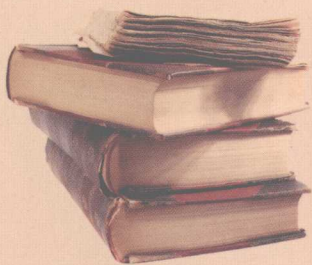


汉字构形属性

历时演变的量化研究



周晓文 著

汉字构形属性

历时演变的量化研究



周晓文 著

任何一门科学，

只有当它充分运用了数学时，

本书在教育部国家人文社会科学重点研究基地北京师范大学民俗典籍文字研究中心立项并获中心出版资助

中国广播电视出版社
CHINA RADIO & TELEVISION PUBLISHING HOUSE

图书在版编目(CIP)数据

汉字构形属性历时演变的量化研究 / 周晓文著. —北京:
中国广播电视出版社, 2008. 2
ISBN 978 - 7 - 5043 - 5577 - 5

I. 汉… II. 周… III. 汉字—结构—演变—研究
IV. H122

中国版本图书馆 CIP 数据核字(2008)第 036961 号

汉字构形属性历时演变的量化研究

周晓文 著

责任编辑 高子如
封面设计 丁琳
责任校对 张哲

出版发行 中国广播电视出版社
电 话 010-86093580 010-86093583
社 址 北京市西城区真武庙二条9号
邮 编 100045
网 址 www.crtpt.com.cn
电子信箱 crtpt8@sina.com

经 销 全国各地新华书店
印 刷 涿州市京南印刷厂

开 本 787 毫米×1092 毫米 1/16
字 数 150(千)字
印 张 8.25
版 次 2008 年 2 月第 1 版 2008 年 2 月第 1 次印刷

书 号 ISBN 978 - 7 - 5043 - 5577 - 5
定 价 19.50 元

(版权所有 翻印必究·印装有误 负责调换)

“任何一门科学，
只有当它充分运用了数学时，
才能成为真正的科学。”

——卡尔·马克思

汉字构形系统的共时描写与历史比较

——代本书序

王 宁^①

周晓文的这本书，是我们 20 世纪 90 年代设计的“汉字构形史”课题的一个子课题，她是计算机专业的硕士，以后又专攻汉字学博士，这本书是用数理统计的方法，对已经形成的历代汉字构形属性的数据加以处理，从而对已经形成的汉字构形系统论的思想加以证实。这是一个文理科交叉的课题，又是一个历史比较的课题。她的成果不但得到了一些计算机信息处理专家的肯定，而且也得到汉字构形史其他分课题作者的认同，这说明了 she 选取这个数理统计的角度来面对汉字符号的做法是成功的。

系统论是从自然科学中总结出的哲学方法论，体现了唯物辩证法“联系的普遍性”的思想，我们在这个思想的指导下，借鉴自然科学的方法，规定了一套操作程序来描写文字符号、解释普遍的文字现象，产生了一批统计数据；但是，这些数据是针对不同时期、不同字体的文字材料统计出来的，取字的底数也不一样，要把它们综合到一起，给数理统计提出了一些新的问题。汉字是一种人文符号，受社会因素的影响极大，分类、分级有时并没有精确的界限，常常带有十分模糊的中间地段，这也会给精确计量带来相当的困难。这些问题，本来都是文理科交叉的课题必然碰到的问题，而周晓文对这些问题的解决，表现出她利用自己文理兼修的知识结构优势所产生的创新精神。经数理统计得出的新数据，不是个别数据的简单相加，而是对汉字构形系统的一种新的构建，也是对汉字构形系统思想从新的角度再次的证实。周晓文课题的完成，给汉字构形史的进一步研究，提供了新的思路。

由于这个课题与前面每个具体课题的直接相关关系，我在下面把已经出版的《汉字构形史丛书·总序》略加删改，来代替本书的序言。

汉字是世界上唯一未曾中断使用而延续至今的表意文字系统。可以说，在包括甲骨文在内的每一层汉字的共时平面上，都已经积淀了非常深厚的汉字本体历史，以及

^① 王宁. 汉字构形学讲座 [M], 第 17 页, 上海: 上海教育出版社, 2002.

