

广州市科学技术协会  
广州市南山自然科学学术交流基金会  
广州市合力科普基金会  
资助出版



Computer  
No character of intelligent building  
Chinese characters can also be  
computer information  
计算机

# 无字库智能造字

—— 汉字也可以这样计算机信息化

皮佑国 著



国防工业出版社  
National Defense Industry Press

广州市科学技术协会

广州市南山自然科学学术交流基金会 资助出版

广州市合力科普基金会

# 计算机无字库智能造字

——汉字也可以这样  
计算机信息化

皮佑国 著

国防工业出版社

· 北京 ·

图书在版编目(CIP)数据

计算机无字库智能造字:汉字也可以这样计算机信息化/皮佑国著. —北京:国防工业出版社,2013. 1

ISBN 978-7-118-08345-3

I. ①计... II. ①皮... III. ①汉字信息处理—研究  
IV. ①TP391.12

中国版本图书馆 CIP 数据核字(2012)第 281938 号

※

国防工业出版社 出版发行

(北京市海淀区紫竹院南路 23 号 邮政编码 100048)

北京嘉恒彩色印刷有限责任公司

新华书店经售

\*

开本 710×960 1/16 印张 26 $\frac{1}{4}$  字数 488 千字

2013 年 1 月第 1 版第 1 次印刷 印数 1—2500 册 定价 98.00 元

---

(本书如有印装错误,我社负责调换)

国防书店:(010)88540777

发行邮购:(010)88540776

发行传真:(010)88540755

发行业务:(010)88540717

## 前 言

本书是一本讨论计算机无字库智能造字——汉字计算机信息化的另一种方法的专著,在没有任何组织和个人资助的情况下,我带领多名研究生研究了十余年,终于完成了无字库智能造字的理论研究,现将其研究思路、研究过程和结果予以出版,任世人评说和取舍。

1999年,广州铁路集团客运公司副总经理潘晨光先生试图解决客运过程中的假票问题联系到我,但当时因为忙于解决税务发票抵扣联的符号识别而未能接受他的要求。到2003年春节,发票识别的任务已经接近完成。在偶然的相聚中,潘先生又旧话从提,令我为事隔多年的问题还未解决而感到遗憾和愧疚。我表示已经有时间并愿与其一起解决问题。在解决假车票问题的方案构思过程中,接触到全国的火车站站名,了解了一些汉字的计算机处理过程,这就是我发现汉字字库存在弊端的原始机缘。有趣的是,假火车票的项目因为铁路体制改革、解散客运公司而“胎死腹中”,而对汉字信息化另一种方法的研究竟然花费我和一些学生们十余年的光阴。

我不是研究中文的,更不是研究汉字的,甚至因为历史的原因,未能受到应有的包括汉字在内的文化教育。要研究汉字信息化,还必须补文化课。整个研究工作可以分为准备、预研和实验三个阶段。

第一个阶段(2003年—2005年)为准备阶段,在这个阶段中,主要是学习和构思。首要任务是学习汉字的相关知识,因为汉字信息化的目的和任务是汉字应用,汉字计算机信息化是为汉字应用服务的,是汉字应用中书写工具和存储介质的变革。二是要学习认知心理学,因为汉字的认读和书写都是一个认知过程,需要从这个心理过程去提炼汉字认知的模型。三是必须寻找合适的数学工具,对汉字造字进行数学描述。经过分析,选择了拓扑数学作为计算机造字的数学工具。其他还有如计算机信息处理、计算机图形学、人工智能等相关的知识,好在这些与本专业多少有些联系,学习起来相对容易一些。在具备相关基础知识以后,就是研究计算机造字的相关理论和方法,制定项目研究方案。这一过程是艰苦的、漫长的,也是激动人心的。到2006年上半年,已经形成了较为明确的研究思路,就等实验验证了。

第二阶段(2006年—2008年)为预研阶段。在这个阶段,我开始和研究生一起进行局部实验验证。参加这一阶段研究工作的有博士研究生梁添才,硕士研究生彭晶、王锐、朱朝华、谌扬帆、岳荣、邱志文、王素利等。其中,博士研究生梁添才作为我的助手带领硕士研究生们做了大量的测试实验,分别、逐项地验证我们对汉字的各项假设,为项目形成系统的理论假设、研究方法和研究路线奠定了坚实的基础。

