# 汉字部件拆解方案新探

# 潘泰

(武汉大学 国际教育学院, 湖北 武汉 430070)

摘 要:拆分汉字时只要相交就不拆解,或者只要不同就分立为不同部件的做法并非总是突出汉字字形特征的最佳方案。对现行汉字字形的系统性拆解应该符合人们的认知规律,尽量体现字形的区别性差异点,满足字形的拓扑相似性等要求。在汉字字形的拆解中可以兼顾字源理据,但并非总是必须如此。

关键词:汉字;部件;系统性;认知

# 一、引言

在汉字的学习中,部件的辨认(recognize)和复写(reform)是其中的关键环节之一。不过,现有的汉字部件拆解方案却不太适用于人们习得汉字的场景。我们以法国基础阶段的汉语教学大纲所包含的1555个基础汉字为研究对象<sup>[1]</sup>,以《现代常用字部件及部件名称规范》<sup>[2]</sup>(以下简称《常用字规范》)和《信息处理用 GB 13000.1 字符集汉字部件规范》<sup>[3]</sup>(以下简称《信息用规范》)这两个规范为参照,对这1555个汉字进行了全部拆解。具体情况如表 1 所示:

表 1 两个规范切分出的汉字部件数量统计

汉字范围	   汉字   数量	常规分别 规分别 同数 一件数	汉字 与部 件的 I	信息用 规分出同 时间的部 是	汉字 与部 件的 比值 II
"法纲" 汉字	1555	450	3. 46	377	4. 12
《常用字 规范》	3500	514	6. 81		
《信息用 规范》	20902			560	37. 33

可见,这两个规范在处理最初的 1555 个基础汉字时,分别使用了 450 个和 377 个不同的部件,每个部件的平均利用率只有 3 至 4 次,似乎很难达成"以部件促整体"的效果。反而是在渡过最初的难关之后,汉字和部件的比值才得到显著提升,部件的效率也才得到较好的体现。

不仅如此,对于这 1555 个基础汉字,这两个规范拆分出来的部件的使用率还非常不均衡。使用频率最高的部件"口"分别有 237 次和 273 次;只使用了1次的部件却分别有 102 个(占 22.67%)和 73 个(占 19.36%)之多,只使用了 3 次或 3 次以下的部件高达 226 个(占 50.22%)和 168 个(占 44.56%)。如果把这两个"规范"拆解出来的所有部件按使用频率从高到低依次放在横坐标上,再用纵坐标代表它们的使用次数,就能得到下面的折线图(只针对这 1555 个基础汉字)(见下页):

图 1 直观地提示我们,无论采用何种办法,汉字部件的常用度都是一种"断崖式"的分布:高频部件非常集中,就那么几十个部件(如使用频率最高的 1 号部件"口");大量的部件使用频率都非常低,但这样的部件却是绝大多数。在对汉字字形的部件拆解过程中,"一字一例"的现象非常突出,这也是造成汉字难认、难记、难学的主要原因之一。

基金项目:孔子学院建设和汉语国际教育 2018 年度课题重点项目"法国孔子学院初中级汉语教学大纲研发"(18CI01B) 作者简介:潘泰,男,武汉大学国际教育学院副教授,语言学及应用语言学博士。

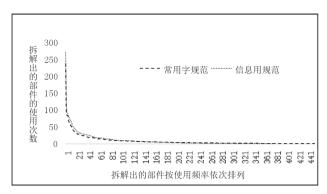


图 1 两个规范拆解出的汉字部件使用频率图示

与此同时,这样的统计结果所反映出来的现状, 也促使我们在汉字下级单位的研究中采取另外的路 径,一条必须与大多数中国人习得汉字的认知规律保 持一致的路径,否则,势必会极大地抑制汉字习得的 效果。

## 二、汉字字形拆解的系统性

将存在于现行汉字系统中的部件规律,尽可能多地通过部件拆解体现出来,这是我们对"系统性"的主要理解。同理,将"黑象禹禺革重鬼""平半夹乎伞类""用曲曲曲"等汉字或汉字构成部分都不予拆解,似乎有点儿回避问题的嫌疑,也未能满足系统性的要求。

