

叙词表词间等级关系构建与思考

鲍秀林, 吴雯娜, 高岩

(中国科学技术信息研究所, 北京 100038)

摘要: 词间关系构建是将叙词表中彼此分散的词语链接成一个完整的、网状的术语系统的重要手段, 分为等同关系、等级关系、相关关系。其中等级关系构建是叙词表编制中最重要、难度最大、工作量最多的工作。本文根据笔者在《汉语主题词表》(工程技术版)中的编制实践, 对等级关系的构建规则及常见错误进行分析总结, 以期提高叙词表的编制质量。

关键词: 叙词表; 等级关系; 词族构建

中图分类号: G254.242

DOI: 10.3772/j.issn.1673-2286.2015.04.003

1 引言

分类法和叙词表是国内外使用最多的两种规范化的可控的检索语言。前者主要用于族性检索, 即用于用户检索指定学科或指定学科领域的信息资源; 后者主要用于特性检索, 即用于用户检索指定主题概念的信息资源。这两种检索工具的使用特性决定了它们特有的、不相同的编制规则。分类法的全部类目, 通常遵循类目的学科从属派生关系设置, 叙词表的词间等级关系(亦称属分关系、词族), 是遵循叙词的概念属性(学科属性、本质属性)构建的。并且叙词表词间等级关系构建是编制叙词表中最为重要的工作, 难度最大、工作量最为繁重。

本文是笔者根据3年多来在“科技部院所公益基金项目”《汉语主题词表》(工程技术版)(以下简称《汉表》)的编制实践, 对叙词表等级关系的构建规则、常见错误和取得的经验教训进行分析论述, 期望能为提高叙词表编制质量有所帮助。

2 构建等级关系的难度

叙词表的等级关系, 包括属种关系、实例关系等。

属种关系, 也称概念包涵关系, 是指一个概念的部分外延与另一个概念的全部外延重合的关系。这就是说, 在概念a和概念b的关系上, 如果所有的b都是a, 但有的a不是b, 那么a和b这两个概念之间就是属种关系^[1]。我国叙词表编制标准规定的等级关系主要是属种关系。在《汉表》编制实践中发现建立词间等级关系时存在如下难度:

(1) 对叙词表给定等级关系的构建原则理解不透, 掌握不准。

等级关系应该建立在一对概念间, 当一个概念的范围完全包含在另一个概念的范围时, 它应该表示二者是上位和下位的关系, 上位概念表示一个类或整体, 下位概念指的是它的概念外延^[2]。但哪个是“上位”, 哪个是或哪些个是“下位”则不易理解透彻, 因而在实际工作中出现一些错误。

(2) 正确判断叙词概念事物与方面的关系, 也具有一定难度, 易与学科分类方法构建类目上下位等级关系相混淆。

事物的整体与其部分或事物与方面之间, 在概念的外延上不存在包含关系, 因此一般不构成属分关系。如在特殊情况下, 为满足族性检索的需求, 特定的整部关系规定做属分处理, 如: ①表示行政区域、地

理区域的叙词之间；②表示人体、生物体的系统与器官的叙词之间；③表示某些组织机构之间的叙词；④表示学科及其分支，事物及其组成部分（通常用“某结构”、“某构件”、“某组件”等作为上位词）的叙词^[3]。除这四种情况之外的事物整部或事物与方面成族皆不应成立。

(3) 像《汉表》这类大型的综合性叙词表涉及的学科和技术领域广泛、专业性强，参编人员的知识素质参差不齐，受专业知识水平的限制，做到真正确切判别词间概念等级关系有一定困难。

3 构建等级关系过程中易于发生的典型错误

等级关系是叙词表区别于词典的标志之一。它主要借助上位词和下位词的划分，对词表进行组织，服务于扩检和缩检。参编人员对学科知识、主题概念以及对叙词表构建等级关系的规则理解程度的差异，都会导致在实际词族构建中的种种问题。下面举例介绍几种在词族构建中易发生的典型错误。

