

戰國秦漢簡牘文字語料庫的設計與使用

張再興*

目 次

1. 語料的收集範圍
2. 語料的加工特點與加工層次
3. 語料庫的功能與使用

戰國秦漢簡牘文字內容重要，數量巨大，目前尚缺乏便捷的引得、詞典等工具書，查檢十分不便。而海內外剛剛起步的資料庫建設無論在材料種類還是在內容加工、功能開發上都存在很大的發展空間。本文將介紹我們的戰國秦漢簡牘文字語料庫建設情況。

1. 語料的收集範圍

語料庫所收集的語料包括所有已經出版的戰國秦漢簡牘文字語料。目前已經發表的簡牘語料主要有：

1. 戰國楚簡：主要包括包山楚簡、長沙子彈庫戰國楚帛書、郭店楚墓竹簡、九店楚簡、上海博物館藏戰國楚竹書、曾侯乙墓竹簡、新蔡竹簡、望山楚墓竹簡、長沙五里牌楚墓竹簡、信陽長臺關、楊家灣楚墓竹簡、仰天湖楚墓竹簡等。

* [中國] 張再興，華東師範大學中國文字研究與應用中心教授
(zxzhang@zhwx.ecnu.edu.cn)。

2. 秦簡：主要包括睡虎地秦簡木牘、天水放馬灘秦簡、龍崗秦簡牘、楊家山秦簡、王家臺秦簡牘、關沮周家臺秦簡牘、里耶秦簡、嶽麓秦簡等。

3. 漢簡：主要包括敦煌漢簡、懸泉漢簡、額濟納漢簡、居延漢簡、居延新簡、肩水金關漢簡、武威漢簡、尹灣漢簡、張家山漢簡、阜陽漢簡、馬王堆漢簡帛書、孔家坡漢簡、銀雀山漢簡、江陵漢簡、長沙東牌樓漢簡、香港中文大學藏漢簡等。

語料庫具備開放性，後續發表的相關簡牘材料將不斷予以補充。

2. 語料的加工特點與加工層次

與現代漢語語料庫相比較而言，戰國秦漢簡牘語料庫的最大不同是簡牘語料的原始書寫形式不是現代楷書。即使是用隸書或草書書寫的漢代簡牘，字形也與楷書存在很大差異。在建庫過程中，需要有一個從實物文字到電腦系統楷書文字轉寫的過程。語料庫的基礎是語料的真實性。在文字轉寫過程中保持基礎語料真實性是出土古文字類語料庫建設首先必須解決的問題。

爲了解決這一問題，我們把戰國秦漢簡牘語料庫的語料本身分爲兩個層次。第一層次是簡牘文獻發表時整理者的基礎釋文，這一釋文是語料庫建設的基礎，但是語料庫加工者未進行進一步的加工，因此可以稱之爲“生語料”。第二層次是語料庫加工者根據文獻發表後學術界的研究成果，對基礎釋文進行校訂，形成“精校語料”。由於釋讀研究的不斷進展，釋讀現狀的把握關乎語料的真實性。這一層次的語料文本除了語料庫加工者對考釋研究成果的選擇結果之外，還顯示未被選擇的諸家考釋結果。而對於古書類簡牘語料而言，還可以顯示可資對照的傳世古籍文本。

就語料庫語料的屬性加工而言，則還有第三個層次，就是對語料進行分詞及詞性、詞義標注等加工，形成“精加工語料庫”。這是語料庫加工的核心層面。

除了古文字方面的特徵外，戰國秦漢簡牘作爲出土文獻還有出土文

獻方面的特徵。作為戰國秦漢時期的語料，還有古代漢語方面的特徵。為了兼顧這些特徵，以便更好地發揮語料庫的作用，除了一般語料庫所必需的加工外，戰國秦漢簡牘語料庫的加工還需要注意相關屬性資料的標記。

1. 相應的原始圖版。有整簡圖版和單字圖版兩種形式。
2. 相關出土文獻屬性。這一方面的屬性有不少，如：
 - (1) 考古形態特徵：清晰度、殘損度、散亂度等。
 - (2) 書跡性質：正常文字、圖畫、習字等。
 - (3) 簡牘形制：簡、牘、椀（觚）、楊、封檢、削衣等。
 - (4) 內容分類：古書、文書、遺冊等。
 - (5) 書體：篆書、隸書、楷書、行書等。

3. 文字釋讀研究方面的屬性資料

- (1) 殘斷簡的綴合與簡的編聯。
- (2) 文字用法的標注，包括通假字、古今字、訛字等。
- (3) 重文、合文等書寫特徵。它們往往只是書寫形態問題，無關其功能與意義。

