中国图书馆学报(双月刊)1996 年第 6 期
ZHONGGUO TUSHUGUANXUEBAO

●戴维民

汉语叙词表轮排索引编制技术论析

ABSTRACT In the light of his own practices, the author discusses the principles, rules, the optimization of structure, the improvement of functions and the techniques of compilation. 11 refs.

SUBJECT TERMS Chinese thesauruses - Stricture - Rotated indexes - Compilation.
CLASS NUMBER G254.241

1 前言

轮排索引技术现已较广泛应用于叙词表,成为叙词表结构体系的一个重要组成部分。这对改进叙词表的结构功能,改善叙词表编制方法具有重要意义。

《叙词表指南》报道的 219 部英语词表中有轮排索引的 39 部,占 18%。《联机数据库中应用的叙词表:分析指南》报道的 122 部词表中有 22 部,有轮排索引的也占 18%。

我国叙词表编制中长期不重视轮排索引技术,在一般叙词表结构体系设计时也不考虑编配轮排索引。张琪玉教授认为其原因有三:(1) 认为轮排索引的价值不大;(2) 会增加词表印刷成本;(3) 编制过程还没有计算机化,手工编制比较麻烦。但近年来随着计算机在词表编制中的广泛应用,以及对轮排索引功能的新认识,一些词表开始引入轮排索引,如中国艺术研究院编制的《艺术科学叙词表》就编配了题内关键词式的轮排索引。这是一个良好的开端。近年在我们编制词表的实践中,也逐渐重视了对轮排索引技术的研究和应用,《社会科学检索词表》、《教育主题词表》、《军队档案常用主题词表》以及《音像资料叙词表》中均编制了轮排索引。我们感到轮排索引是叙词表的一个重要组成部分,对词表的利用和编制能起到重要作用。本文就结合编制轮排索引的体会和心得,系统论述叙词表轮排索引的结构功能和编制的技术方法。

2 轮排索引的基本原理

轮排索引的核心原理是词素轮排方法,即将具有多个词素的语串中的每个具有检索意义的词素置于查词入口位置,以提供多途径查检功能。

叙词表的轮排索引是将全部叙词按检索意义切分后的词素再按字顺排列起来的索引类型,使含有同一词素的叙词能集中显示于一处,从字面成族角度,达到族性检索的效果。

叙词表轮排索引的编配是基于叙词的先组度和字面成族原理。

2.1 叙词的先组度

当今叙词表为达到专指性和表达性,均有一定的先组度,一个叙词常常包含几个具有独立语义的词素。有人认为,现代叙词表的

先组词达到 50%。据兰开斯特的资料，英语和法语叙词表的每个叙词包含 1.5～2.0 个词素，德语叙词表的每个叙词则包含 1.1～1.2 个词素。又据对国外 22 部词表的调查，二元词、三元词和四元词约占总词汇的 45%，四元以上的先组词就很少了。我国《汉语主题词表》总词量为 108568 条，而使用先组词达 65757 条，因此其先组度为 65757÷108568×100%＝60.6%。而国内绝大多数叙词表的先组度都在 60%左右，有的则更高。我们对《军用主题词表》的先组度作了一份抽样统计调查，得知其先组度为 80.4%，每个叙词平均包含 2.40 个词素。由于叙词是经优选的规范化词汇，因此每个词素都有实质意义，但在字顺表中只有词首的词素(严格说是首字)才具有查词入口功能，经过词素轮排后，所有独立词素都可作为检索入口，从而大大增加了检索途径。

2.2 字面成族原理

具有相同词素的叙词在语义上都具有一定的内在联系，这种联系在某种程度上就是“族性”关系。而这些相同的词素所处位置可能是词首、词中，或是词尾，叙词表的其他显示方式都难以全面显示这种字面族性关系。轮排索引则可理想地满足这一功能需要。

3 多词素叙词的切分规则

3.1 叙词切分需要制定切分规则

编制轮排索引首先必须对全部正式叙词和非正式叙词进行切分。叙词切分一般采用人工或人机结合的切分方法，即在可作为入口的词素前加入机器可读的切分符号。为了切分的规范统一，应有一套行之有效的规则。

目前，还没有一套成熟的叙词切分规则。国家标准《信息处理用现代汉语分词规范》(GB13715)声称适应于汉语信息处理各领域，其他行业和有关学科可参考使用。但该规范主要考虑分出词的独立性，而轮排索引分词是更多地考虑词干、词素独立的语义特征、查词入口及字面成族聚类功能。下面是《规范》中的分词单位在编制轮排索引时的不同分词方法：