第三阶段(2009年—2011年)是研究工作的主攻阶段。在这个阶段中,投入了大量的人力和物力,进行了大量的实验。参加这一阶段工作的核心成员主要有博士研究生刘明友,硕士研究生黄坚、段聘森、潘海东。其中,刘明友和黄坚开展了从基元遴选、结构和编码等实验,开发了计算机造字平台以及输入法;段聘森开发了智能造字专家系统;潘海东则通过插件解决在字库系统下的生僻字输入,还对联想功能进行了初步尝试。开发出上述实验工具和平台后,往往是全体同学都参加实验。2010年春节,是我平生第一次也是最后一次干预同学的假期,要求全体同学提前10天到校进行汉字编码实验。参加编码实验的同学有22人,分别是刘明友、黄坚、梁伟林、汪剑、谢宇、贺玲、崔志芳、黎永华、张世雄、刘永平、段聘森、廖文志、林凯迪、潘海东、潘宇、王烈虎、吴红军、张友斌、方培潘、黄忠、卢建平。

从研究过程可知,参与本书研究的人数较多,为防止署名人员过多,出现遗漏和意见不同,经过协商,决定只署我一人之名,实际上我只是这个团队的一个代表。

本书初稿完成后,又请梁添才、刘明友、张碧陶、黄坚、段聘森、骆仁波等同学进行了校对,骆仁波为本书整理了全部参考书目。

对于上述同学的理解、支持、参与和在其中做的贡献,在此表示谢意。

在长达十余年的时间里,我的朋友周兴成、王廉、温思美、张岳恒、马晓明、刘正祥、何宪、周延学、刘新林、范晓东、刘书玲、宋振国等,都从不同的方面给予了鼓励、支持和帮助,在此也表示深深的谢意!

本书分为7章,第1章,汉字及其发展,根据一些文件整理介绍汉字的产生和发展,强调汉字文化既是中华文化的基石,也是人类文化的瑰宝。第2章,汉字的计算机信息化,介绍在信息时代文字信息化的重要性、汉字信息化的艰难历程以及现有字库方式存在的科学问题。第3章,智能造字理论基础,介绍传统造字理论、认知心理学理论、拓扑学基础、计算机图形学基础以及智能系统基础。第4章,计算机智能造字理论,包括计算机智能造字模型、汉字基元理论、汉字结构理论、汉字降熵理论和汉字编码方法等。第5章,无字库汉字智能造字技术,介绍汉字基元拓扑变换技术、知识组织技术、无字库智能造字技术。第6章,人

机接口技术,包括汉字输入法评述、皮氏输入法的系统设计及安装与使用等。第7章,无字库智能造字系统,介绍汉字编码辅助系统、计算机造字系统和无字库智能造字系统,介绍实验平台的搭建和实验过程及其结果。

本书的学科跨度大,涉及汉字学、心理学、数学、人工智能和计算机等学科。我想,对于一般的文科学者,大概从第1章~第4章的大部分就可以获得所需要的信息了,而对于对计算机有兴趣的,则可从第2章开读。对于利用拓扑数学来进行理论分析的部分,都在章节号后面标注了“\*”号,可以直接跳过而阅读后续章节,不影响对本书的系统理解。

皮佑国

2012年6月

# 目 录

|                           |     |
|---------------------------|-----|
| 第 1 章 汉字及其发展              | 1   |
| 1.1 文化及文字                 | 1   |
| 1.2 汉字的起源                 | 4   |
| 1.3 汉字的规范和发展              | 9   |
| 1.4 汉字的传播                 | 12  |
| 1.5 屈辱世纪汉字的艰难历程           | 19  |
| 1.6 汉字重振辉煌                | 25  |
| 第 2 章 汉字的计算机信息化           | 28  |
| 2.1 信息化的历史潮流              | 29  |
| 2.2 语言文字的信息化              | 31  |
| 2.3 汉字面临的信息化考验            | 32  |
| 2.4 汉字在计算机中的表示            | 34  |
| 2.5 汉字字库存在的弊端             | 40  |
| 第 3 章 智能造字理论基础            | 49  |
| 3.1 传统造字理论                | 50  |
| 3.2 认知心理学及其相关理论           | 57  |
| 3.3* 点集拓扑学简介              | 65  |
| 3.4 计算机图形变换简介             | 96  |
| 3.5 智能专家系统简介              | 107 |
| 第 4 章 计算机智能造字理论           | 113 |
| 4.1 汉字认读和书写的心理过程          | 114 |
| 4.2 把计算机比作人脑:汉字的计算机智能造字模型 | 116 |
| 4.3 汉字基元理论                | 117 |

