

秦系简帛文字国际编码的初步设想

沙宗元

(安徽大学中文系 安徽 合肥 230039)

摘要 古汉字的计算机国际编码对于古籍整理研究、文字学研究、中华传统文化在信息时代的传播和弘扬等都具有十分重要的意义。目前,关于甲骨文、小篆、金文等类型古汉字的整理和国际编码工作已经陆续展开,并且已经取得了相当的进展。作为秦文字的重要类型之一,秦系简帛文字的整理及编码目前还没有开展,相关的研究工作还比较薄弱。本文介绍了秦系简帛文字国际编码的背景和意义、秦系简帛文字的研究资料,并着重对编码工作的实施提出了一些初步的设想,认为字形的收集整理、字形属性的系统描述、异体的认同和别异、字形的筛选等这几个方面是编码工作的重点。

关键词 古汉字; 秦系简帛文字; 国际编码

近年来,随着汉字信息处理技术的不断发展,古汉字的计算机国际编码问题不断被提上议事日程。由于古汉字自身形体结构方面的特点,其编码与现代汉字有着较大差别,与古汉字编码相关的许多问题还需要进行深入研究和探讨。为保证古汉字编码的质量,我们必须充分考虑古汉字本身的特点、古汉字字库的应用特点以及国际标准化组织对字符编码的要求,对各类古汉字进行全面的字样整理。^①本文将对秦系简帛文字国际编码的背景和意义、秦系简帛文字的资料、编码的初步设想等相关问题进行初步的探讨,以求正于各位方家学者。

一、编码的背景和意义

目前,信息交换用现代汉字编码字符集的国家标准已经陆续颁布。随着一系列字符集(包括基本集、辅助集和扩充集)和相关国家标准的颁布,可以说汉字信息处理已经取得了令人瞩目的成绩。与此同时,古汉字的信息处理问题也引起了人们越来越多的关注,相关的研究已经取得了不少成果。各类古文字字库的建设有的已经取得了重要进展,有的正在纳入计划之中,积极做好启动的准备工作。例如,台湾中央研究院史语所与资讯所合作的古文字信息系统,北京师范大学民俗典籍文字研究中心的小篆教学与研究系统,华东师范大学中国文字应用与研究

^① 张再兴:《古文字标准编码中的金文原形字字样整理》,《中国文字研究》第5辑,广西教育出版社2004年版。

中心的金文、小篆系统，香港理工大学的古文字检索系统等等，都为这一工作做出了贡献。建立“信息交换用古汉字编码字符集”，对于古籍整理及文字学研究等的现代化建设有重大意义，是实现该领域现代化的必要条件。^①这一点已经成为许多学者的共识。

2002 年以来国家教育部、国家语委多次组织召开语言文字和信息处理专家研讨会，研究古汉字的整理与国际统一编码问题。在 2003 年 11 月召开的国际标准化组织表意文字工作组（IRG）的第 21 次会议上，中国代表团提出“关于中国古汉字进入国际标准字符集的提案”，建议把中国古汉字放入国际标准字符集进行编码。会议接受了中国代表团的提案，并成立了古汉字编码兴趣小组。在 2004 年 11 月召开的国际标准化组织表意文字工作组（IRG）的第 23 次会议上，IRG 接受了古汉字兴趣小组的报告（N1022），同意成立古汉字编码专家组，并通过了古汉字编码专家组的工作计划。在历次两岸四地中文数字化合作论坛（CDF）的议题中，古汉字的编码问题一直都是与会学者较为关注的论题。

古汉字的编码是一项规模庞大的系统性工程。目前，关于甲骨文、小篆、金文等类型的古汉字的整理和国际编码工作已经陆续展开，并且已经取得了相当的进展。但是，作为秦文字的重要类型之一，秦系简帛文字的整理及编码目前还没有开展，目前的相关研究工作比较薄弱，仅仅停留在文字学意义上的资料整理和少部分疑难字形的考释阶段。因此，对秦系简帛文字进行系统的整理研究，并在此基础上，按照 IRG 的标准和规范对它进行汉字的国际编码，是我们当前古汉字编码工作的一项重要环节。

二、文字资料及价值

20 世纪 70 年代以来，秦系简帛文字资料陆续出土，引起了历史学界和文字学界的普遍关注。这些珍贵的出土文献成为研究秦代历史、文化和文字使用状况的第一手资料。目前发现的较为重要的秦系简帛文字资料主要有以下几种：^②

（一）云梦睡虎地秦简

1975 年 12 月，在湖北省云梦县睡虎地挖掘的 12 座战国至秦代的墓葬中的第 11 号墓出土，计有 1155 枚竹简及残片 88 枚；另外在第 4 号墓则发现木牍二件，其内容为家书。最引人注目的是第 11 号墓出土的 1155 枚竹简。内容有 10 种，即：《编年记》、《语书》、《秦律十八种》、《效律》、《秦律杂抄》、《法律问答》、

