

## ※「數位時代的出土文獻」專輯※

# 「上博簡字詞全編資料庫」的設計和製作

程少軒\*

二〇〇六年起，復旦大學出土文獻與古文字研究中心開始承擔國家社科基金重點項目「上博簡字詞全編」。該項目計劃收錄上博簡的全部字形、辭例並加以分析和解釋。由於上博簡字形眾多，僅《上海博物館藏戰國楚竹書》前六冊就有兩萬多字，字頭兩千三百多個<sup>1</sup>，如以手工分選字形、詞彙和辭例，不但工作效率不高，還很容易造成收集字詞的重複或遺漏。為配合項目撰寫工作，我們從二〇〇七年開始研發「上博簡字詞全編資料庫」。

「上博簡字詞全編資料庫」是一個以上博簡字詞為核心內容的數據庫檢索系統，它由數據庫、數據庫附屬文件和檢索程式三部分組成，可以對上博簡字形、詞彙、辭例和相關論著進行基本檢索、高級檢索和提取數據等操作。在本文中，我們將介紹「上博簡字詞全編資料庫」的設計和製作，並簡單總結製作過程中的經驗教訓。

## 一、設計思路

我們設計的「上博簡字詞全編資料庫」由數據庫、數據庫附屬文件和檢索程式三部分組成。

---

\* 程少軒，復旦大學出土文獻與古文字研究中心博士研究生。

<sup>1</sup> 初步統計字頭為 2370 個，但考慮到字頭尚須調整歸併，這一數字恐非最終數據。

## （一）數據庫

數據庫分為七份表單，即表 A：字頭數據表、表 B：字詞對應數據表、表 C：全字形數據表、表 D：詞義數據表、表 E：短句辭例表、表 F：段落辭例表以及表 G：論著目錄表。

### 表 A：字頭數據表

字頭數據表將上博簡所有篇章的隸定文本進行篩選，提取出所有不重複的字形，再將提取出的每個隸定字形添加詳細的信息，如下表所示：

| ID | A1 | A2 | A3 | A4 | A5 | A6 | A7 | A8   | A9 | A10    |
|----|----|----|----|----|----|----|----|------|----|--------|
| 1  | 諸  | 言者 | 者  | 言  | 章  | 魚  | 15 | zhu1 | 3  | 031215 |

A1為隸定字形。A2為該字形所含全部構件。A3為該字聲符。A4為該字部首。A5為該字上古聲紐。A6為該字上古韻部。A7為該字隸定字形筆畫數。A8為該字現代讀音。A9為該字在《說文》中的卷數。A10為該字在《上博簡字詞全編》中的編號；《上博簡字詞全編》字頭按《說文》排序，03指《說文》第三卷，12指「言」為《說文》該卷第十二個部首，15指「諸」為該部首第十五個字。這樣的一行就是一個字頭的數據，每行數據分配一個ID號碼。

字頭數據表中的數據項可按學術需要擴充。該表是資料庫的字形索引，列出上博簡中出現的所有字形，依此表可按各種信息檢得所需字形。該表在整個檢索系統中作為數據庫的樞紐，尤其在高級檢索時起到重要作用。同時，該表也為《上博簡字詞全編》編定和排列字頭。

### 表 B：字詞對應數據表

字詞對應數據表將對上博簡所有篇章的釋文進行篩選，提取出全部不重複的嚴格隸定字形與該字所表示之詞（反映為文字一般即該詞在古書中的通行寫法）的配對組合。該表提供的是上博簡中字詞對應關係的信息，將是撰寫《上博簡字詞全編》的重要參考，同時也在高級檢索中起到重要作用。以〈緇衣〉的一段簡文為例：

𠂔（好）顏（美）女（如）𠂔（好）紃（緇）衣，亞 = （亞亞—惡惡）女

(如) 亞(惡) 衛(巷) 白(伯)。

提取數據如下表所示，B1 為字，B2 為詞，B3 為備注：

| ID | B1  | B2 | B3 |
|----|-----|----|----|
| 1  | 𠂔   | 好  |    |
| 2  | 頤   | 美  |    |
| 3  | 女   | 如  |    |
| 4  | 紂   | 緇  |    |
| 5  | 衣   | 衣  |    |
| 6  | 亞 = | 惡惡 | C  |
| 7  | 亞   | 惡  |    |
| 8  | 衛   | 巷  |    |
| 9  | 白   | 伯  |    |

上博簡中有不少特殊的文字現象，而且並非所有的字都有確定的釋讀，因此該表需要再加注數據 B3，對這些情況加以說明——如 A 為錯別字，B 為合文，C 為重文，D 為釋讀存疑等等。