《规范》分词单位	轮排索引分词方法
工农业	工|农业
非金属	非|金属
超声波	超|声|波
加速度	加|速度
离退休	离|退休
原油	原|油
讲课	讲|课

3.2 叙词切分总则

轮排索引的编制与一般的信息处理有所不同。根据我们编制轮排索引的体会，叙词切分的总原则是：

(1) 对有几元词组成的先组性词组，均切分成几个词或词素；

(2) 具有成族性的词素统一予以切出；

(3) 有些虽是多元词，但切分以后不具备成族功能和查词入口功能则不予切分；

(4) 切分符号以“|”为标志。

(5) 考虑到汉语词汇切分的难度，切分时只考虑切出入口词，而不考虑所切词长，因此称之为“左切分”或称“切左不切右”，这样可以回避抽词难度。根据这一特点，我们建议并推荐汉语叙词表的轮排索引应主要采用题内关键词索引形式。

3.3 词组性叙词切分规则

从切分的角度来说，我们把叙词分成三类：单元词(或单个词素的词)、词组、缩略词。单元词不切分，在轮排索引中排一次。词组和缩略词切分则有所差异。

词组性叙词是根据四种结构来进行切分。其主要规则是：

(1) 联合词组是由两个或两个以上并列词素构成，每个词素均予切出；

(2) 动宾词组是动词词素和名词词素构成，名词词素应予切出；

(3) 主谓词组是由名词词素(或代词)和

动词、形容词或其他修饰性词素构成，后者如能独立表达某一概念，应予切分，否则不切分；

（4）偏正词组是由形容词等修饰性词素和名词词素构成，名词词素应予切出；

（5）带有接头、接尾或虚化成分的词组，在词头的应切出其后词素，在词尾的则可将接尾词素切出；

（6）多种结构复合的词组应作多次切分。

右表是各种结构词组的切分示例：

结构类型	叙词词组	切分示例
联合结构	扶贫帮困 拥军爱民 一专多能	扶贫|帮困 拥军|爱民 一专|多能
动宾结构	查帐 统一祖国 建房	查|帐 统一|祖国 建|房
主谓结构	学生流失 学生贷款 学生自主择业	学生|流失 学生|贷款 学生|自主|择业
偏正结构	地方武装 监视卫星 国防工程	地方|武装 监视|卫星 国防|工程
接头接尾	副教授 超导体 管理处	副|教授 超|导体 管理|处
复合结构	反坦克武器 建党方针 反坦克炮营	反|坦克|武器 建|党|方针 反|坦克|炮|营

3.2 缩略词切分规则

为了使词表的词汇简明，较广泛地采用了缩略词；有些作为正式叙词，有些则作为非叙词，但在轮排索引中均应处理。

语词缩略的方式主要有四种：缩合、节略、简代和统括。

缩略方式	释义	全称词	缩略词	切分示例
缩合	把原来词语分拆成几个词素，然后缩掉词素中的某些部分，最后把保留的部分凝合起来，构成缩略形式。这是缩略语的主要方式，它具有多种具体形式。其保留词素具有独立入口意义应予切分。	科学技术 扫除文盲 香港澳门 预防潮汛 教师学生 中医西医 食品疗法	科技 扫盲 港澳 防汛 师生 中西医 食疗	科|技 扫|盲 港|澳 防|汛 师|生 中|西医 食|疗
节略	把原语词中的某一部分略去，只保留其中间部分，分略尾和节头两种。多数是作为单元词处理，如果保留部分是复合词素，则可作切分处理。	中国人民解放军 农村人民公社 地震烈度	解放军 公社 烈度	解放|军
简代	这是一种"代用"形式的缩略，它是长期以来人们约定俗成的结果。分代称和简称两种。一般不需切分。如尾词具有聚类作用可切出。	上海 中华人民共和国 亚细亚洲	沪 中国 亚洲	 亚|洲
统括	把构成成分较繁的原语词的几部分中共有的成分抽取或概括出来，再在它前面标上与原构成项数相等的数字面构成的一种缩略形式，分"取字"和"取义"两种。这种略语多为正式叙词，根据词素构成情况作必要切分。全称词为非叙词也作切分。如全称词作注释项，也可切分后轮排。	百花齐放、百家争鸣 农业现代化、工业现代化、科学技术现代化、国防现代化	双百方针 四个现代化 四化	双百|方针 四个|现代化 四|化