作为汉字存在背景的社会文化历史。历史是现实的一面镜子，了解汉字的现在，必须有历史做参照，才可能准确。时至20-21世纪，汉字问题已经成为中国文化建设极为重要的问题，并且得到世界各国的普遍关注，但是，汉字历史的脉络仍然没有梳理得十分清楚，致使一些重要问题，特别是汉字发展到今文字阶段的问题，因为没有得到历史的印证而难以得出明确的结论。比如，发展到隶楷阶段的汉字是不是表意文字？它的发展趋势是不是纯粹的简化？历代字书贮存汉字的数量不断增多是不是汉字必然发展的结果，是不是汉语的需要？在社会使用层面上的汉字是否具有系统性？汉字的规范是不是必要？怎样规范才是符合科学的？是否需要保护汉字的构形理据？如何保护汉字的理据才符合汉字的发展而不复古？诸如此类的问题没有对汉字历史的深入研究都是难以回答得具体、准确而周全的。

研究汉字的历史，必须从汉字发展的事实入手，真正的历史首先是符合事实的历史，一切先有了概念，规定了某种观点再去附会的办法，不可能总结出真正的历史。全面的历史来源于对事实的系统考察，一切只凭个别例子支撑而不作全面考察的办法，在总结历史时，不可能消除片面性。考察汉字的历史首先要考察汉字本体的状况，词汇史代替不了汉字史，文化史代替不了汉字史，字典史（字典收录汉字的历史）也代替不了汉字史。汉字史不是与上述这些领域的历史没有关系，但不能由它们来代替，汉字史必须是汉字本体发展的历史。

汉字史的研究应当以科学、系统的理论为基础，这首先是因为，有了对汉字的理性认识，才能使我们总结的历史是客观的、概括的，而不是感想的或猜测的。其次还因为，汉字史的研究必须倚赖各个历史层面的比较，历时的比较又必须倚赖共时的描写。描写和比较又都要有科学的方法，这种方法还必须具有操作性。为了汉字史结论的科学全面而有说服力，我们必须首先创建可操作的描写汉字学。

二

此前的汉字学，已经是一门内容非常丰富的学科。历代的汉字学，包括汉字形义学、汉字文化学、古汉字考释学、汉字个体字形演变学……都取得了大量的成就。但是，对汉字本体的研究，特别是关于各个共时层面上汉字总体系统的描写，一直没有得到足够的重视。

什么是汉字的本体？汉字是记录汉语的视觉符号，它的音与义来源于汉语，字形才是它的本体。在中国，把字形作为汉字的中心来探讨，从理论上研究其内在的规律，必须首先克服传统汉字观造成的两种固有的积习。这两种积习，都是早期汉字研究的实用目的带来的。

中国古代的文字学被称作“小学”，因“周礼八岁入小学，保氏教国子，先以六书”这一制度而得名。它的目的开始是起点很低的识字教学，两汉经今古文斗争以后，因古文经学家的推崇，一下子上升为考证、解读儒家经典的津梁，也就是讲解古

代书面文献的工具，而有了崇高的地位。汉字所以能成为解读古书、考证古义的依据，不仅因为它具有记录汉语的功能，还因为它始终是表意文字系统，据义而构形，从字形上可以探求词的本义，以本义推导引申义，再以语音为线索找到本字而辨明假借义。由词义而句意，由句意而章旨，由章旨而知文献所传播的思想。这种十分明确的解读文献的实用目的，造成了“小学”固有的形、音、义互求的传统方法，这种方法无疑是十分正确的，是完全适合汉字的，是中国文字学的精华；但是由于这种方法最初的目的只是释读古代经典，因此又必然是以义为出发点又以义为落脚点的。“六书”是传统文字学分析汉字构形模式的凡例与法则，但是，“六书”的前四书虽可以较好地涵盖《说文》小篆的构形类型，后二书却与构形没有直接关系。细究“六书”的意图，很大成分是着眼在探求形中的意（造字意图）和义（构字所依据的词义）。只有兼从“释字之法”而不是单从“造字之法”的角度，也就是汉字形义学的角度，才能准确理解“六书”；而通过造字意图来探求词义，已经跨越到语言学范畴，并非单纯的汉字本体研究了。“字”在“小学”家心目中常常是“词”的同义语。正是因为他们忽略作为语言载体的文字相对的独立价值，才经常弄得“字”与“词”混淆，文字学与训诂学划不清界线。