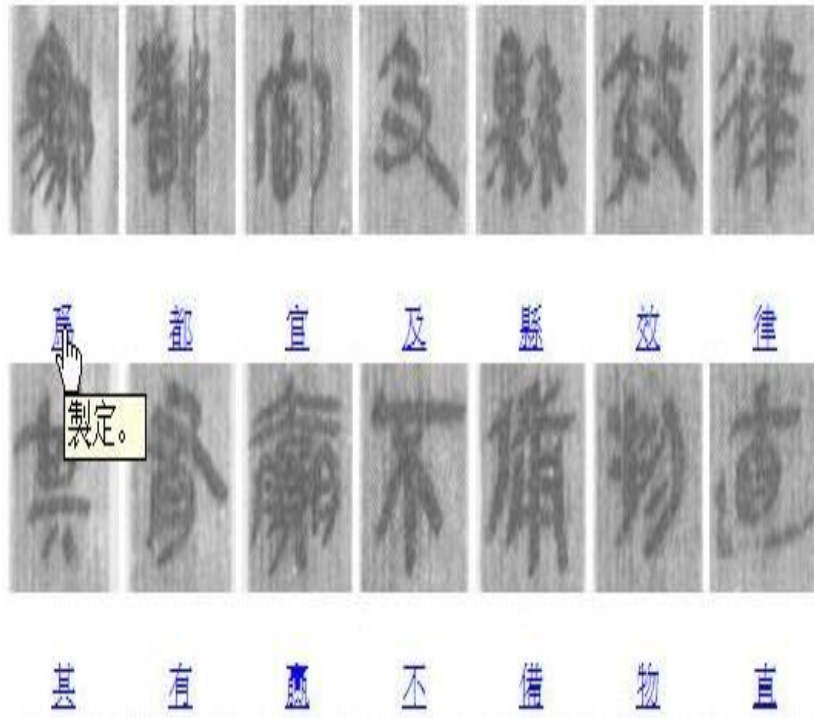
- (4) 文字單位的認定。文字單位的認定就是判定一個字形究竟屬於哪個字，並在語料庫中加以標記。這一工作關係到文字、詞彙等頻率統計的準確性。文字單位的認定不能僅憑外在的字形，而要結合字形的歷時發展及在語料中的意義綜合考慮。由於戰國秦漢時期是漢字動態發展的重要時期，簡牘語料又是民間手寫的文字，通假字、異體字等廣泛存在，戰國秦漢簡牘語料中的字形關係相當複雜，從而給文字單位的認定帶來相當大的困難。字形歸屬及源流關係尚不明確者宜按照從分不從合的原則確定文字單位。

3. 語料庫的功能與使用

戰國秦漢簡牘語料庫在內容上具有相當的完整性和高度關聯性。語料庫提供著錄、文例、字形、參考文獻、考釋信息、新造字字形等六大類型的檢索，基本囊括了對戰國秦漢簡牘進行語言文字研究所需要的各

種信息。語料庫中的簡牘單位、字詞單位、釋讀成果之間通過超鏈結互相關聯。如簡牘單位所列各字的字形和釋文下分別鏈結各字的全部字形及各家的考釋成果。再如字形檢索、文例檢索的結果中，全部的字形、例句均可以直接跳轉到各自的完整語境——簡牘單位頁面。

語料庫在功能上兼顧專業的學術檢索與大眾的學習閱讀。學術檢索功能將在後面詳細介紹。學習閱讀功能旨在為非專業研究的讀者提供淺顯方便的簡牘閱讀形式。例如，原形圖片與隸定楷書逐字對照的形式提供極其方便的字形識讀途徑。詞義即點即現的方式則十分便捷地顯示出簡牘字詞的意義。圖一顯示當鼠標停留在睡虎地秦墓竹簡第四篇《效律》第一簡正面第一字“爲”上時，顯示其語料庫標注的意義為“制定”。



圖一

此外，對於無法用普通輸入法輸入的語料庫專用新造字，包括新造隸定字和原形字，還提供“新造字形檢索”功能。該檢索可以從筆劃、

構件、構件拼音、部首等多種檢索方法進行檢索。檢索結果分兩欄顯示。左側為資料庫中使用的新造字形，可供語料庫檢索及個人寫作使用。右側為相應的字形圖片，供參照及個人寫作使用。

