

《汉语主题词表》的理论和应用

刘湘生 编著



书目文献出版社

1980 北京

《汉语主题词表》的理论和使用

刘 湘 生 编 著

书 目 文 献 出 版 社 出 版

1 9 8 0 · 北 京

内 容 提 要

本书是刘湘生同志根据参加编制《汉语主题词表》工作的认识和体会，从理论、方法、结构和使用等方面对该书进行的论述。全书内容包括：绪论，叙词法的理论探讨，《汉语主题词表》的体系结构，《汉语主题词表》的编制和选词，《汉语主题词表》的使用，主题目录和主题词表的管理共七章。它有助于在标引和检索图书资料时参考。

本书可供图书馆和情报单位的有关人员，大学图书馆学系的师生参考。

《汉语主题词表》的理论和使用的

刘 湘 生 编 著

书目文献出版社出版

北京文津街七号

秦皇岛市第二印刷厂印刷

新华书店北京发行所发行 各地新华书店经售

1980年5月北京第1版 1980年5月北京第1次印刷

印数1—16,000册 定价：0.58元

图书分类号：G254·23 统一书号：7201·14

目 录

第一章 绪 论	(1)
第一节 图书情报工作的深刻变革.....	(1)
第二节 《汉语主题词表》的提出.....	(5)
第三节 对《汉语主题词表》的简评.....	(7)
第二章 叙词法的理论探讨	(10)
第一节 叙词法.....	(10)
第二节 叙词的概念性质.....	(11)
第三节 叙词的规范化性质.....	(14)
第四节 叙词的组配性质.....	(20)
第五节 叙词的语义性质.....	(26)
第六节 叙词的动态性.....	(33)
第七节 叙词法的特点.....	(35)
第三章 《汉语主题词表》的体系结构	(40)
第一节 叙词表的定义和体系结构.....	(40)
第二节 《汉语主题词表》的主表和附表.....	(50)
第三节 《汉语主题词表》的辅助索引.....	(53)
第四章 《汉语主题词表》的编制和选词	(69)
第一节 《汉语主题词表》的编制.....	(69)
第二节 《汉语主题词表》的选词.....	(77)
第五章 《汉语主题词表》的使用	(84)
第一节 主题词表的作用和使用范围.....	(84)

第二节	什么是标引工作·····	(89)
第三节	标引的方法步骤·····	(91)
第四节	标引规则·····	(101)
第五节	组配规则·····	(105)
第六节	计算机检索·····	(116)
第六章	主题目录·····	(123)
第一节	主题目录的特点和作用·····	(123)
第二节	组织主题目录的依据·····	(125)
第三节	主题目录的类型·····	(131)
第四节	组织主题目录的若干规定·····	(133)
第五节	主题目录的物质形式·····	(136)
第七章	主题词表的管理·····	(141)

第一章 绪 论

第一节 图书情报工作的深刻变革

目前,电子计算机已逐步被许多国家普遍采用,从而引起各个领域的工作发生了一场伟大的深刻的变革。在科学的意义上讲,当今的世界已经是电子计算机的时代了。电子计算机使用的程度、数量和水平是衡量一个国家科学技术、经济文化、军事国防等各项事业发达与否的重要标志之一。这场变革犹如十八世纪中叶至十九世纪中叶英国产业革命一样,具有划时代的深远影响。

图书情报工作也正处在这场深刻变革的历史时期。由于电子计算机在图书情报工作中的广泛应用,使图书资料检索工作的面貌发生了巨大的变化。图书情报工作的发展与提高,有赖于科学技术的发展与提高,同时又产生巨大的反作用,不断促进科学技术的前进,图书情报工作由手工检索向电子计算机检索急速发展的这场变革,并不是偶然的现象,而是历史发展的必然趋势。

首先,由于现代科学技术的发展,导致了图书资料数量的急剧上升。据报道,全世界每年估计可产生数百万项科研成果,每年大约要产生数千万页的技术资料。同时,每年出版的期刊、连续出版物达10万种,其中比较重要的约2万余种,而且常以7~8年翻一翻的速度向前增加。加之图书资料的类型复杂、出版分散、文种多样、重复交叉严重。在这种现实情况下,传统的手工检索体系越来越不能适