|             |                            |            |
|-------------|----------------------------|------------|
| 4.4         | 汉字结构理论 .....               | 146        |
| 4.5         | 智能造字降熵理论 .....             | 168        |
| 4.6         | 汉字编码方法 .....               | 177        |
| 4.7         | 汉字计算机智能造字理论 .....          | 200        |
| 4.8         | 对字库存在问题解决讨论 .....          | 201        |
| <b>第5章</b>  | <b>无字库汉字智能造字技术 .....</b>   | <b>205</b> |
| 5.1         | 汉字基元的拓扑变换 .....            | 206        |
| 5.2         | 智能造字中的知识组织技术 .....         | 223        |
| 5.3         | 无字库智能造字技术 .....            | 234        |
| <b>第6章</b>  | <b>人机接口技术——皮氏输入法 .....</b> | <b>246</b> |
| 6.1         | 汉字输入法述评 .....              | 247        |
| 6.2         | 皮氏输入法的系统设计 .....           | 252        |
| 6.3         | 皮氏输入法的安装与使用 .....          | 270        |
| 6.4         | 智能造字中的输出技术 .....           | 280        |
| 6.5         | 在有字库系统中的生僻字输入 .....        | 281        |
| 6.6         | 对皮氏输入法达成设计目标的分析 .....      | 286        |
| <b>第7章</b>  | <b>无字库智能造字系统 .....</b>     | <b>291</b> |
| 7.1         | 计算机汉字编码辅助系统 .....          | 292        |
| 7.2         | 无字库计算机造字系统 .....           | 357        |
| 7.3         | 无字库智能造字系统 .....            | 379        |
| <b>参考文献</b> | <b>.....</b>               | <b>401</b> |

# 第1章

---

## 汉字及其发展

### ■ 本章重点

- ◆ 文化、语言、文字及其古文字被发现和破译的历史
- ◆ 汉字起源的传说及其相关科学的研究成果
- ◆ 汉字的统一和规范
- ◆ 汉字在各地的传播
- ◆ 屈辱世纪汉字的艰难历程
- ◆ 汉字重振辉煌

## 1.1 文化及文字

### 1.1.1 文化与语言

文化是包括知识、信仰、艺术、道德、法律、语言、习惯等和作为社会成员所获得的任何其他能力及习惯的复杂体。语言不仅以一个子系统的身份参加文化整体运作,而且必须充当所有子系统多种特征、属性、内容的有声符号物化表达形式。也就是说,民族文化中的一切文化现象,在其母语中都必须有相应的有声表达符号。从这个意义上说,语言这种文化现象具有概括整理、表达说明文化中一切文化特点的功能。

### 1.1.2 文字

文字是语言的书面符号系统,是一组附着于语言的基本因子。文字的功能是信息的储存和表达,文字不受时空的限制。因此,文字是文化跨时空传播和传承的载体,是各民族文化积累和交流的重要工具。

语言是先于文字而产生的,可以说语言是随着人的产生而产生的。远古人类有语言而没有文字,信息的交流靠器官传递。即便现在,世界范围内的语言多达数千种,但文字种类的数量就少得多了,许多民族至今仍然没有自己的文字。

### 1.1.3 古文字的发现和破译

世界上,最古老的和曾经产生较大影响的文字是四大文明古国的文字,即古埃及象形文字、古巴比伦楔形文字、古印度哈拉本文字以及古代中国的甲骨文。距今约四五千年时,四大文明古国的先民们,几乎不约而同地都萌发了记录与交流的愿望,于是他们依靠环境和经验,分别用美丽的图画将自己的所思所想记录下来。可是,以追求永恒为目的的文字,却和远古人类一样,弱不禁风。如今,四大文明孕育的古文字中,有三种都被深埋在了岁月的风尘中,只有汉字一枝独秀,传承至今。古代文字的发现和破译也都是近几百年内的事。

