

收稿日期:2021-02-01

《通用规范汉字表》独体字统计与思考

邵霭吉

(盐城师范学院 学报编辑部,江苏 盐城 224002)

摘要:《通用规范汉字表》中有多少个字可以认定为独体字,是一个值得研究的问题。依据《信息处理用GB 13000.1字符集汉字部件规范》《现代常用独体字规范》《现代常用字部件及部件名称规范》可得到3个不同的独体字字数。为适应汉字信息处理、辞书编纂和基础教育的需要,期盼国家语委早日制定和颁布“通用规范汉字独体字规范”,给出更为合适的独体字定义、更为适用的独体字认定标准和更为适当的《通用规范汉字表》中独体字的字数。

关键词:《通用规范汉字表》;独体字;部件;《信息处理用GB 13000.1字符集汉字部件规范》;《现代常用独体字规范》;《现代常用字部件及部件名称规范》

中图分类号:H124.2

文献标识码:A

文章编号:1003-6873(2021)03-0033-11

作者简介:邵霭吉(1947—),男,江苏盐城人,盐城师范学院学报编辑部编审,主要从事汉语言文字学研究。

DOI:10.16401/j.cnki.ysxb.1003-6873.2021.03.034

《通用规范汉字表》的8105个规范字中有多少个字是独体字,目前众说纷纭:万森《通用规范汉字笔画部首结构字级笔顺手册》标出《通用规范汉字表》的独体字290个^[1],说词解字辞书研究中心《学生全功能笔顺笔画部首结构规范大全》标出《通用规范汉字表》中独体字为285个^[2],李行健《现代汉语应用规范词典》标出的《通用规范汉字表》中的独体字为288个^[3]。在科学研究领域,分歧更大,李瑛《通用规范汉字独体字研究》认为,《通用规范汉字表》中的独体字有315个^[4];侯冬梅《通用规范汉字构形属性研究》认为,《通用规范汉字表》中的独体字有351个^[5]¹⁴⁶;柳建钰、王晓旭《基于字料库的通用规范汉字构形属性调查研究》认为,《通用规范汉字表》中的独体字共220个^[6]。笔者认为,国家颁布的《信息处理用GB 13000.1字符集汉字部件规范》《现代常用独体字规范》《现代常用字部件及部件名称规范》等三部语言文字规范,可以作为认定独体字的依据。本文拟依据这三部语言文字规范,重新统计《通用规范汉字表》中的独体字的字数,给出三个跟上述字数不同的、有国家颁布的语言文字规范作依据的统计结果,同时提出自己对这一问题的看法和建议。

一、基于国家语言文字规范的《通用规范汉字表》独体字字数统计

(一) 依据《信息处理用GB 13000.1字符集汉字部件规范》统计

1997年12月国家语委发布的《信息处理用GB 13000.1字符集汉字部件规范》，对GB 13000.1字符集20 902个汉字作部件拆分，共得出基础部件“560个”，这些“基础部件”是“最小的不再拆分的部件”^[7]。既然它们具有“最小”和“不再拆分”的属性，那么，它们当中那些能够独立成字的基础部件，所成之字就自然是独体字了。

据我们统计，该规范的560个基础部件中，有230个基础部件所成之字，对应于《通用规范汉字表》的规范字。也就是说，依据该规范，可以统计出《通用规范汉字表》8 105个规范字中有230个独体字。这230个独体字是：