应科研人员的检索要求。如果没有现代化的技术装备，读者势必像大海捞针，很难在这图书资料的海洋中既迅速准确，又全面地查得自己所需要的资料。往往会因手工检索体系的局限，致使情报不灵、资料不全。甚至重复别人研究的成果，造成时间、人力和经费上的极大浪费。这是引起图书情报工作进行变革的根本原因。

第二、电子计算机的产生和发展是实现这一变革的技术条件和物质基础。

自四十年代中期世界上第一台电子计算机问世以后：五十年代初就有人开始研究电子计算机在图书资料工作中的应用问题；六十年代一些单位的图书馆(如美国国会图书馆)就已正式使用电子计算机处理图书资料的编目、阅览外借、期刊管理和其他工作；并产生了机器可读目录MARC I和MARC II；七十年代不少国家已经普遍地研究和应用电子计算机解决图书资料工作的自动化、网络化和缩微化的问题。

一些科学技术和工业发达的国家，如美国、苏联、英国、法国、西德、日本、加拿大等都已实现了电子计算机检索的自动化、网络化。其自动化的程度和水平，也随着计算机的规模和性能的发展而越来越高，处理能力也越来越强。从电子计算机检索本身来说，已从最初的单机单终端检索发展到了多机多终端检索；从检索的形式来说，已从早期的脱机成批检索发展到了人机对话式的联机检索；从检索的范围来说，已从一个单位的检索发展到了地区性和国家间的网络检索。

第三、电子计算机与光学记录技术相结合，使图书资料工作更上一层楼，迈向了新的高度。在七十年代初期，电子

计算机还只能做到二次情报的检索，即只能检索出图书资料的线索(除检索图书资料的名称和著者等著录事项外，还能检索文摘和提要等)，到了七十年代晚期由于光学记录技术的发展，解决了高密度信息存贮的技术难关，尤其是激光缩微技术的发展，可以使图书资料高度的缩微化，促使图书资料的线索检索，提高到本文检索。据了解，国外一张105毫米×148毫米大小的超缩微平片，可以存贮3280页的图书资料。一屈目录盒的超缩微平片，就可以存贮上十万册的藏书(指全文)。像具有一千万册藏书的北京图书馆，如果采用超缩微平片进行缩微处理，则只需要一百屈目录盒即可容纳全部馆藏。如果采用最先进的激光打点记录技术，美国会图书馆的五千多万藏书，只需要四盘缩微胶卷，就可容纳全部馆藏。显然，图书资料的这种缩微化，一方面大大解决了图书馆和情报单位馆藏空间十分紧张的状况，另一方面为电子计算机进行全文(本文)检索创造了条件。如《纽约时报》社、国际商业机器公司技术情报检索中心、巴梯尔纪念研究所、麻省理工学院等单位，都已实现计算机的全文检索。如果需要还可以通过计算机的输出复制设备，为读者提供全文复制。当前，图书资料的计算机缩微输入(简称CIM)和计算机缩微输出(简称COM)，是情报图书检索工作发展的又一尖端技术。

第四、电子计算机与现代通信技术相结合，使图书资料情报的交流和检索展现了新的面貌。由于在电子计算机检索系统中采用了现代通信技术(包括专用通信技术和卫星通信技术)，从而实现了自动化检索的网络化。也就是说解决了远距离检索，包括多单位、多中心、多系统之间的联机网络检

索问题。

如果在传统的手工检索方式中，要实现馆际之间、国际之间的查询和借阅那还是相当困难的。少则几天、多则一两个月才能得到查询和借阅的结果，非常费力、费时和费钱。但在实现了电子计算机网络化的国家，这就不再成为一个问题了。电子计算机检索的网络化是通过通信设备把分散于各地区各国家的电子计算机及其终端设备，所存贮的资料联接起来而形成一个网络。读者可以在网络内的任何一台计算机及其终端，检索到另一台计算机资料档的图书资料。如美国医学图书馆在1975年联接有500个终端，1976年增加到1,000个终端，其中利用通信卫星与英国、法国、西德、加拿大、巴西等国家设置的200个终端进行联网检索。美国俄亥俄大学图书馆中心，在1977年夏，参加该检索中心的有41个州的1,068个图书馆，联接了1,668个终端。这种网络，一方面促使了图书资料情报的交流，加速了检索速度，另一方面达到了各单位、各地区、各国家对图书资料资源进行共同利用的目的。

