

收稿日期:2022-05-31

《通用规范汉字表》不成词语素字统计与思考

——兼谈汉字三记说和汉字的性质

邵 霽 吉

(盐城师范学院 文学院,江苏 盐城 224002)

摘要:统计表明,《通用规范汉字表》8 105 字中有非语素字 1 605 个,不成词语素字 771 个,可成词语素字 5 729 个。汉字具有记录汉语单音节词、记录汉语合成词中单音节语素、记录汉语多音节单纯词中的音节等三项功能。如果着眼于“用一个汉字记写什么”,则汉字的性质可以表述为:汉字是记录汉语单音节词、合成词中单音节语素、多音节单纯词中音节的书面符号系统。如果不局限于一个汉字记什么,则汉字的性质也可以表述为:汉字是单字记录汉语单音节词、多字组合记录汉语多音节词的书面符号系统。

关键词:《通用规范汉字表》;非语素字;不成词语素字;汉字性质

中图分类号:H109.4

文献标识码:A

文章编号:1003-6873(2022)05-0068-08

作者简介:邵 霽 吉(1947—),男,江苏盐城人,盐城师范学院文学院教授,主要从事汉语言文字研究。

DOI:10.16401/j.cnki.ysxb.1003-6873.2022.05.067

中国社会科学院语言研究所词典编辑室《现代汉语词典》第 5、6、7 版“凡例”皆指出:“单字条目在现代汉语中成词的标注词类,不成词的语素和非语素字不做标注。”^{[1]凡例⁵}由此可知,汉字有“成词的”“不成词的语素”和“非语素字”之分。笔者在《〈通用规范汉字表〉非语素字统计与思考》^[2]一文中,已经对《通用规范汉字表》中的非语素字进行了一次统计,本文接着对《通用规范汉字表》中的不成词语素字进行统计,进而得出《通用规范汉字表》中可以“成词的”字的数量,最后对汉字的性质提出自己的一些看法。

一、不成词语素字和不成词语素

不成词语素字跟不成词语素是两个既有联系又略有区别的概念:它们都是“不成词”的,但不成词语素字以“字”为考察对象,而不成词语素则以“语素”为考察对象。

我们所说的“不成词语素字”,其内涵有三:第一,它必须是一个字;第二,它必须有义项,因而是一个语素;第三,它的所有义项都不能够独立成词。

我们以《通用规范汉字表》中 8 105 个字为考察对象,以《现代汉语词典》中的单字条目及其释义为依据,除去被我们认定为非语素字的字,其余剩下的字,凡在《现代汉语词典》中作为条目

给出了义项的,不管它只有一个义项还是拥有几个义项,只要其全部义项都不可以独立成词,我们就把它认定为一个不成词语素字。如果它的义项中有一个可以独立成词,那我们就不认为它是不成词语素字。例如:

“氛”有 1 个义项,这个义项不可以成词,所以,它是一个不成词语素字。

“浩”有 2 个义项,这两个义项都不能成词,所以,“浩”也是一个不成词语素字。

“情”有 6 个义项,都不可以成词,所以,“情”也是一个不成词语素字。

“谈”有 3 个义项,其中义项①③可以独立成词,义项②不能独立成词,由于其中有成词的义项,所以,我们不把它归入不成词语素字。

“民”有 5 个义项,义项①②③④都不能独立成词,仅义项⑤可以独立成词,因而我们也认为它不是不成词语素字。

“朝”有 2 个读音,一读“cháo”,一读“zhāo”,读“cháo”时有 7 个义项,读“zhāo”时有 2 个义项,总共 9 个义项,由于读“cháo”时有 3 个义项可以独立成词,综合考虑,我们认为“朝”是可成词语素字,而不是不成词语素字。

“略”在《现代汉语词典》中被分立为“略¹”“略²”“略³”3 个条目,其中“略¹”可以成词,“略²”(用于“方略、策略、谋略”等词中)和“略³”(用于“侵略、攻城略地”等词中)不可以成词,但“略”在《通用规范汉字表》中是一个字,我们认为它是可成词语素字,而不是不成词语素字。

