

文章编号: 1003-0077(2013)02-0074-05

## 关于两个汉字部件规范的一点思考

王道平<sup>1</sup>, 黄文丽<sup>2</sup>

(1. 西华师范大学 文学院, 四川 南充 637002;

2. 中国地质大学 机械与电子信息学院, 湖北 武汉 430074)

**摘要:**《信息处理用 GB 13000.1 字符集汉字部件规范》和《现代常用字部件及部件名称规范》这两个文件对规范汉字部件具有重要意义。但在实践中, 它们存在着部件过多以及没有可行的拆分规则这两个最突出的问题, 其原因是规范的制定缺乏系统和深层次的思考。要从根本上解决这些问题, 就必须重新审视其中的拆分规则, 将拆分规则和部件制定融为一体, 并立足于有利于汉字输入、汉字教学和汉字检索这三者统一的层面来完善汉字部件规范。实验证明, 这样做行之有效。

**关键词:** 汉字部件; 拆分规则; 汉字输入; 汉字教学; 汉字检索; 规范

中图分类号: TP391

文献标识码: A

## Thoughts about Two Chinese Character Component Standards

WANG Daoping<sup>1</sup>, HUANG Wenli<sup>2</sup>

(1. School of Chinese Language and Literature, China West Normal University, Nanchong, Sichuan 637002, China;

2. Faculty of Mechanism & Electronic Information, China University of Geosciences, Wuhan, Hubei 430074, China)

**Abstract:** *Chinese Character Component Standard of GB 13000.1 Character Set for Information Processing and Specification of Common Modern Chinese Character Components and Components Names* are two important documents for Chinese character component standardization. In practice, they are defected in hard-to-remember large number of coding components and lack of convenient and reliable disassembly rules. To deal with these issues, we need to re-exam the component disassembly rules, combine them closely with components, so as to improve the standards for the joint benefit of Chinese character input, teaching and indexing together.

**Key words:** Chinese character component; disassembly rule; Chinese character input; Chinese character teaching; Chinese character indexing; standard

### 1 引言

国家语言文字工作委员会于 1997 年发布了《信息处理用 GB 13000.1 字符集汉字部件规范》<sup>[1]</sup> (以下简称《规范》1), 针对 GB 13000.1 字符集中的 20 902 个汉字, 制订了包含 560 个部件的《汉字基础部件表》, 给出了汉字拆分的原则和方法, 对引导汉字形码输入法向尊重汉字的形和义方向发展具有重要意义<sup>[2]</sup>, 也为后来的众多形码输入法研究提供了宝贵参考, 大大缩短了其研究进程。但于实际应用

中, 它在一定程度上使本来混乱的局面更加难以收拾。时隔 12 年, 国家语言文字工作委员会又于 2009 年发布了《现代常用字部件及部件名称规范》<sup>[3]</sup> (以下简称《规范》2), 针对现代汉语 3 500 个常用汉字制订了一个包含 514 个部件的《现代常用字部件表》, “纠正”了《规范》1 中的某些错误, 但又不恰当地继承了《规范》1 的“交重不拆”等“失误”, 将《规范》1 的负面影响, 从汉字输入领域进一步扩大到了汉字教育和辞书编纂等领域。本文撇开两个《规范》的贡献不谈, 重点讨论它们在实际应用中最突出的问题、产生的原因及其解决方法, 旨在同时请

收稿日期: 2011-11-08 定稿日期: 2012-06-18

作者简介: 王道平(1977—), 男, 硕士研究生, 主要研究方向为汉语言文字学和计算机辅助教学; 黄文丽(1985—), 女, 硕士, 主要研究方向为信息工程、知识产权法。