0001 一 0002 乙 0003 二 0004 十 0005 丁 0006 厂 0007 七 0008 卜 0009 八 0010 人
0011 入 0012 儿 0013 匕 0014 几 0015 九 0016 刁 0017 了 0018 刀 0019 力 0020 乃
0021 又 0022 三 0023 干 0024 于 0026 工 0027 土 0028 士 0029 才 0030 下 0031 寸
0032 大 0033 丈 0035 万 0036 上 0037 小 0038 口 0039 山 0040 巾 0041 千 0043 川
0045 个 0046 夕 0047 久 0051 丸 0052 及 0053 广 0054 亡 0055 冂 0056 丫 0058 之
0059 尸 0060 己 0061 巳 0062 已 0063 弓 0064 子 0066 也 0067 女 0069 飞 0070 习
0073 乡 0074 丰 0075 王 0077 井 0079 夫 0083 专 0084 丐 0087 木 0088 五 0091 不
0092 犬 0096 歹 0100 车 0101 巨 0102 牙 0103 屯 0104 戈 0108 瓦 0109 止 0111 日
0112 日 0113 中 0114 贝 0116 内 0117 水 0118 见 0120 牛 0121 手 0122 气 0123 毛
0127 长 0130 片 0137 斤 0138 爪 0141 父 0150 月 0151 氏 0155 丹 0161 文 0163 方
0164 火 0165 为 0170 户 0174 心 0175 尺 0177 丑 0178 巴 0184 予 0188 书 0190 玉
0192 未 0193 末 0194 示 0205 甘 0206 世 0210 本 0216 石 0220 戊 0221 龙 0225 东
0229 凸 0231 业 0236 目 0237 且 0239 甲 0240 申 0242 电 0244 田 0245 由 0248 史
0249 央 0258 冉 0259 皿 0260 凹 0262 四 0266 乍 0267 禾 0268 丘 0275 白 0278 斤
0279 瓜 0280 乎 0287 乐 0290 册 0302 立 0321 必 0324 永 0327 民 0328 弗 0336 皮
0343 矛 0345 母 0369 耳 0373 亚 0380 臣 0381 吏 0384 西 0392 而 0403 夷 0423 虫
0424 曲 0443 肉 0444 年 0450 竹 0462 臼 0477 自 0484 舟 0489 兆 0520 衣 0530 羊
0533 米 0535 州 0690 求 0691 甫 0693 更 0694 束 0697 两 0698 酉 0721 里 0736 串
0761 我 0786 身 0826 言 1038 事 1041 雨 1064 非 1079 果 1125 垂 1134 秉 1166 金
1315 承 1403 革 1442 束 1552 重 1573 鬼 1575 禹 1589 食 2361 象 2655 黑 3119 熏
3501 乂 3502 乚 3503 兀 3504 弋 3505 孑 3506 孑 3507 彡 3509 韦 3510 廿 3511 丐
3512 卅 3524 尹 3525 夂 3526 月 3527 毋 3564 耒 3592 曳 3596 缶 3635 聿 3636 艮
3688 豕 3740 豸 3955 史 4195 禹 5127 戢 6502 尢 6503 彳 6507 毌 6509 戈 6584 龙

这与谢泽荣《独体字四论》的统计结果相同。该文指出，《信息处理用GB 13000.1字符集汉字部件规范》中的“《汉字基础部件表》里的成字部件（不包括繁体字、异体字、旧笔形字）共230个”^[8]。柳建钰、王晓旭《基于字料库的通用规范汉字构形属性调查研究》认为，《通用规范汉字表》8 105字中的“独体字共220个”^[3]。这跟我们依据《信息处理用GB 13000.1字符集汉字部件规范》统计出来的230个独体字，字数比较接近。

由于《信息处理用GB 13000.1字符集汉字部件规范》是国家发布的语言文字规范，也由于这

一规范涵盖了20 902个汉字,基本上涵盖了《通用规范汉字表》中的规范字,因此我们完全可以这样说,依据该规范的基础部件来统计,《通用规范汉字表》的独体字就应该是 230 个。

(二)依据《现代常用独体字规范》统计

2009 年 3 月教育部、国家语委发布的《现代常用独体字规范》,是到目前为止国家颁布的唯一一部关于独体字的语言文字规范。这部规范,以“语言文字规范”的身份,给出了独体字的定义和“现代常用独体字的确定规则”,“在现代汉字的范围内确定了 256 个现代常用独体字,形成了《现代常用独体字表》”^[9]。

经我们逐一对比核查,《现代常用独体字规范》确定的 256 个独体字,全都出现在《通用规范汉字表》当中。也就是说,依据《现代常用独体字规范》来统计,《通用规范汉字表》中的独体字至少有这 256 个。

在这 256 个独体字中,有 191 个也在依据《信息处理用GB 13000.1字符集汉字部件规范》所统计出来的结果之中,另外 65 个是依据《信息处理用GB 13000.1字符集汉字部件规范》所统计出来的结果中没有的。

这 65 个独体字是:

0034 与 0048 么 0050 凡 0057 义 0065 卫 0068 刃 0072 马 0076 开 0078 天 0081 无
0082 云 0093 太 0098 尤 0106 互 0110 少 0119 午 0124 壬 0125 升 0126 夭 0152 勿
0157 鸟 0160 六 0166 斗 0181 办 0195 击 0198 正 0211 术 0213 丙 0222 平 0226 卡
0263 生 0264 矢 0265 失 0283 用 0284 甩 0289 匆 0296 鸟 0300 主 0307 半 0310 头
0330 出 0382 再 0389 百 0393 页 0402 夹 0445 朱 0477 自 0479 血 0516 亦 0522 产
0524 亥 0558 农 0671 严 0707 来 0711 鹵 0790 鹵 0854 弟 1303 肃 1305 隶 1455 面
1651 首 2056 兼 2958 鼠 3611 鹵 3764 羌

依据《现代常用独体字规范》来认定《通用规范汉字表》中的独体字,有一个问题是不能忽视的。那就是,《现代常用独体字规范》只是针对“常用字”^①这个范围来认定的独体字,除了其中“凶、羌”两字外,并没有涉及常用字之外的通用规范汉字,即《通用规范汉字表》的《二级字表》《三级字表》中的字。虽然从总体上说,独体字的绝大部分都存在于常用字中,但为数不多的、在常用字范围之外的独体字也是不能无视的。

