

# 大数据中的中国世界文化遗产： 500年国际知名度分析

陈云松,孙 艳,严 飞

[摘要] 文章利用谷歌图书的百万书籍大数据,以中国世界文化遗产近500年来英语书籍中出现的词频来展示和分析其国际知名度的变迁及其特征。研究发现,长城、故宫、丝绸之路、元上都、布达拉宫、莫高窟、周口店、明清皇陵、孔庙孔府孔林和庐山依次为近500年来中国世界文化遗产国际知名度前十名。不在名录之中的南京大报恩寺塔,在17至19世纪中知名度仅次于长城。中国世界文化遗产国际知名度与国家政治命运、跨国文化交流、国际旅游业发展及城市国际知名度等多种因素相关。

[关键词] 大数据;国际知名度;文化遗产;文化传播

[作者简介] 陈云松,南京大学社会学院副教授,硕士生导师,牛津大学社会学博士,江苏,南京210046;孙艳,无锡职业技术学院讲师,江苏 无锡 214028;严飞,斯坦福大学亚太中心博士后,美国,加利福尼亚州,帕拉阿图 94305

[中图分类号] C919:G05

[文献标识码] A

[文章编号] 1004-4434(2015)12-0092-07

## 一、导 言

世界遗产是指被联合国教科文组织和世界遗产委员会确认的人类罕见的、目前无法替代的财富,是全人类公认的具有突出意义和普遍价值的文物古迹及自然景观。世界遗产主要包括世界文化遗产、世界自然遗产、世界文化和自然遗产、人类口述和非物质遗产代表作4类。世界文化遗产在《世界遗产公约》中的定义包括文物古迹、建筑群和遗址。截至目前,中国拥有48处世界遗产,遗产总数仅次于意大利,居世界第二。其中,中国被联合国教科文组织和世界遗产委员会确认的包括长城、故宫、丝绸之路、秦始皇陵兵马俑等在内的世界文化遗产共有34项,分布于我国的北京市、山西省、云南省和澳门特别行政区等多个省、自治区和直辖市。

中国世界文化遗产的国际知名度不仅体现了该文化遗产本身在国际层面的知晓度和关注度,也体现了该文化遗产所在地乃至中国在国际层面的知晓度和关注度。更为重要的是,它是衡量整个中华民族文化国际知名度和国际影响力的一个重要组成部分。纵观世界文明发展史,一个国家的文

语权不仅与它的经济实力挂钩,更与它的文化影响力密不可分。党和国家高度重视中华文化国际影响力的提升,因此,研究中国的世界文化遗产的国际知名度与关注度,是国家层面提升我国文化国际知名度和影响力的题中应有之义。

衡量文化遗产的价值,并没有放诸四海而皆准的标尺。但是,“国际知名度”无疑是其中最重要的测度之一。不过,梳理以往对文物和文化遗产的研究,我们不难发现:绝大部分研究都是基于文物保护、文化遗产建设和保护的思路进行,从文化遗产的国际知名度这一视角对文化、城市以及国家发展进行分析的研究很少。究其原因,主要是因为用传统的问卷方法在人力财力物力上难以实现对国际知名度的抽样调查,而对历史上的知名度更无能为力。不过,大数据的应运而生,为我们对国际知名度历史轨迹进行精准测量提供了重大机遇。

## 二、国际知名度的测量和数据

我们认为,语言学中的语料库词频分析方法可以作为测量文化遗产国际知名度的工具。在本研究中,我们用英语书面语言中使用、提及有关中国世界文化遗产的频次,作为衡量一项中国世界文化遗

产国际知名度的测度。具体而言,我们将分析 500 年以来的英语书籍数据,以此对中国的世界文化遗产“国际知名度”进行概念操作化,使得我们可以在前所未有的时间跨度(本研究中为 500 年)和空间跨度(全球英语使用地区)上对中国各世界文化遗产的国际知名度进行一次全景式的探索。考虑到每年书籍中的词汇量不尽相同,因此我们用词频除以当年的词汇总量作为当年的国际知名度测量。