4 轮排索引的显示结构

轮排索引的类型通常有三种形式：

（1）题内关键词索引；

（2）题外关键词索引；

（3）倒置轮排索引。

上述三种索引形式中，第二种形式使用较方便，但占篇幅较大；第三种形式采用较少。第一种版面醒目程度和篇幅都比较理想。现代叙词表轮排索引形式采用比较多的是第一种。

各种不同的索引类型及同一类型在不同词表中应用时，具体的著录格式和项目也有所不同。

（1）题内关键词索引的入口词素与其他

词素之间空一格，以示醒目。题外关键词索引入口词素单独一行顶格著录，整个叙词则另行缩格著录。入口词素首字第一次出现用醒目字体显示(如黑体)。

(2) 有些词表著录"用"项(Y 或 USE)，对非叙词可用不同字体显示(如汉语的楷体，外文的斜体)。

(3) 为了查词后便于词表的翻检，在款目前或后著录范畴号或分类号以迅速查检范畴表或分类表，著录叙词序号以迅速查检字顺表，也可以著录字顺表和范畴表(分类表)的页码和栏数。

(4) 轮排索引的版面一般分双栏，少数采用单栏。

下面是中外几种词表轮排索引的片断，可以显示轮排索引的基本形式。

10713		**PHYSICS**	
30200		**PICTURE**	
24210	**MOTION**	**PICTURE**	**USE FILM**
22000	**NEWS**	**PICTURE**	
25320		**PICTURE PHONE**	**USE VIDEOPHONE**
24230	**ANAMORPHIC**	**PICTURES**	
10742		**PILOTPROJECT**	
25400		**PIRATE STATION**	
35240		**PLAN**	
32230	**DEVELOPMENT**	**PLAN**	
35240		**PLANNING**	
34300	**COMMUNICATION**	**PLANNING**	
32230	**DEVELOPMENT**	**PLANNING**	
33240	**EDUCATIONAL**	**PLANNING**	
37540	**FAMILY**	**PLANNING**	
35240	**LOCAL**	**PLANNING**	
35240	**NATIONAL**	**PLANNING**	

《大众传播叙词表》轮排索引片断
(题内关键词索引)

FOREIGN
ARMED FORCES (FIREIGN)
FOREIGN BODIES
FOREIGN POLICY
FOREIGN TRADE
FORK
TUNING FORK GYROSCOPES
FORM
FORM FACTORS
JORDAN FORM
FORMALDEHYDE
PHENOL FORMATLDEHYDE
FORMATION
ENERGY OF FORMATION
HEAT OF FORMATION
ICE FORMATION
FORMHYDROXAMIC
FORMHYDROXAMIC ACID
FORMIC
FORMIC ACID

《NASA 叙词表》轮排索引片断
(题外关键词索引)

尊干	爱兵		[14A01]	09793
	爱兵模范		[15E01]	00001
	爱国人士		[23A]	00002
	爱国卫生工作		[34D01]	00003
	爱国卫生运动		[34D01]	00004
	Y 爱国卫生工作		[34D01]	00003
……	……			
拥军	爱民		[0ID02]	08517
拥政	爱民		[22B,0ID02]	08520

《军队档案常用主题词表》轮排索引片断
（题内关键词索引）

	分类	Z155	00450
图书	分类		01988
文献	分类	YH25	02050
学习	分类		02343
	分类法		00451
教育目标	分类学	AB118	01075
目标	分类学		01440
教育	分流	DA2	01011
毕业	分配	JP642	00063
定向	分配	JP642.3	00342
统一	分配	JP642.1	01980
学校经费	分配		02394
	分歧	Z157	00452
	分析	Z159	00453
错误	分析	BD546	00215
精神	分析		01215

《教育主题词表》（分面叙词表型）轮排索引片断
（题内关键词索引）

索引天头著录首字和汉语拼音起止音节。

5 轮排索引的排序方法

5.1 轮排索引排序范围

根据词表规模及成本因素，编制轮排索引时，确定非叙词、单元词（单个词素的词）词组性叙词第一个词素是否参加轮排。从理想角度来说，一般的方法是：

（1）正式叙词和非正式叙词均切分后参加轮排；

（2）单元词全部参加排序；

（3）词组性叙词第一词素参加轮排。

这种所有的词素都参加轮排的索引称之为全轮排索引。

5.2 排序规则

题外关键词索引和倒置法各词素排序方法都只有一种，即从左至右顺排。题内关键词在排序时，同一入口词素的其他词素的排序方法有两种：（1）先顺排入口词素右边，再顺排左边词素；（2）先顺排入口词素右边，再逆排左边词素。其样式如下：