^① 周晓文、李国英：《建立“信息交换用古汉字编码字符集”的必要性和可行性》，《北京师范大学学报》2006年1期。

^② 陈文豪，二十世纪的秦汉简帛概述，<http://www.docin.com/p-193203.html>。

《封诊式》、《为吏之道》、《日书》甲种、《日书》乙种。睡虎地秦简的释文版本目前超过 10 种，以文物出版社 1990 年出版的《睡虎地秦墓竹简》一书较佳。

（二）青川木牍

1979 年 2 月至 1980 年 7 月间，四川省博物馆与青川县文化馆的考古工作人员，对青川县白井坝郝家坪的 72 座古墓进行发掘和整理，在第 50 号墓中发现木牍二件，残损较少，字迹清晰，计有 121 字。经由简文中有“二年十一月己八朔朔日”句考证，系秦武王二年。再加上简文提及“丞相茂”，推知系指甘茂，故此牍为秦牍。此木牍所记载的内容，主要涉及秦的《秦律》，将之与睡虎地秦简有关的秦律结合起来研究，是了解秦土地制度、阡陌和封埧形制的主要材料。

（三）天水放马滩秦简

1986 年 4 月在甘肃省天水市北道区党川乡放马滩秦代墓葬中出土，共有 460 枚。内容有甲种《日书》73 枚、乙种《日书》280 枚、《墓主记》7 枚。《日书》系继睡虎地秦简后的又一次重要新发现，二者可相互比较。《墓主记》是描述人死而复生的故事，对研究秦人的生死观及宗教观有很大的帮助。因此这批竹简是研究秦文化的珍贵材料。每简最多容 43 字，一般在 25—40 字之间。

（四）云梦龙岗秦简

1989 年 10 月至 12 月在湖北省云梦县龙岗出土，计有木牍一块及竹简 150 余枚。内容为秦法律文书，可分成五类：即有关禁苑的法律、有关驰道管理的律文、关于马牛羊管理的律文、有关田赢方面的律文、其他无法归类的律文^①。

云梦龙岗秦简的释文，先后有《云梦龙岗六号墓及出土简牍》^②所载及《云梦龙岗秦简》^③二者。

（五）江陵王家台秦简

1993 年 3 月，于湖北江陵县荆河镇郢北村王家台第十五号墓出土，计有竹简 800 余枚。主要的内容为《效律》、《日书》及《易占》。另有木牍一枚，因残损过甚，字迹模糊，内容不详。全部释文尚未见公布。

（六）周家台秦简

湖北省荆州市周梁玉桥遗址博物馆（原湖北省沙市市博物馆）于 1993 年 6 月发掘清理周家台 30 号秦墓，出土竹简 387 枚，可分成 3 组：甲组 242 枚，内容为二十八宿占、五时段、戎磨日占、五行占和秦始皇三十六年（前 211 年）、三十七年（前 210 年）月朔干支及月大小等；乙组 75 枚，为秦始皇三十四年全年历谱；丙组 70 枚，有医药病方、择吉避凶占卜及农事等内容。另有木牍一件，

^① 刘信芳、梁柱：《云梦龙岗秦简综述》，《江汉考古》1990年第3期。

^② 载《考古学辑刊》，第8辑，87—121页。

^③ 刘信芳、梁柱：《云梦龙岗秦简》，科学出版社1997年版。

正面为秦二世元年十二个月的月朔日干支及月大小，背面书该年十二月份的日干支等。这批竹简，目前只公布了秦始皇三十四年历谱。

（七）里耶秦简

2002年6月湖南龙山里耶秦代古城一号井出土了大量的竹、木简牍，总数近3万枚。这批埋藏了2200多年的秦代简牍，纪年从秦王政二十五年至秦二世元年，记事详细到月、日，十几年连续不断。而在此之前，中国古代正史中关于秦朝的记录不足千字。专家认为，里耶秦简是极为重要的百科全书般的日志式实录，它是继兵马俑以后秦代考古的又一惊世发现，蕴涵着巨大的学术价值。

秦简的发现为汉字发展史研究提供了重要参考资料。自古相传，秦代开始使用隶书，但是秦隶的面貌究竟如何，过去一直不得而知。现在发现的秦简所用的字体同小篆有相当明显的区别。例如，秦简文字字体方面一个非常值得注意的现象，就是左边从“水”的字几乎都把“水”旁写作“𠂇”旁，而不是写成像正规篆文的那样^①。通过秦简，不但可以认识秦隶的真面目，还可以知道隶书实际上在战国时代，至少在战国晚期的秦国，就已经初步形成。“现在我们在秦简上亲见的，正是当时的隶书。”^②所谓程邈作隶书，应该理解为程邈对当时民间流行的这种新字体，作了一些整理总结的工作。睡虎地4号墓家书木牍上的隶书，写得很草率，从中还可以清楚地看到草书的萌芽。