综上所述，当代图书资料工作的巨大变革，是：电子计算机使图书资料工作的自动化，光学记录技术使图书资料的缩微化和现代通信技术使图书资料检索的网络化。

在我国，图书资料工作长期处于传统的手工检索状态，远远不能适应形势发展的要求。与国外主要技术发达的国家相比，非常落后。就以我国藏书丰富、历史悠久、工作人员最多的北京图书馆而言，人家评论时说：有“两个罕见”，一个是收藏古代珍本书籍之多是世界罕见；另一个是图书馆的设备陈旧落后是世界罕见。国内也有不少人对国家图书馆当前

存在的“小、少、慢”现象提出批评。这些看法都是基本合乎客观实际的。面对这一现实，我国图书情报工作如何迎头赶上世界先进水平，这是摆在我们面前的一个刻不容缓的重要任务。列宁早在1921年就曾经指出：“无论如何要继续前进并吸取欧美科学中一切真正有价值的东西——这就是我们第一等的首要任务”。在华主席为首的党中央主持制定的发展我国科学技术的长远规划中，特别强调指出：“现代科学技术研究需要有强大的科技队伍，现代化的实验手段和科技情报工作。”要“建立一批现代化的实验室、实验基地和重大实验设施以及现代化的情报中心和图书馆。”我们应当向世界各国图书馆虚心学习、虚心请教，迅速改变落后面貌，努力实现上述变革，使我国在图书情报工作中达到先进水平，保证能为我国四个现代化建设提供图书资源。

第二节 《汉语主题词表》的提出

1975年周总理在四届人大会议上，向全党全国人民提出了向四个现代化进军的伟大号召。在这一号召的鼓舞下，为了研究和解决电子计算机在我国国民经济、情报检索、新闻报导和出版印刷等方面的应用问题，“汉字信息处理系统工程”（简称“748工程”）正式上马了。如果这一工程的研制能成功，即将对我国在科学技术上赶超世界先进水平，加速实现出版印刷、新闻报导、情报检索和国民经济等各项事业的现代化，具有重大的意义。同时，为电子计算机在我国各条战线上的使用和普及具有关键性的作用。

所谓汉字信息处理系统，就是要解决如何用计算机处理

汉字和汉字信息的系统。也就是说要解决计算机如何用汉字输入,存贮和输出的一整套设备的设计、制造和软件程序的设计、使用问题。这是一种新型的技术设备。它把不同形体的汉字,变成不同的光电信号,通过计算机的控制和运转,可以实现自动制造汉字、自动编排、自动拍照印刷和自动检索图书资料等工作。虽然世界上电子计算机的应用已非常普遍,但是一般都用字母作为信息处理的依据。对于用汉字信息处理的设计,日本虽然已经研究成功,但它是日式的汉字,与我国的方块汉字不尽相同。汉字在我国有着悠久的历史,几千年来在我国历史上一直起着主要的传播作用。目前,外文在我国人民中间的传播作用比较薄弱,外文水平很不普及的情况下,只有汉字信息处理系统得到解决,电子计算机才有可能在我国真正地广泛使用。为了实现四个现代化,汉字信息系统的电子计算机的设计和生 产是一项十分迫切和关键的任务。

在实现我国图书情报检索工作现代化的问题上,存在两条腿走路的趋势和可能。一方面,我们可以直接生产和使用字母信息处理系统的电子计算机,单纯解决外文图书资料的自动化检索问题。它的好处是可以直接利用外国现成的磁带资料,节省人力、物力和时间。但对中文资料的存贮和检索却无能为力。因此,另一方面,则是要求大量生产和广泛使用适合我国汉字信息处理系统的电子计算机。要使用汉字信息处理系统电子计算机进行情报检索,就必须编制一部《汉语主题词表》为上机使用。因此,《汉语主题词表》的编制也就被提到日程上来了。它是作为“748工程”的一个配套项目提出来的。现由中国科技情报所和北京图书馆负责组织和

