

# 现代汉字部件构形特征的计量研究

侯冬梅<sup>1,2</sup>

(1. 河南科技学院 文法学院, 河南 新乡 453003;

2. 华中师范大学 语言与语言教育研究中心, 湖北 武汉 430079)

**摘要:** 基于数据库, 对现代汉字部件构形特征进行了定量分析, 明确了现代汉字成字和非成字部件在外部构形方面的共性和个性特征, 分析了它们在认知方面的优势: 成字部件在组配能力、分布层次两方面比非成字部件容易认知; 非成字部件在分布位置、直接组合部件两方面比成字部件容易认知。

**关键词:** 现代汉字部件; 构形; 定量分析

中图分类号: H195

文献标志码: A

文章编号: 1673-6060(2016)01-0037-05

现代汉字, 又叫现代汉语用字, 是记录现代汉语的字<sup>[1]</sup>。《通用规范汉字表》中的 8105 个字是现代汉字的典型代表。由于汉字以平面作为信息分布形式, 因此汉字字形结构具有复杂性<sup>[2]</sup>。为了便于人们学习和书写汉字, 分析汉字字形是很有必要的。

## 一、现代汉字部件的确定

汉字的数量并不是越多越好。为了提高汉字应用效率、减少汉字学习难度, 有必要对汉字进行定量研究。《通用规范汉字表》就是对现代汉字定量研究的最新成果和规范标准。其一级字表“主要满足基础教育和文化普及的基本用字需要”<sup>[3]</sup>。根据字频统计, 覆盖率超过 99.5%。因此, 笔者主要对一级字表中基础、高频汉字的字形进行分析。

### (一) 现代汉字部件拆分依据和规则

1. 现代汉字字形分析依据——直观构形分析 汉字表意特征突出, 目前多以“字理”为依据, 对现代汉字字形进行分析, 分析汉字的构形单位——部件和笔画。部件是汉字的直接构形单位, 笔画是汉字的间接构形单位。部件在现代汉字构形中居于核心地位, 是连接笔画和整字的桥梁纽带。因此, 笔者重点研究部件在现代汉字构形中呈现出的特征。

汉字经历隶变和“简化”后, 构字理据或发生变

化, 或丧失, 致使以“字理”为依据的现代汉字字形(以 3500 个常用字为拆分基础)分析, 至今难以确定现代汉字部件的数量。笔者不采用这些现有成果, 而是在吸取汉字构形学理论和方法的基础上<sup>[4]</sup>, 从直观角度对现代汉字字形进行“直观构形”分析。因为汉字演变过程中虽然字理会变化、丧失<sup>[5]</sup>, 但其组造规律却深深烙在汉字字形中, 形成了固定的组块或沟壑, 这些组块或沟壑, 不但从直观上是可视的, 而且具有普遍的认知基础。

2. 现代汉字部件拆分规则 从直观构形角度, 对现代汉字进行构形分析的规则如下。

第一, 结构简单, 依字形按沟壑取组块拆分。例如:

江——拆分为“氵、工”, “氵”组配汉字时, 以组块“氵”形式整体组造, 不再拆分。

第二, 结构复杂, 依字形按结构层次, 取整字优先, 从大到小拆分。例如:

绷——第一层次分为“纟、朋”, 第二层次“朋”再拆分为“月、月”; 鸿——第一层次拆分为“江、鸟”, 第二层次“江”再拆分为“氵、工”。

第三, 多沟壑均可拆分时, 依字形按系统拆分。例如:

章——第一层次拆分为“音、十”, 第二层次

收稿日期: 2015-12-13

基金项目: 2015 年国家语委“十二五”科研规划项目“现代简化汉字构形体系理论和应用研究”(YB125-161); 2015 年河南省教育厅人文社科项目“现代汉字部件构形特征计量研究”(2015-QN-081)

