变化不仅带来了不同阶层在生活水平和生活方式上的分野,更使得人们的价值观念和情感心理开始出现“阶层化”特征(马广海,2011)。其中较为典型的是在改革开放后迅速崛起的中产阶级。这一群体不仅在经济生活中表现出与传统阶层截然不同的消费倾向,而且对各项政治议题也持有独特的心理态度(周晓虹,2002)。虽然改革以来中国民众阶层意识的变化与宏观经济发展在实证意义上的关系尚未被证实,但对其他国家的跨国比较研究表明,经济形势的不断走强有助于缓解社会中的紧张情绪,促使人们更加关注与自身阶层利益相关的社会议题(Evans & Kelley, 2004)。另外,中国经济改革在做大“蛋糕”的同时,利益分配的公平性却未得到及时的关注和应对,造成贫富差距问题不断累积,使得当前的阶层结构日益呈现高、中、低比例失调的现象,中下层群体构成了社会结构中的多数(李强,2016)。而且,收入差距的效应进一步蔓延至社会心理层面,集体性的嫉妒和怨恨构成了当代中国社会情绪氛围的基本特征,并在现实层面表现为低收入群体对高收入阶层的“仇富”或草根阶层对政府部门的“仇官”等情绪的滋生和扩大(成伯清,2009)。

其次,作为市场转型的配套制度改革,中国的政治体制转轨的根本目标是恢复和健全在改革开放前遭到严重破坏的社会主义民主政治,这在实践层面表现为促进参与式民主的发展并调适舆论导向与制度转型之间的关系。一方面,改革开放前国家政治生活主要围绕基于阶级话语构建的政治动员体系来运转,在这样的背景下,多数社会成

员的政治实践模式表现为被动“卷入”型,经济、文化、思想等其他领域也都受到以“阶级斗争”为核心的政治导向的波及(郭正林,2003;梁丽萍、邱尚琪,2004)。但随着中国农村和城市社区开始分别引入村民选举制度和社区自治制度,公众参与政治生活的自主性明显增强(胡荣,2008;李晨璐、赵旭东,2012)。这一新的政治参与实践具有两个重要特点:一是参与者在教育程度、收入水平和职业地位上具有较高的相似性(李骏,2009);二是他们日益培养出趋于一致的政治效能感、政治关心和公共责任感(孙永芬,2008)。另一方面,政治体制改革对民众思想意识的影响更直接地反映在主流舆论导向的变化上。改革开放前,阶级类的社会议题在普通民众的日常生活中牢牢占据着中心地位(张济顺,2004)。虽然中国在1978年后迎来了变革传统体制的新时期,但对意识形态体系的改造仍保留了原有的权威性的制度文化资源。此时的舆论导向呈现因时性特征,也即根据政治经济形势的具体变化进行动态调整,具体表现为在“改革”和“稳定”这两种发展导向之间不断转换(陈思,2012)。与此同时,国家舆论导向与社会大众之间的互动关系也开始顺势调整。尽管公众意见表达的自主性在改革活跃期会被官方默许甚至鼓励,但在“维稳”阶段则又会被拉回到官方话语的轨道内(村田忠禧,2002)。这意味着中国民众的阶层意识变迁的动力也可能蕴含于国家舆论导向的变化之中,而且在其变化的同时,阶级意识可能会发生方向相反的变动。

基于以上对中国社会转型期所呈现的一系列事实的回溯和梳理,可以假定制度转型很可能使公共话语中有关社会结构的定义方式发生根本性转变,并且这与宏观经济发展、收入不平等、政治参与以及舆论导向的变动等机制密切相关。本研究将对此进行实证意义上的验证。我们将利用历史书籍大数据来展示1949-2008年间中国社会中“阶级”和“阶层”这两种社会分层的话语定义模式的变迁轨迹,着重分析二者在改革开放前后的阶段性变迁,以便在整体上把握国家意志和公众态度在影响社会结构的话语定义权上的力量变化。在此基础上,我们将依据与中国制度转型有关的多重影响机制,结合长时段的宏观数据进行基于时间序列的因果关系分析,探讨影响社会分层关注度变化的宏观结构性因素。本研究不仅能够弥补传统阶层意识研究的薄弱环节,而且也是在中国社会科学领域首次利用大数据对这一研究问题进行计量模型回归分析。

四、数据、变量和分析策略

(一) 阶级、阶层关注度的数据来源

公众的阶层意识是一种与个体态度存在较大差异的宏观社会现象,这也使得此前的阶层意识研究在此概念的操作化和测量上存在一些缺陷。例如,扬马特(Janmaat, 2013)认为,由于数据和方法的限制,对于某些试图分析受特定文化或结构性因素影响的民众如何看待收入平等问题的研究而言,从中提炼宏观影响机制是一项非常艰难、富有挑战性的工作。近年来,基于超大规模的信息体量以及广泛覆盖时空维度的优势,“大数据”概念正为传统定量研究带来新一轮的方法论革命(陈云松等,2016)。作为人类有史以来最大规模的数字化图书工程项目,谷歌图书(Google Books)大数据语料库的出现为解决这一问题提供了有力支持。考虑到其数据规模性和代表性上的优势,本文将谷歌图书语料库作为分析公共话语中社会结构定义的数据来源。

在电子化传播媒介出现以前,书籍语言是人类进行文化传承和积累的主要方式,涵盖了人类社会中绝大多数的经验、思维和观念,因而是人类社会演变过程中最正式和最权威的知识载体。而且,书籍语言除了表达作者本人的观点和意图之外,还能够在更大范围内反映当时公众的价值取向和社会整体的思潮风尚。如果一个词汇在多种书籍中反复出现,那么不仅可以折射作者群体的核心意图,而且可以展示大众态度的趋势导向,或者说该词具有较高的“文化影响力”,已取得社会的主流共识。总体而言,只要书籍语料库在规模性、时间跨度和代表性方面都获得可靠保证,我们便可以合理地假定某一词汇出现在其中的相对频次能够近似地刻画这个词汇本身及其蕴含的公众关注度(陈云松,2015)。

谷歌图书最新版包含了全世界7种主要语言的800多万种数字化书籍,占到了1500年以来人类全部印刷书籍总量的6%,词汇总量达到8613亿个。这其中包括用汉语(简体)印行的30万种书籍以及269亿个词汇。目前,这一数据已被国内学者探索性地应用于社会科学研究,开展了针对学科发展史、城市影响力传播、文化史以及社会观念变迁等问题的一系列研究,试图探索大跨度历史现象的发展轨迹和变动规律(陈云松,2015;陈云松等,2015;龚为纲、罗教讲,2015;柳建坤等,

2016;张鸿雁、柳建坤,2017;张亮亮等,2016;Chen & Yan, 2016)。因此,本研究将以谷歌图书的汉语简体库数据作为刻画社会分层关注度变迁的分析数据,时间范围限定在1949年到2008年。

(二)阶级、阶层词汇的确定

我们在表1中展示了本文所涉及的阶级类和阶层类各20个检索关键词。确定具体词汇需要考虑两个问题:一是基于谷歌图书语料库所提取的阶级类与阶层类词汇究竟反映的是大众对阶级和阶层议题的关注,或仅仅是一小部分政治社会科学等学术性书籍中的提及率?二是词汇的代表性问题,也即仅依据少量的职业类型能否全面展现中国社会结构自改革开放以来所发生的变化?

