

●袁相琴

三种一体化词表比较研究

ABSTRACT A comparative study is conducted for the facet technology and the change of pre-coordinate terms of the 3 integrated thesauri: the "Faceted Thesaurus", the "Basic Thesaurus" and the "Chinese Classified Subject Thesaurus". The paper explains the limitations in using the integrated thesaurus in process of its development and holds that it is developing now toward the direction of large scale, applicability and usability. 7 tabs 8 refs

KEY WORDS Faceted thesaurus Basic Thesaurus Chinese Classified Subject Thesaurus Integration Comparison

CLASS NUMBER G254.0

60年代以来是第三种情报检索语言^[1],即分类主题一体化语言产生、发展和推广应用的时期。这一时期,国际上相继问世了三部具有代表性的一体化词表:《分面叙词表》、《基础叙词表》和《中国分类主题词表》。

1969年,分类法回归的思想使英国学者艾奇逊创制了世界上第一部分分类法和主题法有机结合的新型词表——《分面叙词表》。它通过对词表的规范化和分面分类技术的改进,实现了分类体系和叙词体系的一体化,成为一种能够满足多功能的检索工具,受到图书情报界的重视和好评,被誉为“首次建立了一个真正实现了分类法和主题法两者合一的系统”;“是对标引理论和实践的一大贡献”。经过10年的实践和经验总结,艾奇逊于1981年又编制了一部颇具特色的一体化词表《基础叙词表》。与《分面叙词表》相比,《基础叙词表》非常熟练地运用了分面分类和词汇控制技术,更重要的是通过宏观结构的简化(比联合国教科文组织叙词表更简化)与微观结构的精化,增强了词表的功能,便利了用

户。1982年第4次国际分类法会议对《基础叙词表》给予了高度评价,认为它是分类法研究的重大成果。

我国图书情报工作者采纳了国际上检索语言研究理论,集我国40多年来分类法、主题法研究成果为一体,于1994年正式出版了《中国分类主题词表》,把具有广泛用户基础的《中国图书馆图书分类法》(简称《中图法》)和《汉语主题词表》(简称《汉表》)有机地结合起来,成为既适用于主题标引,又适用于分类标引的用途多样的我国第一部大型的综合性分类主题一体化的标引检索工具。该词表的问世,把一体化词表的研制从理论研究阶段推向了推广应用的阶段。

笔者认为,这三部词表的编制出版,可以把一体化词表的进展分为3个发展阶段:“拟定方法,建造模型”阶段;总结经验,“完善模型和方法”^[2]阶段,以及推广应用阶段。本文着重从3种一体化词表的先组词和分面结构变化进行比较,说明《中国分类主题词表》推广应用型模式的前景。

1 三种词表的相同性

1.1 宏观结构简化

检索词表的功能决定其结构。一体化词表的整体功能既决定于它的宏观结构,又依赖于它的微观结构。与一般叙词表相比,一体化词表有宏观简化、微观精化的特点。3 种词表宏观结构简明,只有分类表和字顺叙词表两部分。一体化语言的发展,也曾经有过增加基本组成的模式。^[3]《联合国教科文组织叙词表》就编制了词族表等部分。但词族是计算机从字顺叙词表中抽取,而字顺叙词表又是用计算机软件从分类叙词表转换过来的,因此词族的直接来源是分类表。由于一体化词表已经有了详细的分类表部分,可以不用另外编制词族表。为了不重复,《基础叙词表》与《分面叙词表》一样,为二分结构。《中国分类主题词表》是目前世界上最大的一体化词表,宏观结构也只有两部分,比《中图法》、《资料法》、《汉表》的篇幅减少了 54.72%,但整体功能却大于三者之和。这种简明的结构起到了以简驭繁、以少胜多的作用。因此,《中国分类主题词表》无论在特定学科领域,还是在一般学科领域,特别是在综合性的各学科领域都具有很大使用价值。