教有关汉字信息处理和汉字教育方面的专家,将汉字信息处理和汉字教育结合起来,进一步促进《规范》的完善。

## 2 两个规范的不同之处

对比《规范》1 的核心部分第 3、5 部分和《规范》2 的核心部分第 3、4 部分,即两者的术语和拆分方法部分,不难发现它们略有不同。

在术语部分,《规范》2 略去“笔形”、“笔顺”、“笔数”、“字形”条目,增加了“单笔部件”和“部首”的概念,这无足轻重,虽然“部首”概念的新增有将汉字部首融入部件、让部件规范为汉字教育服务的意图。值得注意的是,对“基础部件”的定义,将“最小的不再拆分的部件”修改成了“最小的、按照规则不再拆分的部件”。那么满足什么条件的部件才算“最小的不再拆分的”呢<sup>[4]</sup>? 没有阐述。为此,给出一个“按照规则”来进行“澄清”。但又按什么规则呢? 还是没有阐述。结果《规范》2 较之《规范》1 更让人无所适从。

在部件拆分部分,《规范》1 提出,“对多部件的汉字进行拆分时,应先依汉字组合层次做有理拆分,直至不能进行有理据拆分而仍需拆分时,再做无理拆分。”这里一是要求“层次拆分”,二是允许最后“无理拆分”。汉字拆分最终是要拆分成部件序列,是线性的,拆分正确也就意味着“层次正确”,在这个过程中强调层次显然没有必要,这也恐怕是《规范》2 剔除这一要求的原因。《规范》2 同样允许“有理拆分”,但它以“拆开后的各部分均为非成字或均不再构成其他汉字的,不拆分”来限定之。只是,我们不禁要问,这“均不再构成其他汉字”,一般人是怎么知道的? 可见,其操作性存在问题。

另外,《部件》1 中所说的部件表中的部件不得组成非成字部件使用也是多此一举,因为谁会拿着这些部件组成一个连电脑都显示不出来的非成字部件呢? 倒是部件表中的部件不得再行拆分这个规则比较直截了当而且有用,可惜《部件》2 却默认了。而《规范》2 提出了“因构字造成基础部件相离的,拆分后仍将相离部分合一,保留部件原形”,并指出“裹”拆分为“衣果”,这似乎是在向部首“示好”,有和部首“并轨”的意图,因为“裹”曾经的部首就是“衣”。可惜,新的汉字部首归部规范已然将其归入“亠”部了。

总之,两个《规范》存在着诸多问题,但最突出的

问题是部件数目过于繁多,以及没有给出行之有效的拆分规则。

## 3 实践中最突出的问题

两个《规范》最突出的问题,恰恰就是两者的相同之处。

### 3.1 拆分规则难操作

两个《规范》都在部件拆分部分指出,字形符合理据的要进行有理据拆分,无法分析理据或字形与字理矛盾的依形拆分,并且都“相交不拆”。那么,什么是理据呢? 两个规范亦指出,根据字源或参考字源,从汉字的部件组合中分析出来的造字意图,就是字理。并举例说像太阳(日)从地平线(一)升起是“旦”的理据,而从“木”、“加”声是“架”的理据。那么,这里有一个问题:“章”是要拆分成“立日十”还是“音十”呢? 根据从“音”从“十”的字理,应该拆成后者,但据两个《规范》来看,却要拆成前者,如何是从呢? 原因在于,两个《规范》的部件表中都没有“音”这个部件。但从字理角度看,“音”属最小义符,“章”拆成“音十”是并非“无法分析理据”的。加之其构字能力强、所构成的汉字常见,就更应该成为一个部件,而不宜再拆成“立日”。再比如,“元”如何拆分? 是要拆成“一兀”还是“二儿”? 对一般用户而言,无论从形还是从理拆分都不太好解释。另外,“交重不拆”的规定,直接导致部件过多,使建立在部件基础上的拆分规则更加难以执行。因此,两个《规范》提出的部件拆分规则,只是比较笼统的原则,不能避免拆分的随意性,使拆分难以适从,不具有唯一性。

