

制編表詞叙 冊用手寶

(英) 琼·艾奇逊 合著
艾伦·吉尔克里斯特
侯汉清 黄刚 合译

54.23

山西省图书馆学会

叙词表编制实用手册

(英) 琼·艾奇逊 合著
艾伦·吉尔克里斯特

侯汉清 黄刚 合译

山西省图书馆学会

一九八四年六月

Jean Aitchson Alan Gilchrist
Thesaurus Construction a Practical Manual

叙词表编制实用手册

(英) 琼·艾奇逊 艾伦·吉尔克里斯特 合著

侯汉清 黄刚 合译

山西省图书馆学会出版
山西省图书馆印刷厂印刷
山西省图书馆学会发行

787×1092毫米 1/32开本 4.75印张 100千字
1984年6月出版 印数1—5000册 定价0.85元

目 次

- A 导言
- B 系统设计
 - B1 情报系统方面的考虑
 - B2 操作方面的考虑
 - B3 叙词表方面的考虑
 - B3.1 编制新词表与采用旧词表
 - B3.2 受控语言叙词表与“自由语言”叙词表
 - B3.3 叙词表构件的设计
 - B4 为各种特定的检索系统设计叙词表
- C—G 叙词表的特性
 - C 语词的标准化和管理
 - C1 名词与形容词
 - C2 单数与复数形式
 - C3 直接款目与间接款目
 - C4 拼写、大写和音译
 - C5 缩写、首字母、缩略语
 - C6 标点符号
 - C7 同形异义词
 - C8 注释和定义
 - D 专指度与先组的程度
 - D1 词汇的专指度
 - D1.1 简化词表

- D2 索引词的先组程度
- E 词表结构、词间关系和分类
- E1 词表结构和分类的因素
- E1.1 等同关系和同义词
- E1.2 等级关系
- E1.3 非等级相关词
- E1.4 词束和词集
- E1.5 分面分类法
- E1.6 标记符号
- F 辅助查准措施
- F1 组配
- F2 联号
- F3 职号
- F4 词间关系——关系号
- F5 加权
- G 专用叙词表
- G1 自然语言词库
- G2 可转换的叙词表
- H 叙词表的编排方式
- H1 字顺叙词表
- H1.1 款目设置与参照结构
- H1.2 人口词
- H1.3 字顺排列
- H2 分类叙词表
- H2.1 备有概略分类的字顺叙词表

- H2.2 备有词族图的字顺叙词表
- H2.3 备有族系表的字顺叙词表
- H2.4 备有等级分类表的字顺叙词表
- H2.5 备有概略分面分类表的字顺叙词表
- H2.6 备有详尽分面分类表的字顺叙词表
- H2.7 备有字顺索引的分类叙词表

J 编制技术

- J1 确定主题范围
- J2 确定叙词表的特性和款目设置
- J3 划分基本大类
- J4 选词
 - J4.1 手工选词
 - J4.2 机器选词
- J5 语词的著录
- J6 分类辅助措施
 - J6.1 分类的优点
- J7 现有索引语言的融入
- J8 字顺叙词表的定稿
- J9 分类叙词表的定稿
- J10 与有关学科专家共同鉴定
- J11 分类叙词表的配号
- J12 导言的撰写
- J13 编辑
- J14 试验
- J15 印制

J16 用电子计算机辅助编表

K 叙词表的修订

K1 机器修订

参考书目

本书引用的叙词表和分类表

A 导 言

《简明牛津英语词典》将手册 (manual) 解释为“手边使用的小型书籍”。本书作者正是想写这样一本简明扼要而便于实际操作的书。要尽量缩短这本书的篇幅必须解决两个问题：第一，必须略去大量的理论资料。如果读者打算进一步研究索引语言设计的基本原理，可以参阅兰开斯特 (W. F. Lancaster) 的《情报检索中的词汇控制》*以及吉尔克里斯特 (A. Gilchrist) 的《检索中的叙词表》。第二，由于篇幅有限，J章中并未将所有的叙词表编制技术都加以介绍，而只是对前述的编制理论加以实用的概述。虽然该章中前几部分讨论了诸如手工或机器划分词组、词族等技术问题，但重点却在介绍分面分析方法的使用。

导言之后的B章在此类书中似乎有些不伦不类，作者认为该章特别重要。这一章旨在说明体系设计总体方面的问题，概述了与情报检索系统及其操作有关的一些因素，介绍了不同类型的叙词表。正确地理解这些因素，并能在任何特定情况下检验这些因素，有助于读者选择一种最适用的索引语言，并对编表的工作量作出估计。

