

姜杰 著

汉字行书与篆隶

江苏人民出版社

非外借

本书系国家自然科学基金十三五”规划 2016 年度教育学一般课题“基于数字手写设备的汉字书写自动化教学与评价技术方案设计研究”(BCA160052)的系列成果之一。

汉字书写表征与教育实践

姜杰 著

图书在版编目(CIP)数据

汉字书写表征与教育实践 / 姜杰著. — 南京: 江苏人民出版社, 2023.2

ISBN 978-7-214-27677-3

I. ①汉… II. ①姜… III. ①汉字—书法—研究
IV. ①J292.1

中国版本图书馆 CIP 数据核字(2022)第 222025 号

书 名 汉字书写表征与教育实践
著 者 姜 杰
责任编辑 王 溪
装帧设计 陈 婕
责任监制 王 娟
出版发行 江苏人民出版社
地 址 南京市湖南路1号A楼, 邮编:210009
照 排 江苏凤凰制版有限公司
印 刷 江苏凤凰通达印刷有限公司
开 本 718毫米×1 000毫米 1/16
印 张 16.75 插页 1
字 数 259千字
版 次 2023年2月第1版
印 次 2023年2月第1次印刷
标准书号 ISBN 978-7-214-27677-3
定 价 78.00元

(江苏人民出版社图书凡印装错误可向承印厂调换)

目 录

书写表征篇

第一章 绪论 3

第一节 汉字书写表征研究的背景与意义 3

第二节 现有汉字字形、结构表征体系研究概述 7

第三节 现有汉字笔画识别与标识技术的研究概述 12

第二章 汉字书写表征与表征语言设计 15

第一节 文字书写过程中的意义呈现 15

第二节 书写表征的内容与过程 17

第三节 书写描述语言(WDL)的体系 21

第三章 汉字书写笔画的表征 29

第一节 汉字书写与汉字笔画系统 29

第二节 汉字笔画书写的数字化记录 33

第三节 基于 WDL 的笔画书写表征 39

第四节 笔画表征的技术案例——脱机汉字笔画抽取技术分析 47

第四章 汉字书写笔画间关系的表征 51

第一节 汉字笔画相交关系与 WDL 表征 51

第二节 汉字笔画间的相接关系与 WDL 表征 54

第三节	汉字笔画间的相离、相邻与相对位置关系表征	57
第四节	基于笔画类型与笔画关系表征的书写正确性判定技术	63
第五章	汉字书写部件、整字的表征	68
第一节	汉字部件的概念界定与汉字基础部件表征	68
第二节	基于 WDL 的基础部件与整字的表征	70
第三节	基于部件与整字表征的嵌套结构算子技术	74
第六章	汉字书写笔序的表征	82
第一节	汉字书写笔序的基本概念与内涵	82
第二节	汉字书写笔序的记录与标识	85
第三节	基于笔序表征的续笔表征技术	86
第四节	汉字书写笔顺(规范性笔序)的判断技术与表征	95
第七章	汉字书写笔势的表征	101
第一节	汉字书写中笔势的基本概念界定	101
第二节	书写径迹点集提取与笔势表征	102
第三节	汉字基本笔画规范笔势表征模型	106
第四节	基本笔画规范笔势表征模型的笔势评价技术方案	111

评价方案篇

第八章	汉字书写正确性评价方案	119
第一节	严格的正确性评价方案概述	119
第二节	综合性评价方案	123
第三节	综合性评价方案效果测试	129
第九章	汉字书写工整性评价方案	134
第一节	汉字书写工整性评价概述	134
第二节	汉字书写工整性评价方案与效果	139
第三节	汉字书写跟踪评价技术方案与效果	148

第十章 汉字书写篇章质量评价方案	154
第一节 篇章书写质量评价概述	154
第二节 篇章书写质量评价方案设计与实施	158
第十一章 智能汉字书写教学系统的设计方案	172
第一节 面向个性化的书写评价与指导设计	172
第二节 面向群体管理的班组设计	177
第三节 面向不同群体的字集支持设计	186

