



国家社科基金重大项目 工作简报

课题名称：出土古文献语料库建设研究

批准号：10&ZD118

首席专家：刘志基

[2014]年第1期

(总第5期)

华东师范大学

2014年6月14日

领导批示：

本期目录

1. 编者按····· (2)
2. “偏旁类纂”系列首个成果《包山楚简偏旁类纂》即将面世····· (2)
3. “出土古文献语料库”之“电子诂林”和“电子类书”检索路径测试版上线——简说《中国出土简帛文献引得综录·郭店楚简卷》网络扩展阅读窗口····· (4)
4. 古文字引得类工具书编纂模式又有新突破——即将面世的《中国出土简帛文献引得综录·包山楚简卷》简介····· (5)
5. 项目阶段性成果《中国汉字文物大系》出版后引起社会关注····· (6)



1、编者按

通过中期评估，本课题获得滚动经费资助50万元。由于获得滚动资助，有必要在原项目论证的基础上提高对研究成果的要求，并相应延长课题研究时间。目前，课题组已提出项目结题延迟一年的申请，以保证项目研究以更高标准完成。具体将在如下几个方面来提升项目研究的要求：

一是扩大材料收集的时间范围，将所收各类出土古文献材料的公布时间从2013年延迟至2014年。

二是拓展对各类出土古文献材料的数字化整理加工的层次，增加穷尽分析整理各类古文字偏旁，这将保证项目的最终成果新增一种各种古文字材料偏旁检索的功能，为古文字研究的“偏旁分析法”提供直接的数字化资料检索支持。

三是扩容阶段性成果。本项目的最终成果虽然设定为网络语料库，但是基于语料库对出土古文献资料数字化形态的穷尽贮存，对最新考释研究成果的充分吸收，以及严苛数字化标准的一字一码的精确文字整理，使得本项目在研究建设过程中所形成的数字平台客观上成为古文字资料集成性成果的高效孵化器。依据这样一种优势资源，课题组在完成最终成果的过程中努力编纂出版具补缺、创新意义的古文字工具书作为阶段性成果。这样做的意义表现在两个方面：一方面是有助于提升本课题作为国家社科重大项目对学界的贡献程度和社会影响；另一方面，这又是在以正式出版物的要求来对语料库中的资料释读、标注等研究工作进行检验修正，有助于完善本课题的最终成果研究。项目原论证规定的阶段性成果只是“《战国简帛文献引得》”和“《出土古文献频率词典》”两种，但实际上目前已被入出版计划，而且通过相关出版社申报，得到国家重点出版计划支持的本项目阶段成果就有五种：1、《楚简帛文献引得》：2010年全国古籍整理出版资助项目；2、《中国汉字文物大系》（15卷）：2010国家出版基金资助项目；3、《商周金文综类》（14卷）：2010全国古籍整理出版资助项目；4、《中国出土简帛文献引得综录》：“十二五”国家重点图书出版规划增补项目；5、《古文字考释提要总览（五卷）》：上海市“十一五”重点出版项目、“上海市文化发展基金项目”。除了上述几种阶段性成果外，课题组还计划编纂出版原创性的古文字偏旁类纂类工具书，目前已有《包山楚简偏旁类纂》即将问世。

2、“偏旁类纂”系列首个成果《包山楚简偏旁类纂》即将面世

“偏旁分析法”，可谓古文字研究最重要的方法之一。而这一方法的有效实施，有一个不言自明的前提：“偏旁分析法”的运用者必须能够系统把握相关古文字数据中的“偏旁”。理论上说，对应这种研究方法的要求，早该有一种穷尽式检索的古文字偏旁的工具书以服务于古文字研究了。但现实令人遗憾，迄今为止，我们尚未见到这样一种工具书面世。有两类既有工具书在一定程度上可以帮助人们查找古文字偏旁：一类是涉及古文字材料的部首字典，通常以见于《说文》的部首为材料汇集对象；另一类就是“引得”类古文字工具书。然而，相对于“偏旁分析法”实际应用的要求来说，它们的局限都是显而易见的：前者虽然可以帮助人们找到一些古文字偏旁部首，但却只是举例性的材料搜集，并不能提供穷尽而系统的信息；后者虽然可以给出一定范围内穷尽的文字存在，但对象却不是偏旁。显然，缺少这种古文字偏旁专门工具书，将极大影响“偏旁分析法”在古文字研究中的科学运用。而缺乏一种能够穷尽数据、方便查找的偏旁检索资源系统的支撑，“偏旁分析法”研究成绩的科学性将受到极大局限，因为在这种环境下，研究者不得不囿于各自的资料掌握，只能举例性的罗列一些偏旁来进行“偏旁分析”。而更加负面的影响是，在研究依据的材料和得出的观点



