

漢字語料庫分析叢書

劉志基／著

# 數據庫古文字 研究論稿

元華事陽白水人也高祖孝文皇帝之九嬪盧氏義陽長公主之母謙光柔順播風聲於素宗英清玉粹登樹華而俞額福慶無徵春秋時有八以延昌三年歲在甲午八月丁丑朔十三日巳丑寢疾而薨皇上震悼六宮哀慟使兼大鴻臚奉策即柩退贈元華為九月廿八日癸酉葬於山陵之域敬刊石瑤演述泉宇其辭曰穆矣元華風膏上靈女徽婉婉淑慎其聲承訓素里派光紫蓮金輝玉潤蕭風永蓄浮光未暮秋英中摧朱華流翠組帳凝埃泉宮杳杳深城莫關收勒泉石以裕餘哀

漢字語料庫分析叢書

劉志基／著

# 數據庫古文字 研究論稿

光緒帝陽曰木人也 高祖孝文惠  
帝之九嬪虛氏義陽長公主之母諡  
先柔順播風聲於素宗英清玉粹登  
樹華而俞額福慶無徵春秋卅有八  
以延昌三年歲在甲午八月丁丑朔  
十三日己丑寢疾而薨 皇上震悼  
六宮哀慟使兼大鴻臚奉祭即柩退  
贈元華爲九月廿八日癸酉葬於山  
陵之域敬刊石琤演述泉宇其辭曰  
釋矣元華風實上靈女機婉婉淑慎  
其齊承訓素里派光榮冠金輝玉潤  
靜且永馨浮光未素秋英中揮朱華  
流華組帳凝琰果宮香香添凝莫關  
致勤果石以證餘哀

上海古籍出版社

圖書在版編目(CIP)數據

數據庫古文字研究論稿 / 劉志基著. —上海: 上海古籍出版社, 2019.5

(漢字語料庫分析叢書)

ISBN 978-7-5325-9207-4

I. ①數… II. ①劉… III. ①數據庫—應用—漢字—古文字學—文集 IV. ①H121-39

中國版本圖書館 CIP 數據核字(2019)第 069869 號

漢字語料庫分析叢書

數據庫古文字研究論稿

劉志基 著

上海古籍出版社出版、發行

(上海瑞金二路 272 號 郵政編碼 200020)

(1) 網址: [www.guji.com.cn](http://www.guji.com.cn)

(2) E-mail: [guji1@guji.com.cn](mailto:guji1@guji.com.cn)

(3) 易文網網址: [www.ewen.co](http://www.ewen.co)

常熟文化印刷有限公司印刷

開本 635×965 1/16 印張 26.5 插頁 3 字數 368,000

2019 年 5 月第 1 版 2019 年 5 月第 1 次印刷

印數: 1—1,050

ISBN 978-7-5325-9207-4

H·206 定價: 98.00 元

如有質量問題,請與承印公司聯繫

# 漢字語料庫分析叢書

## 學術委員會

主任：王寧

委員：阿辻哲次（日本）

大形徹（日本）

河永三（韓國）

李圭甲（韓國）

朴興洙（韓國）

阮俊強（越南）

王元鹿

董蓮池

## 編纂委員會

主編：劉志基

編委：白於藍

潘玉坤

姚美玲

郭 瑞

劉 凌

張再興

鄭邵琳

張旭東

臧克和

教育部人文社會科學重點研究基地重大項目

“系列古文字專題數據庫建設”

(項目號: 18JJD740004)

華東師範大學共用交叉基金(人文社會科學)專案

“AI 和 VR 技術介入的漢字應用體系建設”

(項目號: 2019ECNU-GXJC001)

## 出版說明

西方標音文字體系語言學之所以成為衆多領域的領先學科，其中一個重要原因就是從十九世紀到二十世紀初大規模資料庫建設，使舊有的語言學成熟為語料庫語言學。在中國乃至整個表意文字區域，這種情況則相對落後。當二十一世紀已過去十幾年，資料庫的建構及使用，仍是部分學科領域中少數專家的事。中國語言文學專業的教員，直到現在仍然有許多人不太明白電腦裏 word 文本跟數據庫結構有何區別。部分標音文字譯介能力健者，能作無間之郵：從西方語言學那裏浮光掠影一二“方法”，回到東方表意文字語境中努力尋覓二三用例以“匹配”，領異標新三五年。說到底是添了些連自己都不敢相信的混亂。兩種現象有同樣的結果，即導致漢語言文字學進一步滯後與不成體系。