教育实践篇

第十二章 多层次的书写教学实践研究	191
第一节 面向基础教育人群的系列教学实践	191
第二节 面向师范生书写教学的设计与实践	209
第三节 面向汉语非母语群体的教学实践	222
第十三章 多形式的教育应用	232
第一节 基于智能系统的大规模信息化测评	232
第二节 基于智能系统的书写竞赛设计与组织实施	238
结 语	254
参考文献	257
后 记	262

书写表征篇

文字书法作品,由书写而得,从落笔点开始就隐含了书写者对汉字空间布局的构思。起笔、行笔、收笔完成笔画的书写,就蕴含着书写者个性化的笔画形态特征,逐笔完成的笔画之间,呈现出书写者的构字结构风格,最终文字的书写完成,是该字的符号呈现,是汉字构字表意的完成,是书法艺术的展现,亦蕴含着内在的个人风格、素养乃至性格精神。

书写表征篇,探究书写表征的历史文化意义,展示对汉字书写进行数字化表征的对象与过程,表征的方案与方法,以及基于书写表征的部分相关技术说明与验证。本篇的书写表征:微到运笔书写中每个点位的运笔控制笔势模型,小到笔画书写形态的走向码模型与笔画识别提取方法,中到笔画间交、接、邻、离的方位关系模型及其蕴含着的文字正误判断,大到部件方位关系构成整字表征与结构算子方法,将书写表征的思想、原理、技术层面的内容做了具体的呈现。

第一章 绪论

第一节 汉字书写表征研究的背景与意义

一、汉字书写历史与文化内涵

汉字又称作中文字、中国字、国字,属于表意文字的词素音节文字,由华夏族先民发明、使用、改进,经历了史前符号,商朝的甲骨文,周代的金文、大篆,秦朝的小篆,汉代及其之后的隶书、草书、楷书和行书的演变。汉字是中华民族智慧的珍宝,是华夏五千年文明得以传承的最重要的载体。

相对于其他国家使用的拼音文字而言,汉字的构成显得更加丰富和具有内涵,在众多的语言文字中,唯有中国的汉字书写最终演变成为一门艺术^①。这一在中国乃至中、日、韩等汉字文化圈内享有很高社会文化地位的艺术,在欧美却没有相对应的艺术形式。英语中“calligraphy”被翻译成“书法”,但西方通常把“书法”看成一种技艺,是因为他们无法体验蕴含在汉字书写中的历史、文化乃至民族魄力和民族精神。

在华夏民族的文明史中,文字的书写很早就脱离了文字记录的基本技能需求,早在西周时,六艺“礼、乐、射、御、书、数”,中的“书”就是书法一艺^②。后续,文人墨客都通过书法的形式来表达自己的内心世界,以艺术语言来获得美的享受。汉代

① 李志平. 中国书法艺术[M]. 湖南:湖南大学出版社,2004

② 王进. 书源六艺中,博识神采生[J]. 美术观察,2009. 4,35—38

扬雄《法言·问神》中：“言，心声也；书，心画也。”这恰如其分地表达了书写艺术的核心所在，艺术精神所在^①。

二、数字时代文字记录技术与书写困惑

文字书写发展的每个阶段中，导致其变化的因素主要有两个，一是技术，二是使用的方便性。

技术手段的变化，可以提供更佳的书写工具与介质搭配，进而推动新的书写（记录）的形式与形态的产生。文字的基本功能是记录，记录的方便性、实用性是驱动文字书写形式演化的强大动力。随着书写工具与介质的不断出现，更加方便的书写（记录）成为自然的诉求，在纠结及争执中，一种新的状态便出现了。典型的书写工具与介质搭配变迁可以大致分割为：金石时期（如：树枝画于石壁，青铜刀和玉质刀刻画于龟甲兽骨等）、毛笔竹帛时期、毛笔纸张时期、硬笔纸张时期，简单称为金石、竹帛、毛笔、硬笔四个时期。而现在，我们正开始进入第五个时期，即“数字时期”。

考察前四个时期文字记录过程的更替，虽然伴随着书写工具和记录介质的变化，文字依然是各种信息记录的主要载体。在信息记录的过程中，依然保留着传统“书写”动作行为。考察数字时代，视频、图片、语音等多媒体形态信息占据了日常工作、生活、学习、交流等绝大部分应用的场景，文字应用的场景被极大挤压，“书写”这一生成文字的动作行为逐渐消失。