笔画数	<input type="text" value="-请选择笔画数-"/>	构件检索	<input type="text" value="尸典"/>
第一笔	<input type="text"/>	构件拼音	<input type="text"/>
第二笔	<input type="text"/>	部首检索	<input type="text"/>
第三笔	<input type="text"/>	<input type="button" value="提交搜索"/>	

根据您的检索信息 以下是查询结果:

字形	字形图片
𠄎	𠄎

首页 上一页 1 下一页 最后一页 转到 共1页, 当前第 1页, 共1条记录

圖二

下面以戰國楚簡子語料庫為例，說明語料庫的基本使用界面。需要說明的是，由於語料庫的功能在不斷的開發更新過程中，實際使用界面可能會與這裡略有不同，一些新的功能也未在這裡介紹。更準確詳細的信息請見網站語料庫的電子說明文檔。

1. 著錄檢索

著錄檢索提供以簡牘為單位的各種簡的流覽、查找。左側列出資料庫中的簡的類型（見圖三），點開後可以顯示每種簡的各篇及各簡。右側上部可以提供著錄書名、篇名、著錄號的檢索，下部顯示該簡的所有原始字形圖片及逐字對應的釋文。圖四顯示上博楚竹書紂衣第二簡的部分字形，圖五顯示上博楚竹書孔子詩論第二簡的部分字形。

- ④ 《包山楚簡》
- ④ 《吳沙字庫新編楚簡研究》
- ④ 《鄂君啟節》
- ④ 《九店楚簡》
- ④ 《上海楚簡新編楚簡行書》一
- ④ 《上海楚簡新編楚簡行書》二
- ④ 《上海楚簡新編楚簡行書》三
- ④ 《上海楚簡新編楚簡行書》四
- ④ 《上海楚簡新編楚簡行書》五
- ④ 《上海楚簡新編楚簡行書》六
- ④ 《睡虎地秦簡》
- ④ 《新蔡楚簡》
- ④ 《鄂君啟節》

書名:

書名:

冊數:

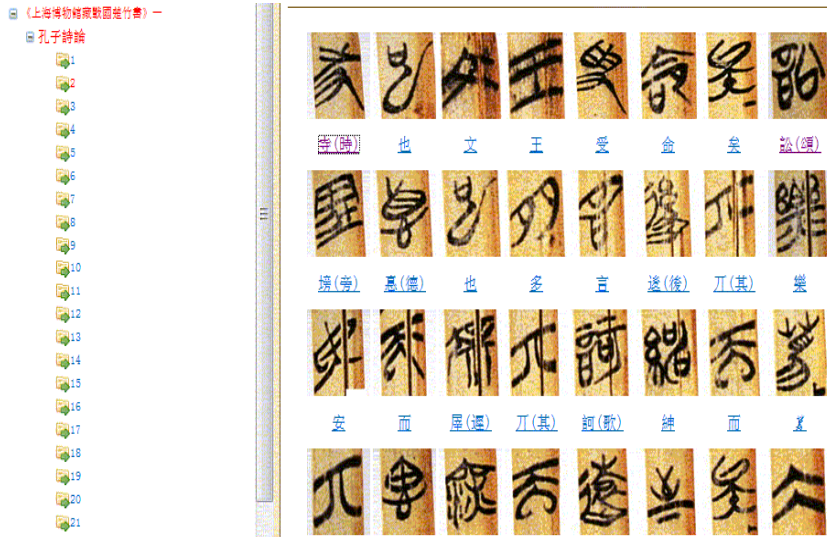


圖三

- ④ 《上海楚簡新編楚簡行書》一
- ④ 孔子論語
- ④ 紉衣
- 1
- 2
- 3
- 4
- 5
- 6
- 7
- 8
- 9
- 10



圖四



圖五

各簡字形及釋文分別鏈結字形表和考釋表。圖六顯示的是從圖五的第一個字“寺（詩）”鏈結的有關該字的考釋內容。圖七顯示的是從圖五第二行最後一個“樂”字鏈結的該字所有楚簡文字字形的第 11 頁。

考釋字形	考釋提要	著錄名	著錄號	篇名	考釋作者
	本簡長五十五·五釐米，上端弧形完整，下端弧形基本完整，上端留白八·七釐米，下端留白八釐米，殘存三十八字。	《上海博物館藏戰國楚竹書》一	2	《孔子詩論》	馬承源
	時，《文王》：“有聞不聰，帝命不時。”《傳》：“不時，時也，時，是也。”《箋》：“天命之不足乎？又是矣。”《文王》為《大雅》之始。	《上海博物館藏戰國楚竹書》一	2	《孔子詩論》	王志平
	時，《文王》：“有聞不聰，帝命不時。”《傳》：“不時，時也，時，是也。”《箋》：“天命之不足乎？又是矣。”《文王》為《大雅》之始。	《上海博物館藏戰國楚竹書》一	2	《孔子詩論》	王志平