并不容易得到真实全面的数据系统的检验的条件下,人们很容易从证明自己的某种观点的需要出发来进行偏旁“举例”,得出未必符合事实的结论。综上,我们迫切需要一种能够穷尽提供古文字偏旁信息的新型工具书。

由于一个多世纪以来古文字材料不断出土,再加原有的传世积累,目前古文字材料已经堪称海量。加之古文字材料类型较多,且各具特点,古文字偏旁工具书的编纂似当分阶段实现:首先当依不同古文字类型分别进行偏旁类纂。其次,对各种材料的古文字偏旁类纂进行整合,形成覆盖各种古文字材料的古文字偏旁工具书。

有鉴于上述种种,我们决定在《中国出土简帛文献引得综录·包山楚简卷》的编纂中增加一个“偏旁类纂”的内容,以期为古文字偏旁工具书的编纂做出尝试和准备。

由于没有成例可援,“偏旁类纂”的各个编纂环节都只能是“摸着石头过河”,在这里进行一点说明。

首先需要说明的是我们对“偏旁”的界定。“偏旁”貌似是个人尽皆知的普通概念,但从古文字偏旁工具书的编纂,更确切地说是从支持“偏旁分析法”的科学应用的角度来看,事情远不是那么简单,因为这种界定并非只需要在理论上成立,更需要支持研究实践的可行和有效。在“偏旁分析法”的视野中,偏旁是一个较为宽泛的研究对象:既可以是严格意义上的“偏旁”,即对合体字进行第一次切分后得到的两个部分,如会意字的会意部件,形声字的声符、义符;也可以是充当严格意义上“偏旁”的特定字符在非严格“偏旁”限定场合的存在,即这种字符作为独体象形、指事字,以及严格意义的偏旁之偏旁,诸如形声字声符之声符、义符之义符,会意字的会意部件之部件的存在形式。这是因为,要深刻认识一个特定偏旁在文字系统中的地位和影响力,只有将其在文字系统中各个层面的存在都系统把握、加以关联才是可能的。故本编纂所辑之“偏旁”即此种宽容认定的“偏旁”。

其次需要说明的是关于“偏旁类纂”之“类”。合理的偏旁分类是科学的偏旁类纂的前提。理论上说,凡是有可能对偏旁分析法实际运用的成败产生影响的偏旁属性都应作为偏旁分类的依据。具体而言,偏旁类纂之“类”,主要指同一偏旁中的不同形体类型。同一个偏旁的形体存在差异,是古文字构形系统所存在的内部差异在偏旁层面的反映,其单字层面的对应则是为人们所熟知的异体字现象。构形系统存在较大程度的内部差异,是古文字的特点,差异可因时间和空间的不同而引发,但即使在同一时代,相同国别地域的文字内部,“文字异形”亦不同程度的存在。对于文字的交际职能而言,同一文字单位却存内部差异将造成释读沟通的障碍,当然也是偏旁分析方法的运用中尤须加以关注的现象。因此,在古文字偏旁工具书的编纂中,有必要全面准确地揭示偏旁的这种内部差异。具体来说,在体例设计的层面,类纂索引的对象不能只是一般意义的抽象掉形体特点的偏旁,而应包含同一偏旁中的不同形体类型。而分类的原则,大致可概括为“不同写法”,具体包括:合体偏旁的构件增减、改换及方位差异;独体偏旁的造字意图变化、构字线条的增减或明显姿态变化等。而出自同一种写法,只是书法风格差异,则不视为分类依据。