#### 1. 古巴比伦楔形文字的发现和破译

1472年,一位意大利人在游览古波斯首都时,在皇宫的遗址里看到了一些刻在墙上的符号(古巴比伦楔形文字),这些符号难道就是上古文字?他把他的疑惑带回意大利,但并没有得到重视。140年之后,另一位意大利人西耶雷卡再次发现这种像星星一样的文字并把它带到了欧洲。那时,欧洲经过文艺复兴,欧洲人开始为发现人类祖先的踪迹,激情澎湃,魂牵梦绕,但并没有取得多大成果。

1835年,年轻英国军官罗林森发现贝希斯顿铭文并制成拓本带回欧洲,经过十多年的苦心钻研,罗林森最终解读了贝希斯顿铭文。湮没了1500年,被发现后解读又困扰欧洲人300年的楔形文字终于通过贝希斯顿铭文一点一点地重获新生。

## 2. 古埃及文字的发现和破译

1799年7月,法国士兵在尼罗河河口附近发现罗塞达碑。1799年9月,法国在与英国争霸的“金字塔”战役中,罗塞达碑又成了英国人的战利品。此后,英、法两国的学者们展开了旷日持久的解读竞争,也许是因为可以时时面对罗塞达碑,英国物理学家托马斯·杨很快便读出了其中几个希腊人的名字。1822年,精通十多门古语法国人商博良正确地判断出古埃及的文字当中既有表音符号,又有表意符号,还有限定符号。商博良因其找到了一个正确的解读方向而成功地解读了罗塞达碑并破译了埃及象形文字。

## 3. 中国古文字的发现和破译

1899年秋天,清朝国子监祭酒王懿荣(1845—1900)从中药龙骨上发现甲骨文,随即派人赶到达仁堂,以每片二两银子的高价,把药店所有刻有符号的龙骨全部买下。后来,他的山东老乡古董商范维卿等人在河南安阳进行搜购,累计共收集了1500多片。一连几天,王懿荣翻遍各种史料、典籍,终于从《周礼·春官》和《史记·龟策列传》中了解到,这就是在秦汉之前的上古文字。王懿荣买下这批龙骨后,找来自己的好友——《老残游记》的作者刘鹗共同研究,很快他们就从这些甲骨文中,读出了中国上古社会许多被淹没已久的秘密。

## 4. 尚未破译的古印度文字

对于古印度印章文字,到目前为止,人们还只能把这些精美的文字当作图画来欣赏,因而古印度的哈拉本文化,只能给人留下一些残缺的印象。

迄今为止,四大文明古国的古文字,已经有三种文字被破译。从破译古文字的过程中会发现一个问题:为什么欧洲人破解楔形文字和古埃及象形文字都是无数专家耗尽了数百年光阴才真相大白,而中国人破解甲骨文,只有一两个人仅仅在几天的时间内就初见成效,而印度印章文字至今还没有被破译呢?

由考古学的知识知道,破译古文字有两种方法:佐证和证前。佐证就是借助其他文字来佐证要破译的文字;证前则是同一文字的以后证前。贝希斯顿铭文和罗塞达碑都是用三种文字书写,三种文字中只要破译了一种就可以相互佐证而破译另外两种。甲骨文则是靠以后证前,以后面的文字去破译前面的文字。以后证前,必要的条件就是只有该文化必须具有继承性。

汉字是目前世界上唯一传承至今的文字,是世界上历史最悠久的文字之一。汉字从古到今传承一贯,是使用人数最多的超级文字。从这个意义上讲,汉字文

明应该属于全人类,保护汉字文化,就是保护世界文化遗产。

## 1.2 汉字的起源

### 1.2.1 关于汉字起源的传说

#### 1. 仓颉造字说

相传,在我国陕西关中有个叫仓颉的人,他生有四只眼睛,披着长发,留着长须,身穿兽皮,手里经常拿着一支笔。他的头形很奇特,头顶高高地隆起,像个奇形的头陀,显得聪颖过人。

