

# 《八家後漢書輯》舊版造字轉碼說明

中研院資訊所文獻處理實驗室  
中央研究院語言所文獻語料小組  
2009/01/07 丁致伶

轉碼主要工作是把檔案中的舊版造字轉換成 Windows XP 能支援的 Unicode 字形，Unicode 目前共收錄漢字 70194 個字，而 XP 只能支援 20902 個字（詳如表一），不支援之字將以構字式表達。例：造字編號 1005 的「堧」字，Unicode 編碼是 213F1，由於 XP 並不支援，仍需使用構字式「土𠂔累」。

表一、Unicode 的字數及編碼區段

Unicode	新增字數	新增編碼區段	總字數	WinXP
1.1 版	20902	4E00-9FFF	20902	支援
3.0 版	6582	3400-4DFF	27484	不支援
3.1 版	42710	20000-2A6D6	70194	不支援

## 一、舊版造字轉碼分析：

《八家後漢書輯》使用舊版造字 108 個，字頻 497 次，這 108 個造字中，95 個可轉成 Windows XP 能顯示的字，字頻 472 次；另外 13 個字必須轉成構字式，字頻 25 次。

轉碼完成製作轉碼分析表，請參考附件一《八家後漢書輯》轉碼分析表，欄位說明如下：

- (一) 編號：Big5 造字空間為 6217 個，編號由 1 到 6217。
- (二) 造字：舊版造字。
- (三) 頻次：舊版造字在文件的出現次數。
- (四) Big5：造字的 Big5 碼。
- (五) Unicode：造字所對應的 Unicode 碼。

(六) WinXP：造字在 Windows XP 的對應字形。

(七) 構字式：Windows XP 無法對應字形改用構字式。

(八) 備註凡例：備註欄中記錄轉碼後字形及修改原因，凡例如下：

1. **異體字問題**：為了使用者查詢和使用的方便，在處理異體字時最主要的方式是以標準字取代，除非是專有名詞或特殊情形，如：人名、地名等。例：造字編號 3796 的「决」字，是「決」的異體字，以標準字「決」取代。
2. **錯字，以程式全部取代**：與原書字形不符，並查詢教育部異體字字典確認非異體字後，皆歸類為錯字，以程式取代為正確字形。例：檔案中使用造字編號 2736 的「𡗗」字，而原書使用之字為「鯁」，因此「𡗗」為錯字，以程式取代為正確字形「鯁」。
3. **部份使用不同字形，手動修改**：檔案中為同一舊版造字，校對時卻發現原書使用兩個以上不同的字形，可能部份舊版造字為錯字或原書使用兩種異體字形，導致同一舊版造字轉碼時對應到兩個以上的字形；將一部份手動修改，剩餘的頻次則用程式取代成對應的 Unicode 或構字式。例：造字編號 4115 的「藁」字頻次為 2，依書將 1 字手動修改為「藁」，剩下 1 字轉為「藁」。手動修改部份請參考附件二《八家後漢書輯》手動取代表。
4. **Unicode 字型呈現差異**：Unicode 字型與舊版造字有些微差異，但只是字體風格差異，實際上仍為同一個字，因此仍取 Unicode 字型。如編號 3978 的「猪」字，Unicode 字型呈現為「猪」，實際上仍為同一字。
5. **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 4747 的「𠂔𠂔童」。

#### 附件一、《八家後漢書輯》轉碼分析表

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
1005	堁	1	8F7E	213F1		土厶累	
1044	儻	1	8FC7	511B	儻		
1067	碁	2	8FDE	68CA	碁		
1077	稟	2	8FE8	7A3E	稟		
1133	瓌	1	9061	74CC	瓌		
1381	穉	1	91DE	7A3A	穉		
1383	穰	1	91E0	7A45	穰		
1513	糧	1	92C5	7CAE	糧		糧，異體字問題。
1530	紉	3	92D6	7CFA	紉		
1565	繇	3	92F9	7DDC	繇		
1668	臈	1	93C3	81C8	臈		
1937	詢	1	9574	8A3D	詢		
1981	讐	4	95C2	8B90	讐		讐，異體字問題。
2472	靄	1	98D6	9741	靄		
2610	鯁	3	99C3	9C53	鯁		
2736	崮	1	9AA4	5CDC	崮		鯁，錯字，以程式全部取