### 表 C：全字形數據表

全字形數據表對上博簡的全部釋文和文字圖片進行梳理，錄入上博簡的所有文字的信息，包括每個字的嚴格隸定字形、所表示之詞、文字字形圖片連結位置、所屬篇章、簡號以及在所屬簡序號。如下表所示：

| ID | C1 | C2 | C3          | C4       |
|----|----|----|-------------|----------|
| 1  | 𠂔  | 好  | 01\02\01\03 | 緇衣 01.03 |
| 2  | 頤  | 美  | 01\02\01\04 | 緇衣 01.04 |
| 3  | 女  | 如  | 01\02\01\05 | 緇衣 01.05 |
| 4  | 𠂔  | 好  | 01\02\01\06 | 緇衣 01.06 |
| 5  | 紂  | 緇  | 01\02\01\07 | 緇衣 01.07 |

|    |     |      |             |          |
|----|-----|------|-------------|----------|
| 6  | 衣   | 衣    | 01\02\01\08 | 緇衣 01.08 |
| 7  | 亞 = | 亞亞惡惡 | 01\02\01\09 | 緇衣 01.09 |
| 8  | 女   | 如    | 01\02\01\10 | 緇衣 01.10 |
| 9  | 亞   | 惡    | 01\02\01\11 | 緇衣 01.11 |
| 10 | 衞   | 巷    | 01\02\01\12 | 緇衣 01.12 |
| 11 | 白   | 伯    | 01\02\01\13 | 緇衣 01.13 |

C3欄表示附屬文件中字形圖片的地址連結，如 01\02\01\03 表示《上博一》的第二篇（〈緇衣〉）的第一支簡的第三個字。該連結用以在檢索程式中顯示圖片和提取圖片。C4 則是字形所屬篇章、簡號以及在所屬簡序號的文字說明。該表直接應用於字形檢索程式。

**表 D：詞義數據表**

詞義數據表將包含上博簡的所有詞的信息，包括詞的嚴格隸定字形、寬式隸定字形和釋義。如下表所示：

| ID | D1 | D2 | D3                                                                 |
|----|----|----|--------------------------------------------------------------------|
| 1  | 𪛗= | 燕燕 | 燕子。《詩·邶風·燕燕》：「燕燕于飛，差池其羽。」孔穎達疏：「此燕即今之燕也，古人重言之。」在上博簡《孔子詩論》中用作《詩經》篇名。 |

該表直接應用於詞義檢索程式。

**表 E：短句辭例表**

短句辭例表收錄上博簡文獻的所有辭例，且辭例以短句子形式出現。該表包含三項數據：E1 是每條短句子辭例的嚴格隸定文本，E2 是每條短句子辭例的通行字文本，E3 是每條短句子辭例的嚴格隸定並加括注的完整文本。如下表所示：

| ID | E1                            | E2                               | E3                                                                |
|----|-------------------------------|----------------------------------|-------------------------------------------------------------------|
| 1  | 子曰為上可齊而智也為下可順而皆也則君不怨丌臣 = 不或於君 | 子曰：為上可望而知也，為下可述而識也，則君不疑其臣，臣不惑於君。 | 子曰：為上可齊（望）而智（知）也，為下可順（述）而皆（識）也，則君不怨（疑）丌（其）臣 = （臣，臣）不或（惑）於君。【緇衣 2】 |
| 2  | 咎員咎人孳 = 丌義不弋                  | 《詩》云：「淑人君子，其儀不忒。」                | 《咎（詩）》員（云）：【緇衣 2】「咎（弔—淑）人孳 = （君子），丌（其）義（儀）不弋（忒）。」【緇衣 3】           |
| 3  | 尹筭員佳尹身及康咸又一惠                  | 〈尹誥〉云：「唯尹允及湯，咸有一德。」              | 〈尹筭（誥）〉員（云）：「佳（唯）尹身（允）及康（湯），咸又（有）一惠（德）。」【緇衣 3】                    |