1.2 类目分面结构化

客观事物具有多重属性和多方面的联系,因此词表类目的设计可以多向成族,分面展示事物的特性,提供多途径检索。无论是完全融合型^[4]的《分类叙词表》、《基础叙词表》,还是对照索引型的《中国分类主题词表》,类目的划分都运用了分面分析技术。《分面叙词表》从二级类目开始按照分面理论予以显示,《基础叙词表》从第 4 级类目开始了这种显示。两种词表的分类显示部分是直接引入一个由“学科—科目—组面—亚面—类目体系”所组成的体系—组配分类法(见表 1)。

《中国分类主题词表》类目体系虽属列举

表 1 用分面分析编制的体系—组配分类表

玻璃工艺	学科或科目
(按原料工艺)	组面
玻璃原料	
(按产品分)	组面
玻璃	
(按成分划分)	亚面
硅玻璃	…… 类目体系
铝玻璃	
磷酸盐玻璃	

式,然而类目属分面结构的数量和范围远胜于前两种一体化词表。因为该词表规模是前两种规模之和的 4 倍有余,系规模巨大所致。根据文献的特点和实际需要,该词表分类部分在基本遵循逻辑划分的原则下,有选择地采用了多重列举法^[5],即分面分类法,对类目上位类进行分面分析,形成几组不同性质的下位类,类似于体系—组配法的编制原理。《中国分类主题词表》分面类目形式主要有 5 种。(实例见表 2)。

表 2 《中国分类主题词表》分面类目实例

例 1	J95	各种电视、电影:按内容分	} 第一层次
	96	各种电视、电影:按表现形式分	
	97	各种电视、电影:按题材分	
	971	依不同行业分	} 第二层次
	972	依时代分	
	973	依年龄性别分	
	974	依思想类型分	
	975	依情节结构分	
	U 459	各种隧道	} 按用途分
	· 1	铁路隧道	
例 2	· 2	公路隧道	
	· 3	地下隧道	} 按位置分
	· 4	穿山隧道	
	· 5	水底隧道	
例 3	I2	中国文学	} 按体裁分
	22	诗歌、散文	
	23	戏剧	
	24	小说	
	……		
	27	民间文学	} 按对象分
	28	儿童文学	
	29	少数民族文学	

例 4	TN 95	雷达	
	951	雷达原理	按事物的方面分
	952	雷达电子电话	
	953	雷达跟踪系统	
	958	各种体制的雷达	按事物的类型分
	959	各种用途的雷达	
例 5	TN 31	半导体二极管	
	311	按工艺分的二极管	
	312	按结构和性能分的二极管	
	313	按作用分的二极管	
	314	按形式分的二极管	
	315	按材料分的二极管	

注: 底下划线的分类标准在词表中没有

——类目分面结构分为两个层次, 相当于完全融合型词表的基本组面和亚面, 如例 1。

——分面类目的划分标准虽然没有注明, 但其隐含的划分原则是明显的, 相当于分面叙词表隐含的“分面标头”, 如例 2。

——在对特定学科进行分面时, 同一类列中多数类目采用同一划分标准, 如“按体裁分”; 少数类目采用另外的划分标准, 如“按对象分”、“按作者民族分”等, 如例 3。

——分类分析从第 5 级类目开始, 同位类使用“事物的方面”和“事物的类型”两个划分标准, 形成一组总论性类目和一组专论性类目, 如例 4。

——同一类列使用多个划分标准, 这种划分形式在“工业技术”大类很普遍, 也是《基础叙词表》常用的方法, 如例 5。

2 微观结构的阶段性特点和趋势

传统而功能单一的检索语言给标引和检索都带来一定困难。而一体化语言分面程度高, 先组度低, 又使推广和使用受到诸多限制。3 种一体化词表模式代表的 3 个发展阶段在这两方面显示了各自的特点和适应能力, 展示了一体化词表的发展方向和普遍推广的前景。

