

DOI: 10.3969/j.issn.1009-8976.2021.01.016

战国楚简文献电子引得建设刍议

楼 兰

(上海海洋大学 海洋文化与法律学院,上海 201306)

摘 要: 伴随计算机技术的普及和发展,出土文献引得的编制工作也正在融入数字化的时代浪潮中。综观已有研究成果,尤其是 21 世纪以来,战国楚简文献电子引得的编制工作取得了长足进步。然而,对于大量的上博简或清华简的引得整理工作仍属空白。呈现战国楚简文献全貌的电子引得建设工作仍有待进一步开展。

关键词: 战国; 楚简; 引得; 出土文献

中图分类号: K231; G272

文献标识码: A

文章编号: 1009-8976(2021)01-0063-04

战国楚简文献是最重要的出土文献之一,它们真实地记录着当时的社会信息和历史风貌,弥补了传世文献的不足,是研究该时段历史、文学、考古等方面的重要材料。日益发达的信息技术,不仅拓宽了战国楚简文献的研究方法,同时也促进了战国楚简文献电子引得的编制。

一、战国楚简文献电子引得的概念

战国楚简文献电子引得,又称战国楚简文献电子索引,它是指使用信息技术对战国楚简文献编制各种类型的检索。从广义上来说,它包括两个方面:一是用计算机编制的各种纸质形态的引得,二是直接用计算机呈现的电子检索系统^[1]。传统编制引得的方式主要是以手工制作卡片为主,这种方式制作步骤繁琐,费时费力,错误率高。因此,随着现代社会信息化水平的逐步上升,在其他学科研究中得到频繁运用的计算机自动分析和数据处理等手段被逐渐应用到战国楚简文献引得的编制工作中。

战国楚简文献是 20 世纪以来出土的重要材料之一。根据统计,目前出土的战国楚简达二十多批,出土地包括河南、湖南、湖北等地。就内容而言,战国楚简文献关涉的内容也十分广泛。在书籍类中,常见的诗赋、散文等体裁都有;在文书类中,公私文书皆有涉猎,此外还出土了不少祭祷记录、遣策等。

要想利用这些浩繁的简帛材料进行学术研究,就必须有一部著录完备、检索便捷的引得。因此,为了方便学者翻阅检索,提高研究效率,海内外学者一直致力于出土简帛文献电子引得的开发与建设工作。

二、战国楚简文献电子引得建设的回顾

中国的索引之学,最早可追溯到明代。明代张士佩编写的《洪武正韵玉键》,在体例上已具备类似检索编排,但此类著作数量很少。真正意义的索引,开始于 20 世纪初。受西学东渐风潮的影响,20 世纪二三十年代中国在胡适、林语堂等人的倡导下,开展了一场索引运动。当时国内出现了一大批有影响力的索引研究论著,比如万国鼎的《索引与序列》(1928)、刘纯的《杂志索引之需要》(1929)、洪业的《引得说》(1932)、钱亚新的《中国索引论著汇编初稿》(1937)等。并建立了多个专业索引编纂机构,比如中华图书馆协会索引检字组(1925)、金陵大学图书馆杂志索引合作社(1926)、哈佛燕京学社引得编纂处(1930)等,编纂了大量索引工具书。其中以哈佛燕京学社引得编纂处的成就最高,共编制了 64 种共 84 册引得,极大方便了学者检索古籍。20 世纪 50 年代,学者们开始在索引开发工作中运用信息化手段。这克服了传统手工编制索引耗时、耗力、耗人的缺点,为索引的编制开辟了一片无限广阔的新天地。20 世纪 80 年代,索引编制理论和技术等都发生了深刻变革,索引载体形式也发生了质的变化,逐渐由纸质文本检索转向电子文本检索。香港和台湾率先开始了索引数字化的开发。1984 年,台湾“中央研究院历史语言研究所”启动了“史籍自动化计

收稿日期: 2021-02-20

基金项目: 上海市哲学社会科学规划一般课题“战国楚简文献电子引得建设研究”(项目编号: 2020BYY017)