为此,我们觉得应该利用依据《信息处理用GB 13000.1字符集汉字部件规范》所统计出来的结果,来弥补仅仅依据《现代常用独体字规范》进行统计的不足。

我们注意到,依据《信息处理用GB 13000.1字符集汉字部件规范》所统计出来的独体字中,有以下 39 个不在《现代常用独体字规范》的 256 个独体字之中:

0194 示 0336 皮 0450 竹 0489 兆 1064 非 1166 金 1589 食 2655 黑 3119 熏 3501 乂
3502 乚 3503 兀 3504 弋 3505 矛 3506 孑 3507 彡 3509 韦 3510 廿 3511 丐 3512 卅
3524 尹 3525 夬 3526 片 3527 毋 3564 耒 3592 曳 3596 缶 3635 聿 3636 艮 3688 豕
3740 豸 3955 臾 4195 禺 5127 戠 6502 尢 6503 彳 6507 毋 6509 戈 6584 龙

所以,只要在《现代常用独体字规范》256 个独体字的基础上,加上这 39 个独体字,就可以得到一个更完善的统计结果:《通用规范汉字表》中的独体字一共是 295 个。《通用规范汉字表》中

① 常用字相当于《通用规范汉字表》的《一级字表》中的字。

的规范字除去这 295 个独体字,都是合体字。

《通用规范汉字表》中有 295 个独体字,这个统计结果跟目前出版的一些辞书所标出来的《通用规范汉字表》独体字的数量,是十分接近的,且做法大致相同。万森《通用规范汉字笔画部首结构字级笔顺手册》标注《通用规范汉字表》独体字 290 个^[1],是在《现代常用独体字规范》规定的 256 个现代常用独体字的基础上,增加了 34 个在常用字之外的独体字。说词解字辞书研究中心《学生全功能笔顺笔画部首结构规范大全》标注《通用规范汉字表》独体字 285 个^[2],采用了《现代常用独体字规范》256 个独体字中的 253 个,另增加该规范之外的 32 个独体字。新增者除“币”为常用字外,另 31 个都是在常用字范围之外的独体字。李行健《现代汉语应用规范词典》标注《通用规范汉字表》独体字 288 个^[3],是在《现代常用独体字规范》规定的 256 个现代常用独体字的基础上,增加了 32 个在常用字范围之外的独体字。

这使我们相信,《现代常用独体字规范》规定的 256 个独体字,是统计《通用规范汉字表》中独体字字数的重要依据。

(三)依据《现代常用字部件及部件名称规范》统计

2009 年 3 月,教育部、国家语委发布的《现代常用字部件及部件名称规范》,是一个关于现代汉语常用字的部件的规范。这一规范,“对现代汉语 3 500 常用汉字逐个进行部件拆分、归纳与统计”,形成《现代常用字部件表》,共有 514 个部件,这些部件“规定了汉字部件拆分的下限,一般不宜再行拆分”^[10]。

由于《现代常用字部件及部件名称规范》给出的“514 个部件”具有“汉字部件拆分的下限”和“一般不宜再行拆分”的属性,所以它如果“独立成字”,那就会是一个不能拆分或者“不宜再行拆分”的规范汉字,也即独体字。因而我们可以依据《现代常用字部件及部件名称规范》中的“部件”,来统计《通用规范汉字表》中独体字的字数。

李瑛《通用规范汉字独体字研究》一文正是“参照《现代常用字部件及部件名称规范》”^[4]来认定《通用规范汉字表》中的独体字的。不过,该文说的是“参照”,而不是“依据”。

本文拟采取比李瑛《通用规范汉字独体字研究》更加严格的做法,完全“依据”而不仅仅是“参照”《现代常用字部件及部件名称规范》,对《通用规范汉字表》中独体字的字数来重新作一个统计。我们的做法是在《通用规范汉字表》8 105 字中,只要是由《现代常用字部件及部件名称规范》中的部件独立所成的字,就算它是独体字。

经过比对,我们发现,在《现代常用字部件及部件名称规范》的 514 个部件中,有 342 个部件可以独立成《通用规范汉字表》中的规范字,也就是说,完全依据《现代常用字部件及部件名称规范》来统计,《通用规范汉字表》中独体字的字数应该是 342 个。

在这 342 个独体字中,有 255 个是《现代常用独体字规范》中已有的,其余 87 个是《现代常用独体字规范》中没有的。在这 87 个独体字中,30 个是依据《信息处理用 GB 13000.1 字符集汉字部件规范》认定的 230 字中已有的;其余 57 字,既是《现代常用独体字规范》中没有的,也是依据《信息处理用 GB 13000.1 字符集汉字部件规范》认定的 230 个独体字中没有的。这 57 个字是:

0049 勺 0143 仑 0144 今 0153 欠 0154 风 0204 去 0291 卯 0313 穴 0325 司 0339 发
0347 丝 0360 老 0387 戌 0410 至 0449 舌 0456 乒 0457 兵 0480 向 0482 后 0493 伞
0508 争 0509 色 0631 赤 0696 豆 0699 丽 0701 辰 0798 谷 0808 龟 0810 免 0814 角
0818 卯 0841 辛 0996 其 1011 直 1056 妻 1077 具 1154 卑 1194 鱼 1205 京 1232 单
1422 南 1466 韭 2021 高 2201 黄 2336 兜 2386 庸 2387 鹿 2617 鼎 3641 丞 3744 奂
4467 鬲 4749 董 6504 叩 6505 爻 6699 隹 6907 段 7081 崔

字之共性,“合体字亦由笔画组成”^[4]。苏培成《现代汉字学纲要》批评说,“由笔画组成”“不是构成独体字的必要条件”^{[11]96};“可以构成合体字”也“不是独体字的必备条件”^{[11]96}。苏培成指出,“承、个、卫、书、凸、再、秉”等19个独体字“在通用字范围内不能构成合体字”。侯冬梅《通用规范汉字构形属性研究》指出,“基于通用规范汉字构形属性数据库的统计,……独体字独有的基础部件有41个”,“这41个基础部件在《通用规范汉字表》中只能构成独体字,不能参与组构合体字”^{[5]147}。柳建钰、王晓旭《基于字料库的通用规范汉字构形属性调查研究》也指出,独体字“承、飞、个、卫、年、伞、书、凸、再、子、乚”等,“只能构成自身形体”^[6]。因此,笔者觉得,重新给独体字一个更为合适的定义,是十分必要的。

(二)关于独体字的认定标准

《现代常用独体字规范》虽然给出了3条“现代常用独体字的确定规则”,但不准确,有遗漏,也不怎么适用。它提出的3条规则,一是“字形结构符合字理和独体字定义的汉字”,二是“符合独体字定义的草书楷化的简化字”,三是“交重结构,不能拆分的汉字”。这三条规则涵盖不了所有的独体字,比如“产、严、业、习”等简化字,就既不符合古文字“字理”,又非“草书楷化的简化字”,也不是“交重结构”,那么,《现代常用独体字规范》又凭什么确定它们为独体字呢?另外,前两条规则在“符合独体字定义”基础上还分别增加了“符合字理”和“草书楷化的简化字”的附加条件,实无必要。既然一个汉字已经“符合独体字定义”了,它不就已经是独体字了吗?因此,应有一个更为适用的独体字认定标准。

笔者认为,独体字是不能拆分和不宜拆分的汉字,也就是说,独体字是“不能拆分的汉字”和“不宜拆分的汉字”的合集。据此我们可以分别从“不能拆分”和“不宜拆分”两个方面来提出一个更为适用的认定独体字的标准。

1. 不能拆分的汉字是独体字

不能拆分的汉字数量不多,很容易确定。它有两个类型。

(1)单笔画汉字,不能拆分。《通用规范汉字表》中只有两个,即:“一、乙”。

(2)字中没有“相离、相接的笔画或部件”的汉字,不能拆分。《信息处理用GB 13000.1字符集汉字部件规范》指出:“相离、相接可拆,交重不拆”^[7]。如果是一个“交重结构”且其中没有“相离、相接”的笔画或部件,自然不能拆分。例如“十、七、九、力、又、丈、丸、也、女、丰、井、车、冉、册、弗、义、乚、韦、卅、尹、毋、聿、卩”等。

2. “不宜拆分”的汉字是独体字

“不宜拆分”的问题比较复杂。“不宜拆分”的意思是可以拆分而不适宜拆分。至于哪些字“宜于拆分”,哪些字“不宜拆分”,不同的学者有不同的见解,有的严,有的宽。例如,《现代汉语常用字表》3500字,晓东《现代汉字独体字与合体字的再认识》认为其中有独体字195个^[12],跟《现代常用独体字规范》的256字相差61个,所差的都是不宜拆分的独体字。同样针对《现代汉语常用字表》的3500字,邢红兵《现代汉字特征分析与计算研究》认为常用汉字中有“独体结构”187个^{[13]45},跟晓东《现代汉字独体字与合体字的再认识》相差8个,跟《现代常用独体字规范》的256字相差69个。苏培成《现代汉字学纲要》还认为《现代常用独体字规范》给出的256个独体字中,有“寸、与、凡、义、习、又、文、方、矛、肉、衣”等“48字不是独体字,应该从《规范》里删除”^{[11]97},删除这48字后只剩下208个独体字。

国家在不同时期颁布的语言文字规范也有不同的处理,例如,《现代常用独体字规范》认为在

常用字中有256个独体字“不能或不宜再行拆分”,而《现代常用字部件及部件名称规范》则认为在常用字中有311个成字部件“一般不宜再行拆分”,两者所不同的59个独体字也都是不宜拆分的。而依据《信息处理用GB 13000.1字符集汉字部件规范》来看,在《现代常用独体字规范》的256个独体字中有“开、无、午、互、办、亥、鼠”等65个是可以拆分且应该进行拆分的,在《现代常用字部件及部件名称规范》的342个成字部件中,有“黄、京、丝、老、戌、至、舌、兵、兵、向、后、伞”等121个是可以拆分且应该进行拆分的。