系统设计这一节之后的C—G节说明了不同的叙词表的各个组成部分及其在检索过程中的作用。此后，H节介绍了叙词表的几种类型。在说明了这些类型之后，还有

***译注：**该书已由科技文献出版社翻译出版，中译名为《情报检索中的词汇规范化》。

各种选择方案——从纯字顺的叙词表到用图表显示的叙词表。

因此,这本手册是介绍性的教科书,而不能自命提供了全部问题的答案,也许甚至未能解决大部分问题。但是,依据这些指示已足够编制一本合格的叙词表了。此外,本书也可提供一些简介,使读者解决自己的问题时能较好地理解其中的困难。

B 系统设计

在编制叙词表的工作开始之前,应先研究使用该叙词表的情报检索系统。只有明确了该检索系统的需求,才能考虑哪种叙词表在特定情况下最为有效。

B1 情报系统方面的考虑

B1.1 首先应考虑检索系统包含的主题范围。确定主题范围*的界限,明确哪些部分需深处理,哪些部分只需浅处理。

B1.2 其次应考虑被标引文献的类型。主要是单本书,还是连续出版物,如报告、期刊或会议文献?单本书一般不必象连续出版物那样详尽地标引。

B1.3 文献数量也是一个应考虑的问题。资料档的规模及文献增长的速度均应予以注意。要标引的文献资料数量愈大,输入的费用就愈少。同样,检索系统的容量愈大,就愈需要一种相当复杂的索引语言来消除误检。

*译注: Subject field有学科范围、主题范围的涵义。本书所译“主题范围”,也包括“学科范围”的意思。

B1.4 情报检索系统的类型也应予考虑。是累积多年文献的回溯性检索系统，还是最新文献通报系统，例如象定题服务系统，定期向用户提供有关资料？由于最新文献通报系统与回溯性资料档无关，不必修订以前的标引作业，叙词表就可经常变动。

B1.5 了解检索系统的用户是谁，也很重要。是由用户自己操作检索系统，还是由专门的情报工作人员操作？如果由用户操作，则索引语言必须尽量简易省时，尽量不采用复杂的方法。

B1.6 不仅要了解用户是谁，还要了解用户的数量及其定期使用检索系统的百分比，这些因素都将对检索需求发生影响。

B1.7 用户将向情报检索系统提出哪些类型的问题（定题服务系统则是提问档），是宽泛的，还是较专指的？宽泛的问题不需要太细的词表，而专指性强的问题则需要确切地限定索引词。每年查询的问题（或提问档）有多少？若提问少，则应削减输入费用而集中力量在检索上，这样比较经济有效。相反，若提问多，较经济的办法则是在编制叙词表时多花力量在输入上，在检索方面可节省一些力量。

B1.8 资料档使用何种编排方式？是在检索时组配语词的后组式，还是在标引时组配语词的先组式？前者可使用手工方式如单元词卡片、比孔卡片或者采用电子计算机方式；也可以采取电子计算机编制、手工使用的双重词表。后者的先组式词汇可以在传统目录中或机器印制的索

引中用作标题词（见F1）。

后组式检索系统的索引语言最好是相当简单的索引词，尽管也可使用较复杂的索引词。既可使用分类号码也可使用概念的词语形式。但是这种索引语言需采取措施提高查准率，以消除词间的虚假联系。

先组式检索系统的索引语言可以使用分类号码或者语词作为索引词。语词宜采用复合词，由索引员组配索引词可以减少词间虚假联系，因而不需要联号了，其它查准措施也不那么要紧了。

B1.9 最后应考虑情报检索系统的人力、物力和财力。

（1）经费来源如何？尽管这个问题常常不属于情报系统的管理工作，但却是一个关键的因素。情报机构的类型和规模大小，是政府办的还是企业或学术团体所办，这些对财源影响颇大。企业单位情报系统比其它情报系统较易受财源波动的影响。财源有限，则必须采取节约的措施。可以采用现有的叙词表，或者使用机检情报系统出的合用的磁带。如果必须搞一个新的叙词表，则可编制不太复杂的词表，即不使用复杂的防误检措施，选用简单的索引词、专指度适中、结构层次最为简单的词表，以节省费用。

（2）使用的工作人员情况如何？如果缺乏工作人员，应设计需要最少的维持费用、最少的标引工作量的词表，以节省经费开支，尤其是文献处理量大的情形更是如此。如果备有电子计算机，则可用计算机打印书本式目录和索引以节省人力。这就需要适合机器操作的索引语言。