作者简介: 楼兰(1982—),女(汉),江西南昌,博士,副教授
主要研究古文字学。

划”,开发了《廿五史〈食货志〉全文语料库》;1988年,香港中文大学集合社会各界力量成立了“汉达古文献数据库中心”,研发了《汉达古籍资料库》,这都加速了相关文史材料的搜集和检索。内地也建立了一批古代经典文献的检索系统。比如深圳大学开发的“《红楼梦》全文数据库”(1986)、四川大学编纂的《全宋文(第一册)》(1988)、中国社会科学院开发的“中国古典文献计算机数据库系统”(1988)等。自20世纪90年代以来,索引数字化的工作更为普及。比如上海人民出版社等发行的文渊阁《四库全书》电子版(1999),是中国第一部既能检索全文中的字词,又能检索书名、作者等多种信息的电子引得。人民邮电出版社发行的《二十五史多媒体全文检索阅读系统》(2000),是中国首套以检索速度快、收录内容全面而闻名的检索系统。栾贵明、田奕的《十三经索引(全四册)》(2004),实现了逐字、分经索引,是《十三经》研究和阅读的重大突破。可见,由于引得学的发展,大量传世古文献的引得问世,这极大方便了学者了解、掌握和利用古籍文献,有助于学者进行更深更广的专题研究。

出土古文献是最接近源头部分的中华历史文献遗产。自甲骨文、金文等大量出土材料问世之后,不少学者也开始对此类文献进行引得工具书的编制。最早可以溯源到罗振玉的《殷虚书契考释》初印本(1914),把有关甲骨卜辞按事类辑录文辞,并列出其出处。其后有不少在甲骨、金文著录书后附加类似的索引。比如中国社会科学院考古研究所编辑出版的《小屯南地甲骨》附册“索引”(1983)、周法高编写的《金文诂林》附册“索引”(1974)等。随着出土材料的增加,甲骨、金文索引工具专书也不断出现。其中,影响较大的是日本学者永田英正和高嶋谦一分别于1968年和1985年编纂的《京都大学人文科学研究所藏甲骨文字索引》和《殷虚文字丙编通检》。或出现汇总多种甲骨、金文著录资料按字类或其它形式进行文辞例编纂的工具书。甲骨文索引工具专书影响比较大的有日本岛邦男的《殷墟卜辞综类》(1967)和姚孝遂的《殷墟甲骨刻辞类纂》(1989)。前者是一部汇总1967年以前出版的63种甲骨著录书而编成的辞例文句检索工具书,该书资料取舍谨慎、编排体例新颖;后者是仿前者编纂而成,在材料的丰富程度、字形的归属、辞条的划分等方面有很大的改进和提高。金文索引工具专书比较有影响的是张亚初的《殷周金文集成引得》(2001)和华东师范大学中国文字研究与应用中心编纂的《金文引得》(2001—2002)。前者可对中国社会科

学院考古研究所出版的《殷周金文集成》进行文字检索,并附有《殷周金文集成》收字情况说明以及所收单字出现频率等方面的内容,是研究殷周这一时期历史、文字等方面的优秀工具书。而后者收录范围则更广,收录了2001年以前所有见于著录的从殷商到战国时期的铭文,并建立了多层次语境的检索系统,实现了字形检索方面的突破,在国内外学术界引起了普遍关注。

自20世纪以来我国出土了大量战国时期楚国的简帛文献,其内容涵盖了中国传统学术文化的多个学科领域,引起了众多学者的极大关注。目前,由于日益发达的信息技术加之学者在古文字数字化方面的持续努力,战国楚简文献电子引得的建设工作取得了丰硕成果。2013年,西南大学张显成主编了《楚简帛逐字索引》,全书4册,共包含逐字索引九种,分别是《〈楚帛书〉逐字索引》《〈望山楚简〉逐字索引》《〈长台关楚简〉逐字索引》《〈九店楚简〉逐字索引》《〈郭店楚简〉逐字索引》《〈曾侯乙墓竹简〉逐字索引》《〈葛陵楚简〉逐字索引》《〈包山楚简〉逐字索引》《散见楚简逐字索引》。该书对上述每一种简帛,都充分吸收了整理小组释文及最新研究成果,并在此基础上编制了逐字索引。2012年和2015年华东师范大学刘志基分别编纂出版了《中国出土简帛文献引得综录》两卷。它们以郭店楚简和包山楚简为材料范围,由“释文”“检字”“索引”三方面组成,其中“释文”“索引”以楷字释文和简帛文真实原形对照出现,有的还辅以网络检索等诸多现代科技手段^[2]。上述这些战国楚简文献引得,是简帛学界的重要工具书,也是百年来楚简文献整理释读的集大成者。它们极大地方便了学界对已出土简帛文献资料的研究和使用,对光大中国古老悠久的传统文化起到了推动作用。但值得注意的是,《中国出土简帛文献引得综录》目前仅有《郭店楚简卷》和《包山楚简卷》两卷发行,而《楚简帛逐字索引》也尚未吸纳数量众多的上海博物馆藏战国楚简、安徽大学入藏的战国竹简、清华大学入藏的战国竹简等。据不完全统计,目前已公布的上博简、安大简和清华简总字数都在30000个左右。就内容而言,它们有不少都是在传世文献中未曾见的佚文。以清华简为例,《系年》这一篇提供了很多前所未知的重要史料,尤其是前四章有关西周晚期与后四章战国早期的史事^[3]。因此,这些涉及中国传统文化的核心内容、文献价值极高的出土文献没有被囊括在战国楚简文献电子引得中,实属遗憾。可见,利用先进技术手段,呈现战国楚简文献全貌的电子引得仍有待进