主持《汉语主题词表》的编辑工作。

第三节 对《汉语主题词表》的简评

《汉语主题词表》的编制，前后经历了整整四年的时间。它是在全部手工操作的情况下，由505个单位、1,378位专业工作者和图书情报工作者克服重重困难，付出了巨大的劳动，而编制成功的。这部词表的出版，为全国主题检索体系的建立，采用电子计算机检索，实现集中统一主题标引，为以后编制多语种主题词表的工作奠定了基础。像《汉语主题词表》这样的编制速度和庞大规模的巨型综合性词表，目前在世界上来说是少有的。它与国外几部重要的词表相比较，有以下几个特点：

1. **范围广、级别深** 《汉语主题词表》收词的范围包括社会科学和自然科学技术有关的各门专业领域。所编制的“范畴分类索引”比国外的词表要深一级，一般都分到第三级。同时，范畴的分类是以《中国图书馆图书分类法》的大类序列为依据的，故大类的确立，主要以学科分类为基础。这是从综合性主题词表出发，考虑到主题词的分类，尽量向学科分类靠拢，以便增强词表族性检索功能。

下表是世界上几个大型词表的情况。

项 目 国 家	词 表 名 词	收 词 范 围	范 畴 类 目
中 国	《汉语主题词表》	综 合 性 的	58大类、675 二级类、1080 三级类

美 国	《DDC》	国防科学技术	未编范畴分类 索引
苏 联	《科技检索词典》	国家科学技术	33大类、 302小类
日 本	《JICST》	理工科学技术	14大类、 173小类

2. **词量大** 由于《汉语主题词表》是综合性的，收词比较多，共有108,000个，比世界上几个大型词表的收词都多，其他的词表为：《DDC》收词52,000个；《科技检索词典》收词18,925个；《JICST》收词33,998个。

3. **英汉对照** 主题词既有英汉对照索引，同时在主表中每个款目主题词下又附有英文译名。因此，《汉语主题词表》既是一部情报检索工具，又是一部综合的大型的翻译参考工具。

4. **划分科学为两大部分** 《汉语主题词表》的主表，辅助索引均划分为社会科学和自然科学两大部分。

由于编表人员的水平和认识所限，加以没有电子计算机检索的实践经验。因而存在的缺点和问题不少。主要有以下几点：

1. **主题词选词不当** 漏选的、不该选的主题词都有。如社会学、思想性、任务等重要概念词都已漏选。这将给标引和检索工作带来一定的困难。此外，各选词单位对入选主题词的标准掌握不一样，有的过细，有的过粗。

2. **体例不一致** 《汉语主题词表》的汇总编辑，是由社会科学编辑小组和自然科学编辑小组分别进行的。人员分

散、工作分散,加之缺乏严格完整的统一筹划和研究,所以在编制体例上出现了某些很不一致的地方。比如“词族索引”自然科学部分只是族首词带有范畴分类号,其下位词均不带此号;而社科部分除了“附表”中的主题词外,不论族首词还是其下位词均带有范畴分类号。又比如自然科学部分编制了主题词部首索引,而社会科学部分则没有编制。两个主表主题词汉语拼音的排列方法也不同,一个按纯字母逐个比较排列,一个则采取按词集中的办法等等。

3. 编排不统一 在编排技术上也缺乏严格的统一性。比如,对《汉语主题词表》的总的说明,对主表、范畴分类索引、词族索引、英汉对照索引等部分的说明及其编排上,出现许多交叉重复,甚至有一些说明解释不一致的地方。又比如对排印的字体上也前后出现两种不同的状况。

4. 有矛盾的地方 在主题词参照关系上,由于手工编校的原因,也可能出现一些矛盾的地方。

5. 索引分散 “英汉对照索引”共分六大块,显得十支离破碎,不利于查找。对英文主题词未加以统一和规范,因此,还不能直接用来组织英文主题目录。

总之,一部主题词表编制的好坏,有待于今后实践的检验,不断修订完善。

第二章 叙词法的理论探讨

第一节 叙 词 法

叙词法 (Descriptor method或Method of descriptor)是十九世纪五十年代随着电子计算机在图书情报工作中的应用发展而逐渐产生的。也是在分类法、标题法、单元词法和关键词法的基础上,适应现代科学技术和图书资料的迅猛发展的需要应运而生的。所谓叙词法,就是以叙词作为标识符号,依据叙词基本性质的理论原则,标引图书资料的主题和检索图书资料的一种新式的检索方法。依照叙词法的原则编制起来的,主要供电子计算机上机使用的表称之为叙词表。由于利用电子计算机进行图书资料检索,具有速度快、存贮量大、自动化程度高、适应性强的特点,所以现在世界上越来越多的国家的图书情报部门,正广泛采用叙词法来编制叙词表,供电子计算机使用。以期实现图书资料检索的自动化。