“不成词语素”的内涵有二:第一,它必须是一个语素;第二,它必须不能单独成词。当然,一个汉语语素也不限于只用一个汉字来表示。

如果一个汉字在辞书中被列为一个条目,一个义项,那么认定它是一个语素,这应该没有不同意见。但如果一个汉字在词典中被列为几个条目,或有多个义项,那么认定它是一个语素还是几个语素,则很可能会有不同意见,有人可能把每一个有义项的单字条目认作一个语素,也有人可能只把单字条目下的每一个义项认作一个语素。

我们暂且把每一个有义项的单字条目视为一个语素,那么,上面说到的“氛、浩、情、谈、民”在《现代汉语词典》中各是一个条目,就各是一个语素,“氛、浩、情”是不成词语素,“谈、民”是可成词语素。这跟我们前面把“氛、浩、情”认定为不成词语素字、把“谈、民”认定为可成词语素字基本上是对应的。而“朝”在《现代汉语词典》中列为两个条目,可能被视为两个语素,读 cháo 的“朝”有义项可以独立成词,所以它是一个成词语素;读 zhāo 的“朝”没有义项可以独立成词,所以它是一个不成词语素。“略”在《现代汉语词典》中被列为 3 个条目,可能被视为 3 个语素,其中“略¹”是成词语素,“略²”“略³”各是一个不成词语素。但我们这里把“朝、略”统认定为可成词语素字。

辞书编纂界早就关注到了不成词语素字。郑远汉《论字典的特殊性质和独立地位》说:“字典通过举例能将代表成词语素的字和不成词语素的字显示出来。”^[3]姜德梧《语文词典词性标注的几个问题》认为:“标注词性遇到的另一个问题是,对于一个汉字来说,有的是词,有的不是词,而是语素字或音节字。……需要研究的是不成词的语素字和语素义项怎样标注。”^[4]

中文信息处理界很关注不成词语素,提出要有一个“不成词语素表”。刘开瑛《中文文本自动分词和标注》在讨论“歧义字段的采集方法”时指出:“对于受分词词典的制约问题,我们认为,对于单汉字只要提出一个不成词的语素表,作为判别单字词的依据。”^[5]许嘉璐、傅永和《中文信息处理现代汉语词汇研究》也认为:“对于单字词,只要提出一个不成词语素表,就可作为判别单字词的依据,而二字、三字以至四字以上词,则需要专门进行筛选,建立实用的分词词典供采集歧义字段使用。”^[6]赵小兵《现代汉语基本词汇自动识别方法研究》给出了一个《CBVE 备选集词语语素过滤的“不成词语素”表》^[7],排在前面的 120 字(左 4 列的上 30 行)是:

CBVE 备选集词语语素过滤的“不成词语素”表(节选)

埃艾碍岸昂傲奥叭爸柏领斑扮膀傍胞贝惫崩彼币毕蔽辨辩飙宾滨濒伐睬参灿仓糙厕又察阐
猖偿敞倡潮彻承诚惩程澄池匙侈斥充憧崇宠畴摧脆粹挫措贷丹耽旦诞淡蹈滴弟帝缔颠典
甸奠叮董督睹杜肚锻兑砝繁仿啡菲废沸纷芬氛奋憤峰锋讽肤伏符袱辐辅腐傳賦縛榦概甘
尴纲糕……

赵小兵的这个“不成词语素表”中都是单字，所以跟我们想象中的“不成词语素字表”非常相像。不过，我们发现，赵小兵这个“不成词语素表”中的不成词语素，有一些并不真的就是不成词语素，比如，依据《现代汉语词典》，“艾、爸、斑、弟”等可以独立成为名词，“扮、贷、叮、兑”等可以独立成为动词，“糙、脆、淡、繁”等可以独立成为形容词，如此等等，它们都是些成词语素。还有些字，连语素字也算不上，比如“砝、啡、尬、尴”等，在《现代汉语词典》中无义项，它们是非语素字，而不是不成词语素。