一般认为,下述几种情形,都是“不宜拆分”的:

(1)如果某字的拆分结果都不是部件,则该字不宜拆分。

比如五画的“民”,前三画可以视为一个笔画组合体,末两画也可以视为一个笔画组合体,好像可以把“民”拆分为两个笔画组合体,但如果这样拆分,则前三画组合体不成部件,末两画组合体也不成部件,所以“民”不宜拆分。又如“艮”字六画,前四画可以视为一个笔画组合体,末两画也可以视为一个笔画组合体,好像也是可以拆分的,但如果这样把两者拆分开来,则前四画组合体不成部件,末两画组合体也不成部件,所以“艮”也不宜拆分。四画的“氏”也一样,如果把前两画跟末两画拆分,则前两画组合体不成部件,末两画组合体也不成部件,所以“氏”也不宜拆分。

(2)如果某字的拆分结果全部都是单笔画部件,则该字不宜拆分。

比如两笔画字“二、丁、厂、卜、八、人、入、儿、匕、几、刁、了、刀、乃”,如果把它两个笔画拆分开来,认为这些字是由两个单笔画部件组成的合体字,肯定不会有人赞同。所以,两画字,除了前面讲的两画相交、绝对不能拆分的“十、七、九、力、又、乂、乚”以外,其余的两画相离、相接的字,都是不宜拆分的。再如三画字中的“三、万、川、夕、久、门、之”,四画字中的“丐、不、瓦、止、片、心、丐、月”,五画字中的“凸、且、皿、凹、丘”,六画字中的“臼、州”,七画字中的“豸”等,如果强行拆分,则结果只能全部都是单笔画部件,所以这些字都不宜拆分。苏培成《现代汉字学纲要》提出的“拆分的下限一般要大于笔画”^{[1]87} 拆分规则,说的也是拆分结果不能全部都是单笔画部件。

(3)如果某字虽有“相离、相接”的笔画可拆,但拆后剩下的部分不成为一个部件(或几个部件的组合体)的,则该字不宜拆分。

如“巾、子、孑、孓”,首笔未参与相交,属于可拆者,但拆出这一笔后,剩余部分不成为一个部件,所以这些字不宜拆分。其他,如“于、及、中、史、申、电、曳”的第一画,“甲”的第一第二两画,“事、吏”的第二画,“屯、丹”的第三画,“世”的第四画,“夫、夊、夷、专、书、弋、戈、戈”的末笔,“长、丑、牙、内、由、曲、我、垂”的首笔和末笔,“串”的第一和第四画,“母”的第三和第五画,“舟、曳、耳、身、必、甫、禺、西、两”等字中的多个笔画,它们虽未参与相交,属于“可拆”者,但拆分出这些笔画后,剩余部分不成一个部件,或几个部件的组合,所以这些字不宜拆分。

(4)如果从某字中拆分出一个类似于部件的笔画组合体后,剩下的部分不成为一个部件(或几个部件的组合体)的,则该字不宜拆分。

例如“乐、东”,最后两笔类似于部件“八”(“兴、具、俊”等字中的部件),可以拆分出来,但余下部分不成为部件,所以“乐、东”不宜拆分。“予”的第一第二两画类似于部件“丩”(“甬、疑、通”中的部件),但若把“予”的“丩”拆分出去,则剩余的部分不成为一个部件,所以“予”不宜拆分。“气、年”两字的第一画第二画的组合体,类似于部件“乚”(“乞、每、复”中的部件),但若把“乚”从“气、年”中拆分出去,所余下的部分不成为一个部件,所以“气、年”不宜拆分。“斥、爪、瓜”三字的前两画,跟部件“厂”(“反、后”等字中的部件)形似,但拆分出这两画后,剩下的部分不是部件,所以不宜拆分。“斤”字也一样,拆分出前两画后,剩下的部分“丌”在《信息处理用GB 13000.1字符集汉字部件规范》被规定为“G列以外汉字专用的部件”,不是中国汉字部件,所以“斤”字不宜拆分。六画字“虫”的前四画类似于部件“中”,但若把它拆分出去,则剩下的部分不成部件,所以“虫”不

宜拆分。“革”若拆分出类似于部件的“廿”，则剩余部分不成部件，所以“革”不宜拆分。

(5)如果在某字的拆分结果中出现连续的单笔画部件，则该字不宜拆分。

这一条规则跟第(2)条有联系，第(2)条说的是拆分结果不能全部都是单笔画部件，这一条说的是拆分结果中出现连续的单笔画部件的，也不行。这一条跟第(4)条也有联系，第(4)条说的是拆分后剩余部分不成部件的不宜拆分，这一条说的是把拆分后剩余的不成部件者再拆分成连续的单笔画部件也不宜。比如“予”不宜拆分为[丿]+[(一)+(丨)]，“气”不宜拆分为[乚]+[(一)+(乙)]，“斤”不宜拆分为[厂]+[(一)+(丨)]，“虫”不宜拆分为[中]+[(一)+(丶)]，等等。