該表直接應用於辭例檢索程式。

**表 F：段落辭例表**

段落辭例表包含上博簡的所有辭例，且辭例以段落形式出現。該表包含三項數據：F1 是每條段落辭例的嚴格隸定文本，F2 是每條段落辭例的通行字文本，F3 是每條短句子辭例的嚴格隸定並加括注的完整文本。

| ID | F1                                                    | F2                                                                   | F3                                                                                                                                                         |
|----|-------------------------------------------------------|----------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------|
|    | 子曰為上可齊而智也為下可順而皆也則君不怨丌臣 = 不或於君咎員咎人孳 = 丌義不弋尹筭員佳尹身及康咸又一惠 | 子曰：為上可望而知也，為下可述而識也，則君不疑其臣，臣不惑於君。《詩》云：「淑人君子，其儀不忒。」〈尹誥〉云：「唯尹允及湯，咸有一德。」 | 子曰：為上可齊（望）而智（知）也，為下可順（述）而皆（識）也，則君不怨（疑）丌（其）臣 = （臣，臣）不或（惑）於君。《咎（詩）》員（云）：【緇衣 2】「咎（弔—淑）人孳 = （君子），丌（其）義（儀）不弋（忒）。」〈尹筭（誥）〉員（云）：「佳（唯）尹身（允）及康（湯），咸又（有）一惠（德）。」【緇衣 3】 |

該表亦直接應用於辭例檢索程式。

表 G：論著目錄表

論著目錄表收錄與上博簡字詞考釋相關的所有論文的數據，包括每篇論文的作者、題目、討論篇目、涉及竹簡、相關字詞、出處、地址等信息。如下表所示：

| ID | G1  | G2                 | G3         | G4    | G5    | G6                 | G7              |
|----|-----|--------------------|------------|-------|-------|--------------------|-----------------|
| 1  | 董珊  | 也說〈容成氏〉「強弱不辭讓」句    | 2008年3月16日 | 2f36  | 絢謁    | 復旦大學出土文獻與古文字研究中心網站 | fdu/2008/1.mht  |
| 2  | 范常喜 | 《上博二·容成氏》簡36「治謁」補議 | 2008年3月15日 | 2f36  | 絢謁    | 簡帛網                | bsm/2008/1.mht  |
| 3  | 陳劍  | 上博簡〈容成氏〉的拼合與編連問題   | 2003年1月9日  | 2fall | A B Z | 簡帛研究               | jbyj/2003/1.mht |

G1是論文作者。G2是論文題目。G3是論文發表時間。G4是論文討論篇目、涉及竹簡，如2f36表示論文涉及《上博二》第f篇〈容成氏〉的第三十六簡，2fall表示論文討論〈容成氏〉全篇。G5是論文討論相關內容，「絢謁」表示論文涉及《上博二·容成氏》簡36中「絢謁」兩字的釋讀，A表示論文涉及〈容成氏〉全篇內容，B表示論文涉及編聯問題，Z表示論文涉及綴合問題。G6表示論文出處。G7表示附屬文件中該論文文件的位置。

該表直接應用於論著檢索程式，且與高級檢索程式相關聯。

## （二）數據庫附屬文件

數據庫附屬文件包括字形圖片、論著文檔、字形索引、釋文文本、造字文件等。

### 1. 字形圖片

字形圖片包括上博簡全部字形的圖片，是按一定標準由上博簡掃描圖片切割出

來的。這些圖片按資料庫的要求存放於特定的文件夾中，每幅圖片的存放位置與數據庫表 C「全字形數據表」所錄圖片連結地址一致。

## 2. 論著文檔

論著文檔是我們收集並按一定要求處理的上博簡研究論著的電子文檔。這些電子文檔也按資料庫的要求存放於特定的文件夾中，每篇論著位置與數據庫表 G「論著目錄表」所錄論著文檔地址一致。需要特別說明的是，由於涉及知識產權問題，數據庫附屬文件「論著文檔」僅能內部使用，即使以後資料庫出版發行，這部分內容也無法公開。

## 3. 字形索引

字形索引是一個 Excel 文檔，該文檔由數據庫表 A「字頭數據表」導出，用以檢索字頭。例如，資料庫的使用者如果不清楚資料庫中到底是用了「宮」還是「宮」作為隸定字形，可以用字形索引搜索一下部首「宀」，或構件「口」，或上古音「見母」、「冬部」，或《說文》卷數「7」，很容易就能找到隸定字形。這在查找生僻字形，特別是當前使用的各種輸入法難以打出的字形時尤為有用。

## 4. 釋文文本

釋文文本是一個 Word 文檔，收錄上博簡的全部釋文。此文檔是使用資料庫的重要參考文件，也可以作為索引使用。例如，想利用資料庫查找某個字的所有字形圖片，不知道隸定字形，但是知道某篇的某支簡上有這個字，或是知道此字的前後語句，只要查看該文檔便可知曉隸定字形。

## 5. 造字文件

造字文件由 Windows 系統自帶的 True Type 造字程式導出。造字文件與特定字體連結後，「上博簡字詞全編資料庫」所有程式和文件的字體才能正常顯示。資料庫默認的連結字體是「方正超大字符集」。