3.1 用学科分类的做法构建词族

受《中图法》类目体系结构的影响，存在构建词族时把分类法类目立类的等级关系与叙词法词族构建的等级关系混同起来的问题。把分类法中所列的某些类目名称，移植到族首词的等级关系中。由于在分类检索语言体系中，用来揭示文献主题内容的标识符是分类号，而不是类目名称（类名）。所显示的每一级分类号都是受学科体系分类等级控制的，而多数情况下，孤立的类名是看不出这些层次关系的。在叙词检索语言体系中，标识符就是叙词。每个叙词概念的等级，都定位在该叙词所在词族的属分关系链中，通过叙词属分关系链就可确定其级别。因此不能照搬分类法的做法来建立词族。

例1：消能

如表1“错例”中“消能”与下分的第一级词不是概念包含关系，如消能方案、消能方式、消能建筑物、消能设计。

正确的处理方式：要视对“消能”（即在泄水建筑物中，为消耗、分散水流落差的能量，防止或减轻水流对水工建筑物及其下游河床、河岸等冲刷破坏而采

取的措施）这个族首词给出的定义来取舍抉择。即应依照定义，把与“消能”有概念包含关系的词，直接属消能（见表1“正例”），其余的词按其属性分别另加处理，例如：“消能方案”属“方案”，参“消能”；“消能设计”属“设计”或属“水工设计”等；“消能方式”与“消能”可建立用代关系。

表1 “消能”错例与正例（示例）

消能——错例	消能——正例
F1 消能方案×	F1 出口消能
F2 二级消能	F1 底流消能
F2 混合消能	F1 洞内消能
F2 扩散消能	F1 洞塞消能
F2 联合消能	F1 对冲消能
F1 消能方式×	F1 二级消能
F2 底流式消能	F1 岸流消能
F2 水垫塘消能	F1 阶梯消能
F2 岸式面流消能	F1 孔板消能
F1 消能建筑物×	F1 联合消能
F2 内消能泄洪洞×	F1 面流消能
F2 消能池×	F1 竖井消能
F2 消能井×	F1 水垫塘消能
F1 消能设计×	F1 水跃消能
F2 消能机理×	F1 挑流消能
F2 消能计算×	F1 泄洪消能
F2 消能结构×	F1 旋流消能

3.2 叙词概念等级混淆

叙词的概念具有两个逻辑特征——内涵与外延。其内涵就是概念所反映的对象本质属性。外延是概念所反映的思维对象的范围。一个概念的内涵完全在另一个概念的外延之中，而且仅仅成为其一部分，那么这两个概念之间构成包含关系也就是属种关系^[4]。叙词概念间等级关系的建立需要严格遵守“全部是一部分是”这一判别规则。否则，将造成等级关系混乱。作为汉语的词，一般可能有多种含义与多种解释，因而在选作族首词时，需按照叙词法这种人为检索语言的规则，在编制词表的过程中应特别注重词义的“人为”规范处理。即组织“主题”的“聚类”，为选定的族首词“选定”其定义，人为规定其内涵与外延，以求能准确概括其下位词。而释义的选定，需视词表的需求而定，这说明了词表的人为性。

例2：建筑材料（见表2）

《汉表》是工程技术领域的一部综合性叙词表，因此“工程材料”定义为“用于机械、车辆、船舶、建

表2 “建筑材料” 错例与正例 (示例)

建筑材料——错例	建筑材料——正例
F1 保温材料×	F1 传统建材
F1 传统建材	F1 化学建材
F1 防水材料×	F2 建筑塑料
F1 隔声材料×	F1 建筑防水材料
F1 工程材料×	F2 桥面防水材料
F1 灌浆材料×	F2 屋面防水材料
F1 化学建材	F1 建筑凝胶材料
F1 混凝土材料×	F1 建筑膜材
F1 节能建材	F1 建筑腻子
F1 结构材料×	F2 防水腻子
F1 金属建筑材料	F2 密封腻子
F1 抗菌建材	F2 抛光腻子
F1 门窗材料	F1 金属建筑材料
F1 密封材料×	F1 抗菌建材
F1 木材×	F1 绿色建材
F1 墙体材料	F1 门窗材料
F1 声学材料×	F1 墙体材料
F1 施工材料×	F1 室内材料
F1 石膏建材	F1 天然建筑材料
F1 室内材料	F1 围护材料
F1 围护材料	F1 新型建材
F1 新型建材	F1 装饰建材
F1 型材×	F2 墙面装饰材料
F1 岩土材料×	F2 装饰木材