(三)关于独体字的字数

前已述及，关于《通用规范汉字表》中的独体字字数，学界众说纷纭。即使依据国家颁布的相关的语言文字规范来统计，也有不同的结果。笔者认为，要严格坚持《信息处理用GB 13000.1字符集汉字部件规范》提出的“相离、相接可拆，交重不拆”的拆分标准，尽量减少独体字的字数，只有“不能拆分”和“不宜拆分”的字，才算是独体字，其余“可以拆分”且不违反“不宜拆分”规定的字统统算合体字。如果这样，《通用规范汉字表》中独体字的数量则可以大大减少，比依据《信息处理用GB 13000.1字符集汉字部件规范》的基础部件认定的230个还要少。独体字数量少，更加有利于汉字信息处理、辞书编纂、基础教育。

1. 把没有被《现代常用独体字规范》认定为独体字的常用字回归合体字

在依据《信息处理用GB 13000.1字符集汉字部件规范》认定的230个独体字中，有“示、皮、竹、兆、非、金、食、黑、熏”等9个常用字没有被《现代常用独体字规范》认定为独体字。《现代常用独体字规范》是晚于《信息处理用GB 13000.1字符集汉字部件规范》颁布的语言文字规范，体现了语言文字学术研究的新成果，我们应该尊重，把这9个常用字分析为合体字应该是有它的道理的，因为这些字都可以拆分出或左右、或上下两部分。这里先说“示、竹、非、金、食”。“示”可以拆分为[二]+[小]，是合体字，[二]和[小]都是《信息处理用GB 13000.1字符集汉字部件规范》中的基础部件。谢泽荣《独体字四论》分析说：“‘示’，依《说文》其构字理据是‘从二，从川(日、月、星)，天垂象见吉凶，所以示人也’，应拆分为[二][小]”^[8]。“竹”可以拆分[(乚)+(丨)]+[(乚)+(丨)]，“非”可以拆分[(三)+(丨)]+[(丨)+(三)]，“金”可以拆分[人]+[(干)+(亠)]^{[5]77}，也可以拆分为[人]+[(王)+(丿)]，“食”可以拆分[人]+[(丶)+(艮)]，这五字应该回归合体字。本文开头提到的三部辞书都把“竹、非”二字标注为左右结构合体字，把“示、金、食”三字标注为上下结构合体字。

2. 仿照《信息处理用GB 13000.1字符集汉字部件规范》拆分“击、出、午、半”的实例，把“缶、肉、父、文、衣、立、米、火”等拆分为两个(或多个)基础部件

该规范将“击”拆分为[丰]+[口]，将“出”拆分为[中]+[口]，“缶”跟“击、出”形近，应该可以拆分为[午]+[口]。依据该规范，“午”还要再拆分为[乚]+[十]，则“缶”应该拆分为[(乚)+(十)]+[口]。侯冬梅把“缶”拆分为[午]+[口]，归入上下结构^{[5]295}合体字是对的。其他如“肉”可拆分为[内]+[人]，“父”可拆分为[丩]+[乂]^{[5]290}，“文”可拆分为[亠]+[乂]^{[5]290}，“巨”可拆分为[匚]+[𠃍]，“臣”可拆分为[匚]+[(丨)+(𠃍)+(丨)]，“衣”可拆分为[亠]+[衣]，“衣”是该规范中“衰、表”等字的部件，“立”可拆分为[亠]+[立]，“立”是该规范中“豆、并、前”等字的部件。该规范把“半”拆分为[丿]+[丰]，因此“米”也可拆分为[丿]+[木]，“火”也可拆分为[丿]+

[人]。“内、人、ハ、乂、亠、匚、冫、衣、亼、ノ、丰、木”都是该规范中现成的部件。

3. 仿照《信息处理用GB 13000.1字符集汉字部件规范》拆分“正、旧、失、主、司”的实例,把一些字拆分成一个单笔画部件跟一个多笔画部件的组合

单笔画部件,有些学者不承认或不全承认,以致他们认定的独体字字数比别人多,而《信息处理用GB 13000.1字符集汉字部件规范》《现代常用字部件及部件名称规定》都是承认有单笔画部件的,只是处理起来还不够彻底,需要加强。

该规范把“正”拆分为[一]+[止],把“无”拆分为[一]+[尢],把“元”拆分为[一]+[兀],那么,“亚”也可以拆分为[一]+[业]^{[5]296},“干”也可以拆分为[一]+[十]。“歹”也可以拆分为[一]+[夕],是上下结构合体字。侯冬梅《通用规范汉字构形属性研究》指出:“‘歹’拆分为‘一夕’,上下关系”^{[5]76}。“兀”也可以拆分为[一]+[儿],上下结构合体字,《通用规范汉字字典》就把“兀”标注为上下结构合体字^{[15]653}。另外,“士、土”也可以拆分为[十]+[一],“豕”也可以拆分为[一]+[豕],“里”也可以拆分为[里]+[一]。“豕、里”分别是该规范中“毅、彖”和“单”的部件。

