

《悲華經》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
2009/03/30 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 4206 的「榆」字，Unicode 編碼是 3BD3，由於 XP 並不支援，仍需使用構字式「木𠂇翁」。

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：

《悲華經》使用舊版造字 39 個，字頻 298 次，這 39 個造字中，38 個可轉成 Windows XP 能顯示的字，字頻 296 次；另外 1 個字必須轉成構字式，字頻 2 次。

轉碼完成製作轉碼分析表，請參考附件一《悲華經》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 4768 的「妬」字，是「妒」的異體字，以標準字「妒」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 3978 的「猪」字，Unicode 字型呈現為「猪」，實際上仍為同一字。

附件一、《悲華經》轉碼分析表

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
1237	稍	1	90EB	77DF	稍		
2506	飭	1	98F8	991D	飭		
2631	鷄	1	99D8	9D44	鷄		
2835	梯	3	9B48	632E	梯		
3805	勅	16	8164	52C5	勅		

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
3814	却	13	816D	5374	却		
3817	廁	6	8170	53A0	廁		廁，異體字問題。
3870	帑	1	81C7	5E0B	帑		
3884	忽	2	81D5	6031	忽		
3940	爵	5	824E	6B1D	爵		
3944	冰	1	8252	6C37	冰		
3978	猪	3	8274	732A	猪		Unicode 字型呈現差異。
4001	痘	1	82AD	7602	痘		
4012	鼓	2	82B8	76B7	鼓		鼓，異體字問題。
4041	豎	4	82D5	7AEA	豎		
4089	脇	2	8346	8107	脇		
4090	脚	2	8347	811A	脚		
4146	覩	5	83A1	89A9	覩		睹，異體字問題。
4173	駝	3	83BC	8EAD	駝		
4176	輒	1	83BF	8F19	輒		

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
4193	鈎	2	83D0	920E	鈎		鈎，異體字問題。
4195	鉢	21	83D2	9262	鉢		
4206	榆	2	83DD	3BD3		木 ㄩ 翁	
4455	劒	1	857A			僉 ㄩ 刀	劒，異體字問題。
4489		17	85BE			口 厶 方	号，異體字問題。
4515	呪	7	85D8	546A	呪		
4561		1	8647			形 ㄣ 厂 曲 比 口	鹿，錯字，以程式全部取代。
4706	瞪	3	86FA	96A5	瞪		
4738	睺	17	875B	777A	睺		
4768	妬	22	8779	59AC	妬		妒，異體字問題。
5133	憇	13	89CE	6199	憇		
5188	捫	2	8A46	62EA	捫		
5300	鬪	22	8AD8	9B2A	鬪		

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
5339	奸	1	8B40	59E7	奸		
5453	涅	89	8BD4	23D40		形: 日 工: □	涅，異體字問題。
5542	鑠	1	8C6E	93C1	鑠		
5588	恡	1	8CBE	6061	恡		
5702	厨	1	8D71	53A8	厨		
5742	澁	2	8DBB	6F81	澁		

二、手動取代表說明：

《悲華經》檔案中出