### 3.2 部件过于繁多

《规范》1 中给出的基础部件多达 560 个,要通过 ASCII 键盘上的二三十个键位,用 500 多个部件来为 GB 13000.1 字符集的 20 902 个汉字编码虽然并不难,但要用户准确和熟练地使用由此产生的输入法,就太难了<sup>[5]</sup>。作为 3 500 个常用汉字的部件,其数目竟也达到 514 个,似乎更让人难以接受。数量本身庞大,加之作为部首而非部件的汉字以及非部首的独体字<sup>[6]</sup>的干扰,人们很难记住这些部件。另一方面,当字符集扩充至 GB 18030-2000 乃至更大的字符集时,许多新增的字要拆分,按照“交重不拆”的原则,势必会产生新的部件。如此一来,部件

的数目就会更加庞大,对用户而言就会使部件表更无可操作性。实践表明,那些曾努力符合这一规范并被全国信息技术标准化委员会或全国中小学计算机教学研究中心向全国推荐使用的表形码、自然码、认知码、表音码、郑码<sup>[7]</sup>,以及后来产生的“千军万码”,几乎都销声匿迹了,反倒是一直不合“规范”的王码五笔 86 版在输入法市场中还占有一席之地,这虽然在一定程度上归因于历史,但近 200 个字根相对于 560 个部件,确实有其先天优势。自《规范》1 制定的 14 年来,无论是国家支持的还是个人自主研发的其他形码输入法,在应用上都没能真正超越难学难用、饱受诟病的王码五笔字型输入法,这一事实足以证明《规范》没有起到预期作用。

#### 4 突出问题产生的原因

《规范》在实际应用中出现诸多问题,其直接原因就在于“交重不拆”这个规则的制定和沿袭。“交重不拆”限制了重叠笔画的拆分,在一定程度上保障了汉字字形的完整,比如“果”不拆成“田木”、“串”不拆成“中中”,在汉字编码学界也引起人们对汉字拆分合理性的广泛关注,但在绝对层面上要使汉字拆分达到“最简单化”,则最终的结果是导致了部件数目过于庞大。面对当时混乱的“万码奔腾”局面,研制者“因噎废食”,来个“一刀切”,认为只有汉字绝对地“交重不拆”才能根治“相交乱拆”的混乱,未免过于心切。“交重不拆”的制定只是迫于形势压力的权宜之计,并没有经过系统的实践和论证。追求目的的手段有多种,有的可以达到,有的无法达到,而“交重不拆”恰恰是无法达到目的的手段。14 年来的实践已然证明“交重不拆”不仅没有制止混乱,反而在某个程度上“乱上添乱”。可见,其深层的原因是,《规范》的整个系统比较零散,其制定应该从更高的角度,至少要立足于汉字输入、汉字教育、汉字检索这三者统一的层面来看待部件的制定和部件拆分规则,更应该将部件的制定和拆分规则紧密地融为一体,从而制定一个可行的拆分步骤,达到汉字拆分的唯一性。从这个意义上讲,汉字拆分虽然乱象横生,但“相交拆分”并非罪魁祸首。

#### 5 解决问题的要求和方法

汉字的字形处理,无论是汉字输入、汉字教育还是汉字检索,在实际应用中几乎人人都遵循着“客观

存在一个作了定量和取码归并处理的部件表图和一个切实可行的拆分方法——按拆分方法选用部件将汉字拆成部件序列——按部件取码来编码汉字”这么一个规律。那么在部件定形定量和拆分规则的制定过程中,必须看到它们彼此的内在联系,将彼此紧密结合起来而不能割裂开,这是解决问题的关键。汉字拆分涉及拆分对象、拆分方法和拆分所用部件三个方面,它是运用拆分方法并选用部件来拆分汉字的过程。离开部件表中的某些部件,拆分方法就难以始终执行下去;离开拆分方法,也无法运用部件表中的部件实现唯一拆分。在汉字拆分过程中,拆分方法应该是一定的,而针对具体汉字的部件选取会呈现必然性的变化,这就意味着,部件的制定在一定程度上更依赖于拆分方法的制定。“交重不拆”规则的制定导致部件数目过于庞大,便是有力佐证。