看来,我们统计一下《通用规范汉字表》中的不成词语素字,从而给出一个以《现代汉语词典》为依据的“不成词语素字表”,还是很有必要的。

二、《通用规范汉字表》中的不成词语素字统计

统计《通用规范汉字表》8 105 字中的不成词语素字，首先要排除掉非语素字。我们在《〈通用规范汉字表〉非语素字统计与思考》^[2]一文中，依据《通用规范汉字字典》，统计到《通用规范汉字表》中有非语素字 1 488 个，排除这 1 488 字以后，《通用规范汉字表》还剩 6 617 字。

我们在依据第7版《现代汉语词典》逐一考察剩余的6617字时，发现其中有117个在《现代汉语词典》上没有义项，《现代汉语词典》对它们的解说是“见下”或“见某页某词”，因此，它们是零义项字，即非语素字：

魅柈孢狈鋪臠算儻藨娘模镔馞穆楂蠶魑蘂熳卤膝炟諤榧痱玢獮斃玞果研骼鞚呱究帽嘴瘊唿
萑圓韙浅鍵脰消襍僅姈爝剗鯈饑洎牘棼範繩繆墺躡躉躉躉躉躉躉躉躉躉躉躉躉躉躉躉躉躉躉躉躉
殍芋筇鰥顱儻洳噲鲨棼紹凇迢鰥般坬韪唔瘡僕嚮巖鷗鷗筅醜姪快熳間嚟黝侯瘞麌終嫿髦

从 6 617 字中排除掉这 117 个“非语素字”，剩下 6 500 字，依据《现代汉语词典》，它们都是语素字。

考察和认定 6 500 个语素字中的不成词语素字，我们使用排除法，在排除掉其中的可成词语素字之后，剩下的就是我们所要的不成词语素字。

(1) 凡是被《现代汉语词典》在其义项上标注了词类的，就是可成词语素字，一律排除。不论该字有几种读音、几个义项，只要它所有义项中有一个义项被标注了词类，就认定它是一个可成词语素字，从而予以排除。

(2)《现代汉语词典》在对“单字条目中的文言义”标注时,讲明了“只给数词、量词、代词、副词、介词、连词、助词、叹词、拟声词标注词类,名词、动词、形容词不做标注”^{[1] 例5},我们的理解是,单字条目文言义的这3类义项虽然没有标注名词、动词、形容词,但《现代汉语词典》已经承认它们是名词、动词、形容词了,只是没有做标注而已,所以,凡是单字条目的义项标〈书〉而没有标注词类的,我们一律视之为可成词语素字,从而予以排除。不仅如此,凡是义项为“古代的……”“古书上指……”等的单字条目,我们也视之为可成词语素字,一律排除。

(3)《现代汉语词典》在给条目注音时,实行“专名和姓氏的注音,第一个字母大写”^{[1]凡例4}的做法。我们根据国家标准《汉语拼音正词法基本规则》“汉语地名中的专名和通名,分写”,认定这些需要首字母大写、需要分写的“地名中的专名”,跟“姓氏”一样,都是可成词语素字,从而把它们

排除。

经过以上 3 次排除, 剩下了 771 个不成词语素字。见下表:

《通用规范汉字表》不成词语素字表

隘鞍案黯邀翱坳懊芭领滂胞悲悖愈币庇婢痹蔽弊壁砭濒疾鬟摒舶箔搏膊怖埠慚灿沧嘈漕惻汉诧诡禪蟾谄闡驟忏羼依猖娼偿慙怅嘲彻澈噴尘碜様讐蛭懲聘匙侈耻敕绸厨畜擒蠹椽榦疵祠匆促纂摧淬瘁悴付措傣怠赕玷玷宕祷蹈棹敌嫡抵蒂缔巔癫淀靛癜凋碉牒牒峒峒犊黠睹妒蠹蝕痏遁惰遏愕噩梵坊妨舫妃绯扉诽吠纷氛焚憤烽峰喙肤麸跗釋俘祫抚脯腐讣妇泔紺烽羔槁塉亘躬孤诂牯腋痼寡褂棺螺馆匱聒裸孩骸駭酣酬鼾涵捍悍撼豪壕浩効閔牿涸闔壑痕江猾寰鬟幻涣豢惶煌恍暎晖微蛔悔卉诲秽婚诨获惑讥击饥肌硯屐畿整疾妓嫉戴虮虮迹寂髻穆笳岬稼箋艰歼鍼柬饯洞践鑒檻疆絳椒蛟跤礁饺剗滔醑桔睫芥誠巾襟烬覲涇睛梗阱儆境鏡韭枢厩咀矩拒炬踞诀蹶攢峻浚竣铠慨刊勘龛戡墈磡磡犒频蹀垦恩窟裨儻筷眶窺跬餽砬嵐瀾覽廊萌痨佬酩薈墀俚沥隶俪砾笠痢怜殓疗僚镣劣猎鬪邻遴伶棂綾齡汎陋矮垆擣碌戮耜履慮牽沦夢逻胴盲茫蟲貿媒楣媚魅们懵麇觅绵婉藐泯冥瞑摹魔陌秣暮漠寘眸逐牡拇躰募衲赧腩昵匿蘖狃奴弩瘞惟懦襦讴殴蟠叛畔堋膨疲呻癖媿僻譬剽频牴枰婆迫魄圃栖凄欺歧脐莫企绮讫泣碛钎谦虔鉗遣歉堑櫬峭穷情併驱躯趋駿龋痊犬裙鬚韧衽妊茸冗猱蠕乳褥蕊锐箬喪嗓齒屢霎芟衫繕膳墒艄奢涉赦娠牲甥尸示勢拭恃逝嗜噬艏授兽绶枢菽淑蔬秋塾暑黍薯曙述墅衰孀瞬烁搠私廝嘶伺祀伺嗣诵嗽嗦溯隧唆蓑琐免汰淮坦袒毯叹溏端叨涛絳滔梯涕惕佻眺霆塊疊屯地豚沱鴟槲宍袜垮亡惆巍违闹梶惟伪苇纬诿痿谓猶慰刎紊倭涡汚诬忤侮杌沵悉惜晰媳臬禧隙侠狎峽狹瑕暇涎廂箱翔鼈霄器滑哮胁谐携屑亵榭懈芯鑫衅猩型兄汹墟婿萱喧绚渲靴勋醺询循迅殉垭蚜涯胭炎硯唁诙焰殃佯佯着钥榔曳液蜃饴疑酏忆屹佚役驿疫谊逸肄毅臆蘢烟浮隱睿櫻佣咏泳踊诱蚴娛漁隅腴愉屿峪浴域諭御蜮援簰垸耘限孕仔赃蚤澡皂責仄憎锃艷柵柵杖帳障嶂瘴沼哲蟄赭蔗侦胗砧診鎮賑拯症肢脂殖址晊室滞艷薨𠵼綈昼蛛囑囑助撰篆妝狀諄苗浊镯咨姿擎髡姊淳毗踪粽族阻昨

我们这个不成词语素字表,跟赵小兵《CBVE 备选集词语语素过滤的“不成词语素”表》相比,两表字数相差不多,都是 700 多字,但是具体内容有好多不同。例如,读音首字母为 A 的,我们表中是“隘鞍案黯邀翱坳懊”8 个,而赵小兵表中则是“埃艾碍岸昂傲奥”7 个,没有一字相同。读音首字母为 B 的,我们表中有“笆颁榜胞悲悖惫币庇婢痹蔽弊壁砭濒殡鬟摒舶箔搏膊怖埠”25 个,而赵小兵表中则是“叭爸柏颁斑扮膀傍胞贝惫崩彼币毕蔽辨辩飙宾滨濒秉摒玻剥膊驳博搏膊捕怖”33 个,两表只有“颁胞惫币蔽濒摈搏膊怖”10 字相同。

三、汉字三记说

《现代汉语词典》说：“单字条目在现代汉语中成词的标注词类，不成词的语素和非语素字不做标注。”^{[1] 凡例5}其中“成词的”“不成词语素”“非语素字”三种情况，讲得非常到位。