### （三）檢索程式

檢索程式分為四個相對獨立的檢索程式：字形檢索程式、詞義檢索程式、辭例

檢索程式和論著檢索程式。每個程式都與數據庫和數據庫附屬文件連接，通過不同指令搜索數據庫，完成所需檢索操作。

## 1. 字形檢索程式

字形檢索程式分為基本功能和高級功能兩部分。

基本功能通過直接檢索數據庫表 C「全字形數據表」實現。可以按字檢索 C1 區，亦可按詞檢索 C2 區，皆顯示所有匹配項目。檢索得到的匹配項目均可提取並複製到指定位置。

高級功能通過數據庫表 A、表 B、表 C 配合完成。

例如，檢索所有帶有「口」的字形：程式將先檢索表 A 的 A2 項，通過 A2 項找到所有匹配的 A1 項，再將檢得 A1 數據依次代入表 C，尋找匹配的 C1 項，輸出結果。

再例如，檢索所有部首為「竹」、可讀為「書」之字的字形：程式將先檢索表 A 的 A4 項，通過 A4 找到所有匹配的 A1 項，再將檢得 A1 數據依次代入表 B，尋找 B1=A1 而 B2=「書」的項目，將匹配的 B1 項代入表 C，尋找匹配的 C1 項，輸出結果。

又例如，檢索《上博二·民之父母》中所有讀為「無」之字：程式將先在表 C 中按 C3 數據鎖定〈民之父母〉區間，然後檢索 C2 項，尋找匹配項目，輸出結果。

總之，以 A、B、C 三表互相配合，檢索程式可實現古文字考釋中所需的多種高級檢索功能，在很大程度上滿足了研究者的需要。

## 2. 詞義檢索程式

詞義檢索程式分為基本功能和高級功能兩部分。

基本功能通過直接檢索數據庫表 D「詞義數據表」實現。可以按字檢索 C1 區，亦可按詞檢索 C2 區，皆顯示所有匹配項目。檢索得到的匹配項目均可提取，並粘貼到指定位置。

表 D 亦可與表 A、表 B 搭配使用，完成高級檢索功能。基本原理大致與先前之「字形檢索程式」相同。

### 3. 辭例檢索程式

辭例檢索程式分為基本功能和高級功能兩部分。

基本功能通過直接檢索數據庫表 E「短句辭例表」和表 F「段落辭例表」實現。可以按字檢索 E1、F1 區，亦可按詞檢索 E2、F2 區，皆顯示所有匹配項目的 E3、F3 區。檢索得到的匹配項目均可提取，並粘貼到指定位置。

表 E、表 F 亦可與表 A、表 B 搭配使用，完成高級檢索功能。基本原理大致與先前之「字形檢索程式」相同。

表 E 與表 F 分別為兩種不同的辭例檢索需求而編製。表 E 辭例短小簡潔，但有時準確理解上下文意或有困難；表 F 辭例具體完整，但冗餘信息較多。檢索程式同時設計兩種表單，供檢索者按需要自行選擇。

### 4. 論著檢索程式

論著檢索程式通過檢索數據庫表 G「論著目錄表」完成檢索功能。表 G 中各項數據皆可檢索。程式可進行條件檢索，例如，要檢索某學者某時間段內撰寫關於某具體篇目的論文，只須將檢索條件一一勾選即可。檢索結果包含論著文件的連結，點擊即可打開閱讀。

## 二、製作流程

製作「上博簡字詞全編資料庫」，可將整個項目拆解成五項任務：論著蒐集、字形圖片切割、釋文製定、數據庫錄入和檢索程式開發。五項任務可同步推進，由中心師生協力完成。

### （一）論著蒐集

論著蒐集要將所見上博簡字詞論著按數據庫表 G「論著目錄表」規定的格式錄入 Excel 文檔。同時將這些論著電子化，格式可以為 pdf、doc 或 mht 等。把電子文檔置於文件夾特定位置，與表 G 的 G7 項保持一致。G5 項「相關字詞」有時會涉及沒有編碼的生僻字，暫以圖片代替，待任務 C「釋文製作」的造字工作完成後，再以有編碼的新造字替換圖片。由於當前上博簡研究是學術熱點，相關研究成果還

在不斷湧現，論著蒐集一部分導入資料庫後，工作仍須繼續。

## （二）字形圖片切割

字形圖片切割需要先将上博簡圖版以 600dpi 的分辨率掃描，再用特定軟體將所有字形一一切下，按規定的方式編號，並存入相應的文件夾。反覆校對，將錯誤的、不清晰的圖版檢出並替換。最後，用特定軟體將這些圖片統一調整尺寸，以符合檢索程式需要。待檢索程式和數據庫製作完成，圖片文件便可與之對接。