首頁 上一頁 1 下一頁 最後一頁 1 轉到 共1頁, 當前第 1頁, 共3條記錄

圖六

字头	書式名	篇名	書式号	字形
樂	《郭店楚墓竹簡》	簡章三	54	
樂	《郭店楚墓竹簡》	簡章三	54	
樂	《郭店楚墓竹簡》	簡章三	59	
樂	《郭店楚墓竹簡》	簡章三	66下	
樂	《上海博物館藏戰國楚竹書》一	孔子詩論	1	

首页 上一页 ... 11 12 13 14 15 16 下一页 最后一页 11 辨别 共16页, 当前第 11页, 共79条记录

圖七

2. 文例檢索

文例檢索提供釋字字頭、隸古定字頭或詞義的檢索。檢索範圍可以進行多重選擇。見圖八。

当前位置: 文例检索

[返回前页] [返回检索类型]

请输入检索条件

释字字头

隶定字头

词 义

全选

《包山楚簡》

《長沙子弹庫戰國楚簡研究》

《郭店楚墓竹簡》

《九店楚簡》

《上海博物館藏戰國楚竹書》一

《上海博物館藏戰國楚竹書》二

文献类型 《上海博物館藏戰國楚竹書》三

《上海博物館藏戰國楚竹書》四

《上海博物館藏戰國楚竹書》五

《上海博物館藏戰國楚竹書》六

《曾侯乙墓》

《戰國楚竹簡匯編》

《新蔡葛陵楚簡》

圖八

文例檢索結果顯示所有文例，見圖九。按簡的出處分列文例。如果語料庫已經經過詞義分析標注，例句將按義項排列。尚未經過詞義分析者，則如普通引得直接排列例句。例句後標記篇名和簡號。單擊簡號，可以直接鏈結到圖四-五所示的相應簡的頁面。

《包山楚簡》(文) 共2例

(1). 姓氏。名詞 共2例

1. 荀誠(蠶)受日州加里子靈(靈)之公文壬百(丙)申之里公百(期)八月 文書 260

2. 訟道目(以)里坪易人文之拘尋(鑿)申(中)易(玄)盤邑人郢 文書 317

《郭店楚墓竹簡》(文) 共7例

(1). 周代國君。名詞 共7例

1. 王堯(刑)文悉(儀) 編衣 803

2. 王文 編衣 835

3. 王惠(德)賧(觀)文獸繕 編衣 838

4. 文 五行 900

5. 王此女(如)見文也之 五行 900

6. 王嗣復(作)文 成之聞之 997

7. 安(焉)聖(刑)王文之美聖(重) 成之聞之 998

《郭店楚墓竹簡》(文) 共1例

(1). 周文王, 人名。名詞 共1例

1. 文也壘(暹)周 辭遠以時 861

《上海博物館藏戰國楚竹書》二(文) 共7例

(1). 共7例

圖九

3. 字形檢索

字形檢索提供楚簡全部字形的檢索。檢索條件可以通過字形構件、釋字字頭兩者進行。檢索範圍也可以進行選擇。見圖十。檢索結果顯示符合條件的全部楚簡字形。圖十一顯示郭店簡、上博一、上博二這三種

材料中的“大”字檢索結果第13頁的字形。通過著錄號，也可以鏈結查看圖四-五所示的該簡全部字形情況。點擊字形圖片，可以鏈結該字形的各家考釋情況，如圖六所示。

当前位置: 字形检索

[返回前页] [返回搜索类型]

请输入检索条件

文字构件

字头

全选

《包山楚简》

《长沙子弹库战国楚帛书研究》

《郭店楚墓竹简》

《九店楚简》

《上海博物館藏戰國楚竹書》一

《上海博物館藏戰國楚竹書》二

《上海博物館藏戰國楚竹書》三

《上海博物館藏戰國楚竹書》四

《上海博物館藏戰國楚竹書》五

《上海博物館藏戰國楚竹書》六

《曾侯乙墓》

《戰國楚竹簡匯編》

《新蔡葛陵楚墓》

提交搜索

圖十

字头	著录名	篇名	著录号	字形
大	《郭店楚墓竹簡》	尊德義	29	
大	《郭店楚墓竹簡》	尊德義	38	
大	《郭店楚墓竹簡》	尊德義	38	
大	《郭店楚墓竹簡》	性自命出	61	
大	《郭店楚墓竹簡》	六德	3	