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
							代。
3780	儁	38	814B	5101	儁		
3787	冢	9	8152	51A2	冢		
3791	况	8	8156	51B5	况		况，異體字問題。
3794	湊	1	8159	51D1	湊		湊，異體字問題。
3796	決	11	815B	51B3	決		決，異體字問題。
3804	効	1	8163	52B9	効		
3805	勅	4	8164	52C5	勅		
3814	却	2	816D	5374	却		
3829	咏	1	817C	548F	咏		
3840	坂	1	81A9	5742	坂		
3849	獎	1	81B2	596C	獎		獎，異體字問題。
3850	姪	4	81B3	59D9	姪		
3858	冤	7	81BB	5BC3	冤		冤，異體字問題。

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
3869	币	1	81C6	5E00	币		
3877	迺	1	81CE	5EFC	迺		
3891	慙	5	81DC	6159	慙		
3930	憵	1	8244	6901	憵		
3966	烟	1	8268	70DF	烟		
3974	牀	6	8270	7240	牀		
3975	牕	1	8271	7255	牕		
3976	犁	1	8272	7282	犁		
3978	猪	8	8274	732A	猪		Unicode 字型 呈現差異。
3984	瑠	1	827A	7460	瑠		
3991	攜	1	82A3	64D5	攜		
3996	疎	5	82A8	758E	疎		
4010	臯	3	82B6	7690	臯		臯，異體字問 題。
4012	鼓	2	82B8	76B7	鼓		
4016	着	6	82BC	7740	着		
4034	穉	13	82CE	7A49	穉		
4035	穽	1	82CF	7A7D	穽		

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
4036	窰	1	82D0	7AB0	窰		
4038	竈	5	82D2	7AC8	竈		
4045	箒	2	82D9	7B92	箒		
4055	糝	1	82E3	7C83	糝		
4080	羣	59	82FC	7FA3	羣		
4081	翱	2	82FD	7FFA	翱		
4090	脚	3	8347	811A	脚		
4091	臯	1	8348	81EF	臯		
4107	葱	2	8358	8471	葱		
4115	藁	2	8360	85C1	藁		部份使用不同字形「藁」，手動修改。
4132		71	8371			血𠄎𠄎	眾，異體字問題。
4134	衛	31	8373	885E	衛		衛，異體字問題。
4139	袴	4	8378	88B4	袴		
4146	覩	1	83A1	89A9	覩		睹，異體字問題。

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
4167	賚	1	83B6	8CEB	賚		
4177	駢	1	83C0	8F27	駢		
4184	迹	5	83C7	8FF9	迹		
4190	邳	1	83CD	90C4	邳		
4193	鈎	3	83D0	920E	鈎		鈎，異體字問題。
4226	凜	2	83F1	98E1	凜		
4232	隲	2	83F7	96B2	隲		隲，異體字問題。
4242	鬪	3	8442	9B2D	鬪		
4250	鷓	1	844A	9C0C	鷓		
4256	鷄	5	8450	9DC4	鷄		雞，異體字問題。
4259	麇	1	8453	9E95	麇		
4278	亘	1	8466	4E98	亘		
4307	罵	1	84A5	99E1	罵		罵，異體字問題。
4378	隲	1	84EC	28F65		陟△馬	隲，異體字問題。