## （三）釋文製定

首先由數名學者共同製作初步釋文，每人負責上博簡的部分篇目。初步釋文完成後，由一位學者按嚴格規定的釋文體例統稿，製成正式釋文。正式釋文為白文，不含任何注釋。經過反覆校對，確保正確率後，利用 Windows 系統自帶的 True Type 程式完成生僻字符的造字工作。造字連結方正超大字符集字體。造字完成後，再進行數輪校對工作，釋文製定工作才告完成。

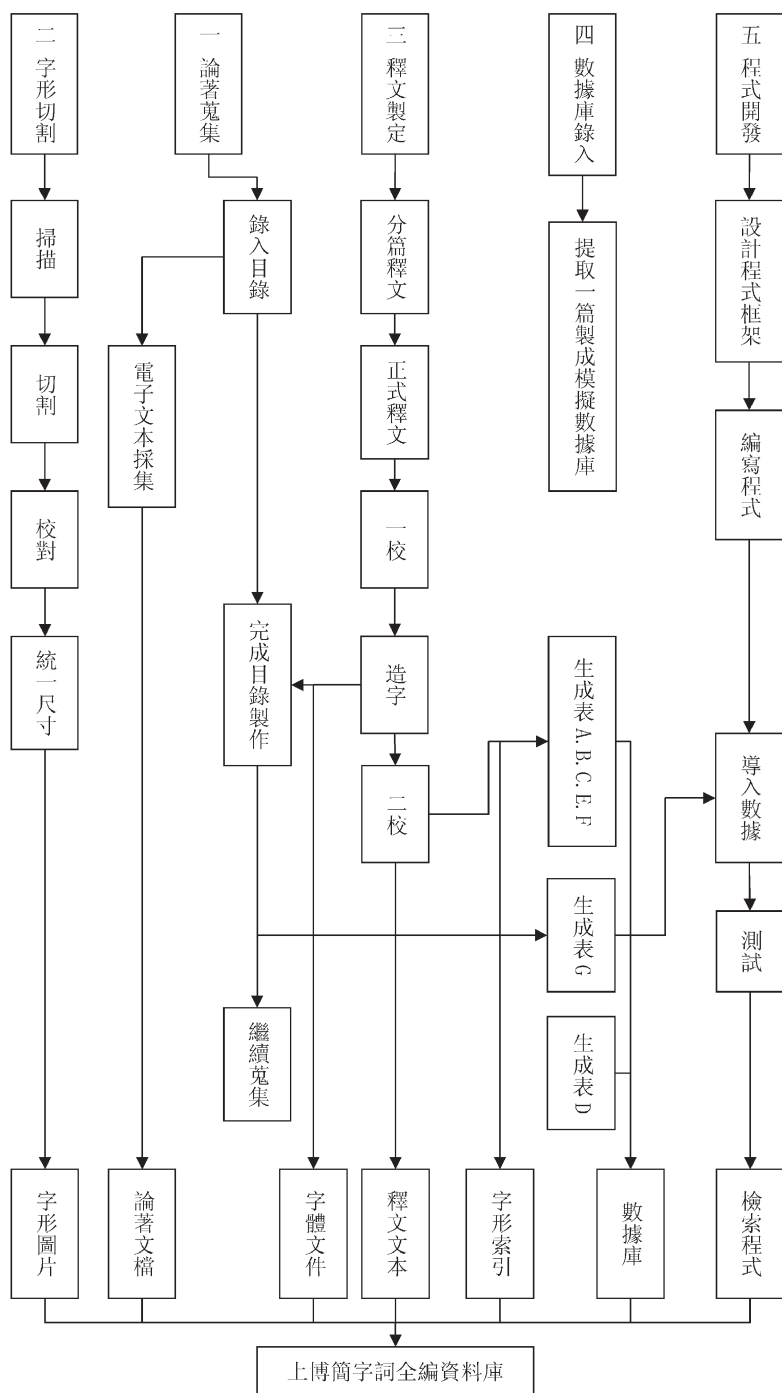
## （四）數據庫錄入

任務 C 完成後，隨即進行數據庫錄入工作。將任務 A 的 Excel 文檔導入數據庫，製成數據庫表 G。將任務 C 的正式釋文複製若干份，分別製成數據庫表 A、表 B、表 C、表 E 和表 F。製作表單的過程中需要進行多次篩選、切割、導入和排序工作。部分工作可以依靠電腦程式自動完成，但有大量工作仍須手動操作。表 D「詞義數據表」是例外，此表要在「上博簡字詞全編」詞條撰寫完成後才能將詞義數據導入。