该方法在科学研究上的严密性和合理性在于:第一,书面语言本身是承载人类观念、意识和价值观的最重要载体,而英语是全球使用最广、最重要的国际通用语言,因此,英语书面语言载体是研究国际影响力的最佳数据库;第二,海量书籍中的语言和词汇既能反映作者、撰稿人的个人观点,又能反映和捕捉社会大众的整体思潮。数个世纪的英语印刷书籍,实际上构成了西方社会知识、观念和经验的总体。只要语料库具有足够的代表性,我们就可以认为一个词汇在书籍中出现的频率,能够近似地反映这个词汇及其相关意蕴的社会文化影响力、知名度,甚至折射出某种社会趋势、风尚或思潮(Twenge et. al. 2012)。实际上,该方法目前已经在国内社会科学界得到应用(陈云松, 2015;陈云松,吴青熹,张翼,2015)。

本研究中所用到的近 500 年来的书籍数据来源于谷歌图书语料库。自 2004 年底开始,谷歌公司陆续对哈佛、牛津等 40 多所顶级大学图书馆藏书及出版社赠书进行了浩大的数字化工程。让·巴蒂斯特·米歇尔等人选择了其中 500 多万种(5,195,769) 识别质量较高的非期刊书籍作为其文化定量分析的语料库,在《科学》杂志上发表了关于文化组学的重要论文(Michel et. al. 2011)。到 2013 年,谷歌已对超过 3000 万种书籍进行了扫描和识别,可供分析的最新版语料库书籍达到 800 多万种(8,116,746), 词汇量更高达 8,613 亿(Lin et. al. 2012)。

谷歌语料库为文化研究、语言学研究、观念史研究等提供了难得的社会科学大数据。戴维斯(Davis, 2011)和乔拉(Joula, 2012)等人研究了谷歌图书语料库的语法结构演变以及文化信息的复杂

度;特温格等(Twenge et. al. 2012)则对美国 20 世纪书籍中个人化用词的趋势进行了大历史的解读;阿瑟比等(Acerbi et. al. 2013)对人类 20 世纪书籍中感情用词的演变以及英式美式英语差异进行了分析;宾利等(Bentley et. al. 2014)研究了 20 世纪美国经济萧条与悲观性词汇之间的关联。这些作为“文化组学”的新的研究案例和方法,为我们从新的角度观察社会学发展史和探索社会学研究新方法提供了启示。

### 三、500 年来中国世界文化遗产的国际知名度

截止 2015 年 7 月,中国共有 48 处世界遗产,其中世界文化遗产 34 项(含文化景观遗产 4 项)。本文图表中所列的中国世界文化遗产数据研究截至 2015 年 1 月,共计 33 处(“土司遗址”于 2015 年 7 月 4 日列入世界文化遗产名录, 暂未进入本文分析),包括长城(北京)、故宫(北京)、周口店“北京人”遗址(北京)、颐和园(北京)、明清皇陵(北京、江苏、安徽)、天坛(北京)、大运河(北京)、元上都(内蒙古)、布达拉宫(西藏)、云冈高句丽(吉林)、丝绸之路(河南、陕西、甘肃、新疆)、莫高窟(甘肃)、承德避暑山庄(河北)、龙门石窟(河南)、登封历史建筑群(河南)、殷墟(河南)、兵马俑(陕西)、平遥古城(山西)、云冈石窟(山西)、孔庙孔府孔林(山东)、苏州园林(江苏)、皖南古村落(安徽)、武当山(湖北)、青城山(四川)、大足石刻(重庆)、丽江古城(云南)、开平碉楼(广东)、土楼(福建)、澳门历史城区(澳门)、庐山(江西)、五台山(山西)、西湖(浙江)、红河哈尼梯田(云南)。

我们发现,在 1500~1700 年间的英语书籍大数据中,仅有长城有所记载,而其他的提及极少。因此,我们分别对近 300 年、200 年、150 年、100 年、50 年和 20 年的时间序列进行单独分析,并将有关时段的世界文化遗产国际知名度进行了计算和排名。我们把排名结果报告在表 1 中(下文表格中为表述方便,部分遗址使用其所在地名作为简称,如登封历史建筑群记作登封、承德避暑山庄记作承德、平遥古城记作平遥等)。

表 1 中国世界文化遗产近 300 年各阶段国际知名度

年 排名	近 300 年	近 200 年	近 150 年	近 100 年	近 50 年	近 20 年
	1700-2000	1800-2000	1850-2000	1900-2000	1950-2000	1980-2000
1	长城	长城	长城	长城	长城	长城
2	丝绸之路	故宫	故宫	故宫	故宫	故宫
3	故宫	元上都	元上都	元上都	元上都	丝绸之路