仓颉经常骑着毛驴,跋山涉水,跨州过郡,去了解民间风俗习惯,收集民间流行的语言符号。他把收集来的资料写在芦苇的叶子上,装在口袋里,让毛驴驮着。后来,毛驴走到今天的陕西省岐山县就累死了。仓颉带着这批资料,没有毛驴不能再行走,就在岐山县住下来。仓颉不断地观察天象,观察山水风雨演变的现象,辨识鸟兽各式各样的脚印,或野兽、车辆等经过后留下的痕迹,分清和区别各种纹理的异同,并开始创造文字。

仓颉创造出文字后,感动了神灵,神灵从天上投下粮食,夜里的鬼怪也惊吓得哭起来了。不久,仓颉受命担任统一文字和领导文字改革的专职官员,还成为黄帝记载史事的史官,部族的重大档案也都由其掌管。

对于仓颉造字这一传说,东汉许慎在《说文解字·叙》里记述得最为全面:“黄帝之史仓颉,见鸟兽蹄远之迹(指鸟兽行走留下的痕迹),知分理之可相别异也,初造书契,百工以义,万品以察。”“仓颉之初作书,盖依类象形,故谓之文,其后形声相益,即为之字。”这个传说已有两千多年了,而且也有文字记载,但是其可信程度却非常低,历代学者普遍认为仓颉有可能是黄帝部族中极有影响力的巫师之类的人物。上古时期的巫师、史官是由一个人担任的,仓颉在这方面做出过贡献,这样容易让人们把搜集、整理文字工作的人误以为是汉字的创造者。鲁迅曾经说过:“要知汉字成就,所当绵历岁时,且由众手,全群共喻,乃得流行,谁为作者,殊难确指,归功一圣,亦凭臆之说也……在社会里,仓颉也不是一个,有的在刀柄上刻一点图,有的在门户上画一些画,心心相印,口口相传,文字就多起来了。”所以,汉字不可能是某一个人造出来的,应该是汉族先人集体创造的结果,是由千千万万个“仓颉”创造的。

#### 2. 结绳记事说

在汉字起源的传说中,“结绳说”是值得我们注意的一种说法。鲁迅《门外杂谈》里曾经提到过:“我们那里的乡下人,碰到明天要做一件紧要事,怕会忘记

时,常常说‘裤带上打一个结’。”鲁迅家乡是浙江绍兴,在那一带农民请别人代办一件事情时,常说这句话,意思是请对方不要忘记。

关于结绳记事,《周易·系辞》记载:“上古结绳而治,后世圣人易之以书契。”《说文解字·叙》中则说:“及神农氏结绳为治而统其事。”虽然没有明确说文字的起源与结绳有关,但确实告诉我们,历史上曾经有过一个结绳记事的漫长时期。所谓“结绳”,是用在绳子上打结的办法帮助人们记忆,处理日常生活中的一些事务。绳子上面大大小小的结表示不同的意思。结绳是原始民族普遍采用的一种记事方法,中国的许多少数民族和国外的一些民族至今仍有用结绳法来帮助记忆事情的。因此有人认为文字就是从大大小小的绳结发展而来的。

但是,从古人采用的结绳记事法看,结绳和文字毕竟是两回事,它们之间的关系不是相生相袭的,即结绳不可能直接发展成为文字。结绳只能帮助记忆或者作为表示某种简单事务的标记,不能用来表达情感、交流思想,不能成为记录语言的工具。因此,结绳只能算是原始的记事方法,同文字还是有根本区别的。

### 3. 刻契起源说

在《周易·系辞》中除了提到结绳说外,还提到了与结绳说性质相近的契刻说。它是原始社会创始的以契刻某种物体,通过物体遗留痕迹,以反映客观经济活动及其数量关系的记录方式,它被原始人所广泛使用。古文字的“契”字,右边是一把刀形,左边的一竖三横表示是用刀在一块小木条上刻下的三个记号,它形象地反映了契刻的记事方法。楷书的“契”字增加“木”旁则表示这种契刻记事是以木条为材料的。后来“木”误写成“大”,就成了今天的“契”字。