作者简介: 侯冬梅(1976—), 女, 河南方城人, 讲师, 在读博士, 主要从事应用语言学的研究。

“音”再拆分为“立、日”；意——第一层次拆分为“音、心”，第二层次“音”再拆分为“立、日”；竟——第一层次拆分为“音、儿”，第二层次“音”再拆分为“立、日”。

第四，拆分出来的基础部件，笔画数一般不小于 2，“一、乙”除外。例如：

引——独体字，不拆分为“弓、丨”；旦——拆分为“日、一”；亿——拆分为“亻、乙”；仁——拆分为“亻、二”。

第五，笔画交叠不拆分。例如：

果——不拆分为“田、木”；串——不拆分为“中、中”。

第六，拆分出来的部件，具有可替换性，否则不予拆分。例如：

虎——拆分为“虍、几”，“七”位置不具有替换性，“虍”不再拆分。

(二) 现代汉字部件的类型

比笔画大的，具有直接或间接组配汉字功能的现代汉字构形单位部件，根据不同的分类标准，可以分为不同的类型。根据复杂程度可以把部件分为基础部件和复合部件，由笔画直接构成的部件是基础部件，有两个或两个以上基础部件构成的部件是复合部件；根据能否单独成字，又可以把部件分为成字部件和非成字部件，能单独构成整字的部件是成字部件，不能单独构成整字的部件是非成字部件；根据组配汉字时部件的分布层次，可以把部件分为一级部件、二级部件、三级部件。例如图 1，《通用规范汉字表》中第 3045 字“缝”所包含的基础部件有“纟、辶、夂、丰”，复合部件有“逢、夆”；成字部件有“丰、逢”，非成字部件有“纟、辶、夂、夆”；一级部件有“纟、逢”，二级部件有“辶、夆”，三级部件有“夂、丰”。

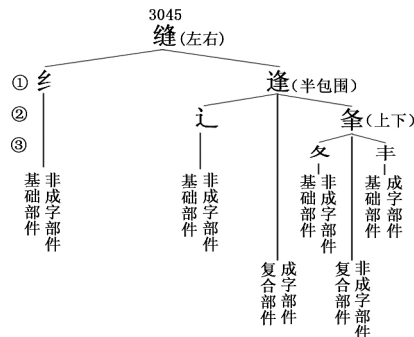


图 1 汉字部件类型

目前学界通常所说的部件，一般专指基础部件；有关部件的研究成果，也多针对基础部件。笔者所研究的现代汉字部件专指基础部件。

(三) 现代汉字部件的数量

从直观构形角度拆分出来的现代汉字部件有

457 个。其中成字部件有 247 个，非成字部件有 210 个。成字和非成字部件数量基本相当，汉字部件构形研究，不仅要重视成字部件研究，也不能忽视非成字部件研究。

部件的构形特征指部件内部构形和外部构形特征。内部构形主要指部件笔画数、笔画位置关系等特征；外部构形主要包括部件的组配能力、分布位置、分布层次、直接组合部件等特征。本文主要对汉字部件的外部构形特征进行研究。

二、现代汉字部件的组配能力及其计量

(一) 现代汉字部件的组配能力

现代汉字部件的组配能力是指部件组配成复合部件或整字的能力。部件组配能力的强弱可以用部件在汉字构形中出现的频次来衡量，如部件出现频次高则部件组配能力强，部件出现频次低则组配能力弱。

(二) 现代汉字部件的组配能力计量

经检验，由于 457 个部件中的成字和非成字部件出现频次，都是  $p = 0.00 < 0.05$ ，因此它们都是偏态分布数据。由表 1 可知，部件之间出现频次有明显差异，单个部件出现频次从 1 ~ 611 不等，成字部件最低出现频次为 2，非成字部件最低出现频次为 1；成字部件累计频次(5765)明显高于非成字部件的累计频次(3781)。由于成字部件、非成字部件都存在  $Mo < Me < X$  的特征，因此它们的出现频次都是右偏态分布数据，存在极大值。成字部件的极大值是 611，非成字部件的极大值是 209。而成字、非成字部件出现频次的峰值( $Mo$ )分别是 3 和 2，因此它们又是呈正偏态分布，峰值都偏向数值小的一侧。由于成字部件的累计出现频次高于非成字部件，且成字部件  $Me$  和  $Mo$  均大于非成字部件，因此成字部件组配能力相对强于非成字部件的组配能力。