漢字語料庫的建設必須邁出沉重的步伐了，二十世紀九十年代末期，經過多方聯繫，我們籌措了百萬元經費，用於包括創刊出版《中國文字研究》、召開首屆國際漢字學術研討會、購買國際品牌硬件系統所構成的一流工作平臺、研發古文字信息系統等。

至今猶記，一九九八年元旦前後，有很多天我們是在廈門、廣州、海口或南寧的國際機場度過的。那是一段艱難的時光，雖然是學校漢語言文字學科帶頭人，但上述所有經費支出包括差旅費，都是在這個過程裏自己去爭取

的,沒有院系的任何資助。我不會忘記當時富有遠見卓識給予慷慨支持的出版家,像廣西教育出版社李人凡總編等。在這套叢書的第一本付梓之際,重提這些淵源,是爲了記住學科建設的那個艱難的起點,這些都直接奠定了後來華東師範大學中國文字研究與應用中心進入教育部高校人文社會科學重點研究基地行列的基礎。日月如梭,這些舊事如今已變得陌生而又疏遠,基地的一切,對於一些年輕人來說,似乎都是天然的存在。這時候,我們要記住,無論世事如何變遷,道德恩義是永遠不能忘記的。

基地建成伊始,就有明確的規劃:以出土文字大資料挖掘爲基礎,發展語料庫文字學,建成以漢字爲核心的表意文字數字化學科體系。

基於如上定位,基地落實“十五”、“十一五”、“十二五”及“十三五”規劃,先後研發了“魏晉南北朝石刻文獻語料庫”、“隋唐五代石刻文獻語料庫”、“出土古文字語料庫”、“今文字實物用字語料庫”,並在此基礎上,完成了“中國文字發展史”、“漢字斷代調查及漢字發展史”、“表意文字系統調研”等一系列重大項目。與此同時,基地歷屆博士研究生攻讀學位論文選題,也大多依託基地出土文字語料庫平臺,進行專題調查統計分析,可以說,在一定程度上呈現出基地固有的學科特色。

根據上述學科建設、研究模式以及所形成的資源特色,基地學術委員會和叢書編委會將每年篩選出當年度相對優秀書目一種(特別集中的年份,會適當考慮增加,但最多不超過兩種),推薦給專業出版社付梓。考慮到出土文字語料庫完全是一個開放系統,書目不追求體系完備,只要或在某個專題上有新意、有價值,或在語言文字的某個類型上有系統、有建樹,甚至是一束有新解、有實證資料的論文集合,都在推選範圍之內。所望博雅君子、專業同好,大家都來參與建設,共同發展吧!

教育部人文社科重點研究基地

華東師範大學中國文字研究與應用中心

漢字語料庫分析叢書編委會

二〇一六年元日

## 前言

二十世紀 90 年代後期，由於承擔《古文字詁林》排印過程中古文字集外字檢索任務，不得不做起了數據庫。這一段故事，我在前一本拙著《中國文字發展史·商周文字卷》“後記”中有過敘述，這裏不贅。後來又由於多個科研項目前後相繼的結項壓力，近 20 年的生涯，幾乎就淹沒在不斷製作各類古文字數據庫的繁忙中。雖然寫論文已經被擠壓到業餘愛好的位置，但生活在當下高校的考核體制下，每年忙裏偷閒地寫三兩篇論文模樣的東西又是不能不幹的事。客觀而言，把寫論文降格到應付差事的位置，並不能完全歸結於做數據庫的繁忙，畢竟並沒有誰在拿著槍逼著你，捫心自問，之所以做這樣的選擇，是自己內心深處的一種學術評估：做成一個有意義的數據庫，比寫一篇論文更重要。至少，對我而言是這樣。當然，這種覺悟，並不是一開始就有的。

當年《古文字詁林》編纂最終不得不靠數據庫來“救命”，如今回想起來，這樣一個具體個案，實在具有某種劃時代的象徵意義：由於時代發展帶來了文字處理方式的進步，以數字化方式來研究古文字必然被提上議事日程，而固守昔日的傳統研究模式，至少在有些時候和有些事情上難免被不斷湧動的時代浪潮拍倒在沙灘上。而依靠數據庫把一盤死棋走活的親歷體驗，則讓我看到了彼時不得不去製作的數據庫，長遠來看，確實有

