

论计算机用中国汉字字形的标准化

石云程 江晓红

〔摘要〕 本文论述了中国汉字字形的特点和历代汉字学家对社会用字所作的统一化、规范化的成就，并且通过一些史例阐述了标准化是中国汉字发展的必然趋势，而计算机用点阵汉字字形标准化又是汉字信息处理技术发展的必然需要。最后本文还介绍了三种国家标准点阵汉字字形风格和特点。

一、中国汉字的特点和它的规范化与标准化

中国汉字是世界上最古老的文字之一，它从殷墟的甲骨文发展到今天的简体字已有三千四百多年的历史了。在这漫长的岁月里，她的生命力仍然是强大的，世代不衰，不因社会政体的变更、地区语言的变化和汉字形体的变革而使它减色。她始终是中华民族团结统一的纽带，是中国人民日常生活中使用的记事工具，她已适应了中国各个朝代的文风和词句的变革。她记录了中华民族丰富多采的文学创作和诸子百家的学术思想，对继承和保留中华文化遗产做出了宝贵的贡献。今天，我们仍能看到秦代篆书，可以说它是中国文字发展史上第一次规范化和标准化的文字。我们也能见到历代的一些碑文题词、诗书手抄本和木刻印刷本，其字体的外形和风格，十分工整雅致，可算是中国最早文字字形规范化和标准化的雏形，尤其是秦代以后发展的楷书，它已成为后代人民书写的样板。

中国的汉字是音、形、义的统一体。它的基本特点是字数繁多、字形复杂、形体变化大，各个汉字的使用频度和构词能力都不相同，同音字多、字体字号多、字义也多、排序和索引没有西文容易。单就楷体而言，同一个字之间，就有几种不同的写法，存在着正体和异体之别，现在还有繁体和简体之分，所以，要使汉字信息处理用的点阵字形规范化和标准化，首先要研究解决汉字字形的标准化和汉字字种的规范化。因此，汉字信息处理在技术上还存在一定的难度和困难，尤其是存在对汉字属性进行规范化、标准化方面的困难。

中国汉字的起源可以追溯到七千年前的历史。由于字体的演变和社会的发展，起初，汉字越来越多，这给字种、字形的整理和字形规范化标准化工作带来了许多不便。就以字种数量来说，从历代字书中可以得知，我国最早的“说文解字”收单字只有9543字，到了宋朝的《集韵》，则收单字53525字了。到了清朝（1716年）的《康熙字典》，收单字已减少到47043个。近百年来，许多古老的字已经淘汰不用了。1915年的中华大字典仅收字44908个，到了1979年再版的《辞海》，就只收单字14872个，而1980年《新华字典》则只收单

字 8500 个。1965 年中国文字改革委员会公布的《印刷通用汉字字形表》也只收单字 6196 个。从上述不同时代产生的几种字典收字情况来看,也能反映出我们社会用字数量的趋向是逐渐减少的。但是根据中国文字改革委员会 1964 年确定的《简化字总表》,又产生了“2234”个简化字。依据汉字简化方案及对繁体字类推简化的原则,1987 年中国标准局公布的计算机信息交换用汉字编码字符集的辅助二集和辅助四集,其中又新生了四千多个简化字,而对应于这些简化字的繁体字和异体字并没有消灭,仍在一定场合使用。因此,中国汉字的总数量是不断增加的,当代社会用字的统一和规范化还需要认真进行大量检查研究和分析探索工作。

二、中国汉字字形发展的必然趋势是走向标准化

我国最早使汉字走向标准化的是秦朝时期。秦代的“书同文”是我们已知的汉字体系中首次系统的全面的标准化的伟大创举。秦始皇在统一了中国之后,采取了李斯的建议,废除其它几国不与秦文相同的异体字,并且简化了字形,从而形成了统一的文字。这种统一了的文字就叫做小篆。小篆,实际上也是对汉字过去长期自然发展的一次总结,使汉字能够进一步做到整齐化和定型化,因而更有利于发挥它作为社会交际工具的职能和作用。现在一般所说的篆书,均指小篆。我们在游览名胜古迹时见到的《泰山刻石》、《琅邪台刻石》、《碣石门刻石》、《绎山刻石》等,据说都是秦始皇叫李斯写的,而且都是标准的小篆样板。秦始皇指出的“书同文”是我国汉字史上首次要求实现汉字字形全面标准化的伟大创举。他提倡的小篆,特点十分鲜明,作到了:①笔画逐渐固定;②方向(向左或向右)渐趋一致;③偏旁部首基本可以分析归类,并开始简化了复杂的偏旁;④直行书写,形体匀圆,并带方块形化、线条化、符号化;⑤增加了形声字。这些特点反映出当时字形标准化的书写原则,充分体现了汉字字形正在向标准化方向前进。公元 174 年,蔡邕倡议把“六经”刻在石碑上,立在太学门前作为示范,开创了正字风气之先,实际上就是推行正确的标准汉字字形,叫世人模写,以防字形混乱。魏正始年间洛阳太学的“三体石经”,唐开成年间长安太学的“开成石经”,清乾隆年间刻于北京国子监的“十三经”都是刻字作为范字,让人模写。为了使文字整齐化,走向定型化,在六朝文字混乱之后,出现了“字样之学”,颜师古作《字样》杜延业作《群书新定字样》,颜元孙作《千禄字书》,欧阳融作《经典分毫正字》……《五经文字》、《九经字样》,直到近世的《字学举隅》都是历代汉字标准化运动的组成部分。