筑、化工、能源、仪器仪表、航空航天等工程领域的材料”，涵盖工程技术领域的所有材料。而“建筑材料（在建筑物中使用的材料统称为建筑材料）”仅仅是“工程材料”的外延分支，其下的各种材料需要建立在工程技术这个综合领域来考虑其概念，而不应仅仅局限于“建筑”这一领域。经分析，木材、保温材料、防水材料、隔声材料、结构材料等都与建筑材料仅仅是一种概念交叉关系，不能构成属分关系，详解见3.5节内容。

在理想状态下，作为一条完整的叙词属分关系链，是应当按叙词概念的等级关系逐级展开的。但在某些属分关系链中出现了越级属分的现象，例如“机车配件”直接列出两个十分专指的“加顶圆盘机车天线”和“加顶垂直折合机车天线”下位词，如果在文献标引中出现了“GPS机车天线”这一新概念，用已有的上位叙词和两个下位叙词都不合适，补充该新词，因词频不高也不妥当，如果有了“机车天线”就恰如其分。在越级属分关系中增补过渡叙词，是为了满足实际的文献信息标引和检索需要。这也是采取逐级属分的主要理由。

3.3 字面成族

字面成族，指将有相同词素或者单词的叙词聚集在一起构建等级关系。虽然也是一种显示概念之间等级关系的方法，而且更加直接，有关的词语标识可以聚集在一起或者附近^[5]。但是其是否成立，前方一致成族、后方一致成族两种字面成族的方式需要区别对待。

利用汉语构词原理，重心后置，后方不一致意味着概念不一致。因此，前方一致构建的词族一般是事物与方面的情况居多，除四种允许整部或事物与方面成族的特例之外，皆不成立。如表3“错例”为前方一致成族，其下位是事物与方面的关系，不能构成属分关系。正确的“交通工程”表3“正例”。

表3 “交通工程” 错例与正例 (示例)

交通工程——错例	交通工程——正例
F1 交通调节×	F1 道路工程
F1 交通对策×	F1 轨道交通工程
F1 交通载荷×	F1 机场工程
F1 交通汇合×	F1 交叉工程
F1 交通监控×	F1 交通机电工程
F1 交通评价×	F1 交通建设工程
F1 交通瓶颈×	F1 交通信息工程
F1 交通设计×	F1 交通运输工程
F1 交通特征×	F1 立交工程
F1 交通效率×	F1 索道工程

例3: 交通工程 (只显示一级下位类)

如上所述，汉语构词原理是概念重心后置，在构建词族时虽然可以考虑后方一致成族，但该种成族方法必须考虑概念的内涵与外延，只有在族首词限定内涵与外延范围内的后方一致成族才是正确的。根据这一点，可以肯定地说，对所设族首词给出“确切的定义”，并据以检验概念外延是否具有包含关系，是保证词族关系构建质量的基本规则。

例4: 线

表4“错例”是典型的用后方一致技巧的字面成族。通常，以单字为族首词构建的词族出现字面成族的几率较大。此例中，关键是未选定族首词的确切定义，违反了一词一义的原则。叙词法是以词为单元来构筑“信息处理”模式、构筑“信息检索”模式并以之作为信息识别本体的一种方式。一词一义是其核心。“线”汉语中至少有六、七种不同的内涵与外延^[6]，因此需首先选定族首词的定义，而后方才需要慎重考虑其下位词的本质

概念是否与之有概念包含关系。

如果选定族首词“线”的确切定义为：“几何学上指一个点任意移动所构成的图形”，那么表4“错例”中的“岸线”（一方以水为界的地区）、“导线”（在器具或设备内部的任何电缆、软线、芯线或导体，既非外接线，也非内装线）、“防线”（为防御敌方进攻，建立的连成片的防御阵地）、“航线”（船舶航行的路线或飞机从某一机场飞往另一机场所遵循的空间轨道）、“纱线”、“射线”、“视线”、“天线”、“明线”、“暗线”，这些用“线”作后缀的词，按限定后“线”的含义来判断均应移出该族，视下位词量情况独立成族或者做为无关联词。余下如图中的“摆线”、“包络线”、“长线”、“曲线”、“直线”等真正几何概念上的“线”，归入“线”一族，方能完全符合规则。