(续表)

排名	年	近 300 年	近 200 年	近 150 年	近 100 年	近 50 年	近 20 年
		1700-2000	1800-2000	1850-2000	1900-2000	1950-2000	1980-2000
4		元上都	丝绸之路	丝绸之路	丝绸之路	庐山	元上都
5		皖南	布达拉宫	布达拉宫	布达拉宫	丝绸之路	莫高窟
6		布达拉宫	莫高窟	莫高窟	莫高窟	布达拉宫	布达拉宫
7		莫高窟	周口店	孔庙孔府	周口店	莫高窟	周口店
8		周口店	明清皇陵	周口店	明清皇陵	周口店	孔庙孔府孔林
9		明清皇陵	孔庙孔府孔林	明清皇陵	孔庙孔府孔林	孔庙孔府孔林	明清皇陵
10		孔庙孔府孔林	庐山	庐山	庐山	明清皇陵	西湖
11		土楼	五台山	五台山	五台山	西湖	兵马俑
12		青城山	西湖	西湖	西湖	五台山	庐山
13		庐山	兵马俑	兵马俑	兵马俑	兵马俑	五台山
14		五台山	大运河	大运河	登封	天坛	天坛
15		西湖	天坛	天坛	天坛	登封	登封
16		兵马俑	颐和园	颐和园	大运河	颐和园	云冈石窟
17		大运河	登封	登封	平遥	苏州园林	苏州园林
18		天坛	平遥	平遥	颐和园	殷墟	颐和园
19		颐和园	承德	承德	龙门石窟	龙门石窟	殷墟
20		澳门	澳门	澳门	高句丽	高句丽	龙门石窟
21		登封	龙门石窟	龙门石窟	殷墟	云冈石窟	承德
22		平遥	高句丽	高句丽	云冈石窟	承德	平遥
23		承德	殷墟	殷墟	苏州园林	平遥	高句丽
24		龙门石窟	云冈石窟	云冈石窟	承德	大运河	青城山
25		高句丽	苏州园林	苏州园林	青城山	青城山	武当山
26		殷墟	青城山	青城山	武当山	武当山	大运河
27		云冈石窟	武当山	武当山	澳门	澳门	澳门
28		苏州园林	丽江	丽江	丽江	丽江	红河
29		武当山	红河	红河	红河	红河	丽江
30		丽江	大足石刻	大足石刻	大足石刻	大足石刻	大足石刻
31		红河	开平碉楼	开平碉楼	开平碉楼	开平碉楼	开平碉楼
32		大足石刻	土楼	土楼	土楼	土楼	土楼
33		开平碉楼	皖南	皖南	皖南	皖南	皖南

从表 1 可以看出,前 10 位文化遗产的平均知名度排序总体上稳定。实际上,近 300 年、200 年、150 年、100 年、50 年和 20 年等六个不同时期的平均知名度前 10 强,集中在 12 项。总体上,500 年以来,长城、故宫、元上都、丝绸之路、布达拉宫、莫高窟、周口店、孔庙孔府孔林、明清皇陵和庐山稳居中国的世界文化遗产国际知名度前十名<sup>①</sup>(西湖在近 20 年曾经进入过前十)。它们在任何一个历史阶段的平均排名都非常靠前。换句话说,中国世界文化遗产国际知名度本身是一个较为稳定的指标,尽管中国幅员辽阔、文化遗产众多,且 500 年来经历了明代、清代、民国和新中国四个政权,但最知名的世界文化遗产基本就是这 10 个。

除了观察平均值之外,我们还有必要对中国的文化遗产国际知名度的总体演变情况进行分析。考虑到图示数量有限,我们在图 1 中用时间序列曲线来展示了 500 年来进入国际知名度前 10 强的世界文化遗产。图 1 最大的特点就是中国世界文化遗产的国际知名度呈现出非常明显的梯次和波动。而这些变迁,富含了大量的历史、政治、经济和社会信息,值得我们进一步分析探讨。尤其是在近 500 年的英语书籍中,中国的世界文化遗产——长城,显示出了长期一枝独秀的稳定局面。1500 年至 1700 年间的 200 年内,仅有长城在英语书籍中被长期提及;1700 年至 1800 年间,除了长城外,仅有孔庙孔府孔林、天坛和丝绸之路被少次提及;