十九世纪末,资本主义在中国开始发达起来,人与人之间的邮电通信频繁增多了,电报通信很快传到了中国,以四个数字表示一个汉字,对精选的常用一万个汉字进行了编码,即今天的邮电四码电报,这种电报码是以电报通信为开端的现代科学技术在汉字领域里的应用,是当时社会用字规范化、数字化、标准化的最新成就。

二十世纪初,中国受外敌的侵略,国民党反动派统治腐败,经济文化受到严重的破坏。当时,印刷行业中的印刷铅字形极不统一,同一个字,在报纸上、杂志上、图书上出现几种不同的笔画结构,在人们的书写文字中,混乱现象更加严重,出版质量大大下降。中华人民共和国成立后,为了整顿这种现象,1955 年中国文字改革委员会成立了“标准字形”研究小组,到 1965 年,中国文化部和文字改革委员会颁发了《印刷通用汉字字形表》,此表包含印刷通用宋体字 6196 个(不包括排印古籍及其他专门用字)标准字形的笔数和笔顺,

对印刷汉字的标准化起了很大的推动作用。随后，国务院又批准了《简化字方案》和《简化字总表》，颁布了认读汉字的《汉语拼音方案》。从此，我国的汉字规范化、标准化开始走向了正规的道路，找到了正确的方向。

三、汉字信息处理技术发展要求汉字点阵字形标准化

随着现代科学技术的发展，国外电子计算机的应用已经很普遍了。在中国，用计算机处理汉字信息是七十年代后期才兴起的事物。近几年来，汉字信息处理技术发展很快，各种类型的汉字智能终端和小型机、微型机的应用，正在广泛展开，汉字计算机将成为各行各业不可缺少的生产工具和办公自动化工具。尤其是计算机已用于汉字编辑排版和印刷期刊书报，其发展的速度表明，印刷行业将取消铅字，大大地缩短排版周期，减少工时劳力，这的确是我国印刷史上的又一伟大的转折和进步。但是，计算机要处理汉字，必须先把成千上万个标准汉字送到计算机里去存储起来，同时，再从计算机存储器中调出来，显示、编辑处理和打印，以供人们对文章的阅读、存档和传送。要实现这一目标，迫切需要的信息交换用汉字编码字符集基本集国家标准 GB2312—80 文本发布了，文本中规定了当前计算机用字 6763 个，并对每一个字给了编码，为制作标准化的汉字库奠定了基础。但是，汉字在计算机里需要以点阵信息形式存入存储器中，才能形成汉字库。八十年代初，许多电脑厂商十分关心这个技术问题，他们希望在微机上配上美观的、统一规范的、标准化的汉字阵字形库。此外，当时中国的汉字识别技术也发展起来了，要求输入的汉字必须是标准化的，如果字库里的点阵字形、笔形不标准化，这样打印出的汉字文件必然不标准化，将这种打印结果送到另一个汉字识别装置中去是不会认识的。特别是在压缩信息方面，更会因字形不标准化而产生新的障碍。由于当时汉字信息处理技术的研制单位很多，制作点阵汉字库的单位也越来越多，作出来的点阵类型自然是多种的，字形点阵各异，重复劳动极大，质量也是低水平的。针对这种情况，电子工业部为了加速发展汉字信息处理设备，准备批量生产标准化的汉字芯片和汉字卡，遂向国家标准局提出了点阵汉字字形标准化的课题。1983 年 9 月国家标准局下达了信息处理 15×16，24×24，32×32 三种点阵汉字字模集和数据集的研制任务。从此，我国开始了信息处理用点阵汉字字模集和数据集的标准化研制工作。

四、低点阵汉字字形规格适合我国目前的需要

汉字是图形字符中的一部分。近年来国内外电脑厂商在汉字显示器和汉字打印机中使用何种类型的汉字图形字符，一直是汉字信息处理最关心的课题。

我国汉字字形复杂，有繁有简，最繁的字形有 30 画，如“囊”，最简的是一画，如“一”。因此，要表现这些不同繁简程度的汉字，最好的方法是把汉字字形置于网状方格内，每格对应存储器中的一位，有笔画的格对应“1”，无笔画的格对应“0”。这就是汉字的点阵表示法。用点阵表示的汉字，技术处理简单，价格便宜，制作容易，修改方便，数字化准确。按字符方式产生的低点阵汉字发生器输出汉字速度快，稳定性好，能较好地表现汉字字形的形态。所以，我们在汉字信息处理中，采用点阵字形，并使其标准化。