表 1 部件出现频次

部件类型	频次特征				
	频次范围	累计频次	中位数 (Me)	众数 (Mo)	均值 (X)
成字部件	2 ~ 611	5765	8	3	23
非成字部件	1 ~ 209	3781	6	2	18

根据表 2，这种正偏态分布的具体情况是：8.5% 的成字部件和 24.8% 的非成字部件出现频次不超过 2；50% 左右的成字部件出现频次在 1 ~ 8 之间；50% 左右的非成字部件出现频次在 1 ~ 6 之间；80% 左右的成字和非成字部件出现频次均在 1 ~ 25

之间;90%左右的成字部件出现频次在1~57之间;90%左右的非成字部件出现频次在1~45之间;95%左右的成字部件出现频次在1~89之间;95%左右的非成字部件出现频次在1~78之间。所以,无论成字还是非成字部件,它们各自80%的个体,出现频次均在25以内,而它们的累计频次仅为2626(部件累计频次为9546)。由此可知,部件的组配能力整体不强。

表2 部件出现频次正偏态分布情况

成字	比例/%	8.5	51.0	70.0	80.2	90.7	95.1
基件	频次	2	1~8	1~14	1~25	1~57	1~89
非成字	比例/%	24.8	54.3	70.0	80.5	90.0	95.2
基件	频次	2	1~6	1~13	1~25	1~45	1~78
	累计频次	—	—	—	2 626	—	—

总之,就整体而言,成字和非成字部件的组配能力都不强:不但峰值极小,分别为 3 和 2,并且各自 80% 的个体出现频次均在 25 以内,其累计出现频次 (2626) 只占部件总出现频次的 27.51%。它们都存在极大值,成字部件“口”组配能力最强,非成字部件“+”组配能力最强。但相对而言,成字部件比非成字部件的组配能力稍强。就其个体而言,部件出现频次分布在 1~611 之间,不同部件出现频次不同,存在差异。

### 三、现代汉字部件的分布位置及其计量

### (一) 现代汉字部件的分布位置

现代汉字部件的分布位置是指在现代汉字构形中,部件和直接组合部件组合时所处的位置。可以用汉字结构方式来对基础部件分布位置进行描述。现代汉字结构方式分为上下、左右、半包围、全包围、穿插等五种基本方式。比如:图1“缝”字部件“纟”的直接组合部件是“逢”,它们之间是左右的结构方式,所以“纟”分布位置就是“左”。

## (二) 现代汉字部件的分布位置计量

总体而言,部件和直接组合部件组配成的复合部件或整字中,上下结构的最多,有4627个,左右结构有3340个,半包围结构有1028个,穿插结构有111个,全包围结构有65个。就部件个体而言,同一部件可能分布在一种或几种结构中,即使分布在同一种结构中,也可能分布在不同的位置上。根据表3可知,部件及其直接组合部件之间最少存在1种结构方式,最多存在5种结构方式;与直接组合部件之间存在1种结构方式的部件最多(230个),存在5种结构方式的部件最少(4个);成字部件与其直接组合部件之间存在的结构方式相对较复杂(为

2~5种),非成字部件与其直接组合部件之间存在的结构方式相对较简单(多为1种)。

表3 部件与直接组合部件结构方式种类统计

结构方式种类	一种	二种	三种	四种	五种
成字部件数量	76	97	56	14	4
非成字部件数量	154	38	16	2	0
合计	230	135	72	16	4

由表 4 可知,只分布在一种结构方式中的 230 个部件,有 205 个的分布位置固定,占有所有部件的 44.86%;其中成字部件 62 个,非成字部件 143 个。分布位置固定的部件,非成字部件明显多于成字部件。