“成词的”“不成词语素”“非语素字”，是《现代汉语词典》对单字条目定性分类的说法。如果改从文字学角度说，文字是记录语言的书面符号系统，那么，词典定为“成词的”的单字条目，它所记录的是“词”（单音节词）；词典定为“不成词语素”的单字条目，它所记录的是“语素”（合成词中的单音节语素）；词典定为“非语素字”的单字条目，它所记录的是多音节单纯词中的一个“章节”。

我们的统计工作是从统计非语素字开始的。在《〈通用规范汉字表〉非语素字统计与思考》^[2]一文中，我们依据王宁《通用规范汉字字典》认定《通用规范汉字表》中有 1 488 字为非语素字，本文又依据《现代汉语词典》认定《通用规范汉字表》其余的 6 617 字中还有 117 字为非语素字，两者

相加,非语素字总数为1605字。本文又依据《现代汉语词典》,统计到《通用规范汉字表》中有不成词语素字771个。依据上述数据,得出《通用规范汉字表》中有可“成词的”语素字5729个。

依《现代汉语词典》“成词的”“不成词语素”“非语素字”的顺序,对《通用规范汉字表》8105字的统计结果如下:

可成词语素字:5729个,占70.69%,记录汉语单音节词;
不成词语素字:771个,占9.51%,记录汉语合成词中单音节语素;
非语素字:1605个,占19.80%,记录汉语多音节单纯词中的音节。^①

由此,我们可以提出一个汉字三记说:汉字具有记录汉语单音节词、记录汉语合成词中单音节语素、记录汉语多音节单纯词中音节的功能。

郑林曦《汉字记写的是汉语的哪个层次》^[8]曾指出,汉字是“记写单音节的词和词素以及音节的文字”,他也说到了汉字的三个记写功能,但是他在“词素”和“音节”前面没有加必要的限定语,没有范围的限制,所以说得还不是很到位。

不过,在学术界流行的,主要还是“一记”(记录汉语语素)和“二记”(记录汉语语素和音节)两种看法。

认为汉字“记录汉语语素”的学者称“汉字是语素文字”。赵元任《语言问题》认为:“用一个文字单位写一个词素,中国文字是一个典型的最重要的例子。”^[9]朱德熙《汉语》指出:“文字是记录语言的。就汉字跟它所要记录的对象汉语之间的关系来看,汉字代表的是汉语里的语素。”^[10]苏培成《语言文字应用探微》指出:“汉字的单字记录的是汉语的语素,所以汉字是语素文字。”^[11]

认为汉字“记录汉语语素和音节”的学者,说“汉字是语素-音节文字”。叶蜚声、徐通锵《语言学纲要》指出,汉字“是一种语素-音节文字”^[12]。李运富《汉字学新论》指出:“我们对汉字性质的看法可以概括为:汉字是用表意构件兼及示音和记号构件组构单字以记录汉语语素和音节的平面方块型符号系统。”^[13]杨润陆《现代汉字学》也认为,从汉字记录语言单位来看,“现代汉字可以称为语素文字或语素-音节文字”^[14]。

虽然我们坚信“汉字三记说”没错,但是,要是真的依据辞书把汉字截然划分为“记词的”“记不成词语素的”和“非语素字”三大类,或者分为“词字”“语素字”“音节字”三大类,这在实际上却是很难做到的。实际情况是:绝大多数汉字都兼有两种或三种功能,很难把它归到哪一类中去。而且辞书也不能把一个字的用例全部列举,依据辞书可定为“记词的”汉字一般也都可以在合成词中“记语素”,依据辞书可定为“记词的”和“记语素的”汉字一般也都可以“记多音节单纯词中的音节”,依据辞书可定为“非语素字”的汉字,有一些也“记词”“记语素”。

