

主题词标引问题探讨

——兼与朱芊先生商榷

文 榕 生

(中国科学院文献情报中心, 北京, 100080)

摘 要 文章论述了检索语言概念, 对其所属类型进行归纳、划分, 并研究了它们之间的相互关系, 尤其对主题词的主要类型——叙词和关键词的定义与 CNMARC 标引等相关问题进行探讨, 澄清了长期以来存在的某些检索语言的术语、用法混乱现象。

关键词 检索语言 文献标引 叙词 关键词
〔文献标识码〕 A 〔中图分类号〕 G254

1 引言

早在 20 世纪初期, 我国图书馆界就开始研究主题词, 限于当时人力、物力、财力与技术设备等, 进展缓慢, 长期处于理论性研究或仅在小范围试验, 直到世纪末, 计算机日益普及, 主题词投入实际应用才逐渐推开。随着研究主题词升温, 人们才较多审视主题词的方方面面, 发现存在的问题或出现不同看法也就显而易见。

朱芊先生“国家图书馆中文图书主题标引格式演变过程的思考”^{〔1〕}(下文简称“朱文”)是一篇应用性代表作品, 是目前相当部分图书馆员认识状况反映。笔者拜读后, 认为“朱文”值得学习处不少, 所总结的五种叙词标引格式显示出他们锲而不舍探索态度, 但限于篇幅, 谨就难以苟同的一些问题进行商榷, 希望得到更多专家学者的指教。

2 术语的规范

名词、术语的规范对于学科建设以及学习、探讨、研究问题等并非无足轻重。长期以来, 种种原因造成检索语言方面不少名词、术语杂乱, 其中尤以“自然语言”与“检索语言”、“主题”、“主题词”、“叙词”、“关键词”、“自由词”等为最。实际上这些名词、术语尽管有部分内涵相同或相似, 但都有各自的内涵外延, 形成特定的概念。当然, 由于划分标准、全称与简称、别称、层次等原因, 也出现同物异名等现象(详见《图书馆》将刊拙文“检索语言的类型与关系”), 检索语言的种类多, 它们之间的关系相互交织(如文示图)。如何认识区分这些名词、术语?

• “自然语言”与“自由词”。二者的共性在都不经过

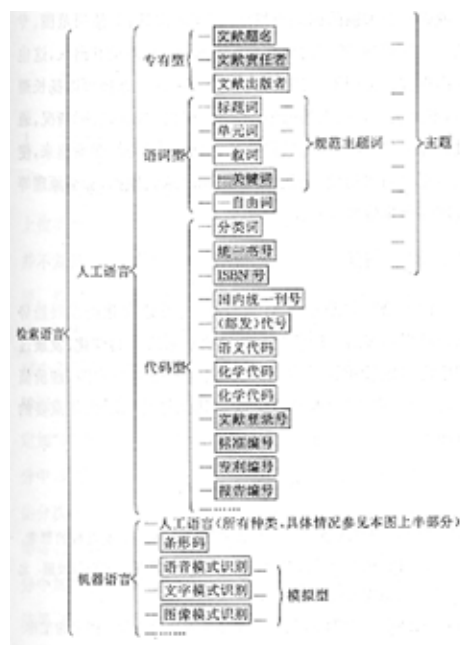
规范, 但区别也很明显。语言的构成单位有字、词、词组、短语、句子等。所谓自然语言是人类在社会活动中用以表达意思、交流思想的工具, 是在自然交往中形成的, 它有别于检索语言是专门用于查找寻找有序化事物, 它包含自由词等所有语言单位, 是属概念。而自由词是不受控制的语词形式, 是种概念。自由词一部分具有检索功能, 一部分是与规范主题词同形, 但这些都不是自由词的主要功能。“朱文”中所谓“自由词”与笔者的尽管名称相同, 实质则有所区别。

• 检索语言。它是各种具有检索作用的语言总称,

是人们利用事物某种或一些特点来查找寻找有序化事物的工具。按不同特点作为标准划分, 检索语言有多种称谓(文示图中还难以囊括)。根据具体检索对象类型不同可形成特定的检索语言, 如: 信息检索语言、情报检索语言、文献检索语言、网络检索语言(搜索引擎)、图书检索语言、档案检索语言……“朱文”是论述“中文图书”, 所提: “情报检索语言发展的过程, 是一个不断自然语言化的过程。……许多书目数据库由原来主要采用受控语言已逐步过渡到自然语言与受控语言结合使用, 其方式是增加了自由词标引。”带下划线的词基本上都是术语, 但它们配合起来形成的语境颇有关公战秦琼意味。“受控语言”应当是包括“前控”与“后控”方式的检索语言, 而“朱文”仅承认前控—叙词; 既然其所谓“自由词”是选取文献中使用, 而词表未收, 可随需要增补。不作为正式标引检索用词, 但可实际用于标引和检索, 具有较大的识别功能的词, 多数是名词或名词性词组, CNMARC 称其为“非控主