对此,本文在词源和词汇两方面进行了努力。首先,在词汇选取的来源上,我们不仅考虑了专业的辞书(斯科特和马歇尔主编的《牛津社会学词典》)和教科书(吉登斯的《社会学》、贾春增的《外国社会学史(第三版)》以及谢立中的《西方社会学名著提要》),更兼顾了面向全社会的一些重要调查报告(如陆学艺的《当代中国社会阶层研究报告》)和以大众为传播对象、作为官方舆论阵地的权威新闻媒体(如《人民日报》)。在表1中,我们还计算了各阶级、阶层类词汇的描述统计量。可以看到,一些具有中国本土特色的词汇(如“农民工”)在语料库中的占比远高于专业性词汇(如“阶层意识”),充分表明本文所选取的词汇具有很高的公众代表性,而并非仅仅反映专业书籍的学术性内容。

其次,需要明确的是,本研究分析时侧重关注改革前后开放阶级话语与阶层话语的变化趋势差异,因而格外关注能够展现改革开放前后阶层结构显著变化的词汇。例如,在农村经济改革和城镇化的背景下,农民工才开始作为一种历史现象在改革开放后的中国社会大规模出现。另外,这些阶层类词汇也具备很好的代表性。这些词汇在本研究所采用的四个词源(专业辞书、教科书、专业调查报告和新闻媒体)中具有使用频率高、重复次数多的特点,而且能够整体性地概括改革开放以来中国职业结构在各个领域所呈现的基本特征。另外,在表2中针对阶层类词汇进行主成分分析后,较高的KMO数值也在实证意义上表明:尽管不能穷举,但可以预期的是,即便加入更多词汇,本文的基本结论也不会受到影响。

(三) 测量和变量构建

1. 词频测量

考虑到每年书籍中的词汇量存在差别,我们参照了在以往类似研究中所采用的“词频比例”方法来实现数据的时间可比效果,也即在时间检索范围限定的 1949 - 2008 年的任一年度中,所选定的阶级类或阶层类的关键词在样本书籍中出现的次数与样本书籍中全体词汇总量的比值。也就是说,任一关键词的词频比例越高,表明公共话语中与此相关的社会分层关注度就越高。对词频测量的结果见表 1。

表 1 谷歌图书(汉语简体)中阶级与阶层词汇的描述统计量结果(1949 - 2008 年)

“阶级”类 关键词	统计量			“阶层”类 关键词	统计量		
	均值	标准差	变异系数		均值	标准差	变异系数
阶级斗争	9.465	12.22	1.2912214	社会地位	.965	.433	.44872819
阶级压迫	.320	.334	1.040815	阶层意识	.000428	.000560	1.307682
阶级地位	.199	.208	1.0454208	社会分层	.000721	.00141	1.9564264
阶级路线	.160	.156	97668124	阶层认同	.00165	.00416	2.5207728
阶级专政	6.947	11.12	1.6005279	阶层身份	.00105	.00140	1.3309394
反革命	86.38	46.26	.53547052	阶层隔离	.000143	.000272	1.8942794
革命	6.220	6.373	1.0245149	阶层冲突	.000885	.00198	2.23863
整风	7.027	6.308	.89774404	精英阶层	.0118	.0217	1.8353715
“左倾”	141.7	103.3	.72843394	中产阶层	.0350	.0667	1.9071928
“右倾”	1.154	.558	.48373463	贫困阶层	.0147	.0195	1.320669
无产阶级	.533	.237	.44492535	管理者	.0183	.0222	1.2119368
工人阶级	.153	.152	.99403927	蓝领	.00931	.0106	1.1357773
群众	.497	.428	.86185872	白领	.0652	.0827	1.2689642
领袖	.585	.411	.70275064	经理	4.325	3.648	.84334498
右派	60.33	34.28	.5681935	公务员	1.419	1.620	1.142004
资本家	3.552	2.418	.68078275	学者	8.338	5.875	.7046421
地主	4.328	3.730	.86179708	农民工	.678	1.590	2.3463173
富农	.113	.0567	.50235407	企业家	.000277	.000458	1.6543546
贫农	.533	.408	.76454341	民营企业主	.584	1.152	1.97408
中农	.276	.168	.6077973	职员	1.207	.585	.48474739

注:为便于阅读,所有词频比例的均值和标准差都扩大了 10000 倍。

2. 因变量:阶层关注度指数

为了进行格兰杰因果检验,我们使用主成分分析法(Principal Component Analysis)构造了“阶层关注度指数”(LC)作为因变量,且在

表2展示了分析结果。Kaiser-Meyer-Olkin (KMO) 和多元相关平方 (Squared Multiple Correlations, SMC) 的检验结果均表明,参与检验的词汇数据适合进行主成分分析。^① 根据主成分的负荷量、特征值以及被解释方差的累积贡献率,可以从20个阶层词汇中提取出两个主要成分,并进一步合成。

表2 阶层类词汇的主成分分析结果

	成分1	成分2		
特征根	16.82089	2.03105		
方差累积贡献率	.8410	.9426		
阶层词汇			KMO	SMC
社会地位	.6763	.5129	.8009	.9968
阶层意识	.9867	.0289	.8353	.9992
社会分层	.9288	-.3595	.7885	.9994
阶层认同	.9136	-.3800	.8315	.9997
阶层身份	.9825	.0197	.8953	.9988
阶层隔离	.9121	-.3615	.8474	.9985
阶层冲突	.9110	-.3728	.7991	.9998
精英阶层	.9800	-.1888	.8467	.9999
中产阶层	.9361	-.3417	.8614	.9993
贫困阶层	.9385	.1363	.9048	.9997
管理者	.9232	.3391	.8211	.9997
蓝领	.9427	.2645	.8796	.9989
白领	.9794	.0421	.8808	.9999
经理	.8264	.5191	.8699	.9991
公务员	.8594	.4329	.8975	.9967
学者	.9898	.1163	.8981	.9998
农民工	.8858	-.4202	.8655	.9998
企业家	.9783	.0436	.9342	.9997
民营企业主	.9528	-.1939	.8900	.9997
职员	.7728	.4234	.8648	.9982

