

《撰集百緣經》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
2009/03/13 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 4856 的「𪛗」字，Unicode 編碼是 3CB2，由於 XP 並不支援，仍需使用構字式「疊𪛗毛」。

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：

《撰集百緣經》使用舊版造字 46 個，字頻 343 次，這 46 個造字中，41 個可轉成 Windows XP 能顯示的字，字頻 273 次；另外 5 個字必須轉成構字式，字頻 70 次。

轉碼完成製作轉碼分析表，請參考附件一《撰集百緣經》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 4193 的「鈎」字，是「鈎」的異體字，以標準字「鈎」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 4561 的「𠄎」字，而原書使用之字為「鹿」，查詢教育部異體字字典確認「𠄎」並非「鹿」之異體字，因此歸為錯字，以程式取代為正確字形「鹿」。
 3. 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 4773 的「姉」字，Unicode 字型呈現為「姉」，實際上仍為同一字。

附件一、《撰集百緣經》轉碼分析表

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
888	捺	59	8EC8	3B88		木𠄎奈	
1530	紉	1	92D6	7CFA	紉		
1668	臙	1	93C3	81C8	臙		
1800	菌	4	94AA	8597	菌		
1981	讐	3	95C2	8B90	讐		讐，異體字問

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
							題。
2042	疏	3	9640	47FD		足厶充	
2250	忉	1	9773	38FC		忄厶刃	
2524	饒	47	994B	994D	饒		
2631	鷄	1	99D8	9D44	鷄		
2931		1	9BCA			此厶白	咎，錯字，以程式全部取代。
3039	覩	18	9C77	89A9	覩		睹，異體字問題。
3769	標	1	8140	3BF9		標厶寸	
3805	勅	58	8164	52C5	勅		
3808	勾	1	8167	5303	勾		
3817	廁	1	8170	53A0	廁		廁，異體字問題。
3850	姪	1	81B3	59D9	姪		
3860	尠	1	81BD	5C20	尠		
3891	慙	2	81DC	6159	慙		
3975	牕	1	8271	7255	牕		

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
4012	鼓	1	82B8	76B7	鼓		鼓，異體字問題。
4041	豎	8	82D5	7AEA	豎		
4089	脇	1	8346	8107	脇		
4176	輒	2	83BF	8F19	輒		
4184	迹	2	83C7	8FF9	迹		
4193	鈎	3	83D0	920E	鈎		鈎，異體字問題。
4195	鉢	65	83D2	9262	鉢		
4226	淪	2	83F1	98E1	淪		
4515	呪	4	85D8	546A	呪		
4561		1	8647			𠃉 [〃] 厂 曲比 [□]	𠃉，錯字，以程式全部取代。
4716	塗	2	8745	57FF	塗		
4738	睺	2	875B	777A	睺		
4768	妬	8	8779	59AC	妬		妒，異體字問題。
4773	姉	1	877E	59C9	姉		Unicode 字型

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
							呈現差異。
4856	疊	6	87F3	3CB2		疊𠂇毛	
5132	愜	5	89CD	6198	愜		
5133	愜	1	89CE	6199	愜		
5179	駟	2	89FC	99C8	駟		
5300	鬪	4	8AD8	9B2A	鬪		
5339	𦍋	1	8B40	59E7	𦍋		
5363	暉	5	8B58	66CE	暉		
5408	絳	2	8BA7	7D4D	絳		
5447	躡	4	8BCE	8E83	躡		
5496	𦍋	2	8C40	7F82	𦍋		
5588	愜	2	8CBE	6061	愜		
5714	瓮	1	8D7D	74EB	瓮		
5742	澁	1	8DBB	6F81	澁		

二、手動取代表說明：

《撰集百緣經》檔案中出現「●」字形，共計27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，請參考附件二。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《撰集百緣經》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。

附件二、《撰集百緣經》「●」字手動取代表

頁碼	檔案原文摘錄	原書字形	備註
p. 203	●馥遍王舍城	馥	
p. 205	過度群●類	𠂇 𠂇朋	待造字
p. 211	●中・出七寶蓮華	𠂇 𠂇肉	
p. 213	●末六兩	𠂇 𠂇鳥	
p. 216	復從索●	𠂇 𠂇 𠂇 丹 𠂇	待造字
p. 218	終不●惜	恪	有 4 處
p. 218	目不●惜	恪	
p. 219	●裂太子	國 𠂇 瓜	待造字
p. 219	都不●惜	恪	
p. 226	●虛遊行	𠂇 𠂇 欠	
p. 226	羸瘦●弱	𠂇	
p. 227	名曰●羅	𠂇 𠂇 遮	
p. 229	●塞三惡道	閑	
p. 230	於牕●中	月 𠂇 庸	
p. 242	不相侵●	𠂇 𠂇 欠	
p. 243	●譬故	𠂇 𠂇 登	
p. 250	尋即命●	𠂇 𠂇 冬	

頁碼	檔案原文摘錄	原書字形	備註
p. 250	●必開我腹	歹 ㄨ 冬	
p. 252	為●●・飲食	齒 ㄨ 齊 口 ㄨ 制	口 ㄨ 制為待造字
p. 252	見為搏●	口 ㄨ 制	待造字
p. 254	水尋●出	漂 ㄨ 寸	
p. 255	還國與●	外 ㄨ 男	待造字
p. 256	●甥舍利弗	外 ㄨ 男	待造字