

●曹宁 仲岩

论中国个人名称标目的区分问题

摘要 中国人的姓氏具有姓氏数量相对于总人口的比例严重短缺、各汉字姓氏在人口中的实际分布严重失衡、取名用字高度集中等特点。在中文编目规范控制过程中对中国个人名称标目进行区分尤为重要。中国个人名称标目的区分问题包括对古代个人名称和现代个人名称标目形式的问题;标目附加成分选取标准的控制原则、附加成分类型的划分和属性的分析等。将文献计量学的概念引入规范控制研究领域,会进一步丰富编目规则修订工作的信息来源,有助于拓展规范控制的研究领域。参考文献 12。

关键词 中国个人名称标目 标目附加成分 文献计量学 同名标目群 编目规范控制

分类号 G254

ABSTRACT The authors think that Chinese characters for family names are comparatively few, family names are unevenly distributed, and characters used for given names account for a small percentage. Therefore, it is very important to differentiate different names in Chinese authority control. In this paper, the authors make a bibliometric analysis of Chinese personal names, discuss forms of modern Chinese names and ancient Chinese names, selection criteria of additions to headings, types and attributes of additions, etc., and propose some recommendations for future revision of cataloging rules. 12 refs.

KEY WORDS Chinese personal name heading. Additions to headings. Bibliometrics. Heading group with same name. Authority control.

CLASS NUMBER G254

对个人名称标目进行有效区分,是编目规范控制的一项重要工作。中国人的姓名具有形式简短、重名率高的特点。这个特点直接影响到编目工作对个人知识责任的准确识别,进而也会影响到用户从责任者角度对文献的有效查找、选择和获取。因而,在中文编目规范控制过程中对中国个人名称标目进行区分的意义就更加重要。本文从分析中国人姓名特点入手,集中讨论对中国个人名称标目进行区分的若干问题。

1 对中国人姓名特点的分析

根据专家考证,我国目前使用的姓氏绝大多数产生于商周时期,姓氏稳定,较少变化。至于取名,从古至今绝大多数中国人均为单名或双名,多字名罕见。姓氏稳定和取名简短是汉族姓名的主要特点。这种姓氏制度,在中国经过几千年的传承,带来如下问题:

(1)据调查,我国目前使用的汉字姓氏约为3000~3500个。按这一数字推算全国平均每个姓大约有40万人;对照日本的情况,现有姓氏在10万以上,平均每个姓约有1200人(日本的权威文献《日本姓氏大辞典》中更收集了29万个姓氏);而美国据1974年的调查结果,全国共有姓氏128万以上,平均每个姓只有200人。可见,我国的汉字姓氏无论绝对数量还是相对于总人口的比例都存在问题。

(2)据1982年全国人口统计的分层随机抽样,李、王、张、刘、陈五大姓氏占全国汉族人口的1/3;常见的120个姓氏集中了全国人口的90%。每个常见姓氏都会拥有几百万、几千

万甚至上亿的人口。汉字姓氏在人口中的分布严重失衡。

(3)1988年国家语委对14个省市的人口普查资料进行了抽样调查,共选取了57万条姓名。结果表明:人名用字仅为4141个,而其中的1505个字就覆盖了99%的姓名。另据统计,当代中国人中90%集中在409个汉字中取名。结论是,中国人取名趋同从众的习惯相当普遍,用字高度集中。

在中文名称规范数据库中,姓名相重的情况非常普遍,我们称之为“同名标目群”。同名标目群的广泛存在是中文名称规范数据的重要特点,也是我们进行个人名称标目区分的主要对象。

2 对中国个人名称标目区分问题的若干思考

在编目规范控制工作中,对同名标目群进行区分的主要方式是选择和确定适合的标目附加成分。换言之,对个人名称标目的区分问题也就是对标目附加成分的选择问题。2005年10~11月,我们对国家图书馆中文名称规范数据库中相关的标目信息,特别是标目附加成分的构成进行了统计分析。

2.1 对中国古代个人名称标目的统计和分析结果

中国古代个人名称,按照《中国文献编目规则》(第二版)(以下简称新版《规则》)的描述,指的是准确卒年或推断卒年在1912年以前的个人名称。截至2005年9月底,国家图书馆规范数据库中,共有古代个人名称标目32240个,其中存在生卒年附加信息的有10806个,约占标目总数的1/3。