^① KMO 越高,表明变量的共性越强。一般认为,KMO 值在0.6 以上就可以接受进行主成分分析。而 SMC 表示一个变量与其他所有变量的复相关系数的平方,也就是复回归方程的可决系数。SMC 越高,表明变量的线性关系越强,共性越强,主成分分析就越合适。

3. 自变量

通过前文对中国制度转型历程的追溯以及理论层面上对其与公共话语中社会结构话语定义之关系的讨论,本文从市场转型、政治参与和舆论导向创新等三个维度设定解释框架,并在经验层面上进行了变量测量和指标构建工作。

第一,针对市场转型的正向结果,本文使用世界银行公布的1978 - 2008 年间中国国内生产总值来衡量此阶段的经济总体走势。考虑到价格变动因素的影响,我们进一步将其转化为可比价格(经过 CPI 校正获得 1978 年可比价),以便对不同时期的经济总量指标进行比较,这一指标用“ GDP_{cp} ”表示。按照一般的做法,我们在后续的图形展示和数据分析中取其对数形式。

第二,对于由市场化改革所带来的收入不平等问题的测量,我们遵循以“基尼系数”为指标的主流做法。但目前中国方面的数据仍不完整,除了国家统计局正式公布的 2003 - 2015 年的官方数据外,其他年份的数据仅零星地见于统计年鉴中。因而本研究采用世界收入不平等数据库(World Income Inequality Database)3.3 版本对缺失数据进行补充,其指标名称为“GINI”。

第三,对于中国公众在 1978 - 2008 年间的政治参与程度,我们使用哥德堡大学和美国圣母大学共同完成的数据库项目 Varieties of Democracy 最新的 6.2 版本中的参与民主指数(Participatory Democracy Index)进行测量,其指标名称为“PDI”。该指数的取值区间为 $[0, 1]$,其中“1”代表政治参与程度最高,“0”代表政治参与程度最低。

第四,根据前文所述,改革开放以来国家舆论导向变化基本上是围绕“改革”与“稳定”两条路径进行转换。因此,本文利用“人民日报图文数据全文检索系统”统计了 1978 - 2008 年每一年标题中包含“改革”或“稳定”的文章报道数量,并以二者的差值作为衡量每年主流舆论导向变动的指标,即“IO”。如果该变量取值为正,表明官方的舆论导向在当年倾向于“改革”;反之,则是“稳定”在舆论导向中占据上风。

图 1 初步展示了 1978 - 2008 年间阶层关注度指数与上述宏观指标之间的时间变动趋势。^① 大体来看,阶层关注度和经济增长、基尼系数都在稳步增长,而代表国家舆论导向变化的曲线波动相对剧烈。在

^① 图 1 中各个时间序列曲线均基于 LC 、 GDP_{cp} 、 $GINI$ 、 PDI 、 IO 变量的标准化形式绘制而成。

整个 20 世纪 80 年代,关于改革问题的报道在官方的新闻媒体中占据核心地位。但在 90 年代,改革导向与稳定导向的差距并不悬殊,二者之间开始出现频繁变动的趋势,进入 21 世纪后,前者在话语体系中的地位逐渐提高。

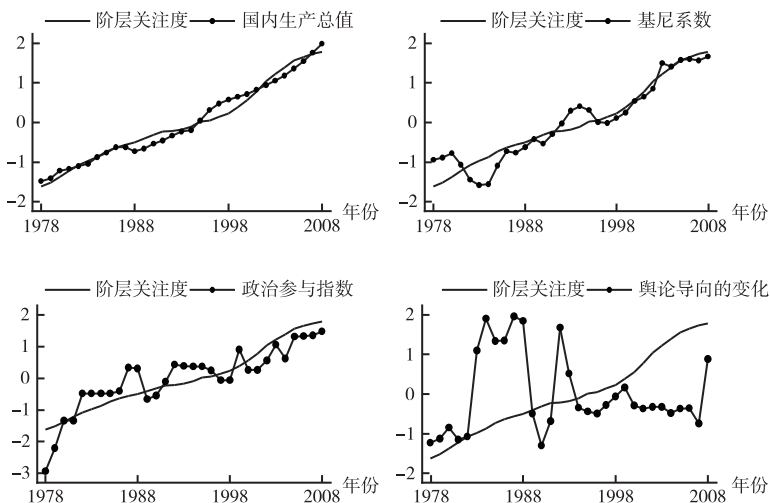


图 1 阶层关注度与宏观经济走势、收入不平等、政治参与和舆论导向之间的时间变动趋势

(四) 分析步骤

本文的数据分析工作主要分为两个部分。第一部分的任务是对从谷歌图书语料库中选取的各阶级、阶层类词汇在 1949 - 2008 年每年度的词频比例及其加总进行可视化呈现,重点分析两类词汇在 1978 年改革开放这一关键历史节点前后的变化,以展示 1949 年以来两种不同意涵的社会分层结构概念在中国社会的变迁轨迹。第二部分的主要工作是利用时间序列回归的方法探索改革开放后公共话语中阶层关注度变化的影响机制。具体而言,本研究使用了条件格兰杰回归的方法进行格兰杰因果关系检验。在计量经济学中,两个时间序列 X 和 Y 之间的格兰杰因果关系被定义为:如果变量 X 有助于解释变量 Y 未来的变化,则认为变量 X 是引致变量 Y 的格兰杰原因 (Granger, 1969)。因此,我们可以使用该方法来识别“阶层关注度”(LC)与“国民经济走势”(GDP_{cp})、“收入不平等程度”(GINI)、“政治参与度”(PDI)和“舆

论走向”(IO)等机制在时间序列意义上的内在关联。此外,如果在不平稳的时间序列直接进行基于F和Wald检验的标准格兰杰因果分析会产生偏误,那么我们首先需要进行单位根检验,即检查各序列的平稳性。如果上述时间序列不存在单位根,我们直接拟合向量自回归模型(VAR模型),并进行格兰杰因果测试;如果存在单位根,则可以先对时间序列的水平值进行差分,直到获得平稳时间序列后再进行格兰杰因果检验。

五、分析结果

(一)“阶级”和“阶层”在公共话语中的历史变迁(1949-2008年)