同样由于解读文献的实用目的，“小学”家对汉字的关注一般是以个体为对象的。汉代以及此后纂集的“小学”专书，大多以音、义为纲来制订编则，以形为纲的《说文解字》尽管包含着十分宝贵的构形系统的思想，但是由于使用它的人解读文献的实用目的太强了，后代人对它的应用，多数都着眼在对单个汉字的考据；而对它的评价，自然也以它提供的每个形体与意义考证的效果为标准，较少注重书中包含的构形系统思想。后代《说文》一类的字书在理论的自觉性上比之《说文》远远不如，绝大多数只是袭用《说文》的框架来罗列字形——而且是不同历史层面的字形——因而很难看出汉字形体所具有的系统性。

汉字本体的研究必须以形为中心，而且必须在个体考证的基础上探讨其总体规律。传统文字学在研究上以形附属于义、着重个体而忽略总体的积习，便无形之中成为这种本体研究的障碍。传统文字学并非完全没有认识到研究汉字总体的重要性，字书对字的类聚本身就表现出“小学”家认识汉字总体的愿望。而且，任何单字的考证都必须借助大量相关字的参照，考据家不可能没有总体与个体关系的意识。但是，历代字书都不区分字形的历史层面，只是一味以多取胜，堆积材料，提供不出一批经过整理的共时系统字料，创建科学的汉字构形学便更加难以起步。

传统文字学不能完成创建汉字构形学的任务，还因为古代哲学与科学的发展，未能给它提供分析总体字形内部结构的科学方法论，因而面对三千多年不断变化又不断积淀的数万汉字，他们难以由纷繁之中见其规律。没有一套统一的能够描写共时汉字构形系统的基础理论与可操作方法，历时的比较就难以进行，汉字史的研究也就难以取得突破性的进展。

三

许慎的《说文解字》经过人为的整理，基本展现了理想的汉字构形系统，使我们看到字与字之间有序的联系，认识到汉字作为一种信息载体，一种被社会创建又被社会共同使用的符号，不可能是散乱无章的，在构形上必然以系统的形式存在。但是，《说文》贮存并整理的是秦代“书同文”后统一的、经过简化的篆字，这批篆字中的主要部分，首先被收入小学识字课本，字形上经过严格的规范。其他扩展部分收入的字形，也经过编著者许慎从他所看到的诸多不同时代的形体中优选并篆化。应当说，小篆是一批经过自觉人为整理的汉字。而且，《说文》小篆之所以能够进行定量分析，是因为在许慎作《说文解字》的当时，这种文字体制就已经不再通行于社会，成为一种历史的文字，许慎在经过对字符的优选之后又以字书的形式把这种文字的数目固定了。于是，这种文字体制的结构内部已不再发生质与量的交换与转化。也就是说，它已是一个封闭的系统。应当说，这种系统的形成带有一定的人为因素。