1800年至1900年间,在英语书籍中提及相对较多的依次为长城、孔庙孔府孔林、故宫、元上都、明清皇陵、澳门历史城区、大运河和布达拉宫;1900年以后,除长城以外的中国世界文化遗产才被普遍提及。毫无疑问,长城和故宫是中国最为知名的文化遗产,而其他遗产与之相比存在较大差距。

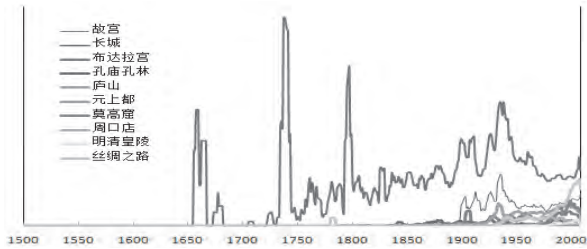


图1 中国世界文化遗产国际知名度前10强(基于书籍大数据)

数据来源:谷歌英语图书,1500-2000年。

此外,1700年后,由丝绸之路、孔庙孔府孔林开始,中国的世界文化遗产在英语书籍中开始普遍提及并且波动上升。例如,孔庙孔府孔林的国际知名度发展比较平稳,没有明显较大的跌宕起伏;丝绸之路的国际知名度自1920年以后一路攀升,近30年来更是呈现飞速上升的发展趋势;莫高窟的国际知名度在改革开放后出现了大幅上升的发展趋势并处在仅次于丝绸之路的较高梯次空间内;布达拉宫在1905年左右出现了历史最高峰值,此后一直起伏较大呈现了明显的波动和梯次;元上都都在1933年左右达到了历史最高点,之后出现了小幅回落并一直稳定在较高梯次范围;而西湖、庐山、周口店等其他几项文化遗产虽然也有不同程度的发展和波动,但大致都处于较低梯次范围内。

#### 四、中国世界文化遗产国际知名度的特征

观察图1,我们将遗产的国际知名度特征总结为以下四个方面:

第一,中国世界文化遗产国际知名度局部“波峰”与国家政治命运休憩相关。仔细观察各个文化遗产的词频曲线,我们会发现,无论是在曲线的高处还是低处,总有不时的尖锐起伏和波动。而这些波动峰值的出现,与中国历史上的“大事件”乃至政治命运有着奇妙的相关度。我们仍以排名前十的文化遗产词频曲线来进行解读。

从图1中我们可以看到,在近500年的书籍文化大数据中,长城凸显了它一枝独秀的独特魅力,一直强势上升并长期维持在一个高位运行状态,自1655年有记录以来,其国际知名度分别显著出现了4次峰值:第一次高峰处于1657~1662年左右,

正值清康熙帝接替顺治帝的权力更替之际,即顺治十八年(1661年)顺治皇帝驾崩,传位于玄烨即康熙皇帝。第二次峰值出现在1735~1740年左右,恰逢乾隆接替雍正的权力更替之际,即雍正十三年(1735年)八月雍正辞世,乾隆继位。这与北京在近300年来英语书籍中被提及频率在这一时段的词频波峰曲线同步。第三次峰值出现在1795~1800年前后,正是嘉庆皇帝接替乾隆皇帝的权力更替之际。第四次峰值出现在20世纪30年代,也正是日军侵华战争爆发时期。换句话说,每逢书籍里提到中国和中国政治,长城是必不可少的文化意象。

从图1中我们还可以清楚看到,1900年以后是中国世界文化遗产被普遍有所记录的开始,在此以前,除了长城的一枝独秀外,只有故宫、元上都、丝绸之路等为数不多的几处有所记载。而1900年后,中国的世界文化遗产在英语书籍中被提及次数大量增多,并且总体幅度有愈来愈高的上升趋势。这与当时中国政治所处的命运更是息息相关,故宫便是一个明显例证。1894年甲午中日战争中国战败,与日本签订《马关条约》,中国国际地位一落千丈。19世纪末,西方列强掀起了瓜分中国的热潮。1900年爆发八国联军侵华战争,给中国人民带来了深重的灾难,中国的政治、经济、文化各方面都遭受了重创。根据中国人民大学清史研究所何瑜教授所述:八国联军侵华时,北京古都遭到了亘古未有的浩劫。珍宝、文物或被抢,或被毁、或被侵略者当场拍卖,实在难以完全统计清楚。相应地,1900年后的故宫——这座晚清遭受重创被欺辱瓜分的皇宫,在英语书籍词频提及率显著升高,并在日军侵华战争时期即1937年前后达到又一高峰。此后直到建国前,中国的世界文化遗产在英语书籍中的提及频次逐渐下降,并于建国前夕降到谷底,这与当时中国的政治命运休戚相关。