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
4458	勑	1	857D	52CC	勑		
4460	勅	27	85A1	52D1	勅		
4747		1	8764			卩厶童	待造字。
4768	妬	3	8779	59AC	妬		妒，異體字問題。
4805	瀆	1	87C0	51DF	瀆		
4825	闢	1	87D4	28D66		門厶俞	
4834	熨	2	87DD	3DE1		形目巨 火	
4907	橐	1	8867	69D6	橐		
5034	弃	1	8949	5F03	弃		棄，異體字問題。
5065	忼	1	8968	5FFC	忼		
5088	夔	1	89A1	244C9		難厶火	
5116	侏	1	89BD	4F37	侏		
5118	敷	3	89BF	22FAD		專厶攴	
5144	禱	1	89D9	8920	禱		
5153	瓊	3	89E2	74C6	瓊		
5154	璿	9	89E3	7468	璿		

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
5156	丕	5	89E5	353B		丕	
5190	褊	1	8A48	463F		衤 ㄩ 屈	
5225	邨	1	8A6B	460F		血 ㄩ 卩	
5292	粵	1	8AD0	65C9	粵		
5355	𡗗	1	8B50	20AE4		𡗗	
5434	湟	1	8BC1	6E7C	湟		
5437	并	22	8BC4	5E77	并		
5439	栢	1	8BC6	6822	栢		
5459	缺	2	8BDA	21647		垂 ㄩ 夬	
5537	离	1	8C69	79BC	离		
5558	龕	2	8C7E	9E84	龕		
5585	寓	3	8CBB	3762		宀 厶 禹	
5588	恠	1	8CBE	6061	恠		
5619	繚	3	8CDD	470C		繚	
5666	汚	1	8D4D	6C5A	汚		汚，異體字問題。
5704	氈	2	8D73	6C0A	氈		氈，異體字問題。
5705	廐	4	8D74	5ED0	廐		廐，異體字問

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
							題。
5706	敘	3	8D75	654D	敘		敘，異體字問題。

## 二、手動取代表說明：

《八家後漢書輯》檔案中需手動修改的情況分為以下幾種：

(一) 《八家後漢書輯》檔案中之舊版造字，因字形部份不同而無法全部以程式取代，皆手動修改成Windows XP能顯示的字或構字式，並製作手動取代表，請參考附件二。欄位說明如下：

1. 編號：Big5 造字空間為 6217 個，編號由 1 到 6217。
2. 造字：舊版造字字形。
3. 取代：實際使用字形與舊版造字不同時，於此欄列出。
4. 檔案原文摘錄：摘錄檔案《八家後漢書輯》中的文句段落，其中紅字底線的部分，為需要手動取代的原文。
5. 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。
6. 備註：記錄修改相同辭句的次數或其他事項。

### 附件二、《八家後漢書輯》手動取代表

編號	造字	取代	檔案原文摘錄	手動取代結果	備註
4115	藁	藁	<u>葬</u> 而	<u>藁</u> 葬而	

(二) 《八家後漢書輯》檔案中出現「●」字形，共計 9 字，比對原書後以 Unicode 字形或構字式手動取代，並製作「●」字手動

取代表，請參考附件三。欄位說明如下：

1. 頁碼：「●」字形所在位置之原書頁碼。
2. 檔案原文摘錄：摘錄《八家後漢書輯》檔案有「●」字形的文句。
3. 原書字形：「●」在原書中之字形，以此字取代檔案原文的「●」。
4. 備註：記錄修改相同詞句的次數或其他事項。

附件三、《八家後漢書輯》「●」字手動取代表

頁碼	檔案原文摘錄	原書字形	備註
P. 45	罷\qd●\Qd\qa 鄭	𠄎𠄎貝卩𠄎	待造字
P. 121	功曹岑●捕子禁	脛	
P. 157	●素清白	躰	
P. 202	出則●升	兩𠄎隻	
P. 203	尹●傳	之△翟	有 2 處
P. 203	尹●字公可	之△翟	
P. 271	長●弓弩勁兵	金𠄎矛	
P. 726	烏●欠頤	口𠄎隹	待造字