(一)依据辞书可定为“记词的”汉字也可以在合成词中“记语素”。例如:“日”在《现代汉语词典》有9个义项,7个义项可以成词,我们可以凭此把它归入“记词的”字,但另两个不成词义项则是在合成词中“记语素”的。单字条目后面附有“日光、日子、日记、日均、日杂、日食”等60多个合成词,“日”在这些合成词中是“记语素”的。这种现象特别普遍,俯拾即是,无需赘举。在《现代汉语词典》中,有些单字条目只有一个“成词的”义项,比如“瞿、祢、芮、邵、余、郤”等,下面没有列出含有它们的合成词,乍看上去它们只能“记词”,不能记别的。但没有列出合成词不等于就没有合成词,倘若在它们后面加上“家”“宅”“氏”“姓”“老”“爷”“兄”“姐”等,或在它们前面加上“老”“小”,就能够构成合成词,它们分别在这些合成词中“记单音节语素”。

^① 施效人《谈同音词和同音字问题》也曾做过类似的统计,他的统计结果是“《现代汉语词典》里,词字有2560个,语素字有2430个,音节字有1104个”。(转引自沈孟卿《现代汉语理论与应用》第169页,南京师范大学出版社1999年版。)

(二)依据辞书可定为“记词的”和“记语素的”汉字,也可以用作“非语素字”去记录多音节单纯词中的音节。例如,“日”在《现代汉语词典》能记词又能在合成词中记语素,但在“日本、日内瓦、尼日尼亚、日德兰、日喀则”等词中则是记多音节单纯词中一个音节。“马”在《现代汉语词典》能记词又能在合成词中记语素,但在“马尔萨斯、萨马兰奇、喜马拉雅、马尼拉、危地马拉、马里、索马里、马达加斯加”等多音节单纯词中,“马”却是一个记写音节的非语素字。周荐《词汇论》把这种用法的汉字称之为“弃义汉字”^[15],意思是说它们原本是有意义的记词记语素的汉字,现在抛弃了意义作非语素字用,去记多音节单纯词中的音节。汉语中,“记词的”和“记语素的”汉字被用作非语素字来记写多音节单纯词中的音节的情况很多,下面提供四个方面的情况。一是用于复姓中的字,比如“欧阳、诸葛、东方、慕容”等复姓中的“欧、阳、诸、葛、东、方、慕、容”。袁义达、杜若甫《中华姓氏大辞典》^[16]收双字姓 4 329 个,三字姓 1 615 个,四字姓 569 个,五字姓 96 个,六字姓 22 个,七字姓 7 个,八字姓 3 个,九字姓 1 个。原本“记词的”和“记语素的”汉字被用在这些复姓和多字姓中记多音节单纯词中的音节。二是联绵词中的字,比如“从容、参差、仿佛、仓卒、瓜葛”等联绵词中的“从、容、参、差、仿、佛、仓、卒、瓜、葛”等。徐振邦《联绵词大词典》^[17]收联绵词 14 000 余条,其中有很多“记词的”和“记语素的”汉字被用作“非语素字”来记写双音节单纯词中的音节。三是汉语音译外来词中的字,比如“俄罗斯、日本、蒙古”中的“俄、罗、斯、日、本、蒙、古”。史有为《新华外来词词典》^[18]收录外来词 20 000 余条。四是多音节拟声词中的字,如“噼里啪啦、丁零当啷、扑通”等词中的“噼、里、啪、啦、丁、零、当、啷、扑、通”等。

(三)依据辞书可定为“非语素字”的字,也有一些能记词、记语素。比如“葡萄”的“葡”,《通用规范汉字字典》《新华字典》《现代汉语词典》《现代汉语规范词典》都没有给出其义项,是把它作为非语素字处理的,但我们发现有“果葡糖浆”一词,百度百科解释说:“它的组成主要是果糖和葡萄糖,故称为‘果葡糖浆’”,“果葡糖浆”中的“葡”显然是个语素,有意义,意义就是“葡萄糖”。另据《辞海》,“葡”还是“葡萄牙”的简称,那么,在“中葡关系、葡方”等词中,“葡”也是一个语素,是语素字。再如“嗨”,我们依据《通用规范汉字字典》《现代汉语词典》认定它是一个“非语素字”,但是我们却常见到“我们去嗨一下”“嗨起来”“太嗨了”“喝嗨了”等说法,这里的“嗨”是个外来词,源自英语 high,意义是“兴奋地大声唱、喊”“兴奋地做某事”“特别高兴、兴奋”“程度高、过分”等。此外,“嗨”还是一个叹词,外来词,源自英语 hi,是当代年轻人互相问候、打招呼的用语。