## （五）檢索程式開發

檢索程式由設計者製定程式框架，由程式員完成編寫。開發過程中，學者與程式員要充分交流，使編定的程式符合學術研究的要求。程式中要留下充足的端口，以備今後添加新的功能。這要求程式的設計者要對相關的學術研究有所了解，對程式的應用前景有正確的估計。在數據庫完成之前，檢索程式應已先行編寫完成，並已利用一個小規模數據庫調試過，確保程式能夠正常運行。數據庫製作完成後，將數據庫和數據庫附屬文件與檢索程式對接，再反覆調試，確保程式和數據庫均無錯誤。

「上博簡字詞全編資料庫」製作流程圖如下：



上博簡資料尚未全部發表，目前資料庫數據僅收集至上博六。由於導入數據庫的操作較為複雜，為節約人力，暫未收錄部分將待上博簡材料全部發表、釋文全面修訂後，再一併導入數據庫。

由於《上博簡字詞全編》撰寫過程中有些高級檢索功能並非亟需，而且程式設計和編寫人員精力有限，因此部分高級檢索程式尚未寫定。我們已在原程式上留下充足的端口，一旦有需要，這些功能隨時可以添加。

### 三、設計優點

「上博簡字詞全編資料庫」的設計力求在學術性和實用性方面有所創新。

#### （一）學術性

為使「上博簡字詞全編資料庫」具有較高的學術水準，最重要的是製定高質量的上博簡釋文。釋文要求嚴格遵循既定體例，充分吸收學術界最新研究成果。

在製作「上博簡字詞全編資料庫」前，我們先制定了「上博簡釋文體例」，對釋文格式、竹簡編號規定了統一格式，對衍文、脫文、合文、重文、通用字、通假字、誤字、殘字等各種文字現象嚴格規定了表示方法。該體例還確定文字隸定的原則，並對某些文字的隸定標準以列舉形式作詳細要求。〈體例說明〉長達十九頁，超過萬字，詳細規定了諸多細節的具體操作。釋文統稿工作也由一名對釋文體例極為熟悉的研究人員獨力完成。這樣細緻的工作使釋文前後標準統一，保證了資料庫數據的正確率。

上博簡釋文還充分吸收了學術界的最新研究成果。這些研究成果包括竹簡分篇、編聯綴合、字詞考釋等許多方面。竹簡分篇方面，如根據〈談談《上博（五）》的竹簡分篇、拼合與編聯問題》<sup>2</sup>一文，將《上博五·鮑叔牙與隰朋之諫、競建內之》兩篇合為一篇，將〈弟子問〉簡 22、簡 3 改入〈君子為禮〉。編

<sup>2</sup> 陳劍：〈談談《上博（五）》的竹簡分篇、拼合與編聯問題〉，簡帛網，2006 年 2 月 19 日，網址：[http://www.bsm.org.cn/show\\_article.php?id=204](http://www.bsm.org.cn/show_article.php?id=204)。

聯方面，如根據〈上博簡〈姑成家父〉一個編聯組位置的調整〉<sup>3</sup>一文，將《上博五·姑成家父》的簡 6-8 調整到簡 2 之後。綴合方面，如根據李銳、福田哲之、梁靜、陳劍等學者的意見<sup>4</sup>，將〈孔子見季桓子〉的多支殘簡綴合。字詞考釋方面吸收的成果則更不勝枚舉。

為保證檢得數據全面完整，資料庫將與上博簡相關的原始材料盡數收入。如〈鬼神之明〉簡 2 背面文字、〈性情論〉所附殘簡、〈中弓〉等三篇的附簡（其中〈內禮〉附簡據福田哲之說改入〈季庚子問於孔子〉<sup>5</sup>）等皆標注特殊編號並導入數據庫。另文增發的《周易·小畜》九五「孚膚女（如）」三字<sup>6</sup>，作為無字形圖片數據收入。香港中文大學文物館藏戰國簡十枚，雖然未藏於上海博物館，但與上博簡為同一批材料，其中數支已經找到篇目歸屬，因此也一併收入。

資料庫嚴格區分「字」和「詞」。在數據庫結構設計上，我們將字和詞設定成兩類不同的數據。在檢索功能上，我們設定字和詞分別檢索，並且通過高級檢索程式實現字詞關係的多功能檢索——通過高級程式的設定，資料庫可以反映與某字或某詞密切相關的字詞關係網絡。這是「上博簡字詞全編資料庫」的一大特色。