在这些标目中,有3190条存在不确定因素(如标记有“?”、“或”、“约”、“以后”、“左右”、“期间在世”等)。这样严格估算下来,在古代个人名称标目中,生卒年附加比较准确的只有7616条,占标目总量的23.6%。

如果更进一步考证,上述标目中也还有不少问题。比如唐代诗人王维,在国家图书馆规范库中的标目形式为:(唐)王维(701-761)。事实上史学界对此形式存在争议。有历史学家指出:按史料记载可以准确推断其胞弟王缙的生年是700年,这样哥哥的生年就出现了问题。再如,唐代名相狄仁杰,目前的标目形式为:(唐)狄仁杰(630-700),其根据来自《新唐书·狄仁杰传》。但另有多种参考信息源,如“国学网”记载其生年为607年,此外还有学者认为其卒年为704年。

出现上述争议的主要原因在于古人的生卒年信息一般都是根据相关史料考证和推断出来的,因为引证史料不同,推断过程又掺杂大量主观判断,所以会出现不同的结论。事实也的确是:中国古代个人名称标目的生卒年附加,相比较而言可靠性较差。如何弥补这个缺陷?一种常见的办法是在生卒年附加信息中完整地记录多种推断结果,以体现信息选取的广泛性和公平性,如:(汉)司马迁(前145/前135-前87?)或(清)曹雪芹(1715/1724-1763/1764)。但这样做带来的另一个问题是标目冗长,使用不便,不是理想的标目形式。

带着上述问题,我们对中国古代个人名称标目中的同名标目群做了进一步分析。统计结果显示在32240个古代个人名称标目中,实际上只有1948人同名(同名率为6%),他们共同使用896个名字(平均同名频率2.17)。也就是说中国古人的同名比例远比我们想象中的要低。经过进一步调查了解,我们发现同一朝代的相同个人名称标目数量为787人,仅占标目总量的2.5%(他们共同使用373个名字,同名频率2.11)。

据此,笔者认为,在中国古代个人名称标目中,真正需要考虑选择附加成分的实际上只有6%;而采用朝代附加(是古人与今人相区别的传统办法)之后,又可以解决同名标目中约60%的区分问题。总之,对中国古代个人名称标目附加成分的选择问题,我们的意见是,古代个人名称比较理想的标目形式是:(朝代)+个人名称。如:(汉)司马迁和(清)曹雪芹就足以准确识别责任人。当同一朝代姓名相同时(统计显示比例为2.5%),我们可以再考虑采用生卒年、生平所处时代以及字号、籍贯等信息做第二种附加成分,其标目形式为:(朝代)+个人名称+(其他附加成分)。

支持上述设想的还有以下两个数据分析结果:中国古代个人名称具有比较可靠的朝代信息的约占99.98%;中国古代个人名称具有比较可靠的字号信息的约占82%。如果中国古代个人名称标目采取上述附加成分,可以基本保证信息的覆盖率和可靠性。

按照上述设想,对中国古代个人名称同名标目群区分方式如下:

现行区分方式	设想的区分方式
(唐)李华(715-774)	(唐)李华
(宋)李华	(宋)李华

(清)李华(女) (清)李华(字敬如)

(清)李华(女,甘泉人) (清)李华(字澹卿)

按照现行区分方式,在一个只有4条标目的同名标目群内就选择了朝代、生卒年、性别、籍贯等4种不同的附加成分;而按照设想的区分方式,则只需要采用两种附加成分就能够准确区分所有标目,这将大大简化对同名标目附加成分的选择问题。

2.2 对中国现代个人名称标目的统计和分析结果

与古代个人名称标目相比,现代个人名称标目的情况更加复杂,主要是现代个人名称标目的同名比例较古代个人名称有大幅度提升。在国家图书馆名称规范库中,截至2005年9月底共有现代个人名称标目421841条,其中同名标目高达102317个,占总数的24.3%(他们共同使用34996个名字,平均同名频率2.92)。在上述同名标目中,2人同名的有23605组,3人同名的有5815组,4人同名的有2234组,5人同名的有1136组,5人以上同名的有2305组(共涉及23252个标目,同名频率10.08)。