“交重不拆”的主观愿望是好的,是为了使拆分直观、简易,但实际上它只是造成了使拆分最简化的假象,满足了人们“拆分必须直观、简易”的心理需求,而没有将“直观、简易”控制在可操作的范围内,致使全面失控,最后既不“直观”也不“简易”。过犹不及,因此汉字拆分既要求直观简易性,也要求对直观简易性必须有一个“度”的把握。“交重不拆”的结果是将诸如“串单电果象豕世事甩禹庸”这些容易被其他部件拆分且拆分结果一目了然的汉字都纳入到部件中来。而一味地进行“交重不拆”,把它们当作部件拿来教学,既不便于汉字教学<sup>[8]</sup>,也只会僵化初学汉字之孩童的大脑,无益于其智力开发。相反,“适当”地“相交拆分”,不仅有利于汉字信息处理,更便于汉字教学,有助于培养青少年的“立体”思维。因此,汉字拆分直观简易性的“度”就是要适当地进行相交拆分,也即是“相交拆分”要有一个“度”,它是和部件的制定紧密结合在一起的。它可以具体描述为:①不违背汉字笔画的形状和数量;②尽量不违背汉字笔画顺序;即使违背,其笔顺也要一目了然;③拆分后的部件还原成该汉字时要比较简单直观;④不为难其他汉字或汉字部分的拆分。

从以往经验看,汉字是平面结构的,很少有人从立体结构的角度来理解或阐释它。首先可能是因为绝大部分汉字都可以在平面上“交重不拆”地一分为几,其次可能是以往相交拆分的实践严重地摧残了汉字,更使人们坚定地认为汉字是平面结构的,惟有“平面”拆分才有利于汉字;再者恐怕是受韩文等的影响。然而,相交拆分并不一定会践踏汉字。如“秉”,从字源来讲就是“禾”和“𠂇(又,即‘手’)”的组

合,无论是依理还是依形都易于拆成“禾𠂔”。在部首检字法中,“肅肅”等字不也是进行“相交拆分”而取“𠂔”部<sup>[9]</sup>吗?立足于汉字信息处理,GB 13000.1 字符集汉字的部件既已达到 560 个,超大字符集<sup>[10]</sup>是不是还要增加几百个?如此“庞大”的数目只会严重影响汉字输入、汉字教学和汉字检索等问题的综合解决。因此,必须从利于汉字健康发展的角度,有选择地对相交对象进行相交拆分。

换个角度来看,汉字的笔顺和结构也是基本一致的。众所周知,一笔一画地书写汉字,不算是尊重汉字结构。那么按照笔顺将汉字分成几个有相交关系的部件,不影响原有笔画的形状、数量和先后顺序,从推理来看,也不应是不尊重汉字结构。若以平面结构论之,则还是尊重汉字框架结构。但只要跳出平面结构的框框,把笔画相交看作是笔画在垂直于视线的立体空间上的动态叠交,那么相交拆分也是可以接受的,因为它被视为立体结构而不仅仅是平面结构。汉字能够发展出独一无二的书法艺术,恐怕也是根源于此吧。可见,相交拆分并非不尊重汉字结构的标志。是否尊重汉字结构,取决于对汉字结构的理解,更取决于相交拆分的“度”。

另外,与汉字拆分方法相联系的部件表最好能够与《汉字部首表》<sup>[11]</sup>结合起来,包含部首表中的所有部首,这对汉字教育、汉字输入和汉字检索都将具有重大意义。对照两个《规范》所列出的部件表和 2009 年发布的《汉字部首表》不难发现,很多部首并不是部件,而很多部件也不是部首,部件表和部首表形成了“两张皮”。其危害是,在接触部件时,必须“忘却”中小学学习过的部首,否则就会造成干扰。而忘却部首,就意味着对汉字基础教育的部分否定。反过来,如果部件包含所有部首,则是完全继承了部首的优势,同时更利于部件在文字信息处理中的运用。如能实现,则基本再无部件和部首之分了。人们只要学习了部首或部件,就可以“一劳永逸”用到老,无须颠来倒去地学了部件再学部首、学了部首再学部件,平白浪费很多人力物力财力。这是社会高效发展的必然要求。