低点阵汉字字形是用多少点合适，要看对汉字质量的要求而定。点阵愈密，字形质量愈

高，对设备的精度要求也高。在过去几年中，国内外汉字电脑厂商和研究机构，按不同用户要求和生产技术条件，设计了多种类型的汉字点阵。如：16×16、17×18、18×18、20×20、22×24、24×24、30×30、32×32、64×64、96×96、108×108等。这样多的五花八门类型的汉字点阵必然给联机通讯的汉字处理和设备制造、资源共享带来许多的麻烦和困难。对此，我们根据当前国内实际使用的汉字字体、字形和字号进行了分析，按现有打印设备和显示设备的精度，首先制定了三种点阵规格。

三种点阵汉字打印结果与铅字号对应表

铅 字 号	字 体 大 小	PR	LBP	
		180 线/英吋	240 线/英吋	300 线/英吋
4 号	12点	32×32		
5 号	10.5点	24×24	32×32	
6 号	7.875点	15×16	24×24	32×32

五、三种国家标准点阵汉字宋体字形的风格和特点

1. 15×16点阵字形力求做到体现宋体结构和宋体风格

15×16点阵栅格可容纳汉字的最大限度竖里是八笔，横里是八笔。由于汉字的结构特性不可能每个字都顶天立地，但又要照顾到笔锋的舒展，特别要尽量避免和上下左右笔画粘连或并线，故实际可用的笔画竖只有七笔，横里也只有七笔。宋体字的最大特点是横细竖粗，要表现粗画需要双画线，而15×16点阵是无法实现的。因此只能采用横竖笔划一样粗细来设计字的宋体结构和宋体风格。由于受栅格的限制，不能字字都做出宋体的风格，有些多笔字没能加上宋体字应有的装饰点。但从整幅字全局来看还是清楚地显示出宋体结构和宋体风格。

2. 15×16点阵字对繁笔字的简化处理较合理

在设计中我们对纵横超过八笔的138个繁笔字参照中国传统书法草书对繁笔字处理的方法，压缩变通笔形，进行了简化。主要设计依据两条原则：

- ①被压缩变通简化处理的繁笔字，保留原字字形的特点。
- ②简化后，字形看起来要依然逼真，不会被认为是其他字。

由于在技术上采取了一定的措施，注意笔形，字形的标准化和字形结构的规范化，所以设计的点阵字形始终保持了宋体风格，用户使用效果很好。

3. 24×24、32×32点阵字形充分体现了宋体字的结构及宋体风格。做到了粗细黑白灵活掌握，克服难度较大的点、撇、捺、勾的处理，做到舒展自然。

中国汉字笔画多少悬殊，笔画多的有三十多画，少的只有一、二笔。在同一点阵方格内，即要使每个字笔画完整，又要体现横细竖粗的宋体风格，还要照顾缩小后的（五号字、四号字）粗细黑白效果，难度是很大的。在设计中只能是灵活掌握粗细黑白，采用次要笔画细处理的措施，达到黑白度适宜。比如“醴”字，只有减细次要笔画，将“酉”中间的二竖和“曲”中间的二竖改为一格，这样处理的结果，缩小后效果很好，而且仍能保持宋体的风格。同时还处理了点阵字的撇、捺笔画，做到了柔和、自然、舒畅。

六、展 望

目前,我国图书、报刊、杂志以及公文文件等常用的字体有宋体、黑体、仿宋、楷体、长宋、扁宋、长黑、扁黑、长仿宋等九种。字号包括从七号到特大号共 16 象。国家已经颁布的三项汉字信息处理标准点阵字形全部采用的是书版宋体,仍不能满足各行各业的需要。为从我国计算机信息处理技术发展的实际需要出发,针对国内现有设备发展水平和文字处理使用现状,制定汉字字形点阵系列标准,已是中外电脑厂商和用户十分关心的一件大事。同时,也是我国汉字处理技术飞速发展中必不可少的规范化标准。在国务院电子振兴办公室和国家标准局的领导下,我国在“七五”期间将全面开展汉字点阵系列标准的研制工作。有关单位正在加紧研究各种字体的点阵系列标准。并决心在发展汉字技术、振兴经济、建设繁荣富强的社会主义中国的伟大事业中,奋发努力,作出更大的贡献。

参 考 文 献

- [1] 汉字基础知识 张之伟 浙江人民出版社
- [2] 楷书汉字字形标准化与中文信息处理 张普 邓采芹 1981.5
- [3] GB 2312—80 文本
- [4] GB 5199、GB 5007、GB 6345 信息处理用点阵汉字字模集和数据集