究其原因，在数字时代相对有限的文字应用场景中，键盘代替了笔，屏幕代替了书本，计算机、平板电脑成为学习与工作主要工具，生成字的“敲击输入”正逐步取代“书写输入”。与此同时，当智能手机这类信息化的移动终端成为人们最常用的通信与交流工具之后，有限的文字信息沟通中的书写行为也被手机屏幕上点点戳戳的动作所替代，抑或通过语音/文字转换功能采用口语输入。书写，作为一种生成文字信息的行为，逐渐变得可有可无。

于是国人的汉字书写能力大幅下降，很多人不仅写不出漂亮的汉字，甚至挂在

^① 闵祥德. 中国书法艺术[M]. 广东: 广东人民出版社, 2007

嘴边的字也经常提笔就忘,几乎患上了“书写遗忘症”。光明网 2010 年 11 月所做的关于“在网络时代如何看待汉字的书写”的调查显示:85.29%的人认同全民汉字手写水平在下降,44.25%的人觉得自己的字不好看,41.52%的人经常提笔忘字,14.23%的人经常写错别字。这就是“数字时期”,文字应用场景萎缩、文字录入行为的改变,给汉字书写带来的最大困惑——“书写能力的遗失”。

三、对书写教育的呼唤与教学困境

在诸如光明日报等多家媒体不断报道国人汉字书写水平下降的事实并呼吁汉字书写教育重视的同时,社会和媒体首先给予了积极的响应,以汉字书写为核心内容的电视节目如《汉字听写大赛》《汉字英雄》等纷纷都在收视的黄金时段推出,被大众所关注和热议。

政府部门则通过制定、发布相关政策来规范汉字书写教育,分别在 2011 年颁发了《关于中小学开展书法教育的意见》,2013 年又发布了《中小学书法教育指导纲要》的文件,这些指导性的文件除了按年龄段要求学生逐步掌握硬笔和软笔汉字的正确书写,还力求书写规范、端正、整洁。《国家中长期语言文字事业改革和发展规划纲要》(2012—2020 年)对语言文字工作提出了要求,倡导现阶段国家重点发展的软实力中应当包括语言和文字实力,着重全民语言和文字能力的提升与和谐语言文字生活的构建,促进语言文字工作的全面开展。

相关政策尤其对国内基础教育学校提出了专项书法课程开设以及专业书法教师配备的明确要求。然而,专业书法教师的匮乏以及无法开展实时书写指导,又严重制约着汉字书写教育的有效开展。

与国内汉字书写教育困境相对应,国际上汉语、汉字教育与推广迎来了更大的机遇与发展。教育部 2021 年底发布统计显示:我国与全球 159 个国家通过中外合作方式设立了 1 500 多所孔子学院和孔子课堂,累计培养各类学员 1 300 多万人。有 180 多个国家开展了中文教育项目,75 个国家通过颁布法令、政令等方式将中文列入国民教育体系,4 000 多所大学设立了中文院系、专业、课程,7.5 万多所主流中小学校、华文学校、培训机构开设中文课程。目前全球正在学习中文的人数超过 2 500 万,全球累计学习使用中文人数接近 2 亿人。

全球学习汉语的热潮,是经济全球化背景下,中国经济硬实力提升的必然结果,更是中国历史、文化、价值观传播的重要途径。同样在2021年,中文正式成为联合国世界旅游组织官方语言,中文在国际交往中的作用日益凸显,中文的国际影响力不断攀升。与此同时,我们都有这样的认知,相对于其他语言,汉语是一门相对难以掌握和学习的语言,令作为表意文字的“方块”汉字体系,令很多学习中文的爱好者困惑。相关教育研究显示:那些非汉字文化圈的汉语学习者普遍对汉字书写有较严重的畏惧心理,汉字书写成为汉语学习者学习中的巨大门槛。