认识了问题的实质后,我们可以循着“客观存在一个作了定量和取码归并处理的部件表图和一个切实可行的拆分方法——按拆分方法选用部件将汉字拆成部件序列——按部件取码来编码汉字”这个规律,首先吸收《汉字部首表》中所有部首为汉字部件,并适当增加部分不是部首的部件。需要说明的是,所有这些部件都是基础部件,如《规范》1 所说,不得

再行拆分。其次,为了有效控制“相交拆分”的“度”,使汉字的拆分妥贴大众直观性拆分心理并具有可操作性,除了少数汉字或部件的拆出比较直观且与之相交的笔画之顺序一目了然外,不让其他笔画逆序组合和拆分,以利于尊重汉字笔画,使拆分最大限度地与笔顺相吻合。为此,作为参考,我们就能以“面向应用,从形出发,力求直观,兼顾字理和笔画”为原则,按以下 7 条规则对汉字进行拆分。

1) 拆分不增减笔画数,不改变笔画形状,不以形似替代。如“果”不拆分为“田木”,“叱”拆分为“口七”而不是“口乚”。

2) 被笔画隔开的“口日木”不按笔顺而单独相交拆出。如:“柬”拆分为“木口ㄣ”,“刺”拆分为“木门刂”,“由”拆分为“日丨”。

3) 常用字“必啄”中的“丿、”不按笔顺而单独相交拆出。如“啄”拆分为“口豕、”,“必”拆分为“心丿”。

4) 非常用字“幾戔”中的“戈弋”不按笔顺而单独相交拆出。如“幾”拆分为“么么戈人”,“戔”拆分为“弋口一彡”。

5) 双挂角、包围结构的围框、穿插结构的干架不按笔顺拆分。如“輿”拆分为“𠂔车一八”,“何”拆分为“亻丁口”,“夾”拆分为“大人人”。

6) 看似半包围结构,实则在笔顺上被其他笔画隔开且未包围分隔它的笔画,此部分严格按笔顺拆分。如“爲”拆分为“𠂔丿ㄣㄣ”而不为“𠂔ㄣㄣ”。

7) 基于上述 6 条,依笔顺,尽量用部件图中笔画数最多的部件,依次将汉字拆分成汉字部件的序列。如“喜”拆分为“土口ㄣ一口”,“产”拆分为“立丿”,“甫”拆分为“一月丨、”,“甞”拆分为“𠂔一一儿”,“龜”拆分为“刀日日ㄣ”。但以下情况例外:

① 如产生不止一个单笔画部件,则按照使部件最少的方式拆分。如“爨”拆分为“火业ㄣ一十耳又”,“派”拆分为“氵厂丿𠂔”,“卸”拆分为“一止口”。

② 如某部分能拆成单笔(特别是“一”)和多笔(特别是“ㄣ”及成字)两个部件,则按后者拆分。如“主”拆分为“丶王”,“失”拆分为“丿夫”,“權”拆分为“木一佳”,“卅”拆分为“一川”,“黎”拆分为“禾丿人水”。