经过实际的描写，小篆的构形系统得到了证实：《说文》小篆具有一批基础构形元素，归纳后大约得到四百多个，我们称之为形位。这些形位都有自己组成字形、体现构字意图的具体功能，它们是构成小篆的基础。我们把全部形位的集合看作构字的储备材料，而把已经进入构字、体现了自身功能的形位及形位的组合称作构件。可以看出，在每一级组合中，随着构件中的形位数不断地增加，结构都发生着质的变化。汉字的结构层次是有序的，改变结构次序也就改变了这个汉字。有序的层次是汉字构形呈现系统性的重要原因。同时，小篆的每一层次的组合，都有固定的模式，这些模式的类型由构件的功能来决定。构件在相互组合中彼此制约，使每一个构件的功能得到体现。每种不同功能的组合都可归纳为一种构形模式。小篆的构形模式可概括为以下六类：（1）全功能零合成：由一个单形位构件自成的汉字。这个单形位构件必定是成字构件，它的功能是表音、表义、表形俱全的。它的组合公式是：单形位成字构件+0；（2）形义合成：由一个表形构件和一个表义构件组合而成的汉字。它的组合公式是：表形构件+表义构件；（3）标形合成：由一个标示符号去指示或区别一个表形构件。它的组合公式是：表形构件+标示记号；（4）会议合成：由两个或两个以上表义构件组合而成的汉字。小篆的会议合成包含两种类型。一种是纯会议合成，另一种是残留图形性的会议合成。后者在组合时构件的位置仍保留着物象的位置。它们的组合公式是：表义构件+表义构件；（5）形音合成：由一个表形构件与一个表音构件组合而成的汉字。它的组合公式是：表形构件+表音构件；（6）义音合成：由一个表义构件与一个表音构件组合而成的汉字。它的组合公式是：表义构件+表音构件。这六种模式，可以涵盖小篆97%以上。进一步归纳，我们就看到了传统的“六书”。对上述六种模式所含小篆的字数加以统计可以看出，小篆中义音合成的字占87%以上。其余五种模式仅占12%左右，而且，它们绝大部分都在义音合成字中充当过构件，因

也可以包含在义音合成字的结构中。义音合成字，即传统“小学”所谓的形声字。它以义符为义类标志，以声符为别词手段。前者为纲，后者为纬，构成了标志鲜明的子系统。再以有序的层次来确定每个子系统中个体字符的相邻相关关系。小篆构形系统就清晰地呈现在我们面前。

四

80年代以来古文字和隶书大量出土，这些出土文字——特别是战国文字和简帛文字——是可以断代和定域的、实实在在被社会使用着的共时汉字字符群体。它们的字形能够经过整理显现系统性吗？根据辩证唯物主义的哲学系统论和结构主义语言学的内在精神，不论是经过整理的、体制单一的《说文解字》小篆共时汉字字符集，还是处于实际使用过程、可以加以整理的共时共域的古文字和今文字，都应当有自己的构形元素，这些元素应当是有限的，元素的组合应当有层次，而且应当能够归纳出几类组合模式。汉字的个体字符既然以这些有限的元素为基础而组合，它的组合又都被几类组合模式所覆盖，那么，它们之间当然不会是孤立的，也不会是散乱的，而是互相关联的、内部呈有序性的符号系统。即使是经过书写变异了的行书、草书，只要它们可以系统识别，也应当是彼此关联的。汉字符号不可能杂乱无章，在共时层面上的汉字总体是有系统的，个体字符的考据只有在整个系统中找到它应有的位置，才能被认为是可信的和合理的。仅仅探讨汉字个体字符的形体变化不能称作汉字史，只有在弄清个体字符形体变化的基础上，考查出汉字构形系统的总体演变规律，并且对这种演变的内在的和外在的原因作出符合历史的解释，才能称为汉字史。

但是，对汉字构形系统性的认识不能只停留在理论上，必须首先实现对共时汉字的系统描写，也就是对汉字构形系统性的理论加以验证。这就需要规定出合理的操作方法，规定出明确的术语系，总结出一种可以为各个历史层面上汉字构形系统的描写和历时层面上汉字构形不同系统的比较使用的汉字构形学基础理论，有了这种基础理论和方法，才使汉字学与汉字史都进一步理性化、科学化。

但是，并非一切历史层面上的汉字都经过类似《说文》这样的科学优选和精密整理。前面说过，《说文》以后的字书往往转相抄录，无法保证字料的共时性。要想验证各个历史层面上的汉字是否以系统的形式存在，只有从这个历史时期用汉字书写的一定量的文献材料中去穷尽地撷取字料。但是，以文本形式存在的使用中的汉字，呈现出三种不同的情况：

第一种，民间书写文本中的汉字。这些文本的书写者是一般的民众，文本的内容社会意义较小，流传范围不大。例如个人书信、账目、便笺、日记、契约，以及不拟流传社会只为个人保存的典籍抄本等等。