一步开展。

三、战国楚简文献电子引得的建设

引得是一种用于文献检索的研究工具,它方便学者在较短时间里获取相关材料。因此,一方面,必须保证对所收出土文献的正确释读,充分反映战国楚简文献原貌;另一方面,在数据库的支持下,通过合理编排,建立简洁有效的多层次语境检索。具体来说,分为收集资料、设计框架和材料数字化三个方面。

(一) 战国楚简文献资料本体的收集与审读

战国楚简文献资料的收集涉及三个方面。第一,收集文献原始形态的影像。这是作为电子引得编制及使用的原始依据。要完成这方面的任务,必须保证文献数据是真实的、穷尽的。真实是指利用一手资料,保持文献资料原始风貌,从而形成战国楚简文献资料的拓片影像。穷尽是指不遗漏所有楚简文献。根据目前正式发表的战国楚简文献材料,范围限定为具有完整语境,且考释研究相对充分的战国楚简文献材料。具体包括《郭店楚墓竹简》《曾侯乙墓》《九店楚简》《望山楚简》《包山楚简》《新蔡葛陵楚墓》《信阳楚墓》《安徽大学竹简(一)》《清华大学藏战国竹简(一~十)》《上海博物馆藏战国楚竹书(一~十)》等。要说明的是,目前只能相对穷尽战国楚简文献,因为有不少已经出土的楚简语料尚未完全公布,比如1987年湖南慈利出土的1000余枚约21000余字等。第二,收集和审读与战国楚简图片相对应的释文,以此作为电子引得检索的对象。编制引得时,应在充分吸收整理小组释文的基础上,跟踪学术界的最新研究成果,从而确定战国楚简文献释文。值得注意的是,由于古文字研究水平的局限,部分战国楚简文字的释读迄今尚存在较多问题和分歧,甚至还有少数文字未得到考释。对于这些歧释字和未识字的正确释读,将是值得关注的内容之一。只有这样,才能在此基础上提供准确、完整的释文总表。第三,录入这些材料各自自然单位的时代、类型、著录出处等属性信息。这些来源属性的信息标注是对战国楚简文献进行检索的重要方式,所以标注的内容应尽可能详细,不仅需要包括原著录书籍篇章名,还需要包括简号、字序号等所有信息。这样有助于实现庞大的战国楚简文献资料的数字化。

(二) 战国楚简文献电子引得的框架设计与程序开发

技术开发主要包括采集与储存数据、设计语料检索系统,并开发相应实现程序。首先,数据采集与

储存包括字符编码、数据关联、数据维护等工作。其中,战国楚简文献的数据维护必须贯穿制作的全过程。编制引得时,要紧跟古文字研究动态,及时吸收最新研究成果,穷尽当下楚简资料,形成反映最新学术水平的释文文本。并在此基础上进行字符编码、数据关联等工作,从而建立具有完整性、准确性和便利性的战国楚简文献电子引得,为学术研究提供方便。其次,设计语料检索系统包括楚简文献语料的自动加工、用户的查询和检索等工作。在设计检索对象时,应以满足多层次、全方位的研究需求为目标。战国楚简文献具有十分重大的学术价值,关涉古文字学、历史学、考古学等多个学科的发展。因此,我们应该根据简帛文献的不同特点,在逐字索引的基础上,进一步开发满足多学科研究需求的检索系统。比如对上博简之类的书籍类文献,除了提供简号、释文、出土地点等信息,还可以设置简牍形状、长度、字体、形制、刻齿等方面的属性检索,使古文书学的专家全面了解每支简牍的特性,从而获取新的研究思路 and 方向。具有多学科特色的电子索引,不仅能使学者节省大量翻阅材料的时间,而且有利于学者开拓新的研究课题,进行更深入的综合性研究。另外,部分研究者对古文字释读可能存在一定的语言障碍。针对此种情况,电子引得可以增加考释检索。这个考释检索可以包括考释者、考释论著名、考释论著出处、考释论著发表时间等考释信息。这样一来,电子引得既有利于研究者对楚简帛文献内容的系统把握,又能方便不同专业研究对楚简帛文献材料的利用。