表 4 分布位置固定的部件统计

部件类型	分布位置固定的部件数量	代表部件
成字部件	62	一血习勺革止凡巨电身长 更及司西朱买垂卑隶万义 叉专
非成字部件	143	㇀㇁㇂㇃㇄㇅㇆㇇㇈㇉ ㇊㇋㇌㇍㇎㇏㇐㇑㇒㇓ ㇔㇕㇖㇗㇘㇙㇚㇛㇜㇝
合计	205	—

总之,部件与直接组合部件之间以上下结构方式组合的最多(4627个),以全包围结构方式组合的最少(65个)。44.86%的部件分布位置固定。非成字部件与直接组合部件之间的结构方式相对简单(73.33%的非成字部件结构方式只有1种),成字部件与其直接组合部件之间的结构方式相对复杂(69.23%的成字部件结构方式为2~5种)。分布位置固定的部件多为非成字部件(68.10%的非成字部件分布位置固定)。

#### 四、现代汉字部件的分布层次及其计量

### (一) 现代汉字部件的分布层次

在现代汉字构形中,部件按照不同的结构方式分层、逐级组配成复合部件或整字,其中部件所在的层次就是部件的分布层次。如图1“缝”的部件“纟、讠、夂”分别分布在汉字构形树的第一、第二、第三层次。

## (二) 现代汉字部件的分布层次计量

现代汉字构形中 整字最多有 5 个层次。由表 5 可知 部件分布层次可以分为简单分布层次(Ⅰ类)和复杂分布层次(Ⅱ类)两大类。就总体部件而言, 155 个部件属于Ⅰ类,分为 A、B、C 三种类型,只分布在第一层次(A 种),只分布在第二层次(B 种)和只分布在第三层次(C 种); 302 个部件属于Ⅱ类,分为 6 种类型,都是分布在 2 个以上的多个层次中,分别

是分布在第一和第二层次( D 种) ,分布在第一、第二和第三层次( E 种) ,分布在第一、第二、第三和第四层次( F 种) ,分布在第一、第二、第三、第四和第五层次( G 种) ,分布在第二和第三层次( H 种) ,分布在第二、第三和第四层( J 种) 次。就成字和非成字部件而言 ,I 类部件中 ,成字部件 96 个 ,包含两种分布类型 ,其中 A 种 94 个 ,B 种 2 个;非成字部件 59 个 ,包含三种分布类型 ,A 种 47 个 ,B 种 11 个 ,C 种 1 个。I 类部件 ,A 种最多 ,有 141 个 ,其中成字部件

94 个 ,非成字部件 47 个;C 种最少 ,只有 1 个非成字部件。II 类部件中 ,成字、非成字部件各有 151 个 ,其中成字部件包含四种分布类型 ,D 种 79 个 ,E 种 52 个 ,F 种 18 个 ,G 种 2 个;非成字部件包含六种分布类型 ,D 种 82 个 ,E 种 41 个 ,F 种 14 个 ,G 种 3 个 ,H 种 9 个 ,I 种 2 个。II 类部件 ,D 种最多 ,有 161 个 ,其中成字部件有 79 个 ,非成字部件有 82 个;I 种最少 ,只有 2 个非成字部件。

表 5 部件分布层次统计

部件分布 层次	简单分布层次( I )			复杂分布层次( II )					
	1 层( A)	2 层( B)	3 层( C)	1~2 层( D)	1~3 层( E)	1~4 层( F)	1~5 层( G)	2~3 层( H)	2~4 层( J)
成字部件	94	2	0	79	52	18	2	0	0
非成字部件	47	11	1	82	41	14	3	9	2
合计	141	13	1	161	93	32	5	9	2
比例 / %	33.92			66.08					