再次需要交代的是关于“偏旁”出处的处理。偏旁既然是字的底层构形成分,则偏旁所出字无疑是首先需要给出的。由于对偏旁的检索需要细化到偏旁的构形类别,这种字出处需要的是特定文献语境中的一个特定字形,即不能仅仅给出一个表达抽象的字单位的隶定字,而必须给出偏旁所出特定字的原形。值得注意的是,现有某些给出原形的古文字引得类工具书,如《殷墟卜辞综类》等所给出的古文字原形是不符合偏旁出处要求的,因为它们不是对原辞原形的忠实摹写,不是古文字文献原初写(刻)手的笔迹,它们反映的是今日工具书的抄写者的笔迹。

字形虽然是偏旁最直接的出处,但偏旁类纂中偏旁出处仅仅给出字形还是有局限的。这是因为,对于一个偏旁的特定形态的形成,仅有所出字形是难以全面呈现语境因由的。偏旁



往往可以受上下文影响而类化，与类化相反，还有一种影响偏旁形态的重要语境因素是“避复”。显然，在这种“类化”或“避复”的情况下，如果仅以偏旁所出字形为出处，就会掩盖导致偏旁形体差异的语境因素。更合理的做法无疑是不但给出相关字形，而且给出这个字形所出文句的所有原态字形。当然，逐字给出文例的各个真实原形的要求，会给偏旁类纂的文字处理带来很大的额外负担，而化解这种负担的一个很好办法，就是将偏旁类纂与引得相关联。从这个意义上说，在实现了引得原形逐字保真呈现的前提下，不关联一个“偏旁类纂”乃是一种资源浪费。而本书的将“引得”关联为偏旁出处的做法，当是“偏旁类纂”的一种合理编纂方式。

3、“出土古文献语料库”之“电子诂林”和“电子类书”检索路径测试版上线——简说《中国出土简帛文献引得综录·郭店楚简卷》网络扩展阅读窗口

本项目最终成果论证有这样的表述：“最终成果，将兼具出土古文献的语料库、电子类书；出土古文字字料库、电子诂林等多种功能。”其中“电子类书”、“电子诂林”，都是将传统小学类工具书形式以数字化载体加以呈现的一种试验，需要既保留传统工具书形式的资料查检特点与长处，又以数字化方式弥补纸质载体检索的不足，以实现其文字传输方式变革后的相应改变。显然，这是需要尝试和摸索的。而最近上线的本项目阶段成果《中国出土简帛文献引得综录·郭店楚简卷》网络扩展阅读窗口中所包含的《郭店楚简电子诂林》、《郭店楚简电子类书》，正是在一定的材料范围内对最终成果的上述设计进行了一种试验，期望通过吸取一定时间的读者使用意见反馈，乃至专家评审意见，对于这一最终成果形式的设计进行完善。

《郭店楚简电子诂林》，检索路径下储存关涉《郭店楚简》字词释读意见 4111 条，各释读分别与相关文献用字关联。具体查询方式分为两种：1，在“篇号-句号”检索栏中输入纸版图书的“唯一句关联号”（即“释文”、“引得”表格中的第一列数据）检索相关字的诸家考释意见；2，在“原出处”检索栏中输入整理者所定的篇名、简号以及简中的字序号检索相关字的诸家考释意见。



《郭店楚简电子类书》，基于《郭店楚简》文献用字的意义分类，提供分义类的郭店简文字检索，基本功能类同于传统类书。其意义分类系统为 11 个大类，34 个二级子类，65 个三级子类，在大、中、小三类的检索栏中输入如下分类表对应列的类别，即可获得郭店简该类



所有字词及其意义描述。

[返回前页]

《中国出土简帛文献引得综录·郭店楚简卷》信息检索

《郭店楚简电子类书》

请选择查询义类：

大类：

中类：

小类：

华东师范大学中国文字研究与应用中心
2013年12月10日

《郭店楚简电子诂林》入口

使用说明：（1）本检索系统与图书《中国出土简帛文献引得综录》之《郭店楚简卷》关联使用。本检索路径下储存《郭店楚简》逐字的语境意义描述及分类标注，实际是一部郭店楚简频率分类字典。读者在大、中、小三类的下拉菜单中选择义类信息（须选至小类），即可获得郭店楚简该类所有字词及其意义描述。