第二种，社会通行文本中的汉字。这些文本或是官方的正式文告，或是流传于社会的典籍抄本，或是名人书写的诗文。雕版与活字印刷发展起来后，刻印文本所用的

汉字也属此类。

第三种，权威规范汉字。历代官方运用政治权力、通过教育与考试制度规定在某些场合必须使用的汉字，或者经书法家写于碑匾、形成字书以为示范的汉字。

《说文解字》所收的汉字，属于第三种。而第一种和第二种，尤其是第一种，却是遵循约定俗成的规律、随着社会种种因素的变化而自发存在的，这些字符群随时都在进行着内部元素与内部关系的建构与破坏。这些字符群是杂乱无章的，还是也以系统的形式存在呢？

为了解决这个问题，首先要对文本中的汉字加以整理。也就是必须对写在文本上实际使用的汉字进行性质不同的三种归纳：

(1) 字样的归纳。这是将重复出现的汉字字样归纳到一起并进行统计的工作，是将一切文本形式的汉字改变为字库形式第一步要做的工作。

(2) 字组的归纳，也就是对结构、功能相同书写略有变化的异写字的归纳。这些字属于同字异写，或称同构异写的关系。

(3) 字种的归纳，也就是对功能相同结构不同的异构字的归纳。这些字或因构件的选择不同，或因构件的增减不同而异形，因而有着不同的构意，但记词的功能是相同的。它们的关系属于同职异字，或称同词异字。

以上情况可以看出，书写与使用汉字纷乱的文本，对以上三种字认同的标准，首先是记词功能的相同，其次是结构的相同，第三是书写字样的相同。经过认同与归纳，文本汉字的纷乱现象减少，但系统尚未显现。进一步的工作是采用优化的原则，在诸多异写字中挑选出一个通行的字样作为这一组字样的信息代码，其他字样则作为它的变体。而在诸多异构字中，也宜优选出一个通行的字作为主形，其余的根据它产生的原因作为通用字、变体字或转写字对待。字样与字形优选的原则，首先应当是减少和统一汉字的基础构形元素，适应汉字的构形模式，使字形的造字意图尽量明确，从而建立与相邻、同类字形的正常关系，找到自己在整个构形系统中的位置。这就是经过人为整理使汉字构形系统显现的过程。

12年来，学科点上十几位博士用统一的术语系和上述操作程序，先后对甲骨文、西周金文、春秋金文、战国楚文字和一部分东方文字、睡虎地秦文字、居延汉简、马王堆帛书、东汉碑刻隶书、魏晋行书、隋唐五代碑志楷书、宋代刻版楷书、明代石刻佛经文字、清代民间手写与宫廷文字进行了认真地整理，对其构形系统进行了一一地描写。为了保证考察的文字材料时代的确定性、体制的一致性，不但是出土文字，就是后代的隶楷文字，我们也没有使用字书，而是选择历代书写的实用文本，一个字一个字地归纳、统计，制成字表，作出完全来源于第一手材料的结论。应当说，每一个考察项目的工作量都是非常大的。为了考验一些基本的原理和方法是否适用于现代汉字，1997年，我把汉字构形学比较完善的操作方法运用于信息处理用 GB13000.1 字符集汉字部件规范；2000年，我又把汉字构形学的基本原理应用于小学识字教学；并在教育部组织的全国识字教学方法交流会上，介绍给第一线的老师。这些都为了证明，

现代汉字也是具有系统性的，汉字构形的学理，是可以在应用领域发挥作用的。在整理每一批共时汉字的过程中，我们都看到，汉字是一种可以人为调整的信息符号，但它的社会约定性又不能违背。所以，汉字构形系统的形成来源于符号系统在社会使用中产生的自组织现象，它不可能是自发的，但也不能违背社会的约定性和符号结构内部的自身规律而一意孤行。在每一个共时系统的描写中，我们似乎都看到了许慎是怎样把小篆整理成《说文解字》的，他的工作成功之处，就在于既尊重了小篆结构和使用的的事实，又正确把握了汉字构形的内部规律，许慎是求实的，又是科学的。