表4 “线”错例与正例(示例)

线——错例	线——正例
F1 岸线×	F1 摆线
F1 暗线×	F2 内摆线
F1 摆线	F2 三尖摆线
F1 包络线	F2 双摆线
F1 长线	F2 外摆线
F1 导线×	F1 包络线
F1 等值线	F2 动态包络线
F1 防线×	F2 破坏包络线
F1 航线×	F2 屈服包络线
F1 回归线	F2 头颅包络线
F1 流线×	F1 长线
F1 明线×	F1 等值线
F1 平面线	F2 等高线
F1 曲面交线	F2等距线
F1 曲线	F2等温线
F1 趋势线	F1 回归线
F1 纱线×	F1 极限线
F1 视线×	F1 平面线
F1 天线×	F1 曲面交线
F1 图线	F1 趋势线
F1 伪线×	F1 图线
F1 直线	F1 直线

3.4 事物整体与部分或事物与方面成族

关于整部关系成族，我国和国际标准上仅允许四种特例成族（见第2节），而美国标准相对而言更为开放，它对整部关系的限定较少，只要叙词间关系本质上是整体部分关系，即不予以限制，均可归入属分关系。用户在检索的时候往往不会太关注词表对整体部分关

系的限定，检索比较主观。显然，美国标准的创新较国际标准与中国国家标准更符合也更顺应检索需求。

例5: 泵结构(摘录)

表5“错例”中涉及到主体实际上是“泵”，但直接提及的是“泵结构”和“泵装置”。在这三个概念中，“泵”是吸排流体机械。“泵装置”是由泵及其附件、管路、储罐组合的总称。“泵结构”是展示泵的各构件（零部件）搭配和排列组合体。在此族中，涉及最多的是泵的零部件，还有由若干泵构件联接完成某种运动的机构，以及实现某种功能的系统。需要把这些各不相同的具体概念，仔细认真分析。

正确处理方法：“泵构件”、“泵结构”、“泵装置”，“泵零件”表示泵的结构部件零件，可以建立同义词。“泵控系统”属“控制系统”，“吸水池”属“贮水池”等，最后处理表5“正例”。

表5 “泵结构”错例与正例(示例)

泵结构——错例	词间关系	泵零件——正例
F1 半螺旋形吸入室	结构与构件(零件)	F1 泵盖
F1 泵径	结构与尺寸	F1 泵管接头
F1 泵膜	结构与构件	F1 泵管支架
F1 泵腔	结构与部分	F1 泵环
F1 泵送机构	结构与机构	F1 泵接合器
F2 泵控系统	机构与系统	F1 泵膜
F1 泵体	结构与构件	F1 泵喷嘴
F2 泵壳	结构与构件部分	F1 泵腔
F1 泵轴	结构与构件	F1 泵体
F1 泵装置	结构与装置	F1 泵筒
F2 水泵装置	装置与具体装置	F1 泵头
F3 水泵叶轮	具体装置与其零件	F2 泵头体
F3 水泵轴封	具体装置与密封零件	F1 泵叶轮
F3 吸水池	水泵装置与吸水源	F2 离心泵叶轮
F1 过流件	结构与构件	F2 水泵叶轮
F1 滑片	结构与构件	F1 泵轴
F2 铰接滑片	构件与具体构件	F1 过流件
F2 压缩机滑片	构件与具体构件	F1 配油盘
F1 配油盘	结构与配油部件	F1 平衡鼓

3.5 并列交叉概念关系的两叙词成族

设有a、b两个概念，如果a概念的外延与b概念外延只有一部分是重合的，即有的a是b，有的a不是b，并且，有的b是a，有的b不是a，那么，a与b两个概念之间的关系就是交叉关系^[2]。例2中“建筑材料”与“防水材料”，它们同属“材料”，部分“防水材料”用在建筑上，部分“建筑材料”是“防水材料”。那么，“建筑材料”与“防水