不过,无论情况怎样复杂,汉字所记都在“三记”当中,或记单音节词,或记合成词中单音节语素,或记多音节单纯词中音节,没有记其他的可能。

四、汉字的性质

前已述及,汉字具有记录汉语单音节词和合成词中的单音节语素以及多音节单纯词中的音节三项功能,那么,如果着眼于“一个汉字记写什么”,则汉字的性质就可以表述为:汉字是记录汉语单音节词、合成词中单音节语素、多音节单纯词中音节的书面符号系统。

或许有人认为,我们对汉字性质的这个表述,可以简称为“词-语素-音节文字说”。不过我们总是觉得,把“单音节词”简称为“词”,把“合成词中单音节语素”简称为“语素”,把“多音节单纯词中音节”简称为“音节”,不怎么适宜。因为“单音节词”跟“词”,“合成词中单音节语素”跟“语素”,以及“多音节单纯词中音节”跟“音节”,其内涵和外延都是不相同的,不能把它们等同起来。

我们认为,在表述汉字的性质时,“记录汉语单音节词”一项不应该被忽视。自古及今,汉字都有这项功能。虽然在现代汉语中,单音节词的总量不及多音节词多,但单音节词使用频率高,在一个具体的语言环境中,单音节词的词次跟多音节词的词次差不了多少。胡裕树在 1981 年版《现代汉语》中曾统计了一个 57 字的段落,共 34 个词,其中单音节的词 14 个,双音节的词 18 个,

三音节和四音节的词各1个^[19],单音节词跟多音节词之比为14:20;1995年重订该书时,胡裕树又统计了另一个段落,也是34个词,其中单音节的词13个,双音节的词17个,三音节的词3个,四音节的词1个^[20],单音节词跟多音节词之比为13:21。所以汉字记录单音节词的功能不容忽视。

汉字“记录汉语单音节词”不应该被忽视,还因为我们认为,汉语在一般情况下是先有单音节词后有合成词再后来有多音节单纯词的。在用“象形、指事、会意、形声”四法造字之时,造出来的字最初就是记单音节词的。甲骨文的“日”是圆圈中加一点,表示太阳,记的就是单音节词。后来有了在词组的基础上形成的合成词,“日”才有了记合成词中单音节语素的用法,例如:日光、日子、日期、日记等。再后来有了多音节单纯词,它才被用于记多音节单纯词中音节,例如:日本、尼日利亚、日耳曼等。

根据文字学的原理,文字是记录语言的符号系统,而语言中最小的能够自由运用的单位是词。文字记录语言,归根结底都必须能够记词。汉语的词有单音节词和多音节词之分,而多音节词又有合成词和多音节单纯词之分。汉字记录汉语词的方法参见下表。

表1 汉字三记功能和汉字记词方法对照表

	记单音节词	记合成词中语素	记多音节单纯词中音节
日	日(出东方)	日期、日记、节日、假日	日本、尼日利亚、日耳曼
马	(一匹)马	马车、马上、马戏、马路	马拉松、马赛克、萨马兰奇
牛	(一头)牛	牛皮、牛市、牛杂、牛毛	牛顿、牛津、牛轧糖
他	他(的书)	其他、他们、他年、他乡	马耳他、维他命、吉他
雅	(很)雅	雅兴、雅致、优雅、雅量	喜马拉雅、雅典、雅加达
拜	拜(了三拜)	拜年、结拜、拜会、拜堂	迪拜、拜伦、阿塞拜疆
拉	拉(我一把)	拉扯、拉杆、拉拢、拉钩	巴拉圭、委内瑞拉、法拉第
大	(很)大	大力、大致、大作、宏大	加拿大、澳大利亚、大阪
太	太(好了)	太空、太平、太爷、太阳	渥太华、李提摩太、迦太基
容	容(我再想想)	容颜、容纳、容许、整容	从容、慕容、雍容
用单字记单音节词		用多字组合的方法记录汉语多音节词	