四、汉字书写表征的作用与意义

汉字的信息化标识与表征,始于汉字编码,如国标码(国家标准信息交换用汉字编码(GB2312—80标准))、BIG5码、Unicode码、ISO2022 CJK码等。汉字编码的目的是按既定的编码规则给每个汉字赋予唯一的代码,以便计算机辨认、接收和处理。与汉字字形相关的编码主要用于计算机字体轮廓特征描述的(函数)编码方案,尤其多用于矢量字体(如:TrueType字体)轮廓描述的编码,即用数学函数描述字体的轮廓外形,含有字形构造、颜色填充、数字描述函数、流程条件控制、栅格处理控制等指令。更进一步的汉字字形表征研究,是基于汉字结构特征的描述(如:CDL等),关注汉字的笔画、部首(部件)、结构特征的分析与表征,主要用来解决汉字字体重构与字库的生成问题。

汉字书写表征是从书写的角度来描述与解释汉字:首先进行基本笔画的书写表征,进而通过描述笔画与笔画之间的关系与结构来完成部首(部件)书写的表征,再通过部件间关系与结构的描述来完成整字书写的表征,并通过书写中笔画间和笔画内的过程描述完成笔序与笔势的表征。

汉字书写是国人的一项基本技能,更是书法艺术的根基。汉字书写表征对汉字静态字形与结构展开了动态的、个性化的书写记录与标识。在汉字书写表征基础之上,汉字书写质量的智能化评价成为可能,面向国内外的数字化汉语书写教育能够更好开展,书写者的笔迹分析与笔迹识别可以更便捷实现,个性化字体字库的自动化生成以及汉字书法艺术的动态展示也能成为现实。

第二节 现有汉字字形、结构表征体系研究概述

汉字表征的研究基本可以归纳到汉字字形和结构表征的研究领域,与其最为相近和值得参考、借鉴的相关研究成果包括两大类:第一类是现有字形、结构描述体系方面的研究成果,主要包含对汉字拆分与表征的层次研究以及结构方法研究。第二类是对既有汉字拆分、分析层次中,最基本分析元素(基元)——“笔段”的识别与提取技术方法研究成果。其中第一类字形、结构描述体系的标准规范以及相关研究成果主要包括国标《汉字部件规范》、Unicode 的表意文字描述序列 IDS 等。

一、国标《汉字部件规范》

在国标《汉字部件规范》出台之前,上海交通大学曾经出版过《汉字信息字典》^①,该字典描述了 7 785 个汉字的结构和部件,并以上下、左右、包容、嵌入 4 种结构加以描述构字部件构成的汉字。

国家语委于 1997 年 12 月 1 日发布,并于 1998 年 5 月 1 日正式实施了《信息处理用 GB13000.1 字符集汉字部件规范》^②(简称《汉字部件规范》)。《汉字部件规范》对 GB13000.1 字符集中 20 902 个汉字逐一拆分、归纳与统计后制定了《汉字基础部件表》。该表收纳了 560 个汉字基础部件,按部件之间的关系,将相关部件归纳成 393 个组。《汉字部件规范》中还给出了汉字部件的拆分规则,以及拆分规则使用说明。《汉字部件规范》已经实施了 20 多年,目前被广泛使用,成为国内汉字权威的部件分类规范。

二、表意文字描述序列 IDS(Ideographic Description Sequences)

2000 年发布的 Unicode4.0 标准提出使用表意文字描述序列 IDS 表达式来描





^① 上海交通大学汉字编码组. 汉字信息字典. 科学出版社. 1988


^② 国家语言文字工作委员会. 信息处理用 GB13000.1 字符集汉字部件规范[Z]. 上海教育出版社, 1999.

述表意文字的结构。其递归语法如下：

$\langle \text{IDS} \rangle := \langle \text{CJK 汉字} \rangle | \langle \text{CJK 部件} \rangle | \langle \text{二元结构类型符} \rangle \langle \text{IDS} \rangle \langle \text{IDS} \rangle | \langle \text{三元结构类型符} \rangle \langle \text{IDS} \rangle \langle \text{IDS} \rangle \langle \text{IDS} \rangle$ 。

Unicode4.0 同时给出了 12 个表意文字结构类型符号(IDC)作为操作符,汉字、部件作为操作数,来描述表意文字的结构。具体有:

- (1) , 左右结构类型符号, 例字: 明、朋、徕等。
- (2) , 上下结构类型符号, 例字: 吕、冒、炎等。
- (3) , 左中右结构类型符号, 例字: 谢、辩、彬等。
- (4) , 上中下结构类型符号, 例字: 竞、章、蒸等。
- (5) , 全包围结构类型符号, 例字: 国、回、园等。
- (6) , 上包围结构类型符号, 例字: 问、同、冂等。
- (7) , 下包围结构类型符号, 例字: 幽、函、凶等。
- (8) , 左包围结构类型符号, 例字: 区、医、匠等。
- (9) , 左上包围结构类型符号, 例字: 压、庞、展等。
- (10) , 右上包围结构类型符号, 例字: 司、句、氨等。
- (11) , 左下包围结构类型符号, 例字: 逃、逢、赵等。
- (12) , 覆盖(穿插)结构类型符号, 例字: 坐、爽、噩等。

以此为基础,IDS 采用结构类型操作符前缀表达式,例如“渐”字,其语法为: 车斤。

IDS 划分了 12 个固定结构类型,但没有固定部件的种类,甚至可以用任意字作为部件,有较大的灵活性。

三、汉语文档处理语言 CPL(Chinese Document Processing Language)^①

CPL 由台湾“中央研究院信息技术研究所”在 1990 年研发,基于 Big5 和汉字部件的字形描述系统,主要为古籍整理服务。CPL 把字根和部件当作汉字的构形单位,汉字可以由有限的字根组成,CPL 确定了 1 千多字根来组合 4 千多部件,总

^① 孙星明,殷建平,陈火旺等. 汉字的数学表达式研究[J]. 计算机研究与发展,2002,39(6):707—711

共可涵盖 5 万多字形。CPL 的组合类型有层次机构和平面结构两种。

层次结构的字由部件分作若干层次累加构成,平面结构的字形可以直接用部件的书写序列加上起点标志和结束标志即可。

其中层次结构的类型有直连 \triangleleft 、横连 \triangleleft ,以及包含 \triangleleft 。

如“明”可字表示为:日 \triangleleft 月。

相对于 IDS,CPL 则是规定了字根、部件类型,没有规定具体结构类型,只说明字根、部件连接方式,故而在组合方式上有较大的灵活性。

四、汉字数学表达式^{①,②}

汉字数学表达式由国防科技大学孙星明,陈火旺等学者提出,它将汉字表示成:部件间位置关系为运算符号,汉字部件作为操作数的数学表达式。汉字数学表达式选定了 505 个部件,其可以通过位置运算组合成汉字。结构关系分为:lr(左右)、ud(上下)、ld(左下)、lu(左上)、ru(右上)、we(全包围)等。

表 1.1 汉字书写表达结构、表示符号

结构关系	表示符	类型	优先级、运算方向
	we	全包围	优先级较高(低于括号);自左向右运算
	lu	左上包	
	ld	左下包	
	ru	右上包	优先级较低;自左向右运算
	lr	左右	
	ud	上下	

① 张问银,孙星明,曾振柄等. 汉字数学表达式的自动生成[J]. 计算机研究与发展,2004,41(5):848—852

② 罗刚,孙星明. 汉字书写表达式开发平台的设计与实现[J]. 计算机工程与应用,2005(5):113—116

五、HanGlyph 汉字分层描述言^①

由香港 Baptist 大学 Candy L. K. Yiu 等提出的一种分层标识汉字的语言。HanGlyph 把汉字分成:上下结构(=)、左右结构(|)、全包围(@)、穿插结构(+),以半包围结构(^)^②,又将半包围分成 8 种形态,用数字 0—7 来标识。HanGlyph 用这些结构符号从笔画级到部件级构,最终构成汉字。