从古代文献记述和西南地区一些少数民族在创造文字以前所存在的刻木记事现象分析,契刻的目的多数是“刻其数”以备忘,有契约、凭证的性质。按照刻契结果所留痕迹的不同,其记事有两种形式:一种形式为在某种物体上契刻或钻孔洞,以缺口或孔洞的数目反映客观经济活动的数量关系;另一种形式为在某种物体上契刻抽象符号,以符号反映客观经济活动及其数量关系。旧石器晚期,距今1.8万多年前的山顶洞人利用禽类腿骨钻孔以记事;距今5000多年的马家窑文化中,原始人使用刻缺口动物骨片以记事,这些都是第一种记事形式的典型代表。这种记事方式以刻契材料的差异而分为刻骨记事、刻木记事和刻竹记事等,但木和竹的质地易朽而无法保存至今。旧石器晚期的峙峪人利用动物骨片刻制划痕;仰韶文化中距今约7000年的半坡人利用陶器刻制符号,这些都是保存下来的第二种记事形式的典型代表。

契刻因其刻在陶器或竹木片上,就有可能逐渐演化成类似青铜器上的族徽文或是竹简木牍这类的文书。从这一点上看,契刻比结绳更具有促进文字产生的条件,但是契刻只能记载一些较为简单的信息,不能满足日益复杂的社会生活

的需要,所以,随着人们传递和记载的信息复杂程度的加深,契刻最终被图画和文字所取代。

#### 4. 书画同源说

在有关汉字起源的几种说法中,“书画同源说”是比较令人信服的。书画同源,就是说汉字起源于图画。汉字是为了记录汉语而出现的一种可以用于在相当大的范围内传递较多信息的手段,必须使接收信息的人和给予信息的人想到同样的内容,写实性的图画可以起到这样的作用。

汉字的起源和图画是密不可分的。图画不仅可以帮助记忆,而且在一定程度上还可以帮助交流思想。不过,图画毕竟只是图画,它与文字之间隔着一条非常深的鸿沟,从带有文字性质的图画转变为早期的象形文字还有一个漫长的过程。随着时间的发展,用图画记录下人类所要表达语言的形式称为图画文字。图画文字并不是真正的文字,用来传递信息的图画只能识其形,而无法读出声音。图画的个体图形和符号不能和语言的词语完全对应起来,它所记录的语言,只能是“近似”。也就是说它还没有同语言建立对应的关系,那么这时它还只是图画。只有当它和语言结合起来,变得可识可读,并且它总体的直观性经过分析,变为一形一词,才可以说它是文字。

#### 5. “河图”、“洛书”起源说

在汉字起源的众多说法中,“洛书”、“河图”的神话传说不可忽视。相传在很久以前,大禹治水时,有条黄龙在洛水中摇摆着尾巴在前面引路,疏通了水道。还有一只玄黑色的大龟,背扛青泥,跟随在黄龙后面,遇到洼地就放下青泥垫平,又把人们居住的地方加高,那些垫得特别高的地方,就成了名山,特别低的地方就成了大川。

还有这样的传说,那只大龟背部还排列着1~9的数字序列,大禹从中悟透秘密,制定了包括五行思想在内的《洪范九畴》,并按照这个排列的次序,划分天下为九州。因负书之龟出自洛水,故该书被称为洛书。

大禹在黄河边观察水情变化时,从河中出现一个白脸鱼身的人,自称是河精,捧给大禹一块光滑的石头。大禹仔细研究,才知是一幅“河图”。这幅河图提供的治水中运用的地理和水文方面的知识,是用图描绘出来的。有了黄龙、玄龟、河精的帮助,大禹治水顺利得多了。

河图、洛书最有名的出处来自于《易传·系辞》中的“河出图,洛出书,圣人则之”。相传在伏羲时代,一龙马从黄河跃出,其身上刻有“一六居下,二七居上,三八居左,四九居右”,此为河图。大禹治水时,一神龟从洛河爬出,其背上数字排列为“戴九履一,左三右七,二四为肩,六八为足,五居中央”,此为洛书。河图和洛书如图1.2.1所示。