题词”)与笔者的“自由词”都是非控词,那就无所谓“受控”问题,其实应当用“受控主题词”才确切,只是笔者与“朱文”所谓的同形术语内涵不同。

• “主题”与“主题词”。主题是事物的核心内容、中心思想的集中反映,相同的主题可以用不同的形式表达。文献的主题可以采用文献题名、分类号、统一书号等多种形式表达的(见文示图)。而主题词则采用简洁的语词形式来表达事物的主题,具有这些特点的语词就有标题词、单元词、叙词、关键词、自由词等(见拙文“‘主题词’刍议”,该文笔者按《著作权法》规定已经先后在《河北科技图苑》2000年第2期与北京图书馆出版社2000年出版的《面向21世纪的情报语言学:全国第三次情报检索语言发展方向研讨会论文集》刊出)。“朱文”论述的所谓“主题词”(实际上应当是叙词和关键词)问题,而不是其题名所谓“主题标引”(因为分类号与主题词等都主要是根据文献主题进行标引)。



图示 检索语言种类与相互关系

注:图中检索语言所标“□”为前控型示意、标“◻”为后控型示意,但“统一书号”和“ISBN号”兼有部分前、后控形式

• “叙词”与“关键词”。二者是目前常用的主题词类型,但叙词长期以来被误作主题词的同义词。尽管“朱文”意识到国家标准已将“文献主题标引规则”(其中就存在“主题”、“主题词”、“叙词”等混乱)更改为“文献叙词标引规则”,但是依然“叙词”、“主题词”混用,这种误导不容忽视。叙词的由来刘湘生早在1978年就专门论述,而不少专家对此语焉

不详,一般图书馆员缺乏独立思考,只能以讹传讹。叙词是意译外来词,又称叙述词、描述词,是一些以概念为基础的、经过规范化的、具有组配性能、显示词间语义关系和动态性的词和词组,是描述文献资料主题的一种标识符号。^[2]笔者将其划归语词型检索语言。关键词是从文献中选取能够集中反映其主题的、用于标引检索的规范性语词。^[3]关键词区别于自由词;后者选取随意性大,又不受控制;而前者通常是取自题名、摘要、目录等的名词或词组,是受控语词。关键词区别于叙词;后者采用前控的词表(含电子版)形式,前者则采用后控形式。上文提到的“受控”,在这种情景下才有实际意义。从“朱文”的“非控主题词是从文献题名、章节、或文献内容中提取对揭示文献起关键作用的名词术语,不加规范,直接标引。因此非控主题词应尽可能选择词义清晰、概念明确的名词或名词性词组”看,尽管称“不加规范”,却有所规定、起关键作用……应当是“关键词”,而不是“自由词”或“非控主题词”。

3 叙词的选择与标引

无论是采用分类号还是主题词标引,都主要是根据文献的内容主题选择恰当的类型或主题词,但二者又有所区别并各有优势与局限。由于分类法的编制着重于强调学科系统性、类目的稳定性以及枝形排列方式等,使得采用分类号标引的文献注重于揭示学科属性;但又因强调属种关系而分散了一些相关的文献。主题词标引是随机性的,侧重于文献反映、论述的重点,可以不受限制地集中相关问题;但又往往因按概念集中而将一些无关的文献包容进来。^[4]二者有一定互补性。采用叙词标引的优势之一还在于通过组配方式(即增加内涵缩小外延的原理),加强叙词的专指性。笔者基本上赞同“朱文”中的“主标题确定原则”,但对具体操作有不同看法。

• 过泛叙词的使用。主要指“研究”、“概况”等类既不可能作为主体因素又区分性较弱的名词,有必要将其作为叙词,进行标引吗?例如:“朱文”标引的《新丝绸之路:陇海、兰新、北疆线漫游》

6060# \$a 铁路线路 \$x 概况 \$y 中国

6101# \$a 陇海线 \$a 兰新线 \$a 北疆线

笔者认为有无“概况”都无所谓,起不到区分作用;如果有更具体的内容,将此类叙词辅助标引,自然可以区分。顺便提到,笔者认为“世界”这类顶级泛指性名词也无必要作为叙词,因为如有组成“世界”的具体空间位置,标引后者就可以起到区分作用。

