

“字音制”純数字汉语拼音著者号碼法(表)初探

朽 予

一 本表的基本編制原則

本表的編制在主观上力求达到三个目标: 1. 号碼力求簡明, 容量力求广大; 2. 使用方法力求簡明了, 便于讀者使用, 便于館員掌握; 3. 有一定的伸縮性、灵活性, 可供大中小型图书馆机动运用。

本表的基本編制原則是: 1. 取消著者号碼第一位的拼音字母, 采用純数字制。大型館使用四位数字, 自 0000—9999 共 10,000 个号碼, 中小型館使用三位数字, 自 000—999 共 1,000 个号碼; 2. 彻底地以字音为基础, 摆脱同音異調、同調異字的束縛, 对“李黎里历”等同音姓氏不加区别, 而只按姓氏音节的使用頻率各分配以相当的号碼。本表在取号时必须依据固定的号碼表, 所以它基本上属于“查号法”系統。但它所采取的“字音制”原則, 則吸取了“拼号法”的一些优点。

現在就来做一些具体的說明。

(一)《卡特表》以及据此原則編制的俄文《哈英金娜表》、日文《植村长三郎表》等, 都是采用字母数字混合制, 采用著者姓氏的首字母作为著者号碼的第一位符号, 我国汉语拼音著者号碼表(以下简称 HP²)的編者們差不多也不約而同地学习了这一点。这样做, 有一个很大的好处, 即把著者号碼的第一位符号由 10 进位改成了 26 进位, 号碼表的容量扩大了一倍以上。但从使用方面着想, 却不如采用純数字制好。我們不要只着眼于号碼表本身, 而要綜合考虑: 分类号中已有了一個拼音字母, 著者号碼再使用一个拼音字母, 而第二著者号、版本号、著作号和其他附加号, 有时也还得使用一、两个拼音字母, 这样一来, 一个索书号中就要有二、三个拼音字母, 这在图书馆員的头脑中必須时刻縈迴着两种不同序列和不同书写形式的符号体系, 在书籍和卡片的排列工作中无疑地要增加精神活动的强度, 降低工作效率, 从讀者要求方面讲, 书号單純些总比复杂些好。因此, 本表才决定采用純数字制。当然, 取消了首位字母以后, 著者号碼表的容量会縮

小一半以上, 这对編表确有困难, 也許正是因为这一点, 現有的各家 HP² 才都走了字母数字混合制的道路。本表就是试图解决这个问题的。

(二) 汉语的最大特点之一, 就是元音和輔音之結合的規律性。汉语每一音节的开头絕大多数都是一个輔音字母, 两个的只有三組, 再一小部分是开头部分根本没有輔音。元音的組合形式, 包括以 n、ng、er 結尾的在內, 总共也只有三十六种。同时, 并不是所有輔音都可与一切元音和元音組相拼, 在現代汉语中, 元音和輔音的拼合, 包括摹声性的汉字在內, 也只有四百一十一个音节。这里有不少是根本不能或极少可能作姓氏使用的音节。这个特点帶給我們的困难, 是同音字多, 同音姓氏多, 常用姓氏音节的使用頻率高。但是这个特点, 如能正确对待和利用, 却是一个很大的优点。我們編制 HP² 所面临的对象, 就是这三、四百个音节, 由此編出的号碼表也必然非常簡明。欧洲人的姓氏能有多少种呢? 誰也无法肯定。譬如, 我們可以規定汉语的 Wen 是 W 48, Weng 是 W 49, 两者之間再沒有其他姓氏了, 而西語的 Wen、Weng 之間, 誰知道会有多少 Wener, Wender……呢? 所以, 卡特制只能分配給各个字母排列区間以一定的号碼, 一个姓氏一个号碼問題, 只有汉语才能作到。在前一段时间內, “拼号法”的 HP² 出現了很多, 就是企图利用汉语这一特点的結果。“拼号法”在实际使用上的不灵活, 是无法补救的缺点, 所以很多人不贊成走这条路。但“拼号法”注意到了汉语語音的特点, 并且主要着眼于字音这一点, 却是值得注意的。(当然不是所有“拼号法”HP² 都是如此。)

另一方面也得承认, 汉语的另一特点——声調問題, 也給我們帶來了一定的困难; 同样的拼音字母书写形式, 因声調不同就代表两个不同的姓氏, 甚至讀音和拼音书写形式完全相同的音节, 人們却要认为它們可能是两个姓氏, 如历、栗等。对待这两个問題必須有所区别, 并且要从发展的观点来看。現在我們的头脑中, 总是把 Zhang 这个音节与“弓