该规范把“旧”拆分为[丨]+[日],把“引”拆分为[弓]+[丨],把“出”拆分为[屮]+[凵]。仿此,“山”也可以拆分为[丨]+[凵],“个”也可以拆分为[人]+[丨],“丫”也可以拆分为[ノ]+[丨],“小”也可以拆分为[丨]+[八],因为“丨、日、人、ノ、丨、八”都是该规范中的基础部件。

该规范把“失”拆分为[丨]+[夫],把“朱”拆分为[丨]+[未],把“血”拆分为[丨]+[皿]。仿此,“牛”也可以拆分为[丨]+[丰],“丰”是该规范中“击、半、用”等字中的部件。“白”也可以拆分为[丨]+[日],“自”也可以拆分为[丨]+[目],“千”也可以拆分为[丨]+[十],“彳”也可以拆分为[丨]+[彳],“禾”也可以拆分为[丨]+[木],“矛”也可以拆分为[予]+[丨],“乡”也可以拆分为[彡]+[丨]。“彡”是该规范中“丝”字中的部件。

该规范把“主”拆分为[丶]+[王],把“义”拆分为[丶]+[乂],把“凡”拆分为[几]+[丶]。仿此,“广”也应拆分为[丶]+[厂],“户”也应拆分为[丶]+[尸],“方”也应拆分为[丶]+[万]。该“规范”把“尤”拆分为[尢]+[丶],仿此,“犬”也可以拆分成[大]+[丶]。该“规范”把“太”拆分为[大]+[丶],“犬”也是“大”字加一点,拆分成[大]+[丶],归合体字。虽然该规范认定“玉”为独体字,但我们认为“玉”可以拆分为[王]+[丶]。

该规范把“司”拆分为[冫]+[(一)+(口)],把“局”拆分为[尸]+[(冫)+(口)],把“勾”拆分为[勹]+[丨],仿此,“习”也应拆分为[冫]+[丨]。陈燕《现代汉字部首法所用单字切分的研究》指出:“‘习’与‘勾’结构相同……拿相类的字形比较就看出能切分,能切分就不宜看做独体结构字”^[14]。苏培成《现代汉字学纲要》说“‘习’……不是独体字”^{[11]96},“习”是上右包围合体字^{[11]99}。北京大学中文系现代汉语教研室《现代汉语》把“习”归入“上右包围结构”合体字^{[16]154}。另外,“已、己、巳”都可以拆分为[冫]+[丨],“丨”是该规范“礼、胤、孔”等字中的部件。

4. 仿照《信息处理用GB 13000.1字符集汉字部件规范》的拆分“庸、殷”的实例,增加少数“只跟某一个特定部件组合的非成字部件”,减少“黑、熏、皮、兆、戠”等独体字

“只跟某一个特定部件组合的非成字部件”在拆分汉字时是应该尽量避免的,但事实上又不能完全排除。《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》都各有几十个“只跟某一个特定部件组合的非成字部件”。比如“庸”字,前者把它拆分为[广]+[甬],“甬”就是一个只跟特定部件“广”组合的非成字部件,它只能跟“广”直接组合成“庸”,而不能跟别的部件直接组合,在“镛、慵、壟、黉、鄯”等字中,它也是先跟“广”组合成“庸”以后再跟

“车、丩、土、鱼、卩”组合的。再如“𠂔”，它只能跟“殳”组合成“殷”，之后方可再跟别的部件组合使用，如“澱”。此外，段字旁、那字旁、制字旁、尧字头、聚字底、可字框等部件，都是该规范中现有的只跟某一个特定部件组合的非成字部件。因此，如果增加一个部件“𠂔”，则“黑”可以拆分为[𠂔]+[灬]；如果增加一个部件“重”，则“熏”可以拆分为[重]+[灬]；如果增加一个部件“𠂔”，则“皮”可以拆分为[𠂔]+[又]。已经有学者提出过把这些字认定为合体字的问题。本文前面提到的三部辞书也都把“黑、熏”标注为上下结构合体字，把“皮”标注为半包围结构合体字。这三部辞书还把“兆”标注为左右结构合体字，作如此分析，也需要增加一个非成字部件。如果是把“兆”的最后两画规定为一个部件，则“飞”也可以分析为合体字。再如“戢”字，前述三部辞书一致把它标注为“半包围”合体字，但我们知道，要把“戢”分析为“半包围”合体字，也必须增加一个只跟某一个特定部件组合的非成字部件。

以上一己之见，是基于“拆分时只根据字形，不考虑字音和字义”^{[11]86}“相离、相交可拆，交重不拆”^[7]等规则的一些大胆的设想，供大家参考。

三、余论：盼望早日颁布“通用规范汉字独体字规范”