在图2中,我们对所有词汇的原始词频比例(图2-1)以及经过标准化处理后的Z值比例(图2-2)绘制了时间序列曲线。两图的趋势变化均表明阶级类和阶层类议题在社会舆论中的地位在1978年前后发生了明显更替。而原始词频数据则显示两类词汇之间存在明显的层次差异,即阶级类词汇在60年间书籍语言中所占比重始终高于阶层词汇。从曲线的阶段性变化来看,这一差距在新中国成立之初就已存在。虽然阶级类词汇的总词频比例在1976年后便大幅下降,但由于阶层类词汇提升有限,造成二者之间的差距直到2008年依旧存在。对此,一个可能的解释是:一方面,在改革开放前中国的经济社会生活长期以“政治挂帅”,包括书籍出版在内的文化宣传工作由官方部门全权负责,社会信息的发布和接收完全在国家掌控之下。因此,此阶段的公共话语基本上反映的是国家意志,社会舆论与官方意识形态处于高度重叠的状态。另一方面,随着市场化转型的启动,在信息传播行业开始出现按照商业规律运作的非官方出版集团,并迅速成为消费市场的主力。在出版业日益商业化、大众化和娱乐化的背景下,不仅书籍的主题和内容得到大幅扩充,而且在以迎合大众需要这一商业运作机制的作用下,一些远离民众日常生活并且带有明显政治性的内容,如阶级议题被海量信息不断稀释。这种“稀释效应”不仅使阶级类词汇不断淡化,而且也造成阶层类词汇虽然在绝对数量上不断增多,但在比例上却显得相对较低(陈云松,2015)。

此外,我们计算了阶级类和阶层类词汇在每一年的词频比例的

加总值。无论是原始词频加总(图3-1)还是加总词频的标准化值(图3-2),都显示阶级类词汇在1949-1976年持续高速增长,在此之后却开始“断崖式”下降,而阶层类词汇在1978年之后的总体运行轨迹处于稳步上升趋势。从图3-2不难看出,20世纪50年代中后期以来,阶级类词汇的总比例开始在全社会的书籍语言中迅速提高,在70年代中期达到顶峰,而同期的阶层类词汇已降至谷底。但从80年代以后,二者在书籍语言中所占的地位发生反转。值得一提的是,进入21世纪,特别是在2002年以后,有关阶层议题的关注度与前一时期出现了明显跃升。^①其中,一些涉及特定群体的阶层定义,如学者、农民工、管理者、白领和公务员等词汇出现了更为显著的增长。

阶层话语在改革期间不但形成了对阶级话语的总体优势,还出现了阶段性的爆发式增长。对于这一现象,我们猜测这可能与当时由著名社会学家陆学艺领衔的“当代中国社会结构变迁研究”课题组所撰写

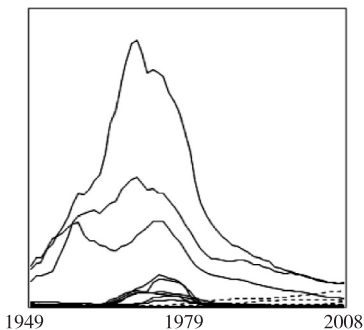


图2-1 全部词汇的原始词频比例

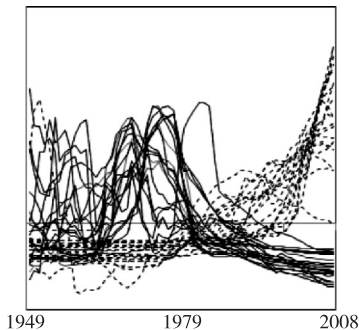


图2-2 全部词汇经过标准化处理后的Z值比例

注:阶级类为实线,阶层类为虚线。

图2 阶级话语与阶层话语的趋势变化(全部词频)

^① 我们以《当代中国社会阶层研究报告》发表的2002年为界限,首先以图形可视化的形式比较了在此前后6年内(1996-2008年)20个阶层类词汇的加总词频比例的总体变化趋势,又进一步通过测量曲线斜率以显示实际的增长幅度。这主要是利用对这一时间序列的一阶差分来实现的。结果表明,在1996-2002年,每一年的加总词频比例增幅为0.000007、0.000008、0.000018、0.000021、0.000025、0.000022;而在2002-2008年,增势仍继续保持,分别为0.000022、0.000025、0.000025、0.000012、0.000009、0.000006。这表明阶层话语在2002年后的公共舆论中明显增强。限于文章篇幅,此处未展示详细的数据结果。

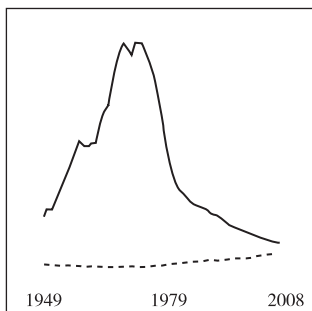
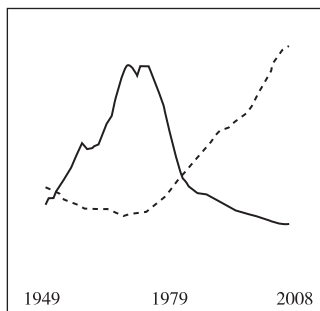


图3-1 原始词频比例加总值

图3-2 经过标化处理后的
词频比例加总值

注：阶级类为实线，阶层类为虚线。

图3 阶级话语与阶层话语的趋势变化(加总词频)

的《当代中国社会阶层研究报告》有着密切联系。这一报告对中国学术界的一个重大贡献是提出了改革开放以来中国社会结构逐渐形成了“十大阶层”的重要观点，并且一经公布，立即在全社会引起了广泛反响，得到了来自官方机构、理论界以及社会舆论的积极反馈。因此，我们认为，“十大社会分层”的论断在推动21世纪后中国社会分层的话语定义向阶层化转变的过程中发挥了至关重要的作用。

上述图形结果可以初步回答在本文提出的第一个问题，即在1949—2008年，中国公共话语中的社会分层的话语定义模式的确经历了从“阶级”向“阶层”的重大转变，其背后反映的则是国家意志在塑造话语体系上的控制力减弱，公众的态度和需求日益成为建构新型话语的重要力量。简而言之，在1949年以来的中国公共话语有关社会结构的话语定义中，改革开放前的阶级化类型反映的是官方意识形态的强势地位，而改革开放后出现的阶层化转向则意味着公众态度开始成为话语建构的主体。而且，这一变迁过程与改革开放这一关键的历史节点存在着紧密联系。但这种直观性描述并不具备实证意义。接下来，我们将对其中所蕴含的宏观变动规律进行详细验证。

(二) 阶层关注度与宏观结构性因素之间的因果关联

本文采用“扩展型迪克—富勒”(Augmented Dickey-Fuller)和“菲利普—帕芬”(PP检验)这两种方法对所有变量进行单位根检验。结

果表明, LC 、 GDP_{cp} 、 $GINI$ 和 IO 均为一阶单整时间序列, PDI 为稳定时间序列。^① 为便于格兰杰因果检验结果的解读, 我们对 PDI 也进行一阶差分。换句话说, 此时我们探讨的是 GDP_{cp} 、 $GINI$ 、 PDI 以及 IO 的变化与变化之间的关联。此外, 为保证所获得的结论在这种多变量条件下的稳定性, 本文采用的是条件格兰杰因果检验, 即在分析阶层关注度与某一变量的因果关联时, 将其他变量作为控制变量纳入分析之中。