在对5人以上的同名标目群进行系统整理和分析的过程中,我们发现,当同名标目超过5个以上时,仅采取生卒年一维标准进行标目区分的现实可能性微乎其微(在2182组同名标目群中,只有1组仅采用了生卒年附加一种成分)。原因主要在于现代个人名称标目生卒年信息的覆盖率只有37%(在421841条标目中,有生卒年附加信息的为156124个)。反映在编目实践中,编目员通过受编文献或其他参考信息源获取个人生卒年信息的情况还不够理想。这样,当编目员需要选择生卒年以外的信息作为标目附加成分时,就出现了多维附加标准的控制问题。

事实上,在国家图书馆名称规范库中,为实现对个人名称中同名标目的区分,往往会过多地使用多种类型的附加信息,而且编目员在确定附加成分描述形式时又有较大的随意性。上文所引古代个人名称标目“李华”的情况就已经暴露出这一倾向。在目前编目规范工作中反映出的诸多问题,其本质是针对多维附加选择的无序化现象而言的。对标目附加成分的选择缺乏必要控制和有效指导,是当前新版《规则》标目法和名称规范控制工作中最薄弱的环节。

附加成分选取的多维标准问题,在实际操作中存在两种情况。一种是在一维标准不能区分的情况下,再通过二维标准进一步限定。也就是说两种标准在逻辑上具有递进关系,在标目限定上是渐进性的。这种区分方式的特点是第一种附加是全局性的,第二种附加是局部性的。在现代个人名称标目中比较典型的情况如新版《规则》22.2.3.3.d中的例1:

杨欣(摄影家1963-)

杨欣(饭店管理1963-)

上述附加成分的选择,从规范控制的逻辑角度讲是比较严密和基本适当的。而我们以下要重点讨论的是规范控制中的另一种情况,即编目员在规范控制过程中对同名标目群中的不同部分分别采用了不同的附加标准。在多维标准之间没有递进性的层次关系而构成了简单的平行关系,其特点是在同名标目群中不存在明显的全局性标准,而同时存在多种局部标准。这种标目区分方式极易出现信息交叉和概念

混淆问题,而且使用的附加成分类型越多,对同名标目群区分的可控性就越差,从而形成“规范控制形式大于实质”的问题。如:

王强(电子商务)
王强(金融法)
王强(排版软件)
王强(1945-)
王强(女,1933.8-)
王强(记者)
王强(主任医师,1941.8-)
(回)王强(1936-)
.....

上述同名标目群很可能是不同编目员在不同时期建立起来的,但最后的结果是为区分而区分,形式大于实质。从每一个标目的形式看都具有唯一性,但从同名标目群整体关系的角度看却相当混乱,对用户的检索也会造成一定干扰。

解决上述问题关键在于要确立关于标目附加成分选择的基本原则和一般标准问题。对于附加成分选择的基本原则,应首先明确附加成分的性质与类型。附加成分包含全局附加和局部附加两种类型。全局附加对标目而言是有则必备的信息,只有在全局附加无法区分的情况下,才在需要区分的局部做进一步限定,这应是附加成分选择的基本原则。按照这一原则,对同名标目群选择全局附加成分应该非常慎重,而且一旦确定就要坚持到底。关于标目附加成分选择的一般标准问题,应古今区别对待。对古代个人名称标目,朝代信息可作为唯一的全局附加成分,而其他各类信息都只能是为解决特殊问题而使用的局部附加。对现代个人名称标目,应考虑两种情况:在同名标目群中所有标目的生卒年信息均完整时,应选择生卒年信息做唯一的全局附加(如果出现生卒年相同,可考虑在相同的局部选择第二种附加信息);而在同名标目群中生卒年信息不完整时(如多条标目的生卒年信息不详),则需要将生卒年信息和学科专长—职业信息共同作为全局附加,若还有标目相同的情况再考虑以其他信息做局部附加。按照上述设想,新版《规则》标目法中所涉及的国别、民族、性别等有则必备的附加成分将不再必备。新的模式下,只有当全局附加成分相同时,它们才应该出现在规范标目当中。这样,标目附加成分选择过程中的随意性将可能大大降低,在同名标目群中出现不适当的附加成分的机率也会随之下降。

按照上述设想,现代个人名称标目的多维附加将通过下述形式实现:

李明(会计师,1960.8-)
李明(经济师,1922.5-)
李明(作家,1962-)
李明(绘画,1966-)
李明(绘画,1973-)
李明(英语教师)
李明(计算机教师)
李明(乐队指挥)
李明(心理学)

李明(矿业工程)

.....

