

《癸辛雜識》舊版造字轉碼說明

中研院資訊所文獻處理實驗室
中研院史語所漢籍電子文獻工作小組
2008/6/13 陳維君 製作

一、《癸辛雜識》一書(癸辛雜識.xml)使用舊版造字 216 個，字頻 981 次，詳如附件一。這 216 個造字中，155 個可轉成 Windows XP 能顯示的字，字頻 584 次；另外 61 個字必須轉成構字式，字頻 397 次。

二、附件一的造字分析表說明如下：

甲、編號：Big5 造字空間為 6217 個，編號由 1 到 6217。

乙、造字：舊版造字

丙、字頻(txt)：造字在「.txt」文件的出現次數

丁、字頻(xml)：造字在「.xml」文件的出現次數



戊、Big5：造字的 Big5 碼



己、Unicode：造字所對應的 Unicode 碼


庚、WinXP：造字在 Windows XP 的對應字形




辛、構字式：Windows XP 若無對應字形，則改採用構字式

壬、備註凡例：

1、校對問題，未確定字形，必須人工取代：在舊版漢籍電子文獻中即存在的錯字，因校對時的疏漏而未更正，持續留存在新版漢籍電子文獻中，必須人工一一比對修改，如編號 309「」，可參考附件二未確定字形「」手動取代表。

2、異體字問題：新版漢籍考量到使用者檢索及使用時的便利性，將用字原則改為除專詞等特殊情形之外，一律改用標準字呈現。如編號 4448 的「糸 厂 墨 」係「纏」字之異體，故以「纏」字取代。

3、待造字：Unicode 及漢字構形資料庫皆未收錄的舊漢籍造字，正在等待補造字中，所以「造字」欄空白無法看到字形。如編號 2992 的「 女」。

三、《癸辛雜識》一書(癸辛雜識.xml)中未確定字形「」共出現 14 次，詳細取代內容可參考附件二未確定字形「」手動取代表。附件二未確定字形「」手動取代表的欄位說明如下：

甲、檔案原文摘錄：摘錄檔案(癸辛雜識.xml)中包含該造字的

- 文句段落，其中紅字底線的部分，為需要手動取代的原文。
- 乙、手動取代結果：變更過後的檔案內容，其中藍字底線的部分，為修改後的結果。
- 丙、備註：註明相同文句出現的次數(未標明者為只出現一次)及其他需註明事項。


附件一、《癸辛雜識》造字分析表















編號	造字	字頻 (txt)	字頻 (xml)	Big5	Unicode	WinXp	構字式	備註
12	陡	1	1	FA4B	28E97		冫 厶 走	陡，異體字問題。
296	厶	47	47	FBEC			厶	
297	厶	66	66	FBED			厶	
298	厶	16	16	FBEE			厶	
299	形	18	18	FBEF			形	
300	口	18	18	FBF0			口	
309	?	14	14	FBF9			?	未確定字形，人工取代，參考附件二。
314	乙	11	11	FBFE			乙	
317	少	22	22	FC42			少	
379	支	7	7	FCA2			支	
538	商	7	7	FDA4			商	
656	音	1	1	FE5B			音	
713	尸	1	1	FEB6			尸	
717	足	1	1	FEBA			足	

編號	造字	字頻 (txt)	字頻 (xml)	Big5	Unicode	WinXp	構字式	備註
725	丨	1	1	FEC2			丨	
729	业	1	1	FEC6			业	
765	冫	2	2	FEEA			冫	
768	艹	11	11	FEED			艹	
770	扌	1	1	FEEF			扌	
775	亠	1	1	FEF4			亠	
777	宀	1	1	FEF6			宀	
778	丶	2	2	FEF7			丶	
780	冫	2	2	FEF9			冫	
783	亠	2	2	FEFC			亠	
784	亻	4	4	FEFD			亻	
1103	曼	2	2	9043	4731		谷△曼	
1364		1	1	91CD			宀△圭	罍，異體字問題。
1518	侏	1	1	92CA	20201		侏△侏△侏	侏，異體字問題。
1700		3	3	93E3			竹△隹△心	𦵏△𦵏△𦵏，異體字問題。
2092	夂	13	13	9672	20646		夂△几	
2188	離	1	1	96F4	4366		離△離	
2189	瞿	1	1	96F5	23C3D		毛△瞿	瞿，異體字問題。

編號	造字	字頻 (txt)	字頻 (xml)	Big5	Unicode	WinXp	構字式	備註
2204	𦉳	1	1	9745	20890		𠃉牙女厶	𦉳，異體字問題。
2368	蟻	5	5	984C	25AC3		立𠃉義	
2373	𦉳	1	1	9851	410D		𠃉𠃉昔	
2426	頴	1	1	98A8	4ABB		弁𠃉頁	
2624		1	1	99D1			舟𠃉義	𦉳，異體字問題。
2992		1	1	9C48			𠃉𠃉女	待造字。
3235		1	1	9DC0			𠃉門𠃉𠃉	𦉳，異體字問題。
3800		4	4	815F			𠃉𠃉𠃉	𦉳，異體字問題。
4132		58	58	8371			血𠃉𠃉	𦉳，異體字問題。
4345		1	1	84CB			口𠃉𠃉	𦉳，異體字問題。
4347	𦉳	3	3	84CD			虫𠃉越	
4387	𦉳	1	1	84F5	20E57		口𠃉哲	
4432		2	2	8563			𠃉𠃉𠃉𠃉	纏，異體字問題。
4448		2	2	8573			𠃉𠃉𠃉𠃉	纏，異體字問題。
4455	𦉳	2	2	857A			𠃉𠃉刀	
4743	𦉳	1	1	8760	22C72		𠃉𠃉𠃉	
4844		2	2	87E7			𠃉𠃉凡	𦉳，異體字問題。
4856	𦉳	1	1	87F3	3CB2		𠃉𠃉毛	

編號	造字	字頻 (txt)	字頻 (xml)	Big5	Unicode	WinXp	構字式	備註
5075	彳	1	1	8972			彳	
5130	藜	1	1	89CB	2A353		攷𠂇麥	
5176		10	10	89F9			氵𠂇冗	沉，異體字問題。
5194		1	1	8A4C			𠂇扌 畝且 𠂇	擅，異體字問題。
5218		1	1	8A64			𠂇宀 寬、𠂇	寬，異體字問題。
5225	卹	3	3	8A6B	460F		血𠂇卩	卹，異體字問題。
5230		1	1	8A70			𠂇土 畝且 𠂇	壇，異體字問題。
5242	𠂇	5	5	8A7C	2064F		𠂇𠂇几	
5591	杞	4	4	8CC1	233CC		木𠂇巳	杞，異體字問題。
5746		1	1	8DBF			每𠂇系	繇，異體字問題。
5852	宀	1	1	C6CB		宀		

附件二、《癸辛雜識》未確定字形「」手動取代表

檔案原文摘錄	手動取代結果	備註
徽宗 	徽宗   一 𠂇 一 𠂇	待造字
孝宗 	孝宗   尸 𠂇 𠂇 𠂇	待造字
寧宗 	寧宗   一 𠂇 𠂇	待造字
理宗 	理宗   𠂇 𠂇 𠂇、𠂇	待造字
稗海本、四庫本作「  」		圖形文字，無法用構字式表達，暫時維持原狀，本句出現2次。
稗海本、四庫本並作「  」		圖形文字，無法用構字式表達，暫時維持原狀，本句出現4次。

檔案原文摘錄	手動取代結果	備註
太祖❓		圖形文字，無法用構字式表達，暫時維持原狀
太祖元押</qn>❓		
太宗❓		
高宗❓		