2.1 类目分面结构函级降低

尽管 3 种一体化词表都灵活运用了分面理论精化微观结构, 强化自身的整体功能, 但各自分面的深度和形式都很不一样, 因此适应能力也就不同。《分面叙词表》率先在工程领域进行试验, 创造了一体化词表的模式和方法, 被誉为标引和检索所用规范化词表的楷模。它把全分面分类法彻底改编成半分面分类法, 一级大类回归到传统的学科, 以鸟瞰学科领域的范围和位置, 在每一个大的学科领域里, 词语按分面的原则予以显示, 缩小了人为因素, 减少了使用难度。《基础叙词表》在此基础上有了大的改进, 分面类目级数降低, 一二三级类目都按传统学科划分, 知识系统性较好, 符合用户思维习惯, 使用范围逐步扩大, 由专业性的领域扩展到综合性的学科。《基础叙词表》是以技术科学为主, 兼以各个知识门类的一部收词为 17300 条的综合性一体化词表。编制社会科学、技术科学等大学科领域的词表, 《基础叙词表》可视为完善的模型。

《中国分类主题词表》基本遵循逻辑划分原则对类目进行层层逐级划分, 是《中图法》体系分类法模式。分面化结构从哪一级开始并没有严格界定, 完全是在体系结构原则下根据学科的发展和实际需要灵活运用分面分类技术。从表 2 所列实例看, 例 1 和例 3 从第 3 级类目开始出现分面结构; 例 2、4、5 则从第 5 级类目开始; 而另一些类目如“TL 4 各种核反应堆”是从第 4 级开始出现分面结构; 还有从第 6 级进行分面分析的, 如“TL 722 放大器”。据笔者统计, 3、4 级以后各级运用分面理论展开的类目较为普遍。5、6 级以后设置亚面的也相当多。

《中国分类主题词表》是两种检索语言即《中图法》(包括《资料法》)和《汉表》两者兼容的新型语言。广大用户群对《中图法》体系中的分面类目的利用已相当熟悉, 因此该词表的推广实施, 就等于用户换用《中图法》第 4

版, 容易进行。因而该词表的问世, 至少在我国, 比《分面叙词表》、《基础叙词表》在英国更有学术价值和适用性。

虽然完全融合型和对照索引型的词表都具有类目分面化的性质, 但标记形式是不相同的。《中国分类主题词表》把分面范畴固定成类表类目, 并配有相应的分类号(见表 3)。如“按作用分的变电所”是一个类表类目, 类号为“TM 631”, “按地点分的变电所”类号为“TM 632”等等。下级同位类作为一个亚面, 也标记有专指类号, 如:

- TM 631+ 1 升压变电所
- TM 631+ 2 降压变电所
- TM 631+ 3 整流变电所

表 3 完全融合型和对照索引型一体化词表分面形式比较

《基础叙词表》分面 范畴形式	《中国分类主题词表》分面 范畴形式
KSM 放大器 (按有源件分)	TM 63 变电所 TM 631 按作用分的变电所
KSM · J 电子管放大器	TM 631+ 1 升压变电所
KSM · K 半导体放大器	TM 631+ 2 降压变电所
KSM · L 介电放大器 (按操作方式分)	TM 631+ 3 整流变电所 TM 632 按地点分的变电所
KSM · N 带通放大器	TM 632+ 1 户内变电所
KSM · O 参数放大器 (按用途分)	TM 632+ 2 户外变电所 TM 633 按结构分的变电所
KSM · R 功率放大器	TM 633+ 1 固定变电所
KSM · T 差动放大器	TM 633+ 2 移动变电所
KSM · U 前置放大器	TM 633+ 3 成套变电所

而《分面叙词表》和《基础叙词表》则把分面范畴以醒目的字体排印在分类表中, 用分面标题标记, 不把划分标准的词语作为独立类目。例如表 3 中“放大器”的分面范畴是“按有源件分”、“按操作方式分”等。这种分面形式具有很大的自由度, 普遍推广使用具有很大的难度。而《中国分类主题词表》固定的类目分面结构, 既保持了《中图法》列举式体系的连