五

汉字构形学的总体思想提出较早，在它的总体思想和最初的创意提出后，遵循初步确立的理论原理和操作规程，进行了多次的共时汉字整理和系统描写。在多次的实践中，完善了汉字构形学的体系。例如，在归纳小篆构形模式时，我们只看到了6种类型，一旦接触到甲骨文和金文，构形模式就增加到11种，而到分析隶书、楷书时，由于理据不同程度丧失的情况比例增多，又有很多新的问题需要面对，必须设置新的术语。理论和方法原则是在实践中丰富完善的。很多参与这项工作的博士都为理论和原则的充实作出过贡献。例如，早在石定果完成《说文会意字研究》^①时，就提出了“关系位”的问题，注意到有一部分会意字构件的相互位置具有区别意义的作用，这启发我们区分会形合成字与会义合成字，区别形合与义合这两种不同的情况；在李国英完成《小篆形声字研究》^②时，针对《说文》中反映出的“右文”现象，首次提出一部分声符可以兼有“示源功能”；过渡部件（即层次结构的整字和形素之间的部件）也有变体，也需要归纳，是李运富最早在他整理战国楚简帛文字时落实的^③；汉字不是都有理据，要增加“记号部件”这一项，而“记号部件”只有构形作用，没有构意功能，和其他有构意功能的部件应当形成总体的对立——这一点，是王立军首次提出的^④；古文字的表形部件，变为今文字时如果不成字而仍有构意功能，是不同于“记号部件”的，应当另列一类，陈淑梅称作“象征部件”^⑤，我们吸收了 this 意见，单设了“象征部件”，作为“表形部件”的附类；汉字发展到楷书阶段，部件的混同现象开始增多，由小篆的线条向今文字的笔画演变时，为了区别的需要，也时有增笔（王—玉）、变笔（刀—刁）等现象发生，这些楷化过程的字形调整现象，与标形、标义的构形模式，很容易相混，是齐元涛首次发现了这一点，提出了辨别它们的

① 石定果：《说文会意字研究》，北京语言学院出版社1996年5月版北京师范大学出版社1996年6月版。

② 李国英：《小篆形声字研究》，北京师范大学出版社1996年6月版。

③ 李运富：《楚国简帛文字构形系统研究》，岳麓书社1997年版。

④ 王立军：《宋代雕版楷书构形系统研究》，上海教育出版社2003年7月版。

⑤ 陈淑梅：《东汉碑隶构形系统研究》，上海教育出版社2005年4月版。

原则^①；我们在区分异构字与异写字的时候，多次发现它们之间的纠葛，但是，将二者的区别和界限厘清的工作，是易敏在分析明代《华严经》石刻时初步完成的^②……上面是我记忆深刻的几条，应当还不只这些。这套丛书理论体系是一致的，术语系基本是统一的，只有个别地方稍有差异；但是，因为每个分课题取字的底数不完全一样，要把他们综合到一起还要做一些艰苦的工作。

六

我们这套丛书中的每本书，是对一个共时的汉字构形系统分别加以描写和研究。前面说过，描写是解释的前提，比较又是探讨演变规律的必要条件。这套丛书的宗旨，是使汉字学与汉字史都进一步科学化。

汉字经过三千多年的变化，有着漫长的历史。汉字的演变绝不是每个个体字符变化的简单相加，而是经历着由个体字符变化累积成整个系统变化——也就是量变到质变——的过程。在这个过程中中的每一个阶段，汉字究竟发生了哪些变化，为什么发生这些变化，其中包含着哪些规律，只有对同时期的汉字构形系统进行比较才能弄清。各个阶段的构形系统既然是可以用统一的方法描写的，相互的比较也就一定可以操作。试以甲骨文与小篆的构形系统的一部分量化数据作比较，可以看到以下几个重要的事实^③：

第一，从1380个已释甲骨文中，归纳出标准形位412个，但它的形位变体就有3498个，非字形位的比例也大大高于小篆。小篆中归纳出的形位是422个，但他的构字总量却是9431个（大徐本实际字数）。这充分说明，与小篆相比，甲骨文字形不固定，形位的可归纳程度很低。