（2）首次使用本检索系统，请下载并安装字体文件：[楚文字](#)

（3）其他检索功能，将在本检索系统的升级版本中陆续实现。

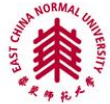
4、古文字引得类工具书编纂模式又有新突破——即将面世的《中国出土简帛文献引得综录·包山楚简卷》简介

本项目阶段性成果《中国出土简帛文献引得综录·包山楚简卷》近期内将由上海人民出版社出版。作为《中国出土简帛文献引得综录》的第二卷，本书在沿袭首卷《郭店楚简卷》的创新编纂格局和基本体例的基础上，又相对《郭店楚简卷》有了若干增益和完善。

“多字语义单位”索引，是本书增益的一个内容。所谓“多字语义单位”，是指包山简文中用字具有固定或相对固定组合性，而意义上有专门所指的多字组合。这种组合一般可以认定为多音节词，但由于古文字文献某些多字组合是否属于严格意义上的“词”不同学者往往有见仁见智的判断，我们姑且将这种多字组合称之为“多字语义单位”。“多字语义单位”索引，拓展了包山简文的检索层次，是对“引得”检索功能的重要补充，这一索引层次的拓展，对古文字类工具书编纂具有借鉴价值。

既有的实现出土文献文句穷尽性检索的工具书，按索引的直接对象，分为仅限于“字”者和兼顾“多音节词”者两类，前者为“引得”，如《金文引得·殷商西周卷》、《金文引得·春秋战国卷》、《殷周金文集成引得》，以及《中国出土简帛文献引得综录·郭店楚简卷》；后者为“综类”或“类纂”，如《殷墟卜辞综类》、《殷墟甲骨刻辞类纂》。虽然两种索引方式各有所长，但其长处的真正实现却需要一定的前提条件：仅以“字”为索引对象，对字形信息有很高要求，如果能够真实呈现字的原形，无论是以时代先后为序，还是以文献种类为序，都可以很直观地呈现字形的断代或文献类型特点，以及文字构形演变发展的轨迹乃至规律，其认识价值不言而喻。但如果不能真实呈现字的原形，而只能呈现字的抽象形体（包括不忠实原出处真实字形的古文字字形摹写），以字为本的上述编纂逻辑就会在很大程度上失去意义。同时，在以字为本的索引体例中，一旦兼顾了多音节词的索引，则以字为本索引所呈现信息的上述内在逻辑势必因被多音节词“搅局”而混乱不清。另一方面，索引对象兼顾“多音节词”，需要的条件并不苛刻，只要对文献语言的多音节词单位的认识达到一定程度就可以做到。

综上，我们对上述既有“引得”、“综类”型工具书的编纂逻辑可以有一个更深入的认识：



甲骨文“综类”，尽管也给出原形字，但因其摹写并不忠实原片刻辞，故仅以字为索引对象无从真实反映字形特点及其变化的逻辑，这种逻辑既不存在，就可以兼顾多音节词索引；《中国出土简帛文献引得综录·郭店楚简卷》，因其实现了简文原形逐字保真的突破，就有意回避兼顾多音词索引所带来的混乱。

在编纂《包山楚简卷》的过程中，我们进一步反思上述编纂逻辑：《中国出土简帛文献引得综录·郭店楚简卷》立足于原形呈现、逐字保真得以突破的前提条件，不在同一索引层次上兼顾字与多音节词，当然是合理的，但这并不意味着不能用不同层次的索引来实现这两种不同语言单位检索的兼顾。毫无疑问，对于古文字文献的索引工具而言，兼顾这两种索引，使用价值将得到提升。本书增加“多字语义单位”索引，就是这种反思的结果。

相对“多字语义单位”索引，“偏旁类纂”无疑是本书编纂中更具新意的内容。关于“偏旁类纂”本身的创新意义，前文有说，这里不赘。而将偏旁类纂与引得相关联，则可视为古文字引得编纂一种新突破：“偏旁类纂”中的偏旁出处，不但需要给出相关字形，而且需要给出这个字形所出文句的所有原态字形。在实现了引得原形逐字保真呈现的前提下，不关联一个“偏旁类纂”乃是一种资源浪费，反之，则是古文字“引得”所提供字形信息的深度拓展。