著不能不去開發研製的理由，因為它是支持古文字研究模式數字化轉型的基礎平臺。由是又逐步形成這樣一種理念：我寫的古文字研究文章，總應是在我製作的古文字數據庫支持下完成的；而我製作的古文字數據庫，總是應該具有支持古文字研究的功能和價值。如此一來，做數據庫自然也就成為主業，或者說常態工作；而寫文章，某種意義上則是依附於數據庫研製的，或者說是數據庫價值的某種體現。

本書所彙集的文章，就是這樣一種基於古文字數據庫而形成的研究文字。因此，放在“數據庫古文字研究論稿”這個題目下，或許是比較合適的。按照內容大類，分為四篇：其一曰“理論方法”，大抵討論數據庫建設基礎上的古文字研究理念、思路乃至操作方式；其二曰“字體研究”，主要彙集基於數據庫的古文字字體研究方面的若干專題文章；三曰“偏旁分析”，收錄數據庫建設基礎上的偏旁分析方面的實踐性專論；四曰“異形探索”，主要依據數據庫的定量數據來發現古文字構形中某些特異現象乃至其認識價值。

然而，本書似乎並不能定義為一種彙集舊作的論文集。

既然積累了一些數據庫，在指導研究生撰寫古文字論文的過程中，自然會利用數據庫建設積累的資源，為學生們營建一個數字化研究的小環境。在這個小環境裏，同學撰寫學位論文，可以根據不同的選題，程度不等地獲得古文字材料數字平臺和古文字考釋研究數字平臺的支持。這樣做的初衷，是試圖取得一石二鳥之功效：既為同學的學位論文撰寫提供一個較高的起點，又為同學的釋讀積累提供查找檢索的方便。而實際的結果則是有喜有憂：可喜者自然是這樣做確實可以提高同學論文撰寫的效率，甚至有助於提升選題的學術內涵；可憂者則是有的同學有了數據庫，便自以為材料都在數據庫裏了，於是就覺得不再有深入解讀的必要，而一旦放棄了這種努力，便失去了發現問題的敏感，不知道如何用數據庫裏貯存的大把材料去解決真正有學術意義的問題，於是寫出的論文便往往是滿紙數據卻不知所云，導致學術認識的平庸化。

然而，這類問題，似乎並不僅僅出現在學生們身上。近年來閱讀一些

從事古文字數字化研究同行的論文，有時也會發現對研究對象釋讀瞭解不夠的問題。比如最近讀到一位同行關於古文字字庫開發研究的論文，文章提出了一種古文字字庫建設的思路：通過建立不同的古文字子字庫的方法，來實現一個古文字的不同字形共用一個碼位的目的。這一方案的初衷，是希望能夠像輸入常用的現代漢字那樣，借助常用輸入法即可在文字處理系統如 Word 中準確、高效地輸入和調用古文字字形，也就是借鑒現行的通用電腦字體技術，如一個字可以轉換宋體、黑體、仿宋一般，來解決古文字的一字多形問題，願望不可謂不美好。然而，這一看似美好的方案，顯然是在很不熟悉方案涉及材料的複雜性的情況下提出的。以一手材料為對象進行整理分析，就會發現，即便以古文字研究最一般的不同字形認定標準，一種特定斷代和特定類型的古文字所分析出的一字多形者，數量多的都可以達到成百上千。比如本書《字體研究篇》中所析出的西周金文“寶”的構形，即達 1 060 個，難道做一個西周金文的原形字字庫，就要做 1 060 個子字庫嗎？如果累加各個斷代和各個類型的古文字，這種數量更會變成天文數字。

此種現實，令我反思一個問題：基於數據庫的古文字數字化研究，固然是一種順應時代潮流的古文字研究模式，相對傳統研究方式，显然有其長處。但是我們還是應該正視這樣一個事實：前數字化時代的古文字研究模式，是歷經千年積累，不斷優化，逐步形成的，因此具有積淀的優勢。與之相對，新的研究模式，只有為時不長的小範圍實驗，因此，雖然有著順應時代發展的合理性，但缺陷的存在、問題的出現也是難免的。就現實情況來看，數字化在帶來數據查詢檢索方便的同時，又具有誘導人們降低深入解讀可能。而在另一方面，數據庫通常會將更多的材料納入人們的視野，而基於數據庫的古文字數字化研究，又是以挖掘大範圍材料的大數據進行定量研究為特徵的，這相對傳統研究方式，往往會要求增加材料釋讀的數量，進而導致材料釋讀力不從心的狀況。