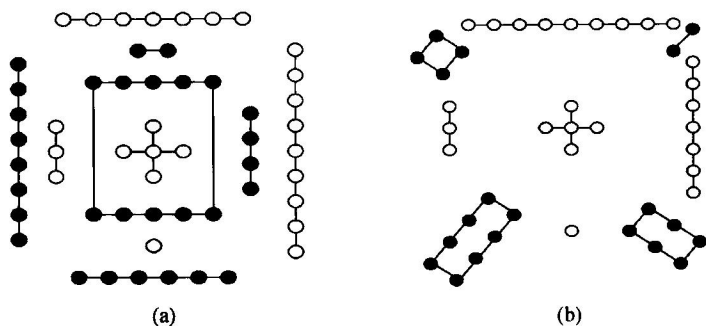


图 1.2.1 河图与洛书

(a) 河图; (b) 洛书。

## 6. 八卦起源说

伏羲氏在河图、洛书的基础上画出了八卦,被认为是最古老的文字符号。文字产生之前,有个过渡阶段,那就是伏羲氏用来表示世界上客观事物的八卦符号,因此八卦符号是创造汉字的来源之一。

八卦符号代表了八种自然现象。它运用的基本符号是“—”(代表阳)、“--”(代表阴)。八卦中“☰”代表天,称为“乾”;“☷”代表地,称为“坤”;“☳”代表雷,称为“震”;“☶”代表山,称为“艮”;“☲”代表火,称为“离”;“☵”代表水,称为“坎”;“☴”代表泽,称为“兑”;“☴”代表风,称为“巽”。任意两卦相叠合就可以得出八八六十四卦,从而揭示了宇宙万物的演变。八卦符号与后来的文字有相似的地方,比如,用“☵”表示水的形象和甲骨文、金文、小篆里“水”的字形是一致的。虽然汉字中确有个别文字或偏旁可能起源于八卦,但汉字与八卦并没有直接的联系,它们是两种不同性质的符号系统。

对于八卦起源说,文字学家唐兰的意见是:“八卦的起源,既是巫者用算筹排列出来的方式,用来做事物的象征,就和文字无关,而且巫术的盛行,恐怕就在殷时,文字已经发生,所以,八卦的卦画,绝不是文字所取代的。”

上述诸多传说都有一定道理,也都不绝对。例如,结绳记事是除了汉族以外许多民族都曾采用过的计数和记事的方法,与记录事情的文字应该存在某种联系,但其方式和水平又与文字相差较大。

## 1.2.2 关于汉字起源的科学研究

汉字的源头究竟在哪里呢?考古发现使人类有可能向自己的祖先无限靠近。根据考古学发现,汉字起源地有黄河下游的大汶口文化、黄河中游的仰韶文

化、长江下游的良渚文化以及长江中游的大溪文化。在这些文化里面发现的史前陶器的刻画符号或者玉器上的刻画符号,都被认为与后来发现的甲骨文系统是可以联系起来的。

1979年,在山东境内的大汶口遗址出土的灰陶制成的刻有神秘符号的酒器中,有一只陶樽的腹部刻着如图 1.2.2 所示的一个符号:上面是个圆圆的太阳;下面是五座连在一起的山峰;对于中间的部分,学者们有着不同的意见。有人认为是火焰,有人认为是一弯明月,还有人认为是一只大鸟。虽然大家对符号的看法不尽相同,但都认为这个符号应该是最古老的汉字之一。因为,除了大汶口,这个符号在山东、莒县、诸城等地出土文物上也出现过,还在南京和安徽出土的文物上也出现过,在良渚文化等不同的文化里也都出现过。也就是说这个图像所代表的含义、所代表的精神方面的内容,在不同的族群和人群里面,大家都是知道的。所以,可以认为,这种符号就可以称为文字。

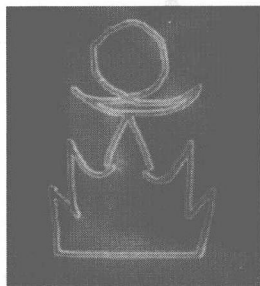


图 1.2.2 酒器上的神秘符号

历代学人的探索都艰难、曲折而扎实。经过半个多世纪的努力,有关汉字产生于距今四五千年前的长江、黄河流域的推论在学术界逐渐成为共识。文化具有从高到低传播的特点,多源头的汉字经过各种渠道的传播,汇集成了主流的汉字源头。