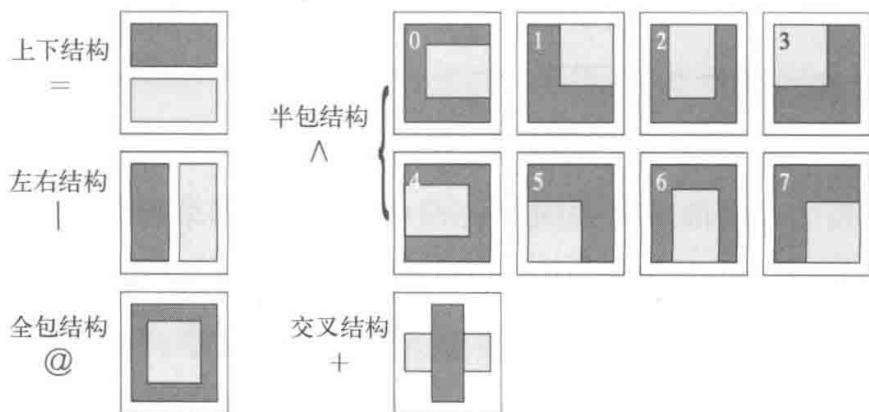


图 1.1 HanGlyph 的结构分类与标识符号

HanGlyph 采用结构描述后缀表达,其构字语法如下:(“/”用来分割表达式)

let(sih){si|h=/:“口”{竖与横折,左右结构;与“横上下结构”}

let(ri_4){sih h@}:“日”{“口”与“横”,全包围结构}

HanGlyph 在描述半包围时,可以在结构符(^)后面跟数字 0—7,来描述具体的包围结构形态,但在描述多重交叉时,HanGlyph 相对比较困难。

六、CDL 字符描述语言(Chinese Character Description Language)

CDL 是由美国加州大学文林研究所 Tom Bishop 和 Richard Cook 提出的一种基于 XML 语言利用笔画和部件进行汉字字形描述的技术。CDL 采用分解的方式将文字拆分到部件甚至到笔画。在笔画层面给出了 39 种精确描述的笔画类

^① Candy L. K. Yiu, Wai Wong. Chinese Character Synthesis Using METAPOST[EB/OL]. 2014-10-9

^② Candy L. K. Yiu, Wai Wong. Summary of HanGlyph(Ver. 1.0)[EB/OL]. 2014-10-9

型集合^①,并可以通过笔画的起点、拐点和终点来控制笔画的放大缩小与变形,在部件层面没有给出完全的限定,可以通过笔画“人为构造”出想要的各种部件乃至各种“字”。对于构造出来的部件和字,CDL 通过外轮廓矩形加以描述,在此基础上通过对矩形的缩放变形进而组合出更多想要的字。

CDL 采用 XML 规范的语法来描述构造的字^②,以“𪛗”(ráng,囊的古体字)字为例,先从整体考察,其部件层面(包含空间坐标定位)的 CDL(基于 XML)描述语法为:

```
<cdl char='𪛗' uni='2714D' points='0,0 * 128,128'>
  <comp char='艹' uni='8279' points='0,0 * 128,16'>
    <comp char='𪛗' uni='21109' points='2,24 * 128,128'>
  </cdl>
```

细化考察其笔画层面,CDL 将整字的每一个笔画标注类型(如,'h'表示“横”),同样通过笔画的外轮廓矩形标记其位置,其采用 CDL(基于 XML)描述语法为:

```
<cdl char='𪛗' uni='2714D'>
  <stroke type='h' points='0,8 * 128,8'>
  <stroke type='s' points='40,0 * 40,16'>
  .....
</cdl>
```

相较于前述的字形描述体系,CDL 对于给定的笔画类型有精确形态与笔向的描述。CDL 通过笔画、部件的坐标定位方式,突破了传统上下、左右、包围等结构类型划分的局限,定位精确而且灵活。CDL 采用基于 XML 的标识方式,虽不便于普通用户直接阅读使用,但更具规范性以及便于检索和开发应用。

七、SCML 结构化字符建模语言^③

SCML(Structural Character Modeling Language)是美国达特茅斯学院 Daniel

① Tom Bishop, Richard Cook. The Set of Basic CJK Unified Stroke Types[EB/OL]. 2014-10-14

② Tom Bishop, Richard Cook. Wenlin CDL: Character Description Language[EB/OL]. 2014-10-15

③ Peebles Daniel G., Balkcom, Devin (Advisor) SCML: A Structural Representation for Chinese Characters, Technical Report TR2007-592 (PDF), Dartmouth College. [EB/OL]. 2014-10-15