关于《通用规范汉字表》中的独体字的字数，本文共提及9种不同的统计数据，可谓众说纷纭。即使依据国家颁布的语言文字规范来统计，也会出现3种不同的结果，这种状况，极不利于汉字信息处理、辞书编纂和基础教育的需要。

为此，我们期盼国家语委早日颁布“通用规范汉字独体字规范”，给出更为合适的独体字定义、更为适用的独体字认定标准和更为适当的《通用规范汉字表》中独体字的字数。有了国家颁布的“通用规范汉字独体字规范”，我们才能明确《通用规范汉字表》中究竟有多少个独体字，是哪些字，是凭什么标准把它们认定为独体字的，使《通用规范汉字表》的独体字问题得以解决。无论是汉字信息处理，还是辞书编纂、基础教育，都能有据可循。

在国家没有公布“通用规范汉字独体字规范”的情况下，如果国家制定和公布了“通用规范汉字部件规范”，这个问题也可以解决。我们可以借助于“通用规范汉字部件规范”中的基础部件，来认定《通用规范汉字表》中的独体字，凡是由该规范的一个基础部件构成的字就是独体字，否则不是。

总之，国家颁布“通用规范汉字独体字规范”或“通用规范汉字部件规范”，很有必要。

参考文献

- [1] 万森. 通用规范汉字笔画部首结构字级笔顺手册[M]. 北京: 商务印书馆国际有限公司, 2016.
- [2] 说词解字辞书研究中心. 学生全功能笔顺笔画部首结构规范大全[M]. 北京: 华语教学出版社, 2018.
- [3] 李行健. 现代汉语应用规范词典[M]. 北京: 语文出版社, 2019.
- [4] 李瑛. 通用规范汉字独体字研究[J], 内蒙古师范大学学报(哲学社会科学版), 2016(3): 90-96.
- [5] 侯冬梅. 通用规范汉字构形属性研究[M]. 北京: 科学出版社, 2017.
- [6] 柳建钰, 王晓旭. 基于字料库的通用规范汉字构形属性调查研究[J], 渤海大学学报(哲学社会科学版), 2019(5): 104-111.
- [7] 国家语委. 信息处理用GB 13000.1字符集汉字部件规范[EB/OL]. (2005-10-27)[2021-01-20]. http://www.moe.gov.cn/jyb_sjzl/ziliao/A19/201001/t20100115_75616.html.
- [8] 谢泽荣. 独体字四论[J], 重庆师范大学学报(哲学社会科学版), 2007(1): 116-120.
- [9] 教育部, 国家语委. 现代常用独体字规范//魏励. 语言文字规范手册[M]. 北京: 商务印书馆国际有限公司, 2014: 142-147.

- [10] 教育部,国家语委. 现代常用字部件及部件名称规范[M]//魏励. 语言文字规范手册. 北京:商务印书馆国际有限公司,2014:149-224.
- [11] 苏培成. 现代汉字学纲要[M]. 3 版. 北京:商务印书馆,2014.
- [12] 晓东. 现代汉字独体字与合体字的再认识[J]. 语文建设,1994(8):28-31.
- [13] 邢红兵. 现代汉字特征分析与计算研究[M]. 北京:商务印书馆,2007.
- [14] 陈燕. 现代汉字部首法所用单字切分的研究[J]. 天津师范大学学报(哲学社会科学版),2006(4):71-76.
- [15] 商务国际辞书编辑部. 通用规范汉字字典[M]. 北京:商务印书馆,2019.
- [16] 北京大学中文系现代汉语教研室. 现代汉语:增订本[M]. 北京:商务印书馆,2012.

The Undecomposable Characters in the *General Standard Chinese Characters Table*

SHAO Ai-ji

(Journal Editorial Department, Yancheng Teachers University, Yancheng, Jiangsu, 224002, China)

Abstract: The number of undecomposable characters in the *General Standard Chinese Characters Table* is worth studying. We have three different numbers according to three different standards, namely, *Chinese Character Component Standard of GB 13000.1 Character Set for Information Processing*, *Specification of the Undecomposable Characters Commonly Used in the Modern Chinese*, *Specification of Common Modern Chinese Character Components and Component Names*. In order to meet the needs of Chinese character information processing, dictionary compilation and elementary education, it is suggested that the State Language Commission should formulate and promulgate the specification of general standard undecomposable Chinese characters as soon as possible, and give a more appropriate definition, a more applicable standard and a more suitable number of undecomposable characters in the *General Standard Chinese Characters Table*.

Key words: *General Standard Chinese Character Table*; undecomposable characters; components; *Chinese Character Component Standard of GB 13000.1 Character Set for Information Processing*; *Specification of the Undecomposable Characters Commonly Used in the Modern Chinese*; *Specification of Common Modern Chinese Character Components and Component Names*

〔责任编辑:何敏敏〕