(3) 使用的设备情况如何? 设备操作的难易将影响资料档组织方式的选择, 进而影响索引语言的设计。没有电子计算机或80列穿孔卡的小型情报系统应使用适于先组式卡片目录的索引语言。用户对手工穿孔卡熟悉、而且掌握较好, 则可为手工穿孔卡选择一定比例的后组式词汇。如果使用电子计算机, 则可用机器进行现期文献报道。在进行回溯性检索的小型情报系统中, 使用计算机编制定期的最新文献的书本式索引比批量检索可能更为经济。为特定的检索系统选择所需的最佳硬件取决于许多相互影响的复杂因素。

B2 操作方面的考虑

情报检索系统的技术操作受不同的标引和检索技术的影响(见兰开斯特著《情报检索中的词汇控制》第13; 15章), 而这些技术的效率又受制于叙词表的特性。在设计检索词表时必须牢记这一点。

B2.1 索引语言与标引工作。索引语言的特性影响索引员的工作能力。例如, 索引语言的专指度低, 索引员标引的精确度就不可能超过索引语言所允许的专指度, 于是标引的精确度就降低了。再者, 索引词定义含混、词义相互交叉, 对于同一概念, 可能标引者选择这一索引词, 检索者则选择另一个。此外, 如果入口词不够, 标引者又未使用广义词去标引, 则可能漏掉某个概念。相反, 索引语言如果展示了结构关系, 则对标引者有所帮助, 引导他使用最适当的专指语词来表达文献的主题。

B2.2 索引语言与检索工作。索引语言的特性对检索工作的影响也很大。如果等级关系和其他关系清楚地显示出来，同义词集中在一起，则可很容易地把检索者从广义索引词引到恰当的专指词。而利用族系表和相关参照又可以扩大检索的范围。索引语言可以节省检索者的精力，使其不必死死记住那些有利检索的密切相关的语词。完善的入口词表是叙词表的另一长处，可以把检索者由高度专指的自然语言概念引向情报系统所用以表达这些概念的语词；没有入口词表，检索者可能因使用不正确的索引词而白费时间，也可能误认为检索系统未收入某个概念。

专指度差的索引语言降低了检索应达到的精确度，而索引词的含混和错误会使得标引和检索无法取得一致。同样，高度先组的语词、职号、联号和参照系统（见 F）使检索程式更加复杂化，这种复杂化可能使标引和检索更难取得一致。

B3 叙词表方面的考虑

情报系统各方面的问题一俟解决，即可考虑索引语言的类型，使其在特定条件下最有效地发挥作用。在仔细考虑叙词表的功能之前，要决定两件事：一是编制一个新的叙词表，还是采用一种现有的索引语言？二是采用受控语言，还是“自由语言（Free language）”？

B3.1 编制新词表与采用旧词表。主题范围确定之后（见 B1.1）应对该主题范围内现有的大量叙词表和分类表进行考察。

美国俄亥俄州克里夫兰市凯斯西部大学收藏有各种英

语索引语言的资料*。波兰中央科技及经济研究所收藏有其它语种的索引语言及其有关资料。这两家都编印各种叙词表的一览表。英国专门图书馆与情报机构协会图书馆拥有叙表的资料档,编印了这些词表的一览表。他们打算定期修订这个一览表(见Aslib Proceedings Vol.23, No6, 1971)。对现有索引语言的鉴定应看它们是否可以:

(1) 全盘采用,只变动极少部分。

(2) 作为骨干骨架采用,其中某些学科领域须作较详细的扩充。

如果这两点都不适合(完全适用的现成检索语言是没有的),则必须有目的地编制新叙词表,以情报检索系统的需要为原则,决定其基本骨架。即使如此,也不能忽视现有的叙词表。其中有的部分可以抽出来用于新的词表。这些部分包括:通用性的有关物质、物理性质和地理区分的一览表;有关边缘学科的现有索引语言中比较专业化的部分,稍加改动后这些都可用于新的叙词表。

B3.2 受控语言叙词表与“自由语言”词库, 计算机化的情报检索系统选择“自由”语言较为经济有效。这种检索系统可以使用文献的全文、文摘或者标题。索引人员可以从文摘、题名和全文中选出自由语言的词汇来。文摘