• 叙词与其他检索点的关系。某些名词如“根”、“英语”等名词,孤立地看,与标题词、单元词、叙词、关键词、题名

等形式相同(如:美国现代小说《根》与植物的“根”,教科书《英语》与名词“英语”),需要综合判断才能确认其实质,CN-MARC 的字段就可以作为判断的参考(题名在 200 字段,叙词在 606 字段,关键词在 610 字段,作为主题的个人名称在 600 字段,等等)。由于 CNMARC 既适宜模糊检索,又适宜指令检索,那么,当同一书目记录中已设置过检索点,还有必要再采用叙词做同样形式的检索点吗?例如:笔者标引《周恩来领导艺术》

600 #0 \$a 周恩来 \$ Azhou Enlai \$ f1898—1976 \$ x 领导艺术

6060 # \$a 领导艺术

就比较简洁。对照以下“朱文”标引,则是采用叙词轮排,似乎更接近手工操作:

600 #0 \$a 周恩来—领导艺术 \$ x 研究

606 # # \$a 领导艺术—周恩来 \$ x 研究

·选词、标引的一致性。叙词的选词应当取专指性的,标引应尽可能一致,尤其是在同一书目记录中。请看“朱文”的《猪鸡鸭鹅快速饲养法与疾病防治》标引:

6060 # \$a 养猪学

6060 # \$a 猪病 \$ x 防治

6060 # \$a 鸡 \$ x 饲养管理

6060 # \$a 鸡病 \$ x 防治

6060 # \$a 鸭 \$ x 饲养管理

6060 # \$a 鸭病 \$ x 防治

6060 # \$a 鹅 \$ x 饲养管理

6060 # \$a 鹅 \$ x 禽病 \$ x 防治

猪与鸡、鸭、鹅分别是家畜与家禽,这些叙词的选择标准与标引方式应当是相似的,遗憾的是“朱文”中除了第一个字段(一个叙词)和第八个字段(三个叙词)与其他字段(两个叙词)不同外,还有标“□”和标“-”的两组叙词不一致。

·复合主题概念的处理。既然复合主题概念是由多个主、次概念合成,理应充分揭示;而叙词标引只有作为主标题(主体因素)的叙词才能作为检索入口,如果不将反映相关主题的叙词按附加原理以重复字段的方式增加检索入口,对查全率是有损失的,笔者认为是不妥当的。

4 关键词的选择

对于关键词的选择,实际上也是一种规范。它的来源是文献的题名、摘要、目录等比较“点睛”位置,而不是漫无边际地寻觅;它的作用与功能是能够集中反映其主题的(“朱文”所谓“对揭示文献起关键作用”)、用于标引检索,而不用离主题较远的冗词赘句;它的词性是简洁的名词,通常不取其他实词,更不采用虚词。尤其是专业人员使用的关键词多为术

语,词频较高(这些也是选择叙词的一些重要条件),相当于待增补的准叙词。而不少人认为关键词没有规范,或如“朱文”认为“不加规范”,未免偏颇。

语言并非固定不变僵死的事物,而是随着时代的发展,人们认识的变化,语言环境的改变……而不断变化、丰富。例如,“电子计算机”是叙词,这是由于它早期只具备计算功能,为了区别“手摇式”而采用“电子”前缀。如今,手摇计算机早已消声匿迹,人们为了区别,已将仅有计算功能的称之为“计算机”,电子计算机简称为“计算机”或“电脑”。实际上,现代的电子计算机不断智能化,纯粹的计算只是它很小的一部分功能,而称之为“电脑”显然就更贴切、简洁。在“计算机”与“电脑”并行的现代,将它们都作为关键词进行后控联系,恐怕要比固守采用叙词“电子计算机”一种模式更加贴近用户,也使操作简便得多。

编制叙词表通常被认为是叙词的规范结果。且不论谁家或哪一级编制的词表更权威,由于理解认识、使用范围、专业特点等差异,不同的词表中的叙词认定、词量有出入,这也是出现叙词标引不一致的原因之一。关键词选择可以扬长避短,不是预先采取人为的限定,而是根据文献的实际情况,通过计算机广泛收集,然后将重复的删除,相关的联系起来,使之有更强的实用性。当然,对于叙词的选词原则、标引原理等合理部分也应当吸收。

