

《菩薩本緣經》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
2009/03/11 丁玫伶

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 1533 的「劇」字，Unicode 編碼是 34DF，由於 XP 並不支援，仍需使用構字式「皮𠂔 丩」。

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：

《菩薩本緣經》使用舊版造字 28 個，字頻 67 次，這 28 個造字中，26 個可轉成 Windows XP 能顯示的字，字頻 63 次；另外 2 個字必須轉成構字式，字頻 4 次。

轉碼完成製作轉碼分析表，請參考附件一《菩薩本緣經》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 4203 的「隣」字，是「鄰」的異體字，以標準字「鄰」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」。

附件一、《菩薩本緣經》轉碼分析表

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
981	豺	1	8F66	72B2	豺		
1237	稍	1	90EB	77DF	稍		
1533	劇	3	92D9	34DF		皮𠄎 丩	
1926	覓	1	9569	8994	覓		
2325	潤	1	97E0	77A4	潤		
3804	効	1	8163	52B9	効		
3805	勅	8	8164	52C5	勅		
3814	却	3	816D	5374	却		

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
3940	鬱	1	824E	6B1D	鬱		
3944	氷	1	8252	6C37	氷		
3966	烟	2	8268	70DF	烟		
3984	瑠	1	827A	7460	瑠		
4041	豎	5	82D5	7AEA	豎		
4089	脇	1	8346	8107	脇		
4109	葛	1	835A	84AD	葛		
4146	覩	2	83A1	89A9	覩		睹，異體字問題。
4195	鉢	1	83D2	9262	鉢		
4203	隣	8	83DA	96A3	隣		鄰，異體字問題。
4260	麇	3	8454	9E9E	麇		
4515	呪	4	85D8	546A	呪		
4561		1	8647			𠃉"厂 曲比𠃉	麤，錯字，以程式全部取代。
4738	睺	1	875B	777A	睺		
4739	鞞	1	875C	97DB	鞞		

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
4768	妬	4	8779	59AC	妬		妒，異體字問題。
5002	鞫	1	88E8	4A6D		鞫	
5339	𦉳	2	8B40	59E7	𦉳		
5447	躡	3	8BCE	8E83	躡		
5453	湴	5	8BD4	23D40		形: 日 工: □	湴，異體字問題。

二、手動取代表說明：

《菩薩本緣經》檔案中需手動修改的情況分為以下幾種：

(一)《菩薩本緣經》檔案中出現「●」字形，共計 15 字，比對原書後以 Unicod 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《菩薩本緣經》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、《菩薩本緣經》「●」字手動取代表

頁碼	檔案原文摘錄	原書字形	備註
p. 55	化民不●萬姓	𠂔	
p. 57	可以蚊●盡海底	此𠂔束	改為此𠂔束

頁碼	檔案原文摘錄	原書字形	備註
p. 57	真珠車●馬璫珊	𠃉王巨木𠃉	
p. 60	犯而橫見●	枉	
p. 61	即將良醫而為●	言𠃉爾	
p. 61	欲●虛我之境	扞	
p. 63	瞋氣●鬱煙	火𠃉孛	
p. 64	闍●神仙	少𠃉兔	
p. 64	帝釋身為●羊	羊𠃉互	
p. 65	鐵●諸鳥挑	口𠃉隼	待造字
p. 65	惡獸之所●食	國𠃉瓜	待造字
p. 65	鈎●・鞅絆拘執	登𠃉斤	
p. 66	蹄如車●	𠃉王巨木𠃉	
p. 68	其●金剛	此𠃉束	改為此𠃉束
p. 69	慰●其心	𠃉ノ友𠃉	改為沃

(二)《菩薩本緣經》檔案中，對於舊版造字以外發現的錯字，以手動修改，並製作錯字修改表，請參考附件三。欄位說明如下：

1. 檔案原文摘錄：摘錄檔案《菩薩本緣經》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。
2. 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。
3. 備註：記錄修改相同詞句的次數或其他事項。

附件三、《菩薩本緣經》缺字外錯字修改

檔案原文摘錄	手動取代結果	備註
常墮坑坎 <u>胃</u> 索	常墮坑坎 <u>𠵼</u> <u>𠵼</u> 索	待造字