新中国成立后,特别是改革开放30年来,随着国力的增强,中国日益以一种开放、自信和宽容的姿态融入国际社会,中华文化也逐渐挺起胸膛。最明显的例子就是,汉语热经久不衰,以传播中国传统文化为己任的“孔子学院”遍布世界。随着中国重新融入世界,中国世界文化遗产在英语书籍中的提及频次才逐渐稳步提升

第二,中国世界文化遗产国际知名度与跨国交流密切相关。文化是一定的历史阶段、一定的地域环境、一定的人类种群的一种生存状态、生活习惯、思维方式的反映。文化是一条历史长河,从远古流到现在,从现在还要流向未来。尽管千变万化,但是



作为特定的文化,总保留着祖先的基因。我们研究文化遗产,就是在研究人的生存状态,研究人的过去和未来。文化遗产往往和一些人类文化现象或重大的历史事件有着密切的关系,更与中国的跨国文化、政治、经济、科技等各方面的交流密切相关。

如排名前三的元上都就是一个很好的例证(以下参考百度百科)。元上都位于今内蒙古自治区锡林郭勒盟,是忽必烈于公元13世纪中叶在中国北方草原上建立的都城,与元大都(今北京市)共同构成元朝两大首都,是一座国际性大都会,在欧亚大陆上有着很大影响。“一座元上都,半部元朝史”,元朝的11位皇帝中,有6位皇帝是在上都登基。著名的“佛道辩论”、创立行省制度等一大批重要的历史事件也发生在上都。元代的许多重大科技活动和决策是在元上都进行的,是当时中外科技交流的中心之一。13世纪马可波罗于此地获忽必烈觐见,并深得器重,在中国为官居住了17年之久。《马可波罗行记》详细记述了元上都的宫殿、寺院、宫廷礼仪、民俗风情,向世界介绍了元上都,让世界了解中国。元上都国际知名度排名如此靠前与成吉思汗和忽必烈这两位举世瞩目的人物一生所创造的重大成就密切相关,也与元朝在跨国文化、经济、政治、军事和科技各方面所享有的影响力密切相关。

在这方面,孔庙孔府孔林也是一个较显著的例证。孔庙、孔府、孔林位于山东省曲阜市,合称“三孔”。是中国历代纪念孔子,推崇儒学的表征,以丰厚的文化积淀、悠久历史、宏大规模、文物珍藏丰富以及科学艺术价值而著称。曲阜孔庙是祭祀孔子的本庙,是分布在中国、朝鲜、日本、越南、印度尼西亚、新加坡、美国等国家2000多座孔子庙的先河和范本,始建于公元前478年,历经2400多年而从未放弃祭祀,是中国使用时间最长的庙宇,也是中国现存最为著名的古建筑群之一。从图1中我们可以清晰看到,虽然孔庙孔府孔林没有特别明显的波动,但它在英语书籍中的记载是仅次于丝绸之路而长期稳定出现的,这与孔子本人及其对世界文化的重要影响密不可分。

第三,中国世界文化遗产国际知名度与城市旅游业发展紧密相关。文化遗产往往包含着某一地区的历史、文化、宗教、产业变迁等元素,这使得文化遗产对旅游者来说具有强大的吸引力(Harrison, 1997; 费孝通, 2001)。由于经济的发展,人们生活质量提高与交通的日趋方便,在全世界范围内出现了前所未有的旅游热潮,每年世界范围旅游者已经达到数十亿人。遗产旅游被认为是最早的旅游

形式之一,早在18世纪晚期的欧洲就已经产生,但一般认为,1975年欧洲的“建筑遗产年”是遗产旅游成为大众消费需求的标志,在这一年里,介绍城市历史的“遗产中心”随处可见(Dower, M 1978)。世界文化遗产由于它突出的文化价值,使它具有超乎寻常的旅游魅力。如英国的遗产被称为“是吸引海外游客的主要力量”(Marwell, S 1997)。同样在中国,几乎所有的中国世界文化遗产都是旅游爱好者的向往之地。