表 3 展示了格兰杰因果检验的具体结果, 从中可以得到以下发现。(1) GDP_{cp} 、 $GINI$ 、 PDI 的变化均为可在时间序列意义上解释的变化 (p 值均小于 0.05)。由于以上变量均是以一阶差分的形式纳入到模型之中的, 准确地说, 前一年的国内生产总值的增加、收入不平等的加剧以及公民政治参与度的提高都可以解释后一年民众的阶层关注度提升这一现象。(2) 就统计显著性而言, 收入不平等 ($GINI$) 对民众的阶层关注度的影响 ($p < 0.01$) 要明显强于宏观经济发展 (GDP_{cp}) ($p < 0.05$)。(3) fd_{IO} 不是 fd_{LC} 的格兰杰原因 ($p > 0.1$), 也即这一原假设在统计上无法被拒绝, 这意味着政府舆论导向并不能解释公众阶层关注度的变动。(4) 我们发现 fd_{LC} 反而是 fd_{IO} 的格兰杰原因 ($p < 0.01$), 即民众对阶层议题的关注会随主流舆论导向发生变化。换句话说, 政府在引导舆论走向的过程中越来越重视民众关于阶层问题的意见。(5) 除了 fd_{IO} 以外, $fd_{GDP_{cp}}$ 、 fd_{GINI} 和 fd_{PDI} 都可以在时间序列意义上

表 3 条件格兰杰因果检验结果 (1978 - 2008 年)

零假设	观察值	Chi ²	p 值
$fd_{GDP_{cp}}$ 不是 fd_{LC} 的格兰杰原因	31	11.382 **	.023
fd_{GINI} 不是 fd_{LC} 的格兰杰原因	31	34.596 ***	.000
fd_{PDI} 不是 fd_{LC} 的格兰杰原因	31	26.347 ***	.000
fd_{IO} 不是 fd_{LC} 的格兰杰原因	31	1.7899	.774
fd_{LC} 不是 $fd_{GDP_{cp}}$ 的格兰杰原因	31	36.731 ***	.000
fd_{LC} 不是 fd_{GINI} 的格兰杰原因	31	20.26 ***	.000
fd_{LC} 不是 fd_{PDI} 的格兰杰原因	31	41.042 ***	.000
fd_{LC} 不是 fd_{IO} 的格兰杰原因	31	13.182 ***	.000

注: (1) “ $fd_{}$ ”表示一阶差分。(2) 根据信息准则 AIC、SBIC 和 HQIC 的结果, 选择 4 阶滞后。(3) * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$ 。

① 限于文章篇幅, 此处不进行详细报告。需要详细结果的读者可以联系作者。

被 fd_LC 所解释 (p 均小于 0.01)。这意味着公众阶层意识的变动不仅会影响官方舆论的走向,而且对宏观经济增长、公民政治参与水平的提高以及收入差距的扩大都会产生特定影响。

(三) 进一步解释

上述统计检验的结果初步验证了在改革的总体进程中阶层话语的地位上升与宏观结构性因素之间存在着逻辑关联。但需要注意到,中国的制度转型在具体推进过程中采取的是一种“渐进式”策略,这使改革的总体趋势因具体形势的变化而在各个时期呈现明显差异。具体而言,自从改革开放在 1978 年正式作为国家决策被提出后,一直到 80 年代末,改革意向在中央和地方政府的施政中都得到了充分体现。但随着国内外经济政治形势在 20 世纪 90 年代的日趋紧张,各领域的改革进程开始受到“维稳”政策的调控。直到 2000 年后,改革才在全球化进程以及日益完善的市场经济环境下获得新的发展动力(渠敬东等,2009)。由此,我们将中国改革的进程划分为三个时段:1978 - 1990 年、1991 - 1999 年、2000 - 2001 年,并在表 4 中展示了在各阶段对阶层关注度与宏观结构性机制进行条件格兰杰因果检验的结果。^① 从中可以获得一些新的研究发现,并对整体模型进行完善。

首先,表 4 的结果表明,虽然发现(1)在总体上验证了制度转型对民众的阶层意识变迁具有显著影响,但二者的关联在改革的不同时期呈现较大的差异性。在 1978 - 1990 年,仅有舆论导向的变化(fd_IO)能够在时间意义上成为阶层关注度变动的格兰杰原因 ($p < 0.001$)。而在 1991 - 1999 年,宏观经济增长(fd_GDP_g)则成为影响民众阶层意识的最显著的因素 ($p < 0.001$)。在进入 2000 年后,上述机制的效应开始消失,但公民的政治参与(fd_PDI)和收入不平等(fd_GINI)开始成为新的影响因素 (p 均小于 0.05)。对于这一新的研究发现,我们可以根据改革进程的具体走势来理解。一般认为,虽然制度创新往往能够在改革初期大刀阔斧地展开,但新思想观念(如产权、市场、竞争、法治)的普及会遇到“文化堕距”的现象,被多数社会成员所接受仍需相当长的时间。而且,由于体制转轨受到“路径依赖”的影响,中央政府在在这一阶段推进改革主要是以舆论的方式对民众进行引导。最具代表