虽然《包山楚简卷》相对《郭店楚简卷》又有了若干完善与增益，但在总体编纂思路上，《包山楚简卷》更充分地践行了“中国出土简帛文献引得综录”系列编纂之初所定的原则，这就是：为最大限度满足新时代古文字研究的需要而实现出土古文献引得类工具书编纂的创新。根据这一原则，《包山楚简卷》目前既定的编纂格局，亦很有可能被本系列后出各卷所突破。其意义首先在于补缺性。

5、项目阶段性成果《中国汉字文物大系》出版后引起社会关注

项目阶段性成果《中国汉字文物大系》（15卷）2013年12月由大象出版社正式出版。虽然面世未久，该书已经得到广泛社会关注，有“人民网”“文汇报”“读书”“文汇报书周刊”等10余家媒体进行了报道，项目负责人还为此书接受中国社科网的专访，节录如次：

用“数字化”激活古汉字宝藏 ——访华东师范大学终身教授刘志基

由华东师范大学终身教授刘志基主编的国家重点工程大型工具书《中国汉字文物大系》系列丛书于近日由大象出版社出版并在沪首发。该丛书为国家出版基金资助项目，是目前国内第一部将“汉字”视为文物进行解读，系统收录历代各类文物上所见文字的原始形态并附具体语境，全面呈现汉字形义源流和书体发展演变轨迹的实物文字资料整理类大型工具书。

该书共15卷，前14卷为正文部分，第15卷收录各种检索及参考文献等。全书按照《说文解字》14卷540部首结构全书，系统收录了自先秦至唐代各类文物上所见文字的原始形态，以及从甲骨文到楷书的具体演变过程。

日前，记者就该丛书的出版及文字学发展的相关问题采访了刘志基教授。

……

中国社会科学网：您怎样看待这套丛书的价值？



華東師範大學

EAST CHINA NORMAL UNIVERSITY

刘志基：关于这套书编纂，我们最在意的是尽可能反映学界关于出土文字考释研究的最新成果。像该书这样汇集历代第一手文字材料，并附语境，加以源流说解的编纂类型，此前尚未有过，它给出了各个研究领域所需要的文字材料应具备的各种信息，因此它的作用也应该是此前各种断代性专类性的文字编所不具备的。

中国社会科学网：据报道，这套丛书是全面运用出土文字语料库、数据库等汉字数字化工程的集大成之作。能否请您介绍一下出土文字语料库、数据库等汉字数字化工程的建设情况，及这两者对该系列丛书起到的具体作用？

刘志基：关于出土古文字数字化，我们是从《古文字诂林》电脑排印发端，开始建设，至今已有 20 年积累。目前从殷商甲骨文到中古唐宋楷书定型的材料均已贮存于数据库。2003 年我等主持开发过第一种先秦金文材料的《商周金文数字化处理系统》和第一种战国古文字材料的《战国楚文字数字化处理系统》作为正式电子出版物出版。本人目前尚在主持国家社科重大项目“出土古文献语料库建设研究”。《中国汉字文物大系》编纂的最大困难是要搜集历代汉字第一手实物材料，而这种材料正是我们的数字平台所贮存的。如果没有这样一个数字平台，从头开始去搜集材料，恐怕几年时间都拿不下来。当然，该书的编纂，对我们的数字化平台建设也是有推动作用的，因为它以正式出版物的要求来促使我们去完善这个平台。

.....

报：全国哲学社会科学规划办公室、上海哲学社会科学规划
办公室

领导同志如有批示或需要调阅成果，请与华东师范大学社科处联系

电话：021-54345118（吴文钰）；传真：021-62604330



華東師範大學
EAST CHINA NORMAL UNIVERSITY

E-mail: skc@admin.ecnu.edu.cn

华东师范大学社科处 [http:// www.skc.ecnu.edu.cn](http://www.skc.ecnu.edu.cn)