平心而論，此種問題，對我而言，同樣是一道坎。看得見看不見是一回事，邁得過邁不過又是一回事。因此，借此機緣，對既有之作加以重新

審視，也是理所當然之事。於是，收入本書各篇，雖多為舊作，但有的重寫，有的大改，至少也做小修。而所有修改，無不基於所涉材料的進一步釋讀、整理、分析而做出。當然，即便如此，這本小書，依然會存在各種各樣的問題，誠摯期待同行方家不吝賜教。

## 凡 例

一、本書撰寫所依託的數據庫，主要是作者所主持國家社科重大項目“出土古文獻語料庫建設研究”之結題數據庫。書中所言及“數據庫”，如無特別說明者，皆係該數據庫。該數據庫的具體情況，可參見本書《理論方法篇》之《簡說古文字數據庫的改善——以新版“文字網”古文字數據庫為例》。

二、本書中所出現古文字原形字，有兩種給出方式，一是直接切割古文字材料拓片或照片而成的圖像字形，另一種是據拓片或照片的字形圖像掃描造字生成字體，進而以字體呈現的字形。前者因為完全保真，主要用於需要嚴格呈現特定材料字形原貌的場合。後者具有提高清晰度和節省篇幅的優點，但一般是經整理歸納而形成的同種文字類型的同種寫法、風格的典型字形，故用於字形寫法、風格具有一般保真要求的場合。至於其保真程度，可通過《理論方法篇》之《簡說古文字數據庫的改善——以新版“文字網”古文字數據庫為例》一文中所涉字形整理歸納個案以見其詳。

三、本書需要量化數據來支持論證，為了保證正文的可讀性，篇幅較大的量化數據材料以附錄的形式附於相關篇目之末。

四、為保證量化數據的可驗證性，本書對所有支持論證的材料皆給

出具體出處。給出的方式一般依次爲著錄名、著錄編號、篇名、字序號等，因各種文字類型的不同而給出方式略有差異，具體可參各篇相關說明。

五、本書對字形分類，一般依據“寫法”。所謂“寫法”，是指特定寫手對於特定文字所持有的書寫方式。一般來講，每個人都有比較固定的寫法，而如果寫手想改變一下自己寫出的字形，那他就需要通過另換一種寫法來實現這種意圖。改變“寫法”的方式大致有：改變字形朝向、以團塊與線條別異、增減構字成分、改變主筆姿態、改變筆劃連接方式、改變偏旁寫法、改變偏旁的方位佈局及構件增減等。

六、本文對古文字字形的斷代分期，主要依據著錄者的分期標注，少數參考近年來在期刊、論文集上發表的各家相關研究，在綜合評估後作出認定。

## 目 錄

前言 / 1

凡例 / 1

### 一、理論方法篇 / 1

數據庫文字學芻議 / 3

數字化與古文字研究新材料 / 18

簡說古文字數據庫的改善

——以新版“文字網”古文字數據庫為例 / 42

古文字數據庫的集外字問題 / 61

### 二、字體研究篇 / 85

微族同文器字體研究 / 87

同銘重見字異寫與金文字體研究 / 168

西周金文“貝”之字體再研究

——兼論斷代分期視角的青銅器銘文字體研究的“字體”界定問題 / 190

### 三、偏旁分析篇 / 209

古文字偏旁數字平臺與數字化環境下的古文字偏旁分析法 / 211

《包山簡文偏旁類纂》簡說 / 229

說楚簡文字中的“𠂔”及其相關字 / 241

也說“噩”字兼論“內”式飾筆之來由 / 255

四、異形探索篇 / 285

殷商文字朝向不定與同辭同字鏡像式避複異寫 / 287

楚簡文字缺邊現象芻議 / 317

上古文字填實構形成分的表敬意義

——以“王”字的分析為中心 / 328

楚簡文字“心”字符的寫法分類與同篇異寫研究 / 370

引書簡稱 / 411

主要參考文獻 / 412

# 一、 理論方法篇