三、主要环节的初步设想

秦系简帛文字的国际编码是一项较为复杂的系统工程，必须认真做好字形的收集整理、字形属性的系统描述、异体的认同和别异、字形的筛选等这几个方面的工作。

（一）字形的系统整理

目前所见秦系简帛文字的字书和其他资料比较散乱，存在的问题较多，不能直接用作字形编码。因此，研究工作的首要任务就是对现有出土的秦系简帛文字进行全面收集整理，穷尽性地采集字形，形成秦系简帛文字的字形汇集群。要做好这一部分的工作，必须做到字形真实准确、搜罗齐备。

为了达到真实准确的目的，所选取的字形必须采自原始简帛照片，一般不使用经过摹写的二手资料。如果字形实在漫漶不清并影响识读，则一般不采用该字形。为了达到搜罗齐备、不出现字形遗漏的目的，我们对字形的搜集整理应尽量

^① 裘锡圭：《文字学概要》，商务印书馆1988年版，68页。

^② 李学勤：《睡虎地秦简文字编·序言》，文物出版社1994年版。

从简帛文字的原始文献中进行，如果仅仅从现有的字表或字书中采集字形的话，就有可能由于字书字表收字未尽而产生字形的遗漏。

（二）字形属性的系统描述

在前面字形搜集整理的基础上，我们必须认真做好文字编码的另一项重要的基础性工作——字形属性的系统描述。字形属性的描述主要包括字形的摹写形式、出处、时代、地域、材质、《说文》部首、隶定形式、对应楷体字形等方面的分析整理。以下试就摹写形式、《说文》部首、隶定形式、对应楷体这几种主要属性分别加以简要说明。

1. 摹写形式的确定

字形的摹写一般有三种情况：一是字形结构清晰完整，摹写形式没有异议的，当然采用通行的摹写形式。二是字形清晰完整，但有不同摹写形式的。在这种情形下，应认真分析比较，采用多数学者赞同的摹写形式。三是字形结构不清晰，无法进行准确摹写的，则不收入该字形。

2. 《说文》部首的确定

在 IRG20 次会议上，古汉字编码小组提出了依据《说文》部首来编排古汉字的提案：“古文字学界一般都按照《说文解字》540 部首的顺序排列古汉字，所以我们建议，给古汉字编码也采用这种方法，所有类型的古汉字编码都按《说文解字》的顺序排序。和《说文解字》中的字没有对应关系的古汉字可以按照《说文解字》的归部原则归入相关的部首。”依照这一排序方法，必须给每个字形确定其部首属性。

3. 隶定字形的确定

字形的隶定与摹写、考释密切相关，对于那些目前隶定和释读仍有疑问的少数疑难字，宜审慎对待，择善而从。例如，《睡虎地秦简·秦律十八种》^①第 27 简有这样一个字形“𠄎”，原书隶定作“𠄎”形，读为“𧈧”，并引《汉书·货殖传》释为小虫之义。陈振裕、刘信芳的《睡虎地秦简文字编》^②从整理者的隶定，但认为该字义未详。王辉在《秦出土文献编年》^③一书中对该字的隶定也表示怀疑。最近有学者认为，该字上面从“尸”应无问题，但是下面不应隶定为“彖”，因为“彖”字在秦简中不作此形，而应该隶定为“𧈧”。所以，第 27 简该字的隶定应为“𧈧”，分析为从尸𧈧声，为“𧈧”的异体字，用为“𧈧”。^④

4. 对应楷体字形的确定

^① 睡虎地秦墓竹简整理小组：《睡虎地秦墓竹简》，文物出版社1990年版。

^② 陈振裕、刘信芳：《睡虎地秦简文字编》，湖北人民出版社1993年版。

^③ 王辉：《秦出土文献编年》，新文丰出版公司2000年版，141页，

^④ 方勇：《读秦简札记两则》，复旦大学出土文献与古文字研究中心网站，http://www.guwenzi.com/SrcShow.asp?Src_ID=400

秦系简帛文字字形与对应的楷书字形关系的确定,大体上有这样三种情况应当引起注意:一是字形结构清晰,尚可进行清晰隶定,但是对应的楷书字形有争议,这种情况下进行字形对应时应当合理判断,审慎抉择。如青川木牍上有字作“𠄎”形,字形可以分析为从二“阜”,从“水”“办”声,到底该字应该对应楷书的哪个字呢?以前有人释作“衍”、“衍”、“涧”等。最近有学者根据文献对勘,认为此字释为“泐”、读为“梁”是比较合适的^①。二是字形结构较为清晰,可以进行清楚地隶定,但是现在却无法找到对应的楷书字形。如《睡虎地秦简·日书乙》30简有一个“从門从牛”的字形,可以清楚隶定,但历代字书未见,无法断定为何字。三是有些字形体怪异或者难以辨识,无法加以隶定,当然也无法找出对应的楷书字形。