在全国科学大会上提出:“八年内要建立起一批文献检索中心和数据库,初步形成全国科学技术情报图书计算机网络。”我国图书情报部门如何贯彻落实,更好更快地实现图书资料的电子计算机检索,建立全国统一的检索体系和网络,以及做到及时、准确和全面地为四个现代化提供图书资料,是摆在我们面前的一项十分迫切和艰巨的任务。目前叙词法和叙词表在我国还比较陌生。因此,如何积极宣传、探讨和掌握它们,具有重要的现实意义。

在涉入本章正文之前，首先就叙词的名称问题作一个简单的交代。目前，我国图书情报界对这个问题存有争论意见。有的同志认为称“叙词”较好，有的同志则主张称“主题词”为宜，各有各的理由。究竟怎样定法？还是二者通用？尚无一定结论。这次出版的《汉语主题词表》的名称也只是暂定，还没有作出统一的规定。依本人一孔之见，还是统一称叙词为好（注：关于这个问题请参看本文作者1978年第三期《图书馆工作》“对叙词及有关几个名词术语的认识”一文。）如果称主题词，似乎难以自圆其说，其中最主要的一个矛盾是对叙词法应如何称呼？叙词法是主题法之一，这一点显然已为大家普遍接受。但是，将叙词改为主题词，那么叙词法又将改称什么法呢？现在还没有任何人把叙词法称之为“主题词法”之说。鉴于此情，本书在论及《汉语主题词表》的有关理论时，只好采取叙词和主题词二者兼用的办法。

叙词法的基本要素是叙词，其理论方法是建立在叙词性质基础上的一种检索方法。下面我们将把叙词及其基本性质的理论探讨一下。

第二节 叙词的概念性质

叙词 (Descriptor) 又称为描述词、叙述词，在国内也叫做主题词。它是用来描述图书资料主题，标引和检索图书资料的一种检索词汇。它和自然语言中的词是有区别的。它是专门为图书资料的标引和检索而设计的一种人为性的，并且是从自然语言的主要词汇中选出的标准化词汇。这种词汇首先是从它的概念性质出发，作为事物概念的表达形式而

存在的。也就是说叙词是建立在概念的基础上，每个叙词都是首先作为某一领域的一个概念而进入叙词表的。

我们知道、任何事物都具有一定的概念。概念是反映客观事物一般的、本质的特征。它是用一定的物质外壳——语言来表达的。因此，语言是概念的表达形式，概念是语言的思想内容，又是客观事物的反映。任何一种检索语言（Retrieval language），包括叙词语言都是在这种“事物、概念和语言”三者辩证关系的理论基础上建立起来的。只是叙词语言对“事物、概念和语言”的关系，较比自然语言具有更严格的一一对应的要求。

在图书资料的标引和检索中，就是用这种和一定概念内容相对应的叙词来描述主题，作为构成主题的一个个概念单元（亦称主题因素）。比如，一篇文献为《法国化学激光器》，经过主题内容的概念分析，得出“法国”和“化学激光器”两个概念单元。这两个概念单元，如果被选入词表，就必须用两个标准化规范化的语言符号来代表，这就是两个所谓的叙词。任何一部叙词表（主题词表）的选词，实质上都是对各学科专业具有检索价值和文献论述的基本概念的选取，即从选择一定概念入手——到选择一定的词来表达。这些概念所要表达的都是一个一个的事物和对象。叙词的这种概念性质，是每个叙词存在的基本条件。对“法国”和“化学激光器”两个叙词来说，如果我们抽去了它们赖以存在的概念性质，也就抽去了它们的思想内容，因而它们就将变成毫无实际意义的符号了。

叙词的概念有内涵和外延之分。所谓内涵，是指概念所反映的事物的一些本质特征，即客观上存在的本质属