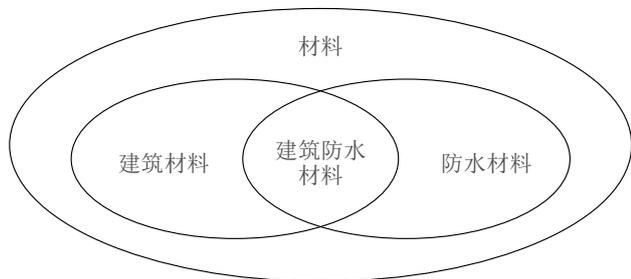


图1 “建筑材料”与“防水材料”概念关系示意图

材料”就是两个并列交叉概念。可用欧拉图表示如图1。

具有并列交叉概念关系的两叙词, 不具备概念真包含关系, 也就不能构成等级关系。如果把并列交叉概念叙词修改为具有属分关系的叙词, 需要对下分叙词特殊处理, 一种是进行专业学科限定, 如在叙词前方加“建筑”限定词, 比如“建筑防水材料”。另一种利用叙词的“注释”项作特定性的说明, 如: 给“防水材料”注释为“特指建筑用的防水材料”。

表2“错例”中类似的有“保温材料”、“防水材料”、“隔声材料”、“灌浆材料”、“结构材料”、“密封材料”、“声学材料”、“型材”、“岩土材料”, 与“建筑材料”都仅是一种概念并列交叉关系, 不能构成属分关系。正确处理方式: “保温材料”、“防水材料”等直接属“材料”, 如需要放在“建筑材料”下, 需要新增一词“建筑防水材料”, 依此类推。

3.6 随意设置“族首词”, 强制成族

编表时, 由于对编制规则理解不当, 为提高入族率, 减少独词数量, 随意设置“族首词”。事实上, 组成叙词的某种集合, 有多种方式可循: 如可以按概念属分规则方式、概念互参方式、词形前方一致方式、以及词形模糊匹配方式等。如果按想当然的方式把一些零散的、按属分规则成族有困难的、在概念上有点相关词语, 强拉硬拽聚集在一起, 为成族而成族, 就不太合适。如强做族首词的词语有“XX工艺”、“XX机理”、“XX结构”、“XX特性”、“XX原理”、“XX理论”等。事实上, 编制规则中允许没有属分关系的独词存在。

例6: 试验工艺

表6中需要考虑其下分词是否是各种具体的“XX工艺”。事实上, 经网上搜索, 该族首词没有找到相关的定义或释义, 概念不是很清晰, 只是有少量题名中含

表6 “试验工艺”错例(示例)

试验工艺×
F1 实验标准差
F1 实验开发
F1 实验模拟
F1 实验评价
F1 实验修正
F1 试验标准
F1 试验程序
F1 试验分析
F1 试验工况
F1 试验环境
F1 试验力
F1 试验模态
F1 试验设计
F1 试验应用

有该词, 这里的“工艺”基本所指的是“程序、规程、流程”。若以此为依据, 很多下位词就不符合要求。“试验应用”的含义难以琢磨, 网上用完全匹配的方式搜索, 仅有一本书译名有此概念: “《试验应用统计: 设计、创新和发现》——阐述了统计方法在试验设计中的应用”。如此, 经分析, 该族正确的处理方式: 如“试验工艺”概念成立, 可考虑归入“工艺”族下。其他下位词, 基本都有相应的词族可入, 如“标准差、开发、模拟、模态、评价、修正……”。

3.7 多属性叙词处理不当

按编制规则要求, 对具有多重属性的复杂概念应归入多个相应的词族, 或在一个词族中具有多个上位关系。表7“错例”中“数控重型卧式车床”一词, 对于“车床”而言, 具有四重属性, 即“数控、重型、卧式、车削”, 因此该词应与具有上述属性的相关概念, 建立相应的等级关系。正确的处理方式是该词可以在“重型车床”、“数控车床”、“卧式车床”下位重复反映, 实现多属(见表7“正例”)。

例7: 车床(摘录)

4 结语与思考

通过《汉语主题词表》(工程技术版)近三年的编制实践, 为保证正确的构建叙词表的等级关系, 有以下三点值得思考:

- (1) 充分、有效地进行人员培训

表7 “车床”错例与正例(示例)