在图1中我们可以看到,在近20年来的世界文化遗产中,位于浙江杭州的西湖也首次进入到了前十强中。“上有天堂,下有苏杭”,而杭州最为著名的文化旅游景点便是西湖,作为中国世界文化景观遗产的西湖在近20年来的英语书籍词频统计中首次进入前十强,并呈现突飞猛进的发展趋势,这与近年来杭州成为旅游热门城市不无关系。根据2015年4月底出炉的中国各地级以上城市的2014年度国民经济和社会发展统计公报及相关公开资料,有关机构制作了《2014年中国旅游城市总收入排行榜》和《2015年中国旅游城市吸引力排行榜》,两榜单分别统计了全国前100名和前50名城市的相关数据进行排名,榜单分别从旅游接待人数、旅游总收入和游客满意度得分三个维度来评价一座旅游城市对游客的吸引力,杭州位于第五位,仅次于重庆、上海、北京和成都,而前五名中,除了北京拥有7处世界文化遗产外,重庆和成都各有一处。除西湖以外,同是中国世界文化景观遗产的庐山和五台山,在近300年来各历史阶段的词频统计中,也多次位于前列。随着全世界人民对世界文化遗产愈加重视以及旅游业发展的良好前景,我们相信,未来二者相互呼应的关系会愈加增强。

第四,中国世界文化遗产与城市国际知名度相互关联。刘易斯·芒福德(1961)在其经典著作《城市发展史》中指出:“城市从其起源开始便是一种特殊的构造,它专门用来贮存并流传人类文明的成果,提供了最广泛的自由和多样性。”城市是在人类历史活动的时空构架中,在文明与人类社会的发展进程中历史地生成的,从某种意义上说,城市本身就是文化遗产。城市不仅是人类为满足自身生存和发展需要而创造的人工环境,也是一种文化的载体和容器,它的变迁和发展,就是与城市有关的人类文化本身的变迁和发展。文化遗产对于城市的意义显而易见,它们是城市灵魂的载体和表现形式。文化遗产只有植根于特定的人文和自然环境——城市中,才能体现其价值。这就使文化遗产与城市有

着极为深刻的情感联系,这种联系在历史发展过程中,逐渐成为文化遗产不可分割的一部分。

利用陈云松等(2015)的《近三百年中国城市的国际知名度》研究成果,我们对照发现,本文所列的33项文化遗产中,有12项所处的城市处在近300年来中国294座地级以上城市国际知名度排名的前20强。尤其是,北京一直强势占据着近300年来中国城市国际知名度的首位。在我们的研究中,位于北京的世界文化遗产就有7项。其中近300年来中国世界文化遗产国际知名度的前十强中,位于北京的就四项,分别是长城、故宫、明清皇陵和周口店北京人遗址,并且长城和故宫长期居于中国世界文化遗产的前两位。遗产排名居于前五强的布达拉宫所处的城市拉萨,位于中国城市国际知名度前十强。此外,尽管敦煌、孔庙、平遥古城所在的地市级(酒泉、济宁、晋中)知名度并不高,但其所处的县级市(敦煌、曲阜、平遥)本身显然更为知名。

我们的研究表明,中国世界文化遗产的国际知名度与中国城市国际知名度是呈现相辅相成、相互呼应、相互促进的态势的。城市既是人类文明的成果,又是人们日常生活的家园。城市中各个时期的文化遗产像一部部史书,记录着城市的沧桑岁月。有时候,城市街区本身就成为了世界文化遗产。例如,澳门本身在国际城市排名中占据第六强的位置,澳门历史城区也名列世界文化遗产。

## 五、失落的文化遗产:南京大报恩寺宝塔

图2为我们展示了南京大报恩寺塔与前十名中国世界文化遗产在近500年英语书籍大数据中的曲线对比图,图中红色曲线代表的是大报恩寺,我们可以清楚地看到,在近500年来的英语书籍大数据中,大报恩寺是仅次于长城而最早出现的。具体说来,长城最早出现于1655年左右,仅次于长城而出现的中国世界文化遗产天坛出现于1746年左右,在近500年来各历史阶段平均排名居第二位的故宫最早出现于1806年左右,排名第三位的丝绸之路最早出现于1780年左右,而大报恩寺在英语书籍中最早提及可追溯至1732年左右。因此,大报恩寺在中国历史乃至西方历史上中享有了重要历史价值、宗教价值和文化价值。我们可以从它悠久的建造、发展、衰毁和重建史来进一步追根溯源。