系统性的含义还可以描述为尽量减少对于同一字形的多种拆解,尽量做到"一字一拆"。第一,现有的两个规范中都有对部件"就大不就小"的说明:"表中部件规定了汉字部件拆分的下限,一般不宜再行拆分。" [2] (P6) "表中部件没有包容关系,不得将大部件拆分为小部件。" [3] (P5) 比如,"京向食言"既然已经作为独立的成字部件,就不应将它们继续拆解为"一口小、丿门口、人、艮、、三口"等更小部件。这个原则还是应该坚持,否则所有的部件就能全部拆解为数量极其有限的笔画,这不符合部件拆解的初衷。当然,对两个"规范"将这几个字看作是独立的部件而不予拆解的作法,我们仍持保留意见。

第二,两个规范中确实还存在一些"一字多拆"的问题。即使是坚持"交重不拆"的原则,有些部件仍然会有可能导致某些汉字或部件的多种拆解。我们认为,这类部件似乎都应该继续拆解或归并进入其他部件为宜<sup>®</sup>。例如以下部件(这里用"→"表示拆解,用"/"表示"或者"):

## 三、汉字字形拆解的认知规律

在认识复杂事物的时候,人们总是会自觉不自觉 地运用比较和分类的方法。也只有当比较和分类比较 充分时,我们对该事物的认知才会更为深入。具体到

①事实上,这两个规范和目前所见的研究成果都没有对"交重不拆"有具体细致的讨论。我们认为,"重"不拆应该坚持,意思是某个笔画或部件不能在字形拆解中"重复计算",就像两个规范中都举到的例子"串"不能拆为"中中",理由是其中的"丨"被重复计算了两次。同理,"果、甫、予"等字中不能拆解出"田、用、子","家、卡、共"等字不能拆为"立豕、上下、廿一",也是"重"不拆的实例。②《常用字规范》无"力"有"戊",这样"成"的拆解就存在一定问题。《信息用规范》有"力",但同时又说明"力"是GB13000.1字符集中G列以外的汉字专用部件,这样"责"的拆解也会存疑。

对汉字字形的认知,至少有以下两个方面的原则应得 到足够的重视。

#### (一) 差异点突出原则

众多的形近部件不仅加大了汉字辨认的难度,更 为关键的是它们模糊了汉字字形的区别性特征。比如, 笔画的数量通常具有区别不同字符的价值, 但也有反 例,"良"和"朗郎"中的"良"就是如此;笔画的形态, 包括是横竖撇捺折中的哪个笔画, 带钩还是不带钩, 相对长短关系, 出头不出头等特征, 通常也具有区别 性价值,但是对于"奏"下半的第一笔、"灬"的第一笔、 "勇"的第四笔、"尖"的竖笔、"周韦"的横笔、"射" 的第二撇等,这些特征又起不到区别不同字符的作用。 这样的情况有很多,可能导致的结果是令人不清楚究 竟哪些是重要的,必须认清楚和写清楚,哪些则是可 以模糊处理的。

我们认为,要想建立汉字字形区别性特征的概念, 分立并非总是最佳方案,有选择地进行进一步拆解, 也应该是解决问题的方案之一。比如,将"子"和"予" 分立为不同部件, 还是未能很好地提示二者的差异, 仍有可能导致写起"予"来,不是少了点,就是忘了 钩, 而将"予"拆分成"一、一一"四个单笔画部件, 似乎能更好地还原母语者的认知过程。同理,将"专 年与带里"等复杂字形拆分为更小的部件,也应该是 有必要的,这样的例子还有很多。

我们对现行汉字系统中存在的高度形近的汉字或 部件进行了归纳,将其大致分为四个大类。下面,就 对这四类分别予以示例说明,其中,带下划波浪线的 不予切分,其他的则可以进行进一步拆解。

# 1. 横竖类<sup>①</sup>

王 丰 千 于 亍 土土工之 生 午 生 丰 王圭壬主 美 羊 \* 羊 **手** 乍 巨 隹 崔 雀 里 重 垂 垩 重

2) 1. 1. 1.