(三) 战国楚简文献材料的数字化

电子引得需要战国楚简文献中的每一个文献用字都能被显示、查询和检索。然而,由于古文字一直被排除在电脑字符集之外,古文字文献资料长久以来面临着电脑输入和电脑检索的困难。因此,要编制电子引得,首先要将所收集的战国楚简文献本体等材料录入数据库,使之变成可作数字化检索的对象。而这一内容的实现,有赖于涉及所收战国楚简文献本体全部用字的符合数字化处理要求的一字一码的精确文字整理,及基于这种整理的字体开发。在开发字体的过程中,需要注意两点。第一,对异体字和异写字的处理。战国楚简文字存在一字多形的现象。其中,异体字对于古文字学的研究具有多方面的研究价值,因此异体字应纳入数字化处理。而对异写字的处理则应分情况讨论。若同一个写手因刻意营造差异产生的异写字,均应纳入数字化处理;而若因手不从心造成的异写字,则不需要作数字化

处理,用典型字形取代即可^[4]。第二,对未识字和歧释字的处理。如前所述,战国楚简文献包含已有定论的字、歧释字和未识字三种。目前,研究考释尚有不少未知或未定论的问题,战国楚简文献中的未识字一般无法隶定,而歧释字如果以隶定字面目出现,就会掩盖其尚存歧释的实际情况。如何在引得中正确反映这些问题,是一个难点。对于这些无法隶定又不能确定读音的未识字和歧释字,应保存原形,并通过考释检索等方式来反映考释现状,从而进行有效检索。

综上所述,编制者首先必须全面系统搜集目前存世的战国楚简文献资料,总结现有研究成果,在此基础上运用古文字信息处理技术,建立战国楚简文献语料库。其次,应对战国楚简文献资料的释文进行全面校订,体现战国文字研究的最新成果。再次,应立足于战国楚简文献语料库全文检索的强大功能,通过合理编排,完成一个多层次语境的检索系统。最后,要实现战国楚简文字字形检索方面的突破,在字体开发的支持下,充分反映战国楚简文献原貌。这样一来,基于数字化整理的电子引得将为战国楚简文献研究提供一部资料翔实、考释准确、反映

最新考释进展的工具书。此外,电脑处理古文字一直是编制电子引得的难题。目前,关于出土古汉字进入国际标准字符集的编码体系问题正在逐步推进^[5],而国内也有王宁教授主持的“中华大字符集创建工程”和裘锡圭教授主持的“中华字库”两大工程。随着今后信息技术的发展,战国楚简文献电子引得编制者应和各类专业人才通力协作,实现通用平台上的楚简文献全文数字化检索。

参考文献

- [1] 毛建军. 两种《古今图书集成》电子版的比较——简谈古籍电子索引的标准与规范[J]. 图书馆理论与实践, 2008(3): 28.
- [2] 刘志基. 先秦出土文献语料类型分析刍议——以《包山楚简》与《郭店楚简》为例[J]. 语文研究, 2015(4): 19.
- [3] 李守奎. 一部古史探新知——清华简《系年》的价值及意义[N]. 人民日报, 2016-01-24(012).
- [4] 刘志基. 数据库古文字研究论稿[M]. 上海: 上海古籍出版社, 2019: 66.
- [5] 周晓文, 李国英. 建立“信息交换用古汉字编码字符集”的必要性及可行性[J]. 北京师范大学学报: 社会科学版, 2006(1): 77.

The Research on the Construction of Electronic Index of Chu Bamboo Slips in Warring States Period

LOU Lan

(College of Marine Culture and Law, Shanghai Ocean University, Shanghai 201306, China)

Abstract: With the popularization and development of computer technology, the construction of the index of unearthed documents is also being integrated into the tide of the digital era. A comprehensive view on the existing research results, especially since the 21st century, the construction of the electronic index of Chu Bamboo Slips in the Warring States Period has made great progress. However, it is still a blank to sort out a large number of bamboo slips from Shanghai Museum or Tsinghua University. The construction of the electronic index which presents the whole picture of Chu Bamboo Slips in the Warring States Period still needs to be further carried out.

Key words: Warring States Period; Chu Bamboo Slips; index; unearthed documents