大报恩寺是中国历史上最为悠久的佛教寺庙,位于南京市,其前身是东吴赤乌年间(238-250年)建造的建初寺,是继洛阳白马寺之后中国的第二座寺庙,也是中国南方建立的第一座佛寺。大报

恩寺琉璃宝塔高达78.2米,通体用琉璃烧制,自建成至衰毁一直是中国最高的建筑,也是世界建筑史上的奇迹,被当时西方人视为代表中国的标志性建筑之一,有“中国之大古董,永乐之大窑器”(语出张岱《报恩塔》)之誉,被称为“天下第一塔”。大报恩寺塔的前身则是阿育王塔,是南京最古老的佛塔,也是中国最早的佛塔之一。2010年6月12日,大报恩寺阿育王塔中的佛顶真骨盛世重光,是世界现存唯一一枚佛祖真身顶骨舍利。

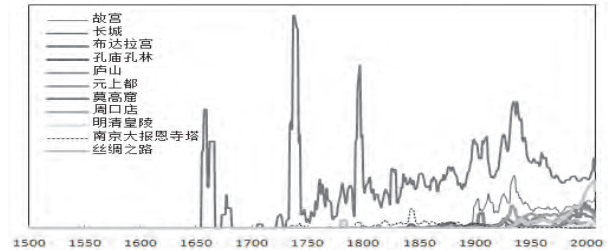


图2 南京大报恩寺塔与中国世界文化遗产(基于书籍大数据)

数据来源:谷歌英语图书,公元1500-2000年。

大报恩寺琉璃塔与罗马大斗兽场、比萨斜塔、中国万里长城等一道被称为中世纪世界七大奇迹,并被西方人视为代表中国文化的标志性建筑之一。不过,直到20世纪前期,中国人自己才知道这样的评价。在某种程度上而言,正是西方对此长期推崇的信息的传入,才重新唤起了20世纪中国人对琉璃塔的历史记忆和现实思考。大报恩寺琉璃塔在欧洲享有的盛誉,首先要归功于荷兰人约翰·尼霍夫及其兄亨利·尼霍夫整理、编辑的《尼霍夫游记》的出版和大量发行,使得大报恩寺塔成为最为欧洲人熟知的中国建筑。到19世纪40年代,尼霍夫的游记和插图仍是欧洲人了解中国的重要知识来源。1839年安徒生就在《天国花园》中提到:“我(东风)刚从中国来——我在瓷塔周围跳了一阵舞,把所有的钟都弄得叮当叮当地响起来。”这里的瓷塔,就是大报恩寺琉璃塔。16-17世纪,随着西方传教士和使团陆续到达中国,琉璃塔逐渐为西方世界所了解。他们惊叹于琉璃塔的壮丽,把它称为中国瓷塔(The Porcelain Tower),并把它与万里长城一道作为中国在西方人心中的表征。

很可惜的是,咸丰四年(1854年),大报恩寺塔被太平天国所毁。它的重建工程已于2004年后积极启动,并作为南京市2010年城市建设“十六个重大项目”之一,2012年11月,南京大报恩寺遗址作为中国海上丝绸之路项目遗产点之一,列入中国世界文化遗产预备名单,根据日程,2015年完成准备工作,2016年正式送交世遗大会审议。我们将中国

文化遗产与大报恩近 300 年来在英语书籍中的词频提及也作一个对比排名,并发现,即便因为在 19 世纪末被焚毁而导致被历史湮没,在 300、200 和 150 年的跨度上,大报恩寺的平均国际知名度仍然是非常靠前的:在近 300 年来英语书籍大数据中的 6 个历史阶段,大报恩寺分别位列第 8、7、7、10、15、24 位,四次进入前十名,只是近 50 年排名才逐渐下降,这充分说明了大报恩寺的国际知名度在中国和世界人民心中的重要位置。同时也更印证了,若非大报恩寺在晚晴时期遭受焚毁,它在中国乃至全世界人民心中所享有的重要历史地位和文化价值,将是仅次于长城的重要文化符号。