[首页](#) [上一页](#) ... [11](#) [12](#) **13** [14](#) [15](#) [16](#) [17](#) [18](#) [19](#) [下一页](#) [最后一页](#) [转到](#) 共19页, 当前第 13页, 共94条记录

圖十一

4. 參考文獻檢索

提供有關楚簡參考文獻的檢索。檢索途徑見圖十二。檢索結果顯示符合條件的所有考釋論著目錄。見圖十三。通過“查看詳細”鏈結，可以鏈結該論著的具體釋字情況，如圖六、圖十五所示。

当前位置: 参考文献检索

[[返回前页](#)] [[返回搜索类型](#)]

请输入检索条件

作者

题目篇名

书名刊名

出版时间期号

圖十二

作者	題目篇名	書名刊名	出版社	出版時間	查看詳細
李學勤	論包山楚簡魯陽公城郭	清華大學學報(哲學社會科學版)		2004年第3期	查看詳細
李學勤	上博楚簡《魯邦大旱》解義	《上博館藏戰國楚竹書研究續編》(上海大學古代文明研究中心、清華大學思想文化研究所編)	上海書店出版社	2004年	查看詳細
李學勤	楚簡《子羔》研究	《上博館藏戰國楚竹書研究續編》(上海大學古代文明研究中心、清華大學思想文化研究所編)	上海書店出版社	2004年	查看詳細
李學勤	釋《詩論》簡“免”及從“免”之字	《北方論叢》		2003年第1期	查看詳細
李學勤	《詩論》說《周禮》等七篇釋義	《齊魯學刊》		2002年第2期	查看詳細

[首頁](#) [上一頁](#) [1](#) [2](#) [下一頁](#) [最後一頁](#) [1](#) [轉到](#) 共2頁, 當前第 1 頁, 共7條記錄

圖十三

5. 考釋信息檢索

可以通過字頭進行檢索。見圖十四。檢索結果以提要方式提供對某個字的釋讀情況。見圖十五。著錄號一欄可以鏈結圖四-五所示該簡的完整信息，考釋作者一欄可以鏈結圖十三所示該作者的論著信息。

當前位置：考釋信息檢索

[返回前頁] [返回搜索類型]

請輸入檢索條件

字頭

圖十四

考釋字形	考釋提要	著录名	著录号	篇名	考释作者
	妄言之：即常语“姑妄言之”。	《上海博物館藏戰國楚竹書》三	10	互先	趙建功
	之，其也，	《上海博物館藏戰國楚竹書》三	10	互先	丁四新
	“之”是代词，指代“名”。	《上海博物館藏戰國楚竹書》三	10	互先	秦梓林 凌瑜
	（庸或）得之，庸或失之：有的能得到，有的卻不能得到。	《上海博物館藏戰國楚竹書》三	13	互先	鹿朴
	“之”读为“时”。	《上海博物館藏戰國楚竹書》三	4	互先	王志平

首页 上一页 1 2 3 4 5 6 7 8 9 10 ... 下一页 最后一页 1 转到 共11页,当前第 1页,共52条记录

圖十五

<參考文獻>

- 駢字騫、段書安，《二十世紀出土簡帛綜述》（文物出版社，2006）
- 劉志基，《出土古文字文獻語料庫建設芻議》，[韓國]《漢字研究》3（2010）
- 張再興，《互聯網時代出土文獻數據庫建設的思考與實踐》，《中國文哲研究通訊》82（2011）
- 程少軒，《“上博簡字詞全編資料庫”的設計和製作》，《中國文哲研究通訊》82（2011）
- 魯家亮、李靜，《數位時代的簡帛研究》，《中國文哲研究通訊》82（2011）

<Abstract>

The content of Warring States, Qin and Han bamboo slips text is huge and of great importance, but there is no handy reference books as index and dictionary yet, so it's very inconvenient for searching. The corpus construction, which is starting up recently domestically and overseas, has great development space in both content-processing and function-developing. Based on the officially published bamboo slips of Chu in Warring States, Qin and Han dynasty, we're now working on the Warring States, Qin and Han Bamboo Slips Text Corpus, with different levels of Raw Corpus, Highly-revised Corpus and Highly-processed Corpus. The corpus has multiple archaeology and character explaining properties, focused on both popularization and researching, and highly connected among all levels.

Key words: Warring States, Qin, Han, Bamboo and Wooden Slips, silk manuscripts, Corpus

投稿日 : 2012.04.28

審查日 : 2012.05.11-20

確定日 : 2012.06.20