笔者认为,对现代个人名称标目而言,学科专长—职业信息在同名标目的区分过程中作用会越来越大。对学科专长—职业信息的规范化描述,将成为规范控制领域内一个值得关注的问题。当前的规范控制工作,不仅亟需有一个科学、严密和系统的《学科专长—职业参照表》,而且还需要进一步深入探讨学科专长—职业信息本身的层次选择问题。在学科专长—职业信息中,比较理想的标目附加形式是具有一定专指度的职业(如电视节目主持人、计算机系统分析员等);其次是学科+职业—职称的形式(如西班牙语翻译、社会学教授等);再其次才是单纯的学科专长(如对外贸易、电子商务等)。这一判断基于以下理由:从概念限定的逻辑角度看,对个人名称进行附加应该优先考虑个人特征,而选择学科专长所包含的文献特征往往要大于个人特征。编目员在实际操作过程中大多是根据文献特征推断著者的学科特征,这种做法存在“以文代人”的片面性因素。如果说这种推断对“work”的责任方式还具有一定意义的话,那么对“expression”的责任方式(比如译者、编者)可靠性就难以保证了。

2.3 关于标目附加成分属性特征的探讨

一般认为,个人名称标目的附加成分可以划分为个人的自然属性和社会属性两种类型。个人的性别、民族、籍贯和生卒年是自然属性,学科、专长、职业等是社会属性。自然属性相对精确和固定,而社会属性则往往含混和易变。新版《规则》标目法 22.2.3.3.f 因此规定:“对个人名称标目附加成分的选择,个人自然属性信息应优先于社会属性信息。”

笔者认为,这种判断有其合理的一面,但也存在认识上的不足。对编目员而言,倾向于自然属性优先的原则是因为自然属性相对精确和稳定,便于标目的准确区分,同时也便于规范数据库的维护。但从 FRBR 所倡导的用户任务角度看,却未必是个人名称最理想的区分方式。比如,我们在谈到“李白”时,绝大多数人的反应过程是:唐代著名诗人,“诗仙”,然后才会是字太白,号青莲居士,701—762 年在世等等。也就是说,我们在识别个人名称时,在习惯上往往会优先选择个人名称的某种社会属性(因其通用和简洁)。因此在编目员所追求的标目附加成分的准确性、稳定性和用户所希望的标目通用性、便捷性之间我们需要权衡。理想的标目附加成分,应该兼顾自然属性和社会属性两个方面。或许古代个人名称标目,朝代附加之所以是一种理想的全局附加,正是因为朝代信息中既包含自然属性的特点又包含有社会属性的要素。新版《规则》22.2.3.2.a 中的第一款规定,“中国古代人物所属朝代的确定,一般以卒年为准……”,这体现了朝代的自然属性特点,而接着又规定,“若不适当……可选择生平主要活动所在的朝代”,就体现出所属朝代这一信息具有某种社会属性因素,比如:(宋)文天祥和(明)郑成功,都是以其社会归属而非准确卒年作为所属朝代划分依据,就是基于这种考虑。

总之,对现代个人名称标目而言,将生卒年信息和学科专长—职业信息同时列为全局附加的优点在于,从某种

程度上讲能够起到兼顾责任者自然属性和社会属性的作用,并且满足编目员的专业编目需要和广大用户的一般检索需要。

2.4 对名称规范数据库维护工作的思考

中文个人名称标目附加成分的选择问题,之所以成为当前规范控制领域内的难点,除了相关原则和标准欠缺,还有一个重要的原因:编目员选择过程中所面对的不仅仅是与在编文献相关的某一个标目,而往往是与这一标目相关的一组标目。在规范控制过程中,对标目附加成分的选择所涉及的对象可能是整个同名标目群。在极端的情况下,编目员有可能为建立一条新的规范数据,而去修改以往建立的所有相关标目形式的规范数据。这种现实的需要会带来难以估量的规范数据库的维护工作和为数更多的与规范记录相关的书目标目的更新工作。比如:

王正(1936-)

王正(1937-)

王正(1930.3-)

王正(1931.4-)

王正(1964-)

王正(1933.1-)

这是国家图书馆规范库中一组罕见的通过生卒年一维附加实现了准确区分的同名标目群。现在,假设我们在新文献中发现了第七个王正(他明显不属于已有标目中的任何一个),而且我们通过任何渠道都无法获得这一新名称的生卒年信息,这时我们就将面临非常复杂的数据更新问题。因为,从逻辑上讲为准确区分全部7个标目,我们需要为6个标目统一增加学科专长—职业信息的附加成分以适应于第七个标目。这项工作如果涉及到需要对以往受编文献重新提阅,规范数据维护的工作量就会进一步加大。