^① 限于文章篇幅,本文并未展示单位根检验的结果,欲获取详细信息请联系作者。

性的就是关于真理标准问题的大讨论在全社会发挥了思想解放的作用,为改革开放的顺利起步奠定了思想基础。

但在 20 世纪 90 年代,严峻的政治经济形势使与阶层有关的敏感话题在社会舆论中受到了一定程度的抑制,但同时建立和完善社会主义

表 4 改革不同阶段中阶层关注度与宏观经济和政治因素之间的格兰杰检验结果

年代	零假设	观察值	Chi ²	p 值
1978 - 1990	fd_GDP_{cp} 不是 fd_LC 的格兰杰原因	13	.11	.742
	fd_GINI 不是 fd_LC 的格兰杰原因	13	.85	.356
	fd_PDI 不是 fd_LC 的格兰杰原因	13	3.39 *	.066
	fd_IO 不是 fd_LC 的格兰杰原因	13	12.13 ***	.000
	fd_LC 不是 fd_GDP_{cp} 的格兰杰原因	13	.78	.377
	fd_LC 不是 fd_GINI 的格兰杰原因	13	.18	.676
	fd_LC 不是 fd_PDI 的格兰杰原因	13	4.90 **	.027
1991 - 1999	fd_LC 不是 fd_IO 的格兰杰原因	13	4.60 **	.032
	fd_GDP_{cp} 不是 fd_LC 的格兰杰原因	9	15.49 ***	.000
	fd_GINI 不是 fd_LC 的格兰杰原因	9	8.88 ***	.003
	fd_PDI 不是 fd_LC 的格兰杰原因	9	4.93 **	.026
	fd_IO 不是 fd_LC 的格兰杰原因	9	17.11 ***	.000
	fd_LC 不是 fd_GDP_{cp} 的格兰杰原因	9	1.28	.258
	fd_LC 不是 fd_GINI 的格兰杰原因	9	5.89 **	.015
2000 - 2008	fd_LC 不是 fd_PDI 的格兰杰原因	9	2.48	.115
	fd_LC 不是 fd_IO 的格兰杰原因	9	1.10	.294
	fd_GDP_{cp} 不是 fd_LC 的格兰杰原因	9	.01	.927
	fd_GINI 不是 fd_LC 的格兰杰原因	9	9.55 ***	.002
	fd_PDI 不是 fd_LC 的格兰杰原因	9	14.06 ***	.000
	fd_IO 不是 fd_LC 的格兰杰原因	9	.21	.650
	fd_LC 不是 fd_GDP_{cp} 的格兰杰原因	9	210.24 ***	.000
fd_LC 不是 fd_GINI 的格兰杰原因	9	147.66 ***	.000	
fd_LC 不是 fd_PDI 的格兰杰原因	9	115.83 ***	.000	
fd_LC 不是 fd_IO 的格兰杰原因	9	55.55 ***	.000	

注:(1)“ $fd_$ ”表示一阶差分。(2)根据信息准则 AIC、SBIC 和 HQIC 的结果,在三个时段内分别进行格兰杰检验需要分别选择 2 阶、2 阶和 1 阶滞后。(3)* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$ 。

市场经济体制也被正式确立为改革的核心主题,经济性议题也因国民经济的良好发展势头而在民众的阶层关注度中占据着重要地位。2000年后,与经济高速增长相伴而来的贫富差距问题逐渐成为全社会共同关注的焦点,构成了影响民众主观阶层意识的重要原因(陈云松、范晓光,2016)。另外,由于群体性的利益矛盾在21世纪初不断激化,各种社会运动开始大量出现。有关实证研究表明,城乡居民不仅在基层社会政治参与中的组织性和凝聚力有了很大提高,而且在意识层面开始对群体性的利益取向以及目标实现手段达成某种共识,从而强化了本阶层的认同感,但也加剧了与其他阶层之间的冲突(刘精明、李路路,2005;陆益龙,2010)。

其二,发现(2)初步验证了收入不平等对于民众阶层意识的影响作用比宏观经济的持续增长更显著。这一结论已在安德森和柯蒂斯(Andersen & Curtis, 2012)针对44个国家数据样本的研究中得到了证实。然而,我们在表4中进一步发现,这一现象主要出现在2000年之后的改革阶段。在这之前,宏观经济发展是提升公众的阶层关注度的主要经济机制。事实上,伊斯特林等人(Easterlin et al., 2012)的一项著名的相关研究发现,虽然中国人均GDP在1990-2010年不断增长,但民众对生活的满意度并没有获得相应提高。他们认为,最有可能的原因是与经济高速增长同时出现的贫富差距问题。在这之后,吴晓刚和李骏(Wu & Li, 2017)针对近年来中国民众的主观幸福感的研究进一步证实了伊斯特林等人的观点。

其三,发现(3)与发现(4)共同反映了改革后国家舆论导向与社会大众之间的关系所发生的深刻变化。一方面,针对主流舆论在引导民众观念上的“失灵”,学者们的解释主要包括两点:第一,随着高度集中的计划经济体制的解体,获得经济自主性的个人开始抽离于高度整合的社会结构,个体化取向逐渐在人们的生活方式、价值取向、行为选择和文化观念上蔓延开来,因而无法与仍保留权威性和支配性特征的某些舆论导向成分建立认同联系(李培林,2005);第二,在市场化改革、科技发展和全球化因素的推动下,公众的信息获取渠道和意见表达途径日益多元化,社会舆论的独立性明显增强(谭伟,2003)。

另一方面,尽管改革以来民间关于阶层议题的讨论已开始影响主流舆论导向的变化,但不同时期其作用方式存在明显区别。改革初期,重大改革政策的出炉总是伴随着公众对某项议题的社会性讨论而出

现。例如在 1978 年的真理标准大讨论之后,中央政府在意识形态和政策设计上的改革导向才日益明晰。但随着改革战略在 20 世纪 80 年代末开始收紧,改革的总体进程开始围绕建立市场经济体制这一主线运行,政策的变动性逐渐降低。同时,随着民意代表机制、政治协调机制、社会监督机制和民意表达机制的逐步确立,民众意见的表达方式开始朝常态化、制度化的方向发展,其社会影响力也不断增强,但也主要是在现有政治制度框架内,很难对国家舆论导向形成强烈冲击。

其四,对于发现(5),即民众的阶层意识也会反过来影响政治经济走势以及舆论导向,我们推测,这可能是由于经济发展和贫富差距是普通民众最为关心的社会议题,并且人们在实践中往往会通过扩大政治参与将其意见纳入主流舆论导向之中,最终转化为实际的经济和政治决策。在这种公众与国家之间重新建构的互动关系中,有两项决策形成机制的作用值得重视。其一,促进决策科学化和民主化是当代中国政治体制改革的目标之一。当前,决策听证制度已在全社会普遍实施,其内容涵盖了价格调整、规章制定和行政处罚等多个领域(彭宗超、薛澜,2000)。其二,网络空间的出现加速了媒介权力结构草根化的趋势。有时在舆情传播和引导过程中,普通民众可以通过设置特定议题进行网络动员和鼓动造势,形成针对政府的巨大舆论压力并迫使其及时回应和处置(杨嵘均,2015)。

(四) 稳健性检验

尽管我们在条件格兰杰分析中控制了多方面的宏观影响因素,但仍可能存在某些干扰变量。特别是词频的变动可能会受到社会科学类出版物变化的影响。依此逻辑,我们之前所获得的研究发现反映的可能并非是公众对社会分层领域问题的关注度,而仅仅是在经济高速增长背景下,学术出版物特别是社会科学出版物不断扩张并最终显示为语料库膨胀的结果。对于这一问题,我们参照陈云松和严飞(Chen & Yan, 2016)采取的以下方法进行验证:引入学科词频,并证明该类词频与经济发展水平无关。具体而言,我们选取了四门学科的词汇,包括可以代表社会科学的“社会学”和“经济学”,以及可以代表自然科学的“物理学”和“生物学”,并将每一门科学在谷歌图书汉语(简体)语料库中的词频比例进行了标准化(分别为 $LC/Sociology$ 、 $LC/Economy$ 、 $LC/Physics$ 、 $LC/Biology$),然后比较它们与经济发展走势(GDP_{cp})的演化轨