1 1 Ш 世 儿니 3) 上 上 止止 正疋 疋 廿 丌 升 开 井 并 # # # # 2. 撇捺类 1) 丛 丛 业 类 シ 上 小小 东 乐 东 ₩, 平 半 夹 乎 伞 立 址 <del>1</del> 严 业 亚 並 釆 小 2) 人 乂 入 义 大 文 太 丈 表 失 失 夹 夭 夬 央 **寿** 美 美 史 吏 重 3) 本 禾 木 本 末 未 朱 東東東 米 来 采 釆 耒 4) 之 廴 丘丘斥丘 乒 乓 K 艮 瓜 二 瓜 1 1 亥 豕 豸 3. 折钩类 1) L L Z 己 己 巳 **土 毛 屯 长** 氏氐 气 乞

- 九 丸 刊 凡 2) 又 叉 久 夂 攴 及 友 发 发
- 3) 戈 支 t 戋 戊 戊 戊 成 戌 威

①"廷"等字中,本应包含"壬",但有的词典(包括《现代汉语词典》[第7版])中的字形包含的是"壬",有的词典(如《中华大字典》 《中文大辞典》)包含的是"王"。"王",徐铉注曰:"人在土上。"

- 4) 
   几
   兀
   元

   九
   尤
   龙

   无
   死
   兄
   冘

- 7) 且 男
- 9) 弖 弗 弗 夷
- 10) [ E J [
- 11) 仁 夕 歹 歺 ӯ
- 12) 点 纟 幺 多マ
- 13) 韦 书 丐
- 4. 框架类
- 1) 山: 山 出 击 缶 缶 口: 中市市币 用甩 甫 用 冉 而 舟 Ŋ 丹 册 内 肉 内 占 崔 首 ,EI, (周) 见 贝 页

之: 足 足 良 尺 严 戶 目 自 艮 五 丑 # 그

2) 旦旦旦迫百百 申

且耳且其曲曲曲

西西西

臼 囟 囱 卤

母 中 田

**坦** 里 电 电 电 車 里

需要指出的是, 在拆分的时候, 一般是选择常见

的保留,不常见的拆解。我们不妨加以举例说明,具体如表 2 所示:

表 2 形近部件的拆解示例

水 ∥ 冰 → ∫ 〉 〈 (绿)	儿 ∥ 九 → 丿乚(免)
手    爿 → 冫   (北) (壮)	月 川 玥 → □‡ (那)
己    巳 → ¬L (包) 已 → ¬L	弓 ‖ 弗 → 弓川 (费) 用→弓   丿 (弟第)
衣   农 → ¬ 化 展 → 尸 比	以    氐 → 丿 レ 、 (低) 氐 → 丿 以 (旅)
タ	天    矢 → J 天 夫 → 二人 失 → J 二人
羊	ム ‖ V → L 、(去) (以) よ → I ー 、(遇) 『 → J レ 、(留) L → L 、(瓜)
木     □     本     →     木       末     →     十八     未     →     木       束     →     木口     束     →     木口(刺)	

在表 2 中," | " 左边的部件不拆解," | " 右边的部件可以继续拆解。需要指出的是," 已" " 已" 的拆分方法一样,但" 已" 不会作为部件出现在其他汉字当中。" 天" 作为部件的常用度要低于" 矢",但" 天" 却是最先学习的基础汉字之一,远比独用的" 矢" 常见得多。" 水" 和" 水" 的道理也一样,大概很少会有人先接触" 水"这个部件,再去习得" 水"这个整字。

这种在部件拆解时的"人为"干预,可能比自然 形成的"分隔沟"更具有认知价值<sup>[4]</sup>。通过拆解,人 为突出形近部件的差异,可以逐步建立部件之间的区 别性特征,符合中国人识别汉字时先轮廓、后细节的 一般心理过程<sup>[5]</sup>。

### (二) 拓扑相似性原则

还必须认识到,大多数情况下,我们的"用户"

是具有高度形象思维能力和抽象思维能力的人,而不是机器。某些部件在表面上有一定的差异,但其实都蕴含着相同的笔画和笔画关系,就像拓扑学所说的"几何图形在连续改变形状时还能保持不变"一样,可以认为是同一个部件。我们把这样具有图形近似性的汉字构成元素,视作是不具有区别性意义的相同部件,而合并为同一部件,同时以其独立使用或最常见的形式作为主部件(在":"左边),其余的作为附形部件(在":"右边)。两个规范在这个问题上处理得有些不够理想,可以进一步归并的至少还有以下部件:

#: #(贲):