## 六、结 语

文化遗产是人类世代代的创造和积累,积淀着各个历史时期的杰出贡献。人类社会的发展是建立在对过去文化、智慧的继承和总结之上的,通过文化遗产,人们才能够知道自己从何处来,知道祖先如何一步步走到今天。由此更加清晰地了解人类的追求,明确我们如何走向明天,走向未来。世界各国都有各自的文化遗产项目 and 对其认识、研究、开发、保护的 tradition。而对于遗产的国际知名度研究,则几乎是一个空白。本文利用大数据的方法,对我国世界文化遗产近 500 年来的国际知名度进行了描述分析。

大数据方法的应用在我国社会科学界尚属起步,其重要原因之一,就是因为大数据难以获得、难以分析。在本文中,我们使用最新版的谷歌图书语料库大数据,对 500 年来的全球数字化英语书籍进行了关键词检索分析,得到了中国世界文化遗产的国际知名度的量化和排名。我们发现,长城和故宫是中国在全球最为知名的文化遗产。17 到 19 世纪中,当时尚未被焚毁的南京大报恩寺塔,则是仅次于长城的中国文化地标。从方法和数据的角度看,本研究表明:第一,从现有的大数据中提取出相关的变量用以描述或分析,是当前利用大数据进行社会科学定量分析的可行途径。第二,词频统计方法是对非结构化的、非为研究定制的文本大数据进行分析的有力武器,其在社会科学领域的用途还将得到进一步拓展。当然,本研究也存在诸多不足:一是用词频代表文化遗产国际知名度的尝试需要在今后的研究中得到进一步验证,以确保这是合理、综合、科学、可信的测度;二是对文化大数据进行检索的精度仍然有进一步提升空间;三是限于时间,我们梳理出的一些现象和规律,有可能只是

挂一漏万,甚至以偏概全。在今后的研究中,我们会加强对上述问题的研究和解决。

## [参考文献]

- [1] 安徒生,叶君健.天国花园[M].上海:上海译文出版社,1978.
- [2] 陈云松.大数据中的百年社会学——基于百万书籍的文化影响力研究[J].社会学研究,2015,(1).
- [3] 陈云松,吴青熹,黄超.大数据何以重构社会科学[J].新疆师范大学学报:哲学社会科学版,2015,(3).
- [4] 陈云松,吴青熹,张翼.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[J].社会,2015,35(5).
- [5] 方晓晓.文化遗产在城市中的双重地位[J].黑龙江史志,2014,(1).
- [6] 单霁翔.留住城市文化的“根”与“魂”:中国文化遗产保护的探索与实践[M].北京:科学出版社,2010.
- [7] 单霁翔.文化遗产保护与城市文化建设[M].北京:中国建筑工业出版社,2008.
- [8] 韦忠生.中国文化遗产的多视角审视[M].厦门:厦门大学出版社,2010.
- [9] 李凯军.世界遗产在中国[M].长沙:湖南科学技术出版社,2008.
- [10] 刘易斯·芒福德.城市发展史:起源,演变和前景[M].北京:中国建筑工业出版社,1989.
- [11] 陈淳.文化遗产研究集刊[M].上海:复旦大学出版社,2009.
- [12] 马奔腾.文化遗产的保护与利用[M].北京:中国社会科学出版社,2014.
- [13] 祁海宁,龚巨平.南京大报恩寺史话[M].南京:南京出版社,2008.
- [14] 邱树森.中国历史大讲堂[M].北京:中国国际广播出版社,2007.
- [15] Acerbi A, Lampos V, Garnett P, et al. The Expression of Emotions in 20th Century Books [J]. PLoS one, 2013, 8 (3): e59030.
- [16] Bentley R A, Acerbi A, Ormerod P, et al. Books Average Previous Decade of Economic Misery[J].2014.
- [17] Dower M. The Tourist and the Historic Heritage [J]. Dublin: European Travel Commission, 1978.
- [18] Lin Y, Michel J B, Aiden E L, et al. Syntactic Annotations for the Google Books Ngram Corpus [A]. Proceedings of the ACL 2012 System Demonstrations[C]. Association for Computational Linguistics, 2012.
- [19] Michel J B, Shen Y K, Aiden A P, et al. Quantitative Analysis of Culture Using Millions of Digitized Books[J]. Science, 2011, 331(6014).
- [20] Twenge J M, Campbell W K, Gentile B. Increases in Individualistic Words and Phrases in American Books, 1960–2008[J]. PLoS One, 2012, 7(7).