迹。此外,我们也对四门学科的词频比例进行了主成分分析,而后合成一个综合性指标“*Discipline*”,与经济发展指标“ GDP_{cp} ”进行双变量的格兰杰检验。分析表明,不仅图形结果显示学科词频的变动与经济走势并不一致,而且统计结果也并不支持经济增长是学科词频扩张的格兰杰原因($p > 0.1$)。^①这样,我们基本上可以排除经济增长通过促进社会科学繁荣进而使阶层类词频所代表的阶层关注度增长的情况。

另一个需要关注的问题是分时段分析的样本较少,可能对模型的稳定性造成影响。不过,我们执行这一分析的目的主要是对前一步的条件格兰杰检验的结果加以进一步解读。而且,对比分时段分析和总体分析的统计结果,可以发现两者存在较高的吻合度,这也从侧面说明不稳定的问题并不严重。另外,我们也采用了基于 T 检验的双变量格兰杰检验重新进行了分时段回归。这是因为这一检验的基本前提——Anderson-Darling Test 检验正态分布所需要的最小样本量为 6,小于分阶段分析的所需的最小样本量 9。从最终呈现的回归结果来看,与采用条件格兰杰回归的分阶段分析的结果基本保持一致,因此可以认为模型结果具有很强的稳健性。^②

六、总结与讨论

基于谷歌图书语料库汉语(简体)中的 40 个阶级类词汇和阶层类词汇,本文对在 1949 - 2008 年中国公共话语中有关社会结构的话语定义进行了全方位考察。我们发现,阶级话语在改革开放前的 30 年中占据着主导地位,但在 1978 年后日益被与阶层有关的议题所取代,从而表现为社会分层的话语建构模式发生了从官方意志到公众态度的重大转型。基于时间序列的格兰杰因果检验为这一变化提供了实证性解释。总体而言,宏观经济的高速增长、收入不平等的扩大以及政治参与水平的提升均有助于解释改革以来公众阶层意识提升的现象。但同样作为市场转型的后果,经济增长所产生的效应远不及收入差距这一负向因素。反过来,民众阶层意识的变化也会对上述制度性变迁产生不

① 限于篇幅,此处并未展示可视化图形和格兰杰因果检验的结果,读者可向作者索取。

② 限于篇幅,详细结果此处不再赘述,请向作者索取。

同程度的影响,尤其是对经济发展和政治参与的影响较大。另外,我们也发现,在改革期间,主流舆论导向对普通民众的影响开始下降,同时也会受到后者的影响。由此,我们提出的第二个问题可以得到确切的答案,也即改革开放以后,制度转型对中国民众阶层意识的提升发挥了重要影响,这主要源于市场化改革和政治体制改革的力量。同时,主流意识形态对建构民众思想意识的强制性影响日益减弱,而基层社会的力量正通过各种渠道影响官方话语体系的走向。

事实上,阶层意识不仅涉及个体对自身客观地位的判断,同时也是衡量“获得感”的重要标准。2015年以来,习近平总书记多次强调要通过改革给人民群众带来更多“获得感”,把“是否给人民群众带来获得感”作为检验改革成效的重要标准。^①党的十九大更明确地指出:“我国社会主要矛盾已经转化为人民日益增长的美好生活需要和不平衡不充分的发展之间的矛盾。”^②因此,本文的政策意涵在于:在改革开放进入深水区的阶段,相较于一味追求经济总量的快速增长,缩小收入差距、提高公民的政治参与积极性和能力,将是缓解阶层矛盾、促进社会流动、防止坠入“转型陷阱”的重要手段。在中国未来的发展过程中,必须不断完善利益分配机制,扩大社会流动的空间和渠道,维护公民的各项合法政治权利,健全利益表达与维护机制。与此同时,国家与民众之间的互动关系也需要进行优化调整,改进公共舆论的引导机制,将公众意见更多地纳入主流舆论导向和重大决策之中,提升国家治理水平与社会现代化程度。

本文在数据和方法层面仍存在一些局限,表现为四个方面。第一,进入21世纪后,互联网数据和社交媒体数据在对公众态度的代表性上可能会逐渐与书籍“分庭抗礼”。第二,以词汇来归纳公众的阶层关注度的方法仍存在一些缺陷。例如,职业类词汇总是无法全面反映改革以来中国阶层结构急剧分化以及职业类型迅速增多的事实。第三,采用格兰杰因果检验方法来确定阶层关注度与宏观结构性因素之间的关联,并不能完全确认为反事实框架下的因果关系。第四,尽管谷

① 参见习近平总书记2015年2月27日、2016年4月18日和7月22日在中央全面深化改革领导小组会议上的讲话(http://china.chinadaily.com.cn/2016-09/25/content_26887181.htm)。

② 参见习近平总书记在2017年10月18日在中国共产党第十九次全国代表大会的报告(http://cpc.people.com.cn/n1/2017/1028/c64094_29613660.html)。

歌图书数字化工程项目的书籍来自全球顶尖高校图书馆和出版社,且双方协议均表明供书方将书籍全部而非选择性地送往项目实施方,但我们也不能完全排除中文图书收藏和出版时的意识形态筛选风险。当然,从国际高校图书馆收藏中文书籍的宗旨和实践来看,这一风险倒不必被高估,因为从根本上而言,基于各种意识形态的书籍都是学术机构收藏和研究的对象。

参考文献:

- 陈光金,2013,《不仅有“相对剥夺”,还有“生存焦虑”——中国主观认同阶层分布十年变迁的实证分析(2001—2011)》,《黑龙江社会科学》第5期。
- 陈思,2012,《社会转型进程中中国政府的“稳定”与“改革”意识》,《中州学刊》第6期。
- 陈云松,2015,《大数据中的百年社会学——基于百万书籍的文化影响力研究》,《社会学研究》第1期。
- 陈云松、范晓光,2016,《阶层自我定位、收入不平等和主观流动感知》,《中国社会科学》第12期。
- 陈云松、吴青熹、张翼,2015,《近三百年中国城市的国际知名度——基于大数据的描述与回归》,《社会》第5期。
- 陈云松、张亮亮、闵尊涛、张柏杨,2016,《大数据机遇与宏观定量社会学的重启》,《贵州师范大学学报(社会科学版)》第6期。
- 成伯清,2009,《从嫉妒到怨恨——论中国社会情绪氛围的一个侧面》,《探索与争鸣》第10期。
- 村田忠禧,2002,《从〈人民日报〉元旦社论看中华人民共和国的历史》,《中共党史研究》第3期。
- 范晓光、陈云松,2015,《中国城乡居民的阶层地位认同偏差》,《社会学研究》第4期。
- 龚为纲、罗教讲,2015,《大数据视野下的19世纪“海上丝绸之路”——以丝绸、瓷器与茶叶的文化影响力为中心》,《学术论坛》第12期。
- 郭正林,2003,《当代中国农民政治参与的程度、动机及社会效应》,《社会学研究》第3期。
- 胡荣,2005,《社会资本与城市居民的政治参与》,《社会学研究》第5期。
- 吉登斯,安东尼,2009,《社会学》,李康译,北京:北京大学出版社。
- 贾春增,2008,《外国社会学史(第三版)》,北京:中国人民大学出版社。
- 李晨璐、赵旭东,2012,《群体性事件中的原始抵抗》,《社会》第5期。
- 李骏,2009,《住房产权与政治参与:中国城市的基层社区民主》,《社会学研究》第5期。
- 李强,1993,《当代中国社会分层与流动》,北京:中国经济出版社。
- ,2016,《中国离橄榄型社会还有多远——对于中产阶级发展的社会学分析》,《探索与争鸣》第8期。
- 李培林,1995,《新时期阶级阶层和利益格局的变化》,《中国社会科学》第3期。
- ,2015,《社会冲突与阶级意识——当代中国社会矛盾研究》,《社会》第1期。
- 李培林、张翼,2008,《中国中产阶级的规模、认同和社会态度》,《社会》第2期。