从 1928 年开始,中央研究院历史语言所在小屯先后进行过 15 次考古发掘,其中 13 次发现甲骨。1936 年 3 月 12 日,考古组在 127 坑发掘出龟甲和手鼓结成的一根高 2 米、直径 1 米、重约 6 吨的甲骨柱,在这里出土了一万七千多块甲骨。一代又一代的考古学家,经过 80 年的整理和辨识,让那些消失了三千多年的岁月复活了。

如果说甲骨文是用来与祖先沟通的话,那么铸造在青铜器上的金文则是为了垂诸后世。1978 年,湖北随县战国时期的曾侯乙墓,出土了 140 件青铜器、65 件青铜编钟以及 4500 件青铜兵器。据粗略统计,这些青铜器的总重超过 10 吨。

通过汉字,自然的血脉关系被提升为神圣的仪式和信仰。中国最初的国家就是在这样的仪式和信仰上建立起来,这正是后世华夏民族文字崇拜的源头。

汉字是多源的,其中黄河流域这个源头发展得最好。文化是从先进文化的地方流到落后文化的地方去,所以源头虽多,黄河流域这一个源头的文化就慢慢流向其他地方,逐步变成了汉字主要的源头了。

## 1.3 汉字的规范和发展

### 1.3.1 汉字的统一

多源头的汉字要实现统一,必然要经历一次又一次的规范和整理的历程。

早期的汉字,在写法上各有各国的特点,一字往往有多种写法。1965年12月,陕西省侯马市东郊,出土了春秋晚期晋国记录诸侯间结盟誓词的官方文书——侯马盟书。在侯马盟书中,“敢”字有九十多种写法,而“嘉”字的写法超过了一百种。为了书写方便,每个执笔者都可以根据自己的喜好简化汉字,而这一点也毫无例外地体现在春秋、战国的钱币和兵器上。中国文字学的鼻祖、东汉学者许慎对当时的描述是:“七个诸侯国各自为政,废弃礼乐,从而导致天下车途异轨,律令异法,言语异声,文字异形。”这种情况下,对汉字的统一提出了迫切的要求,但是,要满足这一要求,既需要巨大的努力,也需要历史机遇。

公元前221年,秦王嬴政借助商鞅变法后一百多年里积攒的力量,并吞六国。江山一统,为汉字的统一和规范提供了必要的政治环境。公元前223年,廷尉李斯奉诏整理六国文字,“罢其不与秦国合者”,规定了统一的书写系统——小篆。这一政令首先对汉字的写法进行规范,把它的写法固定下来。其次,把每一个字的字式即结构固定下来,同时确定某些字和词的搭配关系。这样就使得小篆成为了一种更加严密的文字系统。

公元前223年,在始皇帝诏告天下的文告中,将书同文字列为自己最显赫的功绩之一。“文”与“字”合成一个词第一次出现在汉字的长河中。此后在汉字统一的疆域中,虽然仍并存着数以百计的方言、子方言,但天南地北的人写一样的字,读一样的书,承接一样的历史经验。文字的统一,将不同的族群融合成了一个伟大的民族。随着岁月的变迁,朝代会土崩瓦解,天下会分分合合,甚至汉字的形态仍然会发生变化,但是“书”必同“文”,成为汉字顽强坚持的原则。

在秦始皇统一文字之前,仅秦国的文字也是有很多种写法的,例如,春秋中期,秦国的“秦”字,有的写作从春省从双禾,有的写作从春从三禾。到了春秋晚期,各国文字的差异越来越大,晋国侯马盟书上的草写大篆,要想看到它和西周金文之间的传承关系已经很吃力了,但是,六国文字和秦文字,都是来源于西周金文,是同源的,只是西周金文毕竟还是一种象形意味很强的文字,偏旁部首的位置不统一,形状不统一,这就是后来诸侯文字变异的根源所在。在秦国立国之初,自身文化很不发达,所以,一笔一画地拿来了西周文明。在秦文化拿来主义和保守主义的双重作用下,秦国文字却变成了最规范的文字,而其他六国文字则