+(半/棒): +(那/判);

王: チ (玩);

己: 己(改);

丰: ‡(邦), 孝(寿);

羊: 差(着), 羊(翔);

手: 手(看), 手(拜);

丁:丁(可);

| (临/竖): || (面), || (费), || (师), л (齐/鼻), л (肃):

止:止(延);

木: 木(杂);

→ (买): 「(候/书), ¬(敢), ノ(登), ヿ (丑):

七: 七(化), 弋(民/绕), 七(长), 左(东/练), 七(切), 划(曳);

卩(即): 卩(节), 阝(报);

 $Z: \mathsf{T}(\mathsf{T}/\mathsf{T});$ 

几:几(朵),几(风);

月:目(背), 夕(然);

□ (同): □ (周/用), □ (奂/奥), □ (勇);

<sup>′′</sup> (负): <sup>′′</sup> (尔), <sup>′</sup>⁄′ (久);

日: 曰, □ (冒)。

还有一些部件也是可以归并的:

〈(飞):〈(永/隶/聚),〈(率/兆);

(冰):ン(匀/习/弱),ン(求);

乂(义): 人(更/史), X(凶);

△(合): △(今);

以(展): 反(很);

□(医):□(既/舞/降);

廴: 支(及);

之: 辶(进);

豕: 彖(象)。

同时,作偏旁所导致的字形变化大多也应该归并,如"匕儿工爿巳子穴雨"和"顷顽攻藏顾孩空露"等字中的形近部件。

差异突出和相似归并其实是同一个问题的两个方面,最终的目标仍然是建立系统的汉字部件区别性特征。就此而言,学界同仁今后还应继续努力,以揭示汉字、部件及笔画的哪些特征在哪些场合具有区别不同字符作用的规律。同时,我们所理解的汉字字形的拆解并非是简单的"加减法",需要在拆解的过程中体现辨认(recognize)和复写(reform)汉字的认知规律。

### 四、结语

需要说明的是,将复杂的字形与形近的字形进行 进一步拆解,并非一定会"破坏"汉字的字源理据。 首先,通过字形来获取汉字的音、义信息,并不是一 件容易的事,也许早在许慎创作《说文解字》的时代 就已经是非常困难的了。因此, 在《说文》之外, 还 有以义为核心(以《尔雅》为代表)和以音为核心(从《仓 颉篇》到《切韵》等)的字书传统[6]。这三大传统发 展到如今,反而是以形求音义的《说文》类型最为薄弱。 汉字的符号性也许早就超越了理据性, 现行汉字的字 形理据亟需理论与实际相结合的研究。其次,对于有 些汉字或部件来说, 进一步拆解更能体现其字源理据。 如上文提到的"⇒",还有从"尺、首、金、夂、美" 中拆解出的"尸、自、丛土、夕、羊"等,都能更好 地体现部件提示整字语音和语义的理据。当然,当字 源与字形相冲突时, 应以现行字形为准则, 这也是包 括两个规范在内的大多数人的共识。

基于上述汉字部件拆解的新方案,我们对1555个"法纲"汉字进行了第三次部件切分,结果如表3所示:

表 3 新方案与两个规范切分出的汉字部件数据比较

汉字数量	《常用字 规范》每 个汉字的 平均部件 数	《信息用 规范》每 个汉字的 平均部件 数	新年字的部件数	新案不部总	汉 与 部 的 住 III
1555	2. 4157	2. 6914	3. 0618	199	7. 81

从表 3 可以看出,按照本文所建议的新方案切分出来的每个汉字的平均部件数量(3.0618)较之《常用字规范》(2.4157)和《信息用规范》(2.6914)有一定程度的上升,但并不很大。这体现出新方案切分出的汉字部件的颗粒度变得更为细致,但并不"细碎",没有发生令人担心的与笔画混同起来的现象。最为显著的是,在这 1555 个最基础的汉字中,新方案的每个部件的使用频率(7.81次)比《常用字规范》(3.46次)和《信息用规范》(4.12次)有了非常大的提升,大致翻了一番。同时,新方案部件的使用频率折线图(以频率最高的前 50 个部件为例)也变得平缓了一些。具体如图 2 所示(仅以 1555 个基础汉字为范围):