续性、稳定性, 又具有分面结构的灵活性; 既稳定了广大用户群, 又使兼容的语言拥有使用环境。因此, 该词表具有适应性广、通用性强的特点。研究的目的是为了用, 如果词表规模大、影响面广, 而使用单位亦多, 可作为衡量词表质量高低的首要指标, 那么词表结构的兼容程度, 则应作为第二衡量指标。象阮冈纳赞创制的冒号分类法, 虽是检索语言的一场革命, 但真正照搬使用的不多, 人们只运用了他的分面理论, 推动着检索语言向前发展。

2.2 先组词逐渐增多

冒号分类法彻底抛弃了对知识整体系统的客观追求, 把传统学科的内容分析成一个又一个基本概念, 类目属语义单元, 完全后组式。诚然, 此法优点甚多, 但它违背了传统学科的发展规律和人们的思维习惯。虽问世多年, 国内外使用者很少。但人们重视它的理论价值。艾奇逊第一个意识到, 必须突破冒号分类法不宜采用先组词的框框, 在《分面叙词表》中首次设立先组词。这些先组词由 2、3 个或几个单词组成, 能确切地表达某一主题的含义, 并且是人们经常查找的。在方法上将类目自由组配改为限定组配。可以采用组配的先组词类目用注释加以规定, 并在叙词表部用参照说明构成方式, 同时在非叙词后标注了分类号, 如“微波振荡器”(见表 4)。

表 4 《分面叙词表》先组词的组配形式

分类表部先组词 类目的组配形式	叙词表中先 组词的参照
LL 振荡器 (按频率分) 用频率的分类号组配, 例如: LL/CFQ 音频振荡器 LL/ΔYF 微波振荡器	微波振荡器 LL/ΔYF synth [*] S RT 微波 S BT 振荡器

《基础叙词表》的先组词较《分面叙词表》多, 兼容方式也作了改进。为防止标引员和用户用先组词进行标引和检索, 减少来回翻查

表5 《基础叙词表》的两种组代参照格式和实例

《分面叙词表》和《基础叙词表》都是新编的一体化词表,且收词规模小,用户面窄,先组词数量容易依需要确定。而《中国分类主题词表》则不然,其规模宏大,收词量大,使用单位多,包括了世界上所有的知识领域。更重要的是,《中国分类主题词表》的宏观两部分,即分类号——主题词对应表和主题词——分类号对应表,是我国图书情报单位作为分类标引和主题标引国家试用标准已历时 15 年的《中图法》(包括《资料法》)和《汉表》的新版,《中图法》是高先组式的体系分类法,《汉表》是高后组式的叙词表,为稳定我国 90% 以上的图书情报用户单位,推进我国检索语言的标准化进程,在兼容过程中不容许对《中图法》体系结构和标记制度作大的变动,因此,《中国分类主题词表》先组类目与《汉表》叙词的兼容技巧完全不同于《分面叙词表》和《基础叙词表》。该表对先组词与后组词的兼容采取了两条措施。第一,由于体系分类法类目不

表6 《中国分类主题词表》先组式类目形式

应表, 加上注释和参照系统, 在计算机的辅助下, 进行反向对应, 编制出了其字顺显示部分: 主题词—分类号对应表。这种独特的分类号—主题词对应表形式, 其左栏是《中图法》分类号, 复合类目概念; 右栏是对应的组配标题。由此, 分类表部分 5 万多个类目系统组织了《汉表》(包括新增的 14000 个主题词) 21 万余主题词和主题词串。由于《中图法》和《汉表》是国内通用检索语言, 因此, 分类号—主题词对应表这种特殊兼容形式虽为国际创新, 但却为我国广大用户熟知, 普遍推广使用很容易。第二, 对分类号—主题词中较难查检的先组词复合类目, 以非叙词收入主题词—分类号对应表, 款目以组代参照形式表示(见表 7)。若有一关于“三槽出钢”内容的文献,