- 李培林、张翼、赵延东、梁栋,2005,《社会冲突与阶级意识:当代中国社会矛盾问题研究》,北京:社会科学文献出版社。
- 梁丽萍、邱尚琪,2004,《建国以来中国公民政治参与模式的演变分析》,《学术论坛》第5期。
- 刘精明、李路路,2005,《阶层化:居住空间、生活方式、社会交往与阶层认同——我国城镇社会阶层化问题的实证研究》,《社会学研究》第3期。
- 刘欣,2001,《转型期中国大陆城市居民的阶层意识》,《社会学研究》第3期。
- 柳建坤、吴愈晓、刘伟峰,2016,《中国城市国内知名度的变迁和机制——基于海量书籍和互联网搜索引擎的大数据分析》,《学术论坛》第6期。
- 陆学艺,2002,《当代中国社会阶层研究报告》,北京:社会科学文献出版社。
- ,2003,《当代中国社会阶层的分化与流动》,《江苏社会科学》第4期。
- 陆益龙,2010,《乡村居民的阶级意识和阶层认同:结构抑或建构——基于2006CGSS的实证分析》,《江苏社会科学》第1期。
- 卢汉龙,1996,《城市居民社会地位认同研究》,中国社会科学院社会学研究所编《中国社会学年鉴1992.7-1995.6》,北京:中国大百科全书出版社。
- 马广海,2011,《存在、认同与冲突:转型期我国社会的阶层意识概览》,《山东社会科学》第5期。
- 彭宗超、薛澜,2000,《政策制定中的公众参与——以中国价格决策听证制度为例》,《国家行政学院学报》第5期。
- 渠敬东、周飞舟、应星,2009,《从总体支配到技术治理——基于中国30年改革经验的社会学分析》,《中国社会科学》第6期。
- 仇立平、韩钰,2015,《中国城市居民阶层地位认同偏移研究》,《社会发展研究》第1期。
- 孙立平,1996,《“关系”、社会关系与社会结构》,《社会学研究》第5期。
- 孙永芬,2008,《政治心态与影响因素间变量关系的实证分析》,《政治学研究》第1期。
- 谭伟,2003,《网络舆论概念及特征》,《湖南社会科学》第5期。
- 翁定军,2010,《阶级或阶层意识中的心理因素:公平感和态度倾向》,《社会学研究》第1期。
- 谢立中,2007,《西方社会学名著提要》,南昌:江西人民出版社。
- 杨嵘均,2015,《论网络空间草根民主与权力监督和政策制定的互逆作用及其治理》,《政治学研究》第3期。
- 张鸿雁、柳建坤,2017,《中国城市国内知名度的历史变迁和变动机制的大数据分析——以江苏苏南城市为例》,《新疆师范大学学报(哲学社会科学版)》第1期。
- 张亮亮、张洪、李双龙、吴楷文,2016,《大数据中的中国文化名人名著:500年国际知名度分析》,《学术论坛》第6期。
- 张济顺,2004,《上海里弄:基层政治动员与国家社会一体化走向》,《当代中国史研究》第4期。
- 张乐天,2011,《国家话语的接受与消解——公社视野中的“阶级”与“阶级斗争”》,《社会学研究》第6期。
- 周晓虹,2002,《中产阶级:何以可能与何以可为?》,《江苏社会科学》第6期。
- Andersen, Robert and Josh Curtis 2012, “The Polarizing Effect of Economic Inequality on Class Identification: Evidence from 44 Countries.” *Research in Social Stratification and Mobility* 30

(1).

- Chen, Yunsong & Fei Yan 2016, "Economic Performance and Public Concerns about Social Class in Twentieth-Century Books." *Social Science Research* 59.
- Curtis, Josh 2015, "Mobility and Class Identity in 35 Societies: The Role of Economic Conditions." *The European Sociological Review* 32(1).
- Easterlin, Richard, Robson Morgan, Malgorzata Swite & Fei Wang 2012, "China's Life Satisfaction, 1990 - 2010." *Proceedings of the National Academy of Sciences* 109 (25).
- Evans, Mariah D. R. & Jonathan Kelley 2004, "Subjective Social Location: Data from 21 Nations." *International Journal of Public Opinion Research* 16(1).
- Evans, Mariah D. R. , Jonathan Kelley & Tamas Kolosi 1992, "Images of Class: Public Perceptions in Hungary and Australia." *American Sociological Review* 57(4).
- Granger, Clive W. J. 1969, "Investigating Causal Relations by Econometric Models and Cross-Spectral Methods." *Econometrica* 37(3).
- Hodge, Robert & Donald Treiman 1968, "Class Identification in the United States." *American Journal of Sociology* 73(5).
- Jackman, Mary R. & Robert W. Jackman 1983, *Class Awareness in the United States*. Berkeley: University of California Press.
- Janmaat, Jan Germen 2013, "Subjective Inequality: A Review of International Comparative Studies On People's Views about Inequality." *European Journal of Sociology* 54(3).
- Shirahase, Sawako 2010, "Japan as a Stratified Society: With a Focus on Class Identification." *Social Science Japan Journal* 13(1).
- Scott, John & Gordon Marshall 2005, *Oxford Dictionary of Sociology*. Oxford: Oxford University Press.
- Wu, Xiaogang & Jun Li 2017, "Income Inequality, Economic Growth, and Subjective Well-being: Evidence from China." *Research in Social Stratification and Mobility* 52.

作者单位:清华大学社会学系(柳建坤)

南京大学中美文化研究中心、社会学系(陈云松)

责任编辑:杨 可