5 结 语

检索语言向兼容化、计算机化、自然语言化的发展趋势已被越来越多的人所接受。在目前检索语言的变化、发展过程中应当继续重视对叙词、关键词等的研究与实践,理清概念,探究它们的原理、相互关系及其优势与局限也是重要的方面。

参 考 文 献

- [1] 朱芊. 国家图书馆中文图书主题标引格式演变过程的思考. 见:中国图书馆学会编,世纪之交:图书馆事业回顾与展望. 北京:北京图书馆出版社,1999
- [2] 刘相生. 对叙词及有关几个名词术语的认识. 图书馆工作, 1978 (3)
- [3] 文格生. 检索语言的返朴归真:从叙词到关键词. 图书馆论坛, 2000 (1): 27~30, 64
- [4] 文格生. 揭示文献主题的检索语言. 图书馆界, 1999 (3): 5~7

〔作者简介〕:文格生,男,1951年生,副研究馆员,发表论文160余篇,出版专著2部,获西南西北优秀科技图书一等奖、中科院自然科学二等奖等多项多次奖。

(收稿日期:2001—06—22)

主题词标引问题探讨——兼与朱芊先生商榷

作者: [文榕生](#)
作者单位: [中国科学院文献情报中心, 北京, 100080](#)
刊名: [山东图书馆季刊](#)
英文刊名: [SHAN DONG LIBRARY QUARTERLY](#)
年, 卷(期): 2001 (3)
被引用次数: 7次

参考文献(4条)

1. 朱芊 [国家图书馆中文图书主题标引格式演变过程的思考](#) 1999
2. 刘湘生 [对叙词及有关几个名词术语的认识](#) 1978 (03)
3. 文榕生 [检索语言的返朴归真:从叙词到关键词](#)[期刊论文]-[图书馆论丛](#) 2000 (01)
4. 文榕生 [揭示文献主题的检索语言](#)[期刊论文]-[图书馆界](#) 1999 (03)

本文读者也读过(10条)

1. 吕凤英 [关于网络信息分类的一点思考](#)[期刊论文]-[天中学刊](#)2002, 17 (5)
2. 周潇洒, 南存兰 [国内生物医学期刊主题词标引现状](#)[期刊论文]-[医学情报工作](#)1999, 20 (6)
3. 李玉安 [试论关键词语言在文献检索语言中的地位和发展趋势](#)[期刊论文]-[情报杂志](#)1999, 18 (6)
4. 李爱明, 朱玲玲, Li Aiming, Zhu Lingling [基于关键词的网络信息检索与知识检索比较](#)[期刊论文]-[情报探索](#) 2006 (5)
5. 苗丽娟, MIAO Li-juan [学术论文关键词的标引](#)[期刊论文]-[唐山师范学院学报](#)2007, 29 (4)
6. 劳云飞 [也谈MeSH词表和主题词标引](#)[期刊论文]-[现代情报](#)1998 (3)
7. 陈开慧, CHEN Kai-hui [分面叙词表的关键词法链接模型](#)[期刊论文]-[情报科学](#)2005, 23 (3)
8. 文榕生, Wen Yongsheng [揭示文献主题的检索语言](#)[期刊论文]-[图书馆界](#)1999 (3)
9. 李新, 汪德彪 [情报检索语言研究与《中国分类主题词表》](#)[期刊论文]-[大理学院学报](#)2004, 3 (z1)
10. 郑小莉 [关键词采用“类主题词标引法”之管见](#)[期刊论文]-[中国科技期刊研究](#)2004, 15 (6)

引证文献(7条)

1. 文榕生 [论网络与数字化环境的检索语言发展趋势](#)[期刊论文]-[四川图书馆学报](#) 2002 (6)
2. 吴春玉 [中文全文检索系统中实现主题词标引思路](#)[期刊论文]-[情报杂志](#) 2005 (1)
3. 吴春玉 [中文全文检索系统主题词标引](#)[期刊论文]-[情报科学](#) 2004 (6)
4. 文榕生 [网络环境下的书目数据质量控制探讨](#)[期刊论文]-[江苏图书馆学报](#) 2002 (3)
5. 文榕生 [书目数据的完善与整合](#)[期刊论文]-[图书馆杂志](#) 2002 (2)
6. 文榕生, 闻桦 [兼容与互补——论检索语言的发展趋势](#)[期刊论文]-[图书情报工作](#) 2003 (2)
7. 文榕生 [完善书目功能 推进数字图书馆可持续发展](#)[期刊论文]-[图书馆学研究](#) 2002 (3)

本文链接: http://d.g.wanfangdata.com.cn/Periodical_sdtsgjk200103016.aspx