(三) 异体的认同和别异

出土文献上的文字均为个人手书,因此,一字多形、一字多写、异体繁多等现象极为常见,也是古汉字形体结构方面的一个重要特点。对汉字进行编码时我们不可能把这些众多的异写、异构字统统不加选择地收罗进来,而是应当认真地进行形体的分析和比较,妥善、合理地解决古汉字字形之间的认同和别异问题,所谓认同,就是将所有属于一个字种的不同异体、异写归为一组,所谓别异,就是将虽形体近似,但却不属于一个字种的不同字形区分开来。应该说,这方面同与异的认定区分是古汉字编码过程中必须要认真对待、特别引起注意的问题。那么,什么样的字形之间可看作为一字异体,什么样的字形之间不可以看作为一字异体,而是应该区别为不同的字呢?我们认为,这不仅要看简帛文字的特定的结构特点、构形习惯,也要根据文献中的用法和辞例来确定。比如《睡虎地秦简·日书》中的“夙”字,多数情况下作上下结构,有时也写成左右结构,虽然形体比较罕见,但依据辞例,认同为“夙”字异体当没有问题。

(四) 字形的筛选

对于古汉字来说,同一个字种不同写法的异写字、异构字之间的差异,有的属于形体结构方面的不同(如构字部件的同义换用、数量增减、位置变易),有的属于书写形态和书写习惯上的不同(如线条的长短、粗细、位置的变化)。从汉字编码的角度说,不容许也没有必要将形体区别不甚明显的异写字都纳入字符集中。因此,在编码汉字进入字库之前,就有必要对这些异写字进行归纳,剔除那些形体区分价值不大的字形,精心筛选出结构上具有代表性、字形区分度大的字头单位和形体结构单位,同时也要尽量兼顾到字的意义和用法以及特定字形所具有的时代和地域特征。

^① 周波:《释青川木牍“泐”字及其相关诸字》,复旦大学出土文献与古文字研究中心网站,http://www.gwz.fudan.edu.cn/SrcShow.asp?Src_ID=393

四、余论

古汉字有着特定的形体结构特点，其字形众多，异写、异构纷繁复杂，且与现行的楷书汉字不能完全一一对应，这些都决定了对其进行编码具有相当的工作难度。但是“古汉字由于具有十分典型的表意文字特点，以及所含有的文化内涵，被不断引进文化教育领域，迅速走向普及，成为世界各国了解中国文化的一个重要窗口，因此古汉字进入计算机势在必行。”^①

以上谈了一些关于秦系简帛文字国际编码的初步设想，实际上要真正做好这项工作，还有很多环节需要认真考虑、妥善处理。例如，某些简帛文字标准字形的选定、个别疑难字形的隶定和考释等。

秦系简帛文字的国际编码将以科学文字学理论为指导，结合汉字信息处理的技术和手段，做到专业性与实用性相结合。研究中，我们力求在吸收以往研究与应用成果的基础上，通过对秦系简帛文字材料的通盘整理及属性的全面测查及描述，整理出较全面、科学的秦系简帛文字国际编码字符表。在当今的信息化时代，让古老的秦系简帛文字能够应用于计算机信息处理和信息交换，在汉字的国际编码中取得应有的一席之地。

参考文献

- 陈振裕，刘信芳 1993 《睡虎地秦简文字编》，湖北人民出版社
- 刘志基 2007 《古文字在国际标准字符集中的编码与古文字通用数字化平台建设》，《杭州师范学院学报》第6期
- 裘锡圭 1988 《文字学概要》，商务印书馆
- 睡虎地秦墓竹简整理小组 1990 《睡虎地秦墓竹简》，文物出版社
- 王辉 2000 《秦出土文献编年》，新文丰出版公司
- 张守中 1994 《睡虎地秦简文字编》，文物出版社
- 张再兴 2004 《古文字标准编码中的金文原形字字样整理》，载《中国文字研究》第5辑，广西教育出版社
- 中华书局编辑部编 1981 《云梦秦简研究》，中华书局
- 周晓文，李国英 2006 《建立“信息交换用古汉字编码字符集”的必要性和可行性》，《北京师范大学学报》第1期

^① 王宁：《汉字研究与信息科学技术的结合》，汉字与全球化国际学术研讨会，台北，2005年1月。