## （二）實用性

如前文「檢索程式」一節介紹的，「上博簡字詞全編資料庫」提供的檢索功能

<sup>3</sup> 沈培：〈上博簡〈姑成家父〉一個編聯組位置的調整〉，簡帛網，2006 年 2 月 22 日，網址：[http://www.bsm.org.cn/show\\_article.php?id=219](http://www.bsm.org.cn/show_article.php?id=219)。

<sup>4</sup> 李銳：〈〈孔子見季桓子〉新編（稿）〉，簡帛網，2007 年 7 月 11 日，網址：[http://www.bsm.org.cn/show\\_article.php?id=606](http://www.bsm.org.cn/show_article.php?id=606)。李銳：〈〈孔子見季桓子〉重編〉，簡帛網，2007 年 8 月 22 日，網址：[http://www.bsm.org.cn/show\\_article.php?id=703](http://www.bsm.org.cn/show_article.php?id=703)。福田哲之：〈〈孔子見季桓子〉1 號簡的釋讀與綴合〉，簡帛網，2007 年 8 月 6 日，網址：[http://www.bsm.org.cn/show\\_article.php?id=689](http://www.bsm.org.cn/show_article.php?id=689)。梁靜：〈〈孔子見季桓子〉校讀〉，簡帛網，2008 年 3 月 4 日，網址：[http://www.bsm.org.cn/show\\_article.php?id=798](http://www.bsm.org.cn/show_article.php?id=798)。陳劍：〈《上博（六）·孔子見季桓子》重編新釋〉，復旦大學出土文獻與古文字研究中心網站，2008 年 3 月 22 日，網址：[http://www.gwz.fudan.edu.cn/SrcShow.asp?Src\\_ID=383](http://www.gwz.fudan.edu.cn/SrcShow.asp?Src_ID=383)，又載復旦大學出土文獻與古文字研究中心編：《出土文獻與古文字研究》第二輯（上海：復旦大學出版社，2008 年），頁 160-187。

<sup>5</sup> 福田哲之：〈上博四〈內禮〉附簡、上博五〈季康子問於孔子〉第十六簡的歸屬問題〉，簡帛網，2006 年 3 月 7 日，網址：[http://www.bsm.org.cn/show\\_article.php?id=271](http://www.bsm.org.cn/show_article.php?id=271)。

<sup>6</sup> 李零：〈讀上博楚簡《周易》〉，《中國歷史文物》，2006 年第 4 期。

較多，很多功能是为古文字研究者量身訂做的。資料庫的檢索界面設計也考慮到了學者研究需要。字形檢索程式提供字形預覽，可方便比對檢索出的各種字形。檢索程式還有提取數據功能，可將檢索出的數據複製粘貼。數據庫附屬文件中的字形索引、釋文文本為資料庫的使用提供了完備的索引。

這裏特別介紹一下論文檢索程式的一個創新點。現在各種論文檢索數據庫，其關鍵詞都是由論文作者自己設定。倘學者們對某一特定字形的隸定和釋讀不同，各篇論文的關鍵詞很可能差別極大。況且古文字字形很多都沒有編碼，即便知道隸定字形也無法檢索。這給蒐集古文字字詞考釋論文造成了很大的麻煩。我們在錄入數據庫表 G「論著目錄表」時，G5「相關字詞」一項是以我們制定釋文中的隸定為準的，這大大方便了論文檢索。以上博簡《周易》簡 14 、〈鬼神之明、融師有成氏〉簡 8  為例，學界有釋「適」、「夏」、「遼」、「𡗗」、「逕」、「通」、「𡗗」、「速」等多種觀點<sup>7</sup>，若按這些文章的意見分別標注 G5「相關字詞」，就無法一次性檢索到所有論文。我們在數據庫中統一以「𡗗」標注，檢索者只需通過字形索引和釋文索引找到「𡗗」字形，再檢索「相關字詞」，便可得到所有涉及此字的討論文章。

「上博簡字詞全編資料庫」的數據庫結構和檢索程式有較強的通用性，進一步完善即可作為開發出土文獻資料庫的通用平臺。二〇一〇年十二月，復旦大學出土文獻與古文字研究中心申報成功國家社科基金重大項目「馬王堆漢墓簡帛字詞全編」。有了開發「上博簡字詞全編資料庫」的基礎，我們在研製「馬王堆漢墓簡帛字詞全編資料庫」時就可省去不少相關工作了。

為檢測數據庫的開發效率，二〇一一年一月《清華大學藏戰國竹簡（壹）》公布後，我們嘗試按照「上博簡字詞全編資料庫」中「字形檢索」和「辭例檢索」的開發要求製作「清華簡字形辭例數據庫」。結果前後僅用了七天時間就完成了數據庫的製作，可見數據庫的結構設計較合理，開發效率較高。