第二，在甲骨文的构形模式中，义音、形音合成字只占21%强，而零合成字与会形、会义、形义等非声合成模式却占到78%强。在非声模式中，甲骨文的纯会义合成字不足10个，而小篆没有的会形合成与形义合成字却高达734个，占了绝大多数。这就说明，甲骨文的构件参构时的功能以表形为主体。表音、表义的功能尚未得到充分的发挥。

第三，甲骨文的结构过程87.5%都是平面组合，层次组合只有12.5%，比例相当低。这又进一步说明了，甲骨文不但构件的功能以表形为主导，而且结构的方式也是

① 齐元涛：《隋唐五代碑志楷书构形系统研究》上海教育出版社2007年7月版。

② 易敏：《云居寺明刻石经文字构形研究》上海教育出版社2005年版。

③ 我这里引用的数据，甲骨文的部分是郑振峰测查的，见《甲骨文字构形系统研究》（上海教育出版社2006年8月）小篆部分是齐元涛测查的，见齐元涛硕士论文《〈说文〉小篆构形数据的测查》。但这些都是他们最初测定的，现在，我们还在以集体的力量继续添加材料，核实数据。如果将来正式发表的数据与我的引用有出入，当以新的统计为准。这里，我要表达的只是如何通过比较总结汉字发展历史的问题，这些数据即使还有不够精确的地方，但应当不会影响历史比较的思路和结论。

图形式的。

以上三点说明了殷商至两周阶段，汉字的整个系统处在由表形文字向表意文字发展的过渡阶段。形位数量的固定和归纳程度的加强；构件功能从表形为主到表义为主、表音辅之；结构方式从图形式的平面组合到义音式的层次组合——这三点，就是这一阶段汉字演变的主要表现。

再以小篆与隶书的构形系统作一总体的比较，又可以看出以下几点重要的事实：

第一，隶书，以至更后来的楷书，在标准形位的归纳程度、以义音组合为主体、层次组合占主导地位等方面，完全承袭了小篆。

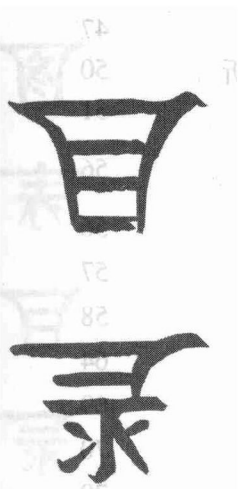
第二，隶书形位变体、构件变体、异写字与异构字的数量大大超过小篆，是因为秦代“大发隶卒，兴役戍，官狱职务繁”，文字的使用范围越来越大，长期缺乏许慎这样的专家进行整理、规范的原因。这不是它与小篆的主要差别。

第三，隶书与小篆的主要差别在于，在隶书构形系统中，原来小篆的单形素构件字大量变形，多形素构件字大量黏合，构意的形成本来是由单形位的末级部件作基础的，隶书却转移到由字符一级拆分得出的直接构件上。例如：“更”在小篆里是音义合成字，从“丙”、从“攴(pu)”，隶书黏合后，“便”、“梗”等字的构意就不能从原来的基础构件去找，而要在直接构件“更”上去找了。“卑”在小篆里是会议合成字，从“甲”、从“又”，隶书黏合后，“碑”、“埤”、“俾”、“婢”等字的构字意图，无法再找到末级的单形位构件上，也只能由直接构件“卑”上去找了。

这说明，隶书的义音化程度比之小篆更加大幅度增强，形体与物象的联系几乎不存在了，在任何一个层次上，构件的表形功能完全被表义、表音功能所替代。这一方面说明汉字总体的性质没有变化，仍是表意文字；另一方面也可看出，构件的义音化给汉字的简化提供了充分的条件。

以上比较的举例告诉我们，只有对每个历史层面上的汉字构形系统作了准确的描写后，经过比较，汉字的发展历史才可能真正弄清楚。这就是我们希望通过这套丛书达到的总结历史的目的。当然，汉字的发展历史不仅仅是靠统计数据 and 描写系统就能完全弄清的，汉字是一种被全社会应用的符号系统，又是与人的自身创造相关的人文性很强的符号系统，要真正了解它的历史，解释诸多现象，还要结合对社会发展和人文环境的考察，这一点，我们在工作的后期已经逐步注意到了，从王立军的《宋代雕版楷书构形系统研究》开始，我们的工作虽然仍以共时汉字整理和构形系统描写为起点，但在论著中，已经不以描写过程为主要篇幅，而用较多的篇幅来对描写的数据进行分析，和对某些现象进行解释了。