第一种	第二种
入口词素	入口词素
↓	↓
(4)(5)(6)(1)(2)(3)	(6)(5)(4)(1)(2)(3)

上述第二种是较为科学的排序方法。这是因为在语言表达时，词组的重心一般在后和逻辑位置邻近限定性较强的原则。

如果是计算机排序，第二种排法只能逐字排，否则需要人工辅助。

下面是《军队档案常用主题词表》轮排索引片断，展示了第二种排序方法：

	（查词入口）
林彪反党	集团案
江青反党	集团案
反党	集团成员
社会	集团购买力
达赖	集团回归祖国问题
	集团军
合成	集团军
机械化	集团军
陆军	集团军
坦克	集团军
	集训
	集训队

采用左逆序法可以使有些相关款目相对集中。如下例：

陆军船艇	部队
导弹艇	部队
布雷舰艇	部队
登陆舰艇	部队
水面舰艇	部队
反潜舰艇	部队
护卫舰艇	部队
鱼雷艇	部队
潜艇	部队
弹道导弹潜艇	部队
常规动力潜艇	部队
核动力潜艇	部队
猎潜艇	部队

此例中的左逆序排序使13条有关“艇”的

部队集中一处，在13条款目中又有相对集中的小排序。《军队档案常用主题词表》中含“部队”一词的叙词共167条，如不采用左逆序排序，上面13条款目则被分散在167条款目之中，不便查检。

《中国农业叙词表》编制了《农业叙词逆序轮排索引》，是完全根据逆序词典方法编制的，其根据是汉语词汇重心在后的原理。这种逆序索引与顺表的顺排起到了互补作用，但与依词素切分和左逆序方法编制的轮排索引相比，其功能是有差别的。

5.3 排序结构的优化

在一定的字顺范围内，可调整引入逻辑排序。例如：

（常规排序）	（调正后后逻辑排序）
八一勋章	八一勋章
二级　八一勋章	一级　八一勋章
三级　八一勋章	二级　八一勋章
一级　八一勋章	三级　八一勋章
八月	八月

又如：

（常规排序）	（调正后逻辑排序）
步 校	校官
大 校	少 校
党 校	中 校
上 校	上 校
少 校	大 校
学 校	步 校
院 校	党 校
中 校	学 校
校官	院 校
校长	校长

现代叙词表在设计字顺排序时大多采用汉语拼音排序，轮排索引也是如此。为配合查检方便，可采用一些辅助手段，包括：增加笔画检字表，在版面的眉头上标明汉语拼音和首字，入口词首字第一次出现时用黑体字。

6　轮排索引的参照方法

为了尽可能地增加查词入口，叙词切分时切得比较细，这就使印刷本词表篇幅大大增加，加大印刷成本。通过参照既不减少入口，又可大大节省篇幅。

下面是《军队档案常用主题词表》轮排索引中的参照一例：

				434页：
边	境→边境	434		边境
因公出	境	[94B]	09297	边境地区
因私出	境	[94B]	09301	边境工作
入	境	[94B]	06919	边境观察
	境内	[98A03]	04795	边境管理
	境外	[98A03]	04796	边境管理部门
	纠察	[08E03]	04797	边境禁区
				边境口岸
				边境事件
				中越　边境自卫反击战

上例左栏中“边境”一词被切分为两个词素，但含有“边境”这一单元词的叙词共有10个，如依同样切分方法又有10个入口。为节省篇幅，采用参照法。左栏中“边　境—边境　434”是一参照款目，表示含有“边境”一词的其他叙词“境”前不再切分，由此省略了9条款目，“边境”作为一个词素在434页轮排。右栏展示的就是434页轮排情况。

由于参照方法的应用，使叙词切分可以更细，并在很多情况下可以切分到单字，在不增加很多款目的情况下而增加入口。

7　轮排索引的功能与使用

（1）充当一种提供多个检索入口的、按词素排列的字顺索引，帮助用户迅速判断词表中有无所需的词组，提高查词速度。

（2）一个比较完善的轮排索引收录范围广泛，包括正式和非正式叙词切分单位，有的轮排索引还收入其他入口词。因此，轮排索引相当于入口词表，提供更多的查词和检索的入口。

（3）将具有同一词素的叙词汇集在一起，可以帮助用户选词，提高标引和检索的准确性。