***译注:** 国际标准化组织颁布的《单语种叙词表编制规则》第二版(ISO—2788)规定,英语词表及包括英语部分的多语种词表的国际收藏中心在加拿大多伦多大学图书馆学系的词表交换中心,其他语种词表收藏中心不变,仍在波兰中央科技及经济研究所。

刊物印刷过程中生产的磁带上带有资料全文或文摘，可以节省输入费用，在“自由语言”检索系统中，著者的用词在转译成规范的索引语言时，不会被歪曲。除非采用了“自由标引”系统，是会出现标引错误和专指度不够的问题的。合成新词可以很快地收入词表，因为新词一出现在文献的题名或正文中，或被“自由语言标引员”选定，它就成为“自由语言”的一部分了。

自由语言的缺点在于检索时比较费力。在宽泛的学科范围内必须设置同义词表、族系表乃至检索词表（见G1）以保证检索用的“自由语言”与文献一致。当然，在狭窄的主题范围内则不必用这些工具。

如属下列情况则可选择“自由语言”系统：备有电子计算机，主题范围狭窄，操作者熟悉词汇；或者是在大型机器检索系统中，有现成的机读输入磁带；或者供大量分散的标引者使用，受控语言的叙词表需要迅速更新，而经费又不许可。

B3.3 叙词表构件的设计。确定了必须编制检索词表之后，应了解叙词表的各个构件，并选定最适合该情报系统的构件，索引语言并非仅由索引词或词表构成，还有一定的“索引语言措施”（Index language devices）不同的索引语言措施对叙词表的功能具有不同的影响。这些措施可分为两类，一类用于保证检索获得最齐全的相关文献，叫做“查全措施”，另一类用以防止带出检索者不需要的文献，叫做“查准措施”。本书对这两种措施均有详细的说明（见C—G），先概述如下：

查全措施包括:

(1) 入口词表, 该表把检索者从一特定概念引向该情报系统用以表达这个概念的索引词, 以提高查全率(见H1.3)。

(2) 词形的规范(包括单、复数形式, 语法形式、词序以及其它)(见C)。用以防止漏检切合需要的文献。这种漏检是由于同一语词在不同词面形式下形成的概念分散所造成的。

(3) 同义词、准同一词的规范。将同义词引向选用的索引词以防止同义词或准同义词分散在多个标目之下(见E1.1)。

(4) 揭示结构关系、聚类关系、成族和不成族关系。通过显示关系密切的词语来扩大检索范围, 提高查全率(见E1.2—E1.5)。

查准措施包括:

(1) 加强索引语言的专指度。叙词表越详尽细致, 则词汇量越大, 而文献的主题被描述得越精确。检索系统的精确度取决于索引语言的专指度。不过也要求更熟练的标引与检索技术(见D1)。

(1) 组配。这是最有力的查准措施。标引阶段或检索阶段均可进行组配以增加词汇量。这样, 查询的概念被定义得更精确, 不需要的资料则被剔除。

(3) 语词的先组程度。这是另一种组配方式, 不过它包含在叙词表之内, 而不是操作系统中的因素。多语词组成的概念保证每篇文献的主题被精确地识别, 以防止文

献的误检（见 D2）。

（4）联号、职号和关系号。通过给相关语词分组标号，说明这些语词的职能，以消除虚假组配和错误的词间关系（见 F2—F4）。

（5）加权。该措施用以区分重要概念和次要概念（见 F5）。

有些措施，包括“提高语言的专指度”，“族系表”和“同义词规范”可以看作整个词表的一个组成部分。其他的措施如“组配”、“联号”或“加权”则独立于词表之外使用，可称为“辅助性措施”（见 F）。

查全率与查准率二者相互矛盾。这意味着提高查全率则降低查准率，而提高查准率则降低查全率。在设计叙词表时，应记住：采用某种查准措施，如提高专指度、使用复合词、联号和职号等，将会降低词表的查全功能；而降低词表的专指度，使用单一词做索引词，或者不用辅助性查准措施，则将提高查全率、降低查准率。

结构复杂、带有分类显示和专指性强的先组式词表的索引语言在输入阶段费用较大，而且要化力量进行修订。如果仅是主观估计的话，应将这种费用与各种输出的价值作比较。辅助措施，特别是职号与联号在标引和检索阶段显然都会增加成本费用。此外，结构很不完备、语词宽泛的叙词表也许比语词较专指，能显示词间关系以帮助检索的叙词表在检索阶段花钱更多。

B4 为各种特定检索系统设计叙词表

图表 1 是五个假想的情报系统的叙词表编制决策的举