总之 ,部件层次分布类型可以分为两大类九小种 ,一个部件最少分布在 1 个层次中 ,最多分布在 5 个不同的层次中。66.08% 的部件属于 II 类 ,部件均有 2 个以上的分布层次。II 类部件中 ,53.31% 的分布在第一和第二层次 ,30.79% 的分布在第一、第二和第三层次 ,10.60% 的分布在第一、第二、第三和第四层次。33.92% 的部件属于 I 类 ,部件均只有 1 个分布层次。I 类部件中 ,90.97% 的只分布在第一层次;97.92% 的 I 类成字部件都分布在第一层次。成字部件分布层次( 两类 6 种) 比非成字部件分布层次( 两类 9 种) 简单。

五、现代汉字部件的直接组合部件及其计量

(一) 现代汉字部件的直接组合部件

现代汉字构形中 ,可以和部件直接组合 ,构成复合部件或整字的部件是该部件的直接组合部件。例如图一“缝” ,其部件“纟”的直接组合部件是“逢” ,部件“辶”的直接组合部件是“辵” ,部件“攴”的直接组合部件是“丰” 。和部件直接组合的部件数量就是这个部件的直接组合部件个数。部件的直接组合部件个数各有不同。

(二) 现代汉字部件的直接组合部件计量

成字和非成字部件的直接组合部件个数 ,都是  $p=0.00<0.05$  ,是偏态分布数据。由表 6 可知 ,部件的直接组合部件个数从 1~192 不等 ,成字和非成字部件的直接组合部件个数最小都是 1;成字部件直接组合部件个数累计 2690 个 ,非成字部件直接组合

部件个数累计 2065 个;由于成字、非成字部件的直接组合部件个数都存在  $Mo < Me < X$  的特征 ,因此 ,它们都是右偏分布数据 ,存在极大值。成字部件直接组合部件个数的极值是 177、峰值是 2 ,非成字部件直接组合部件个数的极值是 192、峰值是 1。由于成字部件直接组合部件累计个数高于非成字部件直接组合部件累计个数 ,且成字部件  $Me$  和  $Mo$  均大于非成字部件 ,因此 ,相对而言 ,成字部件直接组合部件比非成字部件直接组合部件复杂。

表 6 直接组合部件个数统计

部件类型	直接组合部件个数 变动范围	累计 个数	中位数 ( Me)	众数 ( Mo)	均值 ( X)
成字部件	1~177	2690	5	2	11
非成字部件	1~192	2065	3	1	10

根据表 7 ,正偏态分布的具体情况是: 10.5% 的成字部件和 30.0% 的非成字部件直接组合部件数是 1; 50% 左右的成字部件直接组合部件数在 1~5 之间 ,50% 左右的非成字部件直接组合部件数在 1~2 之间; 80% 左右的成字部件直接组合部件数在 1~13 之间 ,80% 左右的非成字部件直接组合部件数在 1~8 之间; 90% 左右的成字部件直接组合部件数在 1~26 之间 ,90% 左右的非成字部件直接组合部件数在 1~19 之间; 95% 左右的成字部件直接组合部件数在 1~40 之间 ,95% 左右的非成字部件直接组合部件数在 1~51 之间。那么 ,90% 左右的成字部件直接组合部件数都在 26 以内 ,90% 的非成字部件直接组合部件数在 19 以内 ,它们直接组合部件的累计个数为 414( 部件累计个数 4755) 。因此 ,部件的直接组合

部件数相对较少。

表7 部件直接组合部件正偏态分布情况

成字 基件	比例/%	10.5	52.2	71.3	80.2	90.7	95.1
	直接组合部件个数	1	1~5	1~9	1~13	1~26	1~40
非成字 基件	比例/%	30.0	49.5	68.6	80.5	90.5	95.2
	直接组合部件个数	1	1~2	1~5	1~8	1~19	1~51
	累计个数	—	—	—	—	414	—