从上表可以看出,汉字记录汉语的方法是:用单字记录汉语的单音节词,用多字组合的方法记录汉语的多音节词。因此,如果不局限于一个汉字记什么,则汉字的性质也可以这样表述:汉字是单字记录汉语单音节词、多字组合记录汉语多音节词的书面符号系统。

参考文献

- [1] 中国社会科学院语言研究所词典编辑室.现代汉语词典(第7版)[M].北京:商务印书馆,2016.
- [2] 邵震吉.《通用规范汉字表》非语素字统计与思考[J].现代语文,2022(7):62-68.
- [3] 郑远汉.论字典的特殊性质和独立地位[M]//《汉语大字典》湖北省编辑部.汉字的源和流.武汉:湖北辞书出版社,1986:15.
- [4] 姜德梧.语文词典词性标注的几个问题[M]//谢小庆,彭恒利.考试研究文集(第2辑).北京:经济科学出版社,2004:63.
- [5] 刘开瑛.中文文本自动分词和标注[M].北京:商务印书馆,2000:62.
- [6] 许嘉璐,傅永和.中文信息处理现代汉语词汇研究[M].广州:广东教育出版社,2006:44.

-
- [7] 赵小兵. 现代汉语基本词汇自动识别方法研究[M]. 北京: 中央民族大学出版社, 2012: 169 - 170.
 - [8] 郑林曦. 汉字记写的是汉语的哪个层次[J]. 语文建设, 1988(2): 32 - 37.
 - [9] 赵元任. 语言问题[M]. 北京: 商务印书馆, 1980: 144.
 - [10] 朱德熙. 汉语[M]//朱德熙. 语法丛稿. 上海: 上海教育出版社, 1990: 198.
 - [11] 苏培成. 语言文字应用探微[M]. 北京: 商务印书馆, 2019: 2.
 - [12] 叶蜚声, 徐通锵. 语言学纲要[M]. 北京: 北京大学出版社, 1981: 178.
 - [13] 李运富. 汉字学新论[M]. 北京: 北京师范大学出版社, 2012: 17.
 - [14] 杨润陆. 现代汉字学[M]. 2 版. 北京: 北京师范大学出版社, 2017: 5.
 - [15] 周荐. 词汇论[M]. 北京: 商务印书馆, 2016: 168.
 - [16] 袁义达, 杜若甫. 中华姓氏大辞典[M]. 北京: 教育科学出版社, 1996: 凡例 19.
 - [17] 徐振邦. 联绵词大词典[M]. 北京: 商务印书馆, 2017: 凡例 19.
 - [18] 史有为. 新华外来词词典[M]. 北京: 商务印书馆, 2019: 凡例 16.
 - [19] 胡裕树. 现代汉语[M]. 上海: 上海教育出版社, 1981: 241.
 - [20] 胡裕树. 现代汉语(重订本)[M]. 6 版. 上海: 上海教育出版社, 1995: 210.

Statistics of and Reflection on the Non-Individual Morpheme Characters of the *General and Standard Chinese Characters Table*: Also on the Functions and Nature of Chinese Characters

SHAO Ai-ji

(School of Chinese Language and Literature, Yancheng Teachers University, Yancheng, Jiangsu, 224002, China)

Abstract: There are 1 605 non-word characters, 771 non-individual morpheme characters and 5729 morpheme characters in the total 8 105 characters of the *General and Standard Chinese Characters Table*. Chinese characters have three functions: recording monosyllabic words, recording monosyllabic morphemes in compound words, and recording syllables in multi-syllabic simple words. From the perspective of “denotation of character”, the nature of Chinese characters can be defined as: Chinese characters are a written symbol system to record the monosyllabic words, monosyllabic morphemes in compound words and syllables in multisyllabic simple words. If it is not limited to the denotation of characters, its nature can also be described as: Chinese characters are a written symbol system in which single characters record monosyllabic words, and multi-character combinations record multisyllabic words.

Key words: the *General and Standard Chinese Characters Table*; non-morpheme characters; non-word morpheme characters; nature of Chinese characters

〔责任编辑:何敏敏〕