长張、立早章”等联系在一起，等将来汉语走上完全拼音化的道路以后，人们的观念是会逐渐改变的。到那时就不会再去区别不论在发音或书写形式上都变得毫无区别的汉字姓氏了，到那时不但張三和張四是同姓，就是章五也必会变成“同宗”。至于同音異調的姓氏，是否将来在书写形式上都要使用調号区别的問題，目前学术界还討論得不多。（这个問題不但是姓氏問題，而且牽涉到是否一切专有名词都要加注調号的問題了。）但不論標調与否，在排檢上一般的是不会照顾四声的。我們应当从卡特制中取得必要的借鉴。卡特制只給一定的字母排列区间分配号码，而我們对书写形式完全相同、仅只声調有别的姓氏却要坚决划分界限，实在毫无必要。为了解决重号問題，我們可以寻求別的方法，不必在同一音节內还要严加区别。现有的各种 HPZ 就是因为死守这一原則，結果給号码的分配造成了很多困难。如果依据“字音制”的原則来編制 HPZ，将会有如下几个优点：

（1）从字音出发，不受字形牽掣，符合汉语拼音化的方向，符合拉丁字母的国际排檢习惯，有利于对目录的組織和利用。依音序組織目录，是我們坚定不移的方向，問題仅在于迟早；目前的館已开始实行了。依音序排列，首先必须是按字母順序排列先后，很难想像先依字母，再依四声，一个音节一个音节地比較着去排列，当然更不会考虑到这个音节代表的是哪个汉字，它是多少笔画、什么笔順等問題了。目前的某些“拼音音序排列法”，如依各个汉字首一拼音字母次序的排列法，或依拼音和四声而又兼顧汉字字形諸因素的排列法等，都只是一种过渡的方法，或者仅适用于某些特殊的需要，不是彻底的音序排檢法，不能成为我們的基本排檢制度。

（2）实行“字音制”，可以打破号码分配上同音異字、異調的界限，互相調剂，灵活使用，从而大大提高号码的利用率。“字形制”著者号码表必須为一切可能用作姓氏的汉字分配号码，即使某些使用頻率极低的字，也不能完全摒置不顾。这就形成了一种很不合理的現象：常用姓氏音节的号码有时不够用，非常用姓氏却占据着大量的号码而很少使用，甚至完全不用。现有的几种“字形制”著者号码表大体都是如此。为了說明問題，現在我們且引袁涌进、周树基两位先生合編的 HPZ 为例（以下簡称“袁表”）。因为它受着“字形制”的限制，所以在号码

分配上不够理想：如“馮”仅占 Feng 音节的 $\frac{6}{16}$ ，“李”占 Li 音节的 $\frac{6}{12}$ ，“馬”占 Ma 音节的 $\frac{6}{11}$ 。其他“字形制”的著者号码表这种情况就更为突出了。北京矿业学院刘德人先生的 HPZ 中，“馬、李”等字仅占 Ma、Li 等音节的 $\frac{26}{81}$ 。事实上除去“范、馮、李、馬、刘”等之外，“丰、逢、楓、風、樊、番、反、翻、里、黎、理、立……”的使用率都是很不高，只要翻翻各館著者目录，立刻就能发现这种情况。

为了对各个音节以及各汉字姓氏的使用情况有个較清楚的了解，我曾对北京大学图书馆的著者目录作了一个粗略的調查。下面就用調查統計資料和“袁表”的具体情况做一些比較研究：

我发现“袁表”的号码分配与北大館的著者目录統計材料出入是很大的。如果北大館采用“袁表”編制著者号码的話，情形将是这样：“袁表”所收的 2,838 个汉字，北大館只用到了 1,248 个（此外有 240 个字是“袁表”所沒有的），即有 56% 以上的字是用不到的。“袁表”共有 1,620 个号码，然而竟有 505 个号码根本沒用到，占全部号码的 31% 以上。在其余的 1,115 个号码中，使用頻率极不平衡，有 305 个号码的卡片数在 5 張以下，有 149 个号码的卡片数在 15 張以下。这两項合計 454 个号码，占总号码的 28% 左右。北大館著者目录共有卡片 90,692 張，減去著作特別多的个人和团体著者的卡片 13,771 張，实余 76,921 張，如果把 5 張卡片以下的 305 个号码各按 3 張計算，把 15 張以下的 149 个号码各按 10 張計算，則两者合計为 2,405 張，占卡片总数的 3.1%。在 1,620 个号码中，31% 的号码閑置不用，28% 的号码只負担了卡片总数的 3.1%，其余的 41% 的号码却負担着卡片总量的 96% 以上，而在这 41% 的号码中（共 661 号），还有 64 个号码負担畸重：每号負担卡片 200—299 張者有 37 号，300—399 張者有 18 号，400—499 者 4 号，500 以上者 5 号，如将上述各号都分別按包含 250, 350, 450, 550 張卡片計算的話，則它們共含 $250 \times 37 + 350 \times 18 + 450 \times 4 + 550 \times 5 = 20,100$ 張卡片，这就是說占总号码 3.9% 的 64 个号码負担着总卡片量的 26% 以上。

“袁表”如此，其他的“字形制”著者号码表也不例外，甚至有的还更严重些。譬如有的坚决排斥“許多字占用一个号码”的原則，要“基本上一字一碼”。为了避免可能出现号码表上沒有的字，竟收录了四千多汉字，并且还“预留空位，以备扩充”。可以想