<sup>7</sup> 相關研究綜述可參看陳劍：〈釋「琮」及相關諸字〉，收入其所著：《甲骨文文考釋論集》（北京：線裝書局，2007 年），頁 273-316。

## 四、經驗總結

在「上博簡字詞全編資料庫」的製作過程中，我們也遇到了很多困難，汲取了不少經驗教訓，這是值得與大家分享的。

在選取數據庫軟體時，我們發現有的數據庫軟體僅支持 GBK 編碼，有的數據庫軟體存在每單元字符數量限制等。在編寫程式時發現，有些程式模板不支持方正超大字符集，或是支持擴展 A 區卻不支持擴展 B 區編碼，必須通過一定技術手段才能解決大字符集的搜索和顯示問題。這都給我們的開發工作帶來了不少困難。所以在製作數據庫檢索程式時，必須先定好數據庫軟體，並熟悉相關軟體的性質。在估算程式的開發時間和開發成本時，要對這些技術障礙有充分估計。

目前「上博簡字詞全編資料庫」數據庫有幾萬條數據，數據量還是比較大的。這就要求編寫程式時儘力優化算法，防止出現檢索程式運行緩慢等情況。檢索結果輸出時，也要考慮到有的匹配項目很多，需要分屏顯示，否則容易造成當機。設定檢索項目之間的邏輯關係時，何處使用「包含」，何處使用「等於」，亦須深思熟慮。編寫程式時，要對數據庫結構了然於胸，以省去反覆核對之苦，也防止了頻出錯誤。

「上博簡字詞全編資料庫」製作過程中的最大教訓是，一定要在檢索程式上添加修改數據庫的功能。由於資料庫數據量很大，錄入偶爾出錯不可避免。許多錯誤很容易在檢索操作時發現。可惜我們在前期設計中未考慮到在檢索程式上添加數據庫修改功能，因此只能另做記錄，再到數據庫文件中逐個尋找、修改，浪費了很多時間。我們會在研發「馬王堆漢墓簡帛字詞全編資料庫」時增添修改功能，以便隨時修改校正有關數據。

## 五、開發前景

由於上博簡資料尚未完全公開，數據庫無法完整錄入，加之高級檢索程式未完全開發好，因此目前「上博簡字詞全編資料庫」仍用於內部參考，暫不能公開發布。

接下來一段時間，我們會利用製作「馬王堆漢墓簡帛字詞全編資料庫」的機

會，將檢索程式修改完善。檢索程式中需要重點改進的有三個方面：

（一）如前文所述，舊有程式中沒有修改數據庫功能，這給資料庫的校對和使用帶來不小麻煩。我們計劃在新程式增添修改功能，讓使用者在操作界面中可以修改數據庫。

（二）資料庫現有造字文件是以附屬文件形式與方正超大字符集連結的。如今字體文件的開發技術已經成熟，我們希望下一步能製作出獨立的字體文件。

（三）目前的數據庫各項數據是分開顯示的，下一步擬開發出字形和辭例能同步顯示的界面，讓研究者使用更便捷。

（四）現有資料庫是單機運行的。我們計劃下一步開發出網路版檢索程式。網路版檢索程式更有利於學術信息的傳播。而且，具備修改功能的網路版檢索程式更有利於提高數據庫的校對效率——我們可以為學界同仁開通修改權限，大家在使用數據庫時，發現錯誤便可隨手訂正，訂正信息先以校記形式出現，只要有多名權限用戶認同便可正式進入數據庫——這可大大保證資料庫的學術質量。當然，這一設想仍有不少細節需要仔細推敲，程式主創人員正在努力構思，希望近期能設計出一個完善的方案。

我們認為資料庫最理想的發布狀態是：我們僅保留數據庫及其附屬文件中學術信息的知識產權，而將檢索程式原代碼向學界公開，讓更多的學者可利用程式製作其他出土文獻材料的資料庫。

希望我們對「上博簡字詞全編資料庫」的介紹能為學界帶來一些有用的信息。歡迎學界同仁向我們提供更多想法，使我們能將資料庫不斷完善。

看校補記：

編輯先生依排版要求，對拙文原稿的字體、標點和術語進行調整，因此文中所引數據庫的字體和標點與實際情況有所不同。術語的更改主要有：「程序」改為「程式」，「鏈接」改為「連結」，「計算機」改為「電腦」。特此說明。

## 附：檢索程式部分功能演示



檢索字形「亓」



檢索讀為「凡」之字

檢索題目含有「三德」  
的論文



檢索含有詞「安」的辭例



提取檢索結果粘貼到 Word 文檔