表7 主题词—分类号对应表中的组代参照

要判断其学科内容,查“平炉炼钢—出钢,三槽”正式叙词即可。对这两条措施的具体步骤,还制订了一套严密的对应规则,包括一般对应规则和特殊对应规则,总的对应规则和

各项具体的对应规则等。其设计科学严密,编制技术精细,保证了对照索引型词表的广泛适应性。

3 结语

(1)《中国分类主题词表》的问世以及它拥有的众多用户,表明一体化语言的研究也走向定型成熟的阶段,即推广使用的阶段。现有分类表和叙词表数量颇多,若采用该词表的模式,不仅能节约编制时间和经费,降低标引难度,而且能充分利用现有分类法和主题词表的成果。

(2)一体化词表兼容的水平是相对的,而不是绝对的。正如奥地利数理逻辑学家哥德尔(Kurt Godel)的不完备性定理断言:任何足够丰富的数学系统,如果是相容的,就一定是不完备的。反之,如果是完备的,就一定是不相容的。因此应该说,用户最多的词表,水平也最高,尽管不完备,但能得到认同和共识。

(3)《中国分类主题词表》的分类号——主题词对应表用主题词或先组标题作类目的注释是该表的独创。这种特殊的注释克服了体系分类法线性排列的缺点。体系分类法是从某一主要方面反映学科性质,而用作注释的主题词或组配标题却是多维地反映学科的性质,隐含着多个微型分面分类表,因而为体系分类法的分面改造创造了按事物集中的条件。从类目对应的主题词和主题词串的多少可以窥见哪些类目学科性质较强,哪些较弱,哪些类目需要彻底改编,进而决定体系分类法分面改造的深度和方式。

(4)若在以类目主题概念为形式的非叙词后标注分类号,《中国分类主题词表》的使用将更加简便。

(5)由比较看出,代表一体化词表三个发展阶段的《分面叙词表》、《基础叙词表》和《中国分类主题词表》模式越向前发展,分面

程度越低,先组度越高,适用价值亦越大。分面程度高了,一些常用的概念在一体化词表中没有固定位置,用户很难接受,也容易产生错误组配。为便于标引和检索,适当增加先组词,减少人为因素是必要的;要使一体化词表容纳更多的新学科概念,《中国分类主题词表》必然要在用户接受的原则下逐版增加分面成份,缩小先组词数量。这样,完全融合型和对照索引型一体化词表在向前发展中会逐渐接近一个交汇点。

(6)一体化词表由新编的完全融合型发展到现在用两种检索语言的对照索引型,增强了新型检索语言的适应能力,并且开辟了广泛研究的前景。我国现有分类法 120 余种,主题词表 75 种^[8]。充分利用《中国分类主题词表》的研究成果,以此作为基本模式和蓝本编制专业一体化词表,可谓是“水到渠成”。这也将是我国一体化词表发展的必由之路。

参考文献

- 1 刘湘生 分类主题一体化是我国情报检索语言的发展方向 北京图书馆通讯, 1987, (4): 12~16
- 2, 3 侯汉清 当代分类法主题法索引法研究 北京: 书目文献出版社, 1993
- 4 尚克聪 试评《中国分类主题词表》 中国图书馆学报, 1996, (2): 68~71
- 5 《中国图书馆分类法》编委会编 《中图法》、《资料法》(第三版)使用手册 北京: 书目文献出版社, 1991
- 6 《中国图书馆图书分类法》编委会编 中国分类主题词表 北京: 华艺出版社, 1994
- 7 张琪玉, 刘湘生主编 中国分类主题词表教程 北京: 华艺出版社, 1994
- 8 胡明 建国以来我国分类法和主题法词表研制的统计分析 图书情报工作, 1994(4)

袁相琴 兰州大学图书馆副研究馆员。通讯地址: 兰州市, 邮编 730000。

(来稿时间: 1996 11. 11。编发者: 李万健)