象得出,在这种情况下,号码闲置的现象必然更为严重。如果一定不放棄“字形制”原則,束縛在汉字和四声的框框里,这种弊病是无法避免的。汉字常用姓氏和非常用姓氏的使用频率相差太悬殊了,簡直使你无法根据其使用频率的比例来分配号码。以 Li 音节为例,“袁表”共收汉字 19 个,分配号码 12 个,其中“李”姓分到号码 6 个,而北大館著者目录中 Li 音节有 3,143 張卡片,其中仅“李”姓就占了 2,760 張,其余 18 字总共不过 383 張。再以 Yi 音节为例,“袁表”收字 30 个,分配号码 8 个,北大館实用字 33 个,共有卡片 582 張,但除了“依、伊、易、一”四字外,其余 29 字才仅有 130 張卡片。如果严格按比例分配号码的話,在 Li 音节的 12 个号码中,“黎”等 18 个字只应分配 1.5 个号码,在 Yi 音节的 8 个号码中,“翼”等 29 个字还占不到 2 个号码。很明显,哪一个“字形制”的著者号码表也不能这样做,否則,哪儿还有什么汉字和四声的区别呢?

当然,如果常用姓氏所拥有的号码已經足够多了,那么非常用姓氏多占些号码也还是无可非議的,但事实上在一些含著作数量較多的类目中,常用姓氏的同姓著者名字的首一字母重复的情况是很严重的。——各种 HPZ 的号码分配的原则和方法不同,各个姓氏所分配到的号码多少不同,因而不能抽象地确定在何种情况下便要造成重号,所以我們在这儿只能研究“著者姓氏相同、名字首一字母也相同”的情况。对于絕大多数 HPZ 来說,名字首一字母相同都要造成重号。(以下把这种情形簡称为“重名”,以表示它与“重号”这一概念約略相当而又有所区别。)

从表面上看来,一般 HPZ 对常用姓氏至少也分到了 26 个号码,大概已够用了吧?事实不然,著者名字首一字母的出現很不平衡,很不規律。以北大館的分类目录为例,在“中国小說——建国后的个人著作”这一类中,“高”氏有著者 9 人,但其中高玉宝、高云覽、高源、高延昌四人的名字首一字母都是 Y,高树文、高士心、高参三人的名字首一字母都是 S,在 9 人中竟有 7 人相重。在“李”姓中:李翼、李月潤、李英儒、李养正、李逸民——Y;李先偉、李新民、李晓——X;其他:W、L 各有 3 人,M、G、Z、Q 各 2 人,31 人中重 14 人。在其他姓氏中,“王”姓 33 人重 17 人,“周”姓 10 人重 5 人,“張”姓 16 人重 6 人,“楊”姓 18 人重 6 人,“刘”姓 13 人

重 5 人,“陈”姓 13 人重 5 人……。其他类目虽好些,但此种现象也不少。各种 HPZ 应付这种困难的办法,不是大量运用附加号(著者号将达 5 位以上),就是改依其他标准取号(这意味着对原定的复分标准的破坏,著者号码将与著者姓名的拼音字母排列不一致。)—这两种办法都不是很好的。最理想的办法,就是再大大扩展一下常用姓氏的号码。

另一方面,就整体看来非常用姓氏所占号码似乎很多,但就每个字来看,所分到的号码仍是很少的。它們的著者数目偶而多了一点之后,其本身的号码也还是不敷应用。如果我們不分“李、黎、立、里、馮、逢、楓、丰……”,一律把它們作为 Li、Feng 来对待,使它們拥有原来各姓氏的那样多号码(比原来的号码少一些也沒关系),把“立高”(Li G)“李古北”(Li G)这样的“重姓重名”問題就象处理“李高”、“李古北”那样来解决,前述弊病便可消除了。这样在一般情况下,常用姓氏有較多的号码可用,而非常用姓氏偶而出現較多的著者时,也可以有大量的号码可用,两者灵活使用,非常便利。

(3)“字音制”不是为几千汉字,而是为三、四百个音节分配号码,这就使号码表大大簡化,它的总量可以大大縮小,有利于編目員熟练掌握,提高工作效率。贊成“拼音法”的同志們批評“查号法”最多的地方,就是它永远离不开一本号码表,时时翻檢,影响編目速度。現在,如果我們按照一些同志所說的那樣,把这个簡短的号码表印成一張表,置于办公桌上的玻璃板下,在編目取号时;向下一看,应当說是不費事的。固然不如“拼音法”几条規則更簡明,但总不会因而“降低工作效率”吧。

(4)“字音制”不区分四声,能比較适合普通話的逐步推行的情况,有利于讀者使用按音序排列的目录,因而也就能促进我們的目录更快地向音序排列的方向发展。使几亿人民都能讲一口标准的普通話,这不是短期内所能达到的目标。方言区的人,正确地讀出汉字的音来比較容易些,但完全讀准四声,还是比较难的。我們常可見到,有些普通話已讲得相当好的方言区的人,对許多字的声調仍然讀不准。在生活中,讀音有点不准,关系并不大,但在目录排列上却絲毫馬虎不得,一張卡片的排列次序,因一字一調之差,就可能“失之毫厘,差以千里”。我們必須預見到这点。如果不計器声,这个問題也就附帶地解决了。(未完)