上述规范数据的维护问题,我们应该有未雨绸缪的解决方案。比如,我们对中文名称规范控制中的常见姓名要进行科学预测,建立常见姓名参照表。该参照表应提供给规范控制应用系统,当编目员所建立的个人名称标目出现表内常用字时,系统应给予提示,指导编目员对标目预先进行多维附加限定。当然,更重要的是,我们是要进行系统的数据跟踪、分析和总结同名标目群全局附加从一维标准扩展到二维标准的规律,从而做到对名称规范数据维护的工作总量和规模频率先期把握,合理配置工作岗位和编目人员。

3 文献计量学在规范控制领域的应用

编目规则的基本原则和具体规定,是用来指导编目实践的,这些原则与规定的产生,往往是对前期编目实践的归纳、总结和概括、提炼。这也是编目规则需要根据编目实践的反馈信息不断修订和完善的根本原因。

在此,我们希望探讨的问题是:在编目实践中所遇到的特殊性问题如何能通过科学合理的筛选途径上升到编目规则修订的视野中来。对已有数据的统计分析是一种将编目实践中的特殊性问题归纳、汇总形成普遍性问题的有效途径,它不仅有利于我们发现问题的,而且有利于我们提出解决

问题的方案。由此,我们提出将文献计量学的概念引入规范控制研究领域的设想,具体说,就是运用数学和统计学的方法来研究名称规范问题,用以指导编目规则标目法以及规范数据著录规则的修订和执行,这将进一步丰富编目规则修订工作的信息来源。同时,文献计量学的研究方法,还有助于我们进一步拓展规范控制的研究领域,比如我们可以通过对名称规范数据与书目数据中被引用标目之间的定量分析,讨论如何建立名称规范数据库核心著者群的概念,进而指导文献采编工作;再如,我们还可以通过对 OPAC 系统用户检索历程的聚类分析(目前许多网络搜索引擎都有类似功能),来讨论规范数据中单纯参照的选取原则问题等等。

总之,将文献计量学的概念引入规范控制领域的意义,在于我们能够为编目理论研究和规范控制的实践决策提供一种量化分析的基础。这将有助于规范控制工作在更加科学、合理的模式下健康发展。

参考文献

- 1 罗伯特·H·伯格著;熊光莹译.图书编目规范工作.北京:商务印书馆,1993
- 2 国家图书馆编.中国文献编目规则.北京:北京图书馆出版社,2005
- 3 IFLA Study Group on Functional requirements for bibliographic records. Functional requirements for bibliographic records: final report. [2006-06-06]. <http://www.ifla.org/VII/S13/frbr/frbr.pdf>
- 4 谭君久.同姓名现象的症结与改革我国姓名制度的思路. [2006-06-06]. <http://sub.whu.edu.cn/jjtan/name.htm>
- 5 李培林.重新崛起的日本. [2006-06-06]. <http://book.sina.com.cn/longbook/1102661624-japanrebuildingup/33.shtml>
- 6 George R. Stewart. From Personal Names. [2006-06-06]. <http://www.ws.zscas.edu.cn/fd/book4/unit13/13reference.htm>
- 7 美国人什么都敢姓. [2006-06-06]. <http://www.blogchinese.com/user/20032/archives/2005/149244.shtml>
- 8 马斗全.三晋历史人物:王维. [2006-06-06]. http://www.sxu.edu.cn/yjjg/mypower/Article_Show.asp?ArticleID=306
- 9 国学网. [2006-06-06]. <http://www.guoxue.com/xiao/0203.htm>
- 10 王京阳,袁宪.唐狄仁杰卒年考辨.故宫博物院院刊,1995(1)
- 11 台湾汉学研究中心.人名权威档个人专长学科领域表. [2006-06-06]. <http://catweb.ncl.edu.tw/sect-2.htm>
- 12 邱均平.文献计量学.北京:科学技术文献出版社,1988

曹宁 国家图书馆文化教育培训部主任,副研究馆员。通信地址:北京中关村南大街33号国家图书馆。邮编100081。

仲岩 国家图书馆采编部工作。通信地址同上。

(来稿时间:2006-06-08)