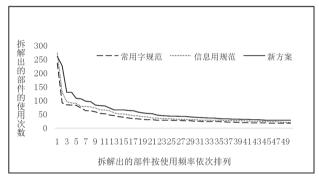


图 2 新方案与两个规范拆解出的汉字部件使用频率(局部)比较

这说明我们用较少的部件组成了较多的汉字,部件的重复利用率得到了较大幅度的提高,高频部件的分布也更加均衡合理,能够助推实现以旧知拓展新知的设想,同时也符合一般性的认知规律。

当然,由于汉字问题的复杂性,不可能一蹴而就 地解决所有的问题。新方案也不是完美无缺的,也许 还有对部件拆分系统性考虑不够周全的地方,但本文 的思路应该更容易得到识字的心理过程的验证,也许 还能起到抛砖引玉的作用,激发更多人的研究兴趣。 此外,还有少量汉字,不能仅仅依靠部件拆解的方式 来促进习得,如"员"和"呗"、"只"和"叭"、"束" 和"杏"、"杲"和"杳"等,这些问题仍需持续深入 研究。 附录:

新方案基础汉字部件表(202个)

1画(15):

ー I J 、 \ ¬ フ Z L J L L

4 4 0

2 画 (45):

一 亻 丿 厂 八 〈 人 乂 ′ ′ ′ ′ ′ L 儿

几 九 (14);

→ / / → i ∨ (6):

三 丰 干 ; 土 士 工 艹 廾 大 扌 オ 寸 (13);

₩ ↑ □ □ 山 巾 (6);

卢川ル 彳 多 丛 个 犭 夕 久 饣 (11);

广 亡 门 氵 ሣ ウ 之 (7);

<del>□</del> ∃ ⇒ 尸 弓 己 也 女 汤 小 以 子

纟 马 乡 幺 (16)

4 画 (43):

丰王丰天下乡井廿五木犬车

戈 牙 瓦 (15):

止 日 中 贝 见 (6);

牛 手 毛 攵 片 斤 爪 爫 月 欠 农(11);

文 方 火 灬 户 ネ 心 (7):

巴 爿 水 小(4)

5 画 (25):

**寿** 甘 石 戊 (4);

目且田由甲申皿四(8);

**车 生 禾 白 鸟 (5)**;

疒 立 头 穴 衤 (5);

皮 孕 母 (3)

6画(13):

耳西页 电虫竹 \*\* 自 日 舟 亥 羊 米 7 画 (6):

豆豕足我身豸

8画(2):

雨 隹

#### 参考文献:

- [1]Ministère de l'Éducation nationale (法国国家教育部). Chinois, classe de seconde générale et technologique (普通及技术类高中第一学年汉语大纲)
  [DB/OL]. https://www.education.gouv.fr/bo/2002/hs7/default.htm,3 octobre 2002,2002-10-03.
- [2] 中华人民共和国教育部. 现代常用字部件及部件名称规范 [DB/OL]. http://www.moe.gov.cn/ewebeditor/uploadfile/2015/01/13/20150113090318445.pdf, 2009-03-24.
- [3] 中华人民共和国教育部. 信息处理用 GB13000.1字符集汉字部件规范 [DB/OL]. http://old.moe.gov.cn/ewebeditor/uploadfile/2015/01/12/20150112165337190.pdf, 1997-12-01.
- [4] 苏培成. 现代汉字的部件切分[J]. 语言文字应用, 1995, (3).
- [5] 胡平. 汉语儿童识字的心理机制及其给教育的启示 [J]. 华东师范大学学报(教育科学版),2000,(1).
- [6] 张滌华.论《康熙字典》[J]. 江淮论坛, 1962, (1).

# A New Way to Dispart Chinese Characters

#### Pan Tai

(School of International Education, Wuhan University, Wuhan 430070, China)

Abstract: In the process of disparting Chinese characters, the best method is not always to persisting in the traditional theory which requires to keep crossing storks or parts in wholes or set different parts as independent ones. A new way to dispart Chinese Characters systematically this paper suggested is to try to show the differences of similar parts if they are valuable to distinguish different characters as more as possible while merge them together if they are not. This new method accords with the cognition of Chinese characters better and probably cannot give consideration to Chinese etymology within every case.

Key words: Chinese characters; parts; systematicness; cognition