总之,成字和非成字部件的直接组合部件个数均呈偏态分布。就整体而言,直接组合部件都不多:不但峰值极小,分别为2和1,而且90%成字部件和非成字部件的直接组合部件个数累计仅为414次,占部件直接组合部件数的8.70%;成字和非成字部件都存在极大值,成字部件“口”直接组合部件最多,非成字部件“丿”直接组合部件最多。但非成字部件直接组合部件相对简单(30%非成字部件直接组合部件个数为1,且90%非成字部件直接组合部件个数在19以内)。就其个体而言,部件直接组合部件个数,分布在1~192之间,不同部件的直接组合部件又有一定差异。

## 六、现代汉字部件构形特征定量研究的意义

通过对现代汉字部件构形特征的定量研究,笔者发现:

(一) 部件在外部构形方面具有以下特征

(1) 部件组配能力不强。部件出现频次分布在1~611之间,呈偏态分布,有极大值,但峰值极小。

(2) 部件可以分布在5种不同的结构中,近一半部件的分布位置固定。

(3) 部件最少只分布在1个层次上,最多可以分布在5个不同的层次上。部件层次分布可以分为两大类九小种,多数部件分布层次复杂。

(4) 部件直接组合部件不多。部件直接组合部件个数分布在1~192之间,呈偏态分布,有极大值,但峰值极小。

(二) 成字和非成字部件在外部构形特征方面存在一定的差异

(1) 组配能力方面,成字部件的组配能力稍强于非成字部件的组配能力。

(2) 分布位置方面,非成字部件与其直接组合部件之间的结构方式相对简单,非成字部件位置较成字部件位置固定。

(3) 分布层次方面,成字部件分布层次(两类6种)比非成字部件分布层次(两类9种)简单。

(4) 直接组合部件方面,非成字部件比成字部件

直接组合部件简单。

(三) 成字和非成字部件特征研究,应受到同等重视,二者不可偏废

正因为在外部构形特征方面,成字和非成字部件比较,既有优势也有劣势。并且成字和非成字部件不是对立概念,而是相对概念。所以,在部件研究中,既要重视成字部件研究,也不能忽视非成字部件研究。

(四) 为未来用定量数值形式评价部件的认知难度奠定基础

由于部件组配能力越强,越容易识记;部件分布位置越简单,越容易识记;部件分布层次越简单,越容易识记;部件直接组合部件越简单,越容易识记。并且成字、非成字部件及其个体之间在组配能力、分布位置、分布层次、直接组合部件等方面存在一定差异。所以,我们可以制作统计量表,在外部构形特征的这四个维度上,分别分成不同的水平,并按水平差异给每一维度上的特征赋值。然后,综合四个维度,从外部构形特征方面计算出每一个部件的认知难度数值。本文的部件构形特征研究结论,可以为每一个部件认知难度的多维度量化评价奠定基础。同时,这种部件认知难度的量化评价方式也为部件认知难度研究提供了一个新的视角。

综上所述,通过对部件外部构形特征的量化研究,我们明确了部件在外部构形方面的特征:组配能力不强,近半数(44.86%)分布位置固定,多数部件(66.08%)分布层次复杂,直接组合部件不多等。但相对而言,成字部件在组配能力、分布层次两方面具有认知优势,非成字部件在分布位置、直接组合部件两方面具有认知优势。因此,汉字部件研究中,既要重视成字部件研究,也不能忽视非成字部件研究。

## 参考文献:

- [1] 苏培成. 现代汉字学纲要[M]. 北京: 北京大学出版社, 1994: 21.
- [2] 周健. 汉字教学理论与方法[M]. 北京: 北京大学出版社, 2007: 16.
- [3] 王宁. 《通用规范汉字表》解读[M]. 北京: 商务印书馆, 2013: 11.
- [4] 王宁. 汉字构形学讲座[M]. 上海: 上海教育出版社, 2002.
- [5] 李轶. 文字与构字偏旁发展的不一致性与文字理据的丧失: 以夕月肉偏旁的演变为例[J]. 中国文字研究, 2012(1): 163-167.

(责任编辑: 王凤玲)