③ 除上述两种情况外,如不相交拆分未产生比相交拆分更多的单笔画部件,则不相交拆分。如“衡”拆分为“彳田大一丁”。

这样一来,就可以完成对汉字的唯一拆分,并进

行汉字信息处理。比如输入“繁”，先拆成部件序列“母女系”，取其编码即可输入该字。再比如输入“為”，按照拆分方法，先拆成唯一的部件序列“丿 ㇀ ㇀ ㇀”（不因含有“勹”形而拆出“勹”作部件），取其编码便可输入该字。比如对“繁”的教学，可先解释它是由“敏”和“系”这两个部分上下结构而成的，其中“敏”由“每”“攴”左右结构而成，“每”又由“艹”和“母”上下结构而成。即便是“為”，也能解释前几笔画的书写顺序和方位，然后在“勹形”下书写一个部件“㇀”。由此掌握 300 左右个部件就能比较轻松地识记几千个汉字。再比如检索“繁”，拆分成部件序列“母女系”后，可按编码在词典中检索，也可像部首查字法那样，先提出部首“系”来，再按先后顺序排列其他部件，构成“系 母女”，用编码来检索。无论哪种方法，都能大大提高检字平均速度。

由此可见，汉字形码输入、汉字教育和汉字字形检索三者是可以统一于同一部件集和同一拆分方法的，部件也可以做到尽量少，而“相交拆分”也不至于将汉字拆得“一塌糊涂”。

## 6 结束语

两个《规范》为汉字编码的继续探索提供了重要参考，也启迪着人们从新的角度去思考汉字部件规范问题。山东大学科学社会主义博士生导师王建民说，“学者们对马克思某些结论的批评并不少见，但怯于触碰他的基本范畴和基础理论。而要实质性地推进马克思主义研究，就应该像马克思本人那样不畏艰险。”<sup>[12]</sup>既然对指导社会主义取得如此巨大胜利的马克思主义的研究都要求敢于触碰其基础理论，假如“交重不拆”确实行不通，那么破除对“交重不拆”的迷信又有什么不可以的呢？拆分方法永远只是手段，达到更好的目的才是关键。“不管白猫黑

猫，会捉老鼠就是好猫”，因此我们有理由寻找更好的解决问题的拆分方法，而不能死抱着“交重不拆”不放。解决汉字信息处理的问题，任重道远，本文在吸收《规范》成果的同时，提出一些质疑并表达一点思考，难免有失偏颇，但出发点是善意的，都是为了促进《规范》在信息处理的应用中不断发展和完善，因此敬请有关专家学者不吝批评指正。

## 参考文献

- [1] 国家语言文字工作委员会. 信息处理用 GB 13000.1 字符集汉字部件规范[S]. 北京: 国家语委, 1997.
- [2] 何克抗. 汉字认知模型与形码方案设计[J]. 中文信息学报, 1995, 9(3): 11-26.
- [3] 国家语言文字工作委员会. 现代常用字部件及部件名称规范[S]. 北京: 国家语委, 2009.
- [4] 张小衡. 《信息处理用 GB 13000.1 字符集汉字部件规范》在输入法应用中的难点讨论[J]. 中文信息学报, 2004, 18(4): 60-65.
- [5] 孙基寿. 汉字输入编码优劣评测方法的探讨[J]. 中文信息学报, 2006, 20(5): 97-104.
- [6] 国家语言文字工作委员会. 现代常用独体字规范[S]. 北京: 国家语委, 2009.
- [7] 王宁, 陈一凡. 谈从理与从形拆分原则——兼论“相离可拆”与“交重不拆”[J]. 计算机世界, 1998(15).
- [8] 《汉字规范码应用于基础教育教学实践》成果发布会在京召开. 中文信息学报, 2001, 15(1): 58.
- [9] 国家语言文字工作委员会. GB13000.1 字符集汉字部首归部规范[S]. 北京: 国家语委, 2009.
- [10] 李宇明. 搭建中华字符集大平台[J]. 中文信息学报, 2003, 17(2): 1-6, 53.
- [11] 国家语言文字工作委员会. 汉字部首表[S]. 北京: 国家语委, 2009.
- [12] 朱又可. 社会主义不应画地为牢——一桩停招科学社会主义博士生的案例[N]. 南方周末, 2011-6-23: D21-22.