2007年12月2日修改



汉字构形系统的共时描写与历史比较 (代本书序)	王宁
绪论	1
第一章 统计学原理与材料的预处理	14
第一节 统计学原理	14
一、统计学基本思想与方法	14
二、统计学基本概念简介	16
三、统计学与汉字构形学	17
第二节 原始数据的来源及数据预处理	19
一、数据选取	19
二、断代汉字构形属性测查概述	22
三、统计数据的预处理	27
第二章 构件	30
第一节 构件集合数学模型	30
第二节 直接构件分析	31
一、构件集的提取	32
二、随机抽样数据处理	33
三、直接构件分布模型的建立	37
四、构件集增长速率及构件集的收敛性分析	40
第三节 基础构件分析	43
一、构件的提取	43
二、抽样数据处理	44

三、建立基础构件分布模型	47
四、构件集增长速率及构件集的收敛性分析	50
第四节 基础构件数分析	51
第三章 结构	56
第一节 构形层级	56
一、断代系统构形层级数统计特性分析	57
二、等容量样本层级数测查	58
三、传承字层级数变化情况测查	64
四、构形层级对系统性能的影响	69
第二节 构形模式	70
一、构形模式定义	70
二、构形模式分析	72
三、构形理据	75
四、造字方法	80
第三节 形式结构	83
一、笔形样式	84
二、构件要素	87
三、组合方式	89
第四章 关联	91
第一节 汉字形体关联度的文字学界定	91
第二节 汉字形体关联数学模型	92
第三节 汉字形体关联度的计算及分析	94
第四节 汉字形体的义关联和音关联	97
第五节 汉字形体离散度	101
一、汉字形体离散度的文字学界定	101
二、汉字形体离散度的数学模型及计算	101
三、汉字形体发展所遵循的基本原则	102
结 语	103
参考文献	106
注 释	112

图 表 目 录

汉字构形属性历时演变的量化研究

图 1	汉字系统运作图	7
图 2	统计学研究步骤	14
图 3	汉字构形树	30
图 4	构件均值散点图	34
图 5	金文拟合曲线	38
图 6	直接构件拟合图	41
图 7	构件增长一阶导函数	42
图 8	随机抽样基础构件均值图	45
图 9	随机抽样 500 字基础构件构字数分布图	46
图 10	基础构件拟合图	48
图 11	基础构件增长一阶导函数	51
图 12	基础数分布图	52
图 13	基础数累积分布图	54
图 14	1000 字层级累积分布图	62
图 15	构形层级数平均值	63
图 16	甲骨文构件要素分布图	88
图 17	直接构件系统关联度	95
表 1	直接构件统计表	32
表 2	直接构件构字数统计表	32
表 3	等容量直接构件数统计表	34
表 4	随机抽样 500 字的直接构件构字数统计	35

表 5	平均直接构件数 95% 置信区间	36
表 6	春秋金文样本观测值与函数拟合值对照表	38
表 7	基础构件统计表	44
表 8	等容量基础构件数统计表	44
表 9	随机抽样 500 字基础构件构字数分布	46
表 10	基础构件增长率	50
表 11	基础数等于 i 的频率表	52
表 12	基础数大于等于 i 的频率表	54
表 13	构形层级统计表	57
表 14	构形层级频率表	57
表 15	等容量层级数统计特征表	59
表 16	前 1000 字累积频率 S_i	61
表 17	春秋—宋楷层级比例	66
表 18	说文—汉隶层级比例	67
表 19	汉隶—宋楷层级比例	68
表 20	构件功能分布	74
表 21	随机抽样 500 字平均关联度	94
表 22	关联数分布表	96
表 23	关联数分段表	97
表 24	随机抽样 500 字表义构件构字数统计表	99
表 25	随机抽样 500 字表音构件构字数统计表	100