车床——错例	车床——正例
F1 仿形车床	F1 仿形车床
F1 精密车床	F1 精密车床
F2 超精密车床	F2 超精密车床
F1 立式车床	F1 立式车床
F2 数控立车	F2 数控立车
F1 落地车床	F1 落地车床
F1 普通车床	F1 普通车床
F1 数控车床	F1 数控车床
F2 车削中心	F2 车削中心
F2 经济型数控车床	F2 经济型数控车床
F2 数控立车	F2 数控立车
F2 数控卧式车床	F2 数控卧式车床
F1 卧式车床	F3 数控重型卧式车床
F2 数控卧式车床	F1 卧式车床
F2 重型卧式车床	F2 数控卧式车床
F3 数控重型卧式车床	F2 重型卧式车床
F1 重型车床	F3 数控重型卧式车床
F1 专用车床	F1 重型车床
F1 自动车床	F3 数控重型卧式车床
F2 半自动车床	F1 专用车床

编制人员充分了解叙词法的基本原理,确切地理解、把握词表编制规则中的“等级关系建立方法”,而不能凭想当然的方法构建等级关系。在词表编制之初,需要充分、有效地进行人员培训。通过上机实际练习、改错加深对编制规则的深入认识和理解。

(2) 对选词质量进行控制和协调

对于综合性词表采用分工协作方法编制时,需要对分编词汇有效地实施质量控制和协调,严格把握叙词的检索意义和词义的准确性,对词族构建过程中出现的共性问题,需要有效地加以研究、分析。对于分编

叙词及构建关系需要不断审核、讨论,确保整表的一致性。在编制过程中发现普遍性的问题要及时研究讨论并作出决定,随即通报参编者要求检查改正。

(3) 要充分发挥叙词表专家的作用

对于成员单位编制的词族,必须经词表专家把关,做到每一个词族审核通过才可提交数据。

总之,我们应该在充分理解编制规则的基础上,认真对待每一个叙词,对收留叙词的概念应有大致的了解。对于不熟悉的叙词,需要勤查勤问,充分运用字典、词典以及网络等工具,还可借助编表专家们的经验,多角度反复查证,在充分弄清叙词概念的前提下建立等级关系。如此,才能建立正确、完善的等级关系,保证词表质量。

参考文献

- [1] 郭永良. 从属种关系看划分、限制与定义[J]. 自然辩证法研究, 2000, 16(zl):50-52.
- [2] ISO 25964-1, Information and Documentation Thesauri and Interoperability with Other Vocabularies Part1: Thesauri for Information Retrieval[S]. GENEVA :INTERNATIONAL ORGANIZATION FORr STANDERDIZATION,2011.
- [3] 中华人民共和国国家标准. GB/T 13190—1991汉语叙词表编制规则[S].北京:中国标准出版社,1992.
- [4] 余军成. 概念及其内涵和外延述评[J].安徽文学(文教研究), 2007(1):121.
- [5] 张琪玉. 字面成族原理与应用[J].高校文献信息学刊, 1997(3):3-5.
- [6] 中国社会科学院语言研究所词典编辑室.现代汉语词典[M].修订本.北京:商务印书馆,1996:1368.

作者简介

鲍秀林,女,1982年生,硕士在读,研究方向:知识组织, E-mail: baox1926@163.com。
吴雯娜,女,1973年生,硕士,中国科学技术信息研究所副研究馆员,研究方向:知识组织。
高岩,男,1981年生,本科,工程师,研究方向:数据挖掘。

Study on Construction of Hierarchical Relationships in Thesaurus

BAO XiuLin, WU WenNa, GAO Yan
(Institute of Scientific and Technical Information of China, Beijing 100038, China)

Abstract: The construction of term relationships, which can join each scattered terms into a full mesh terms system in thesaurus, is an important means. There are three relationships in thesaurus: equivalence relationship, hierarchical relationship and associative relationship. Moreover, hierarchical relationship is important, although it is difficult to compile. In this paper, according to the writer's compiling experiences of Chinese thesaurus (engineering technology edition), compilation rules and frequent faults are analyzed, in order to improve the quality of compiling in thesaurus in the future.

Keywords: Thesaurus; Hierarchical Relationships; Construction of Generic Structure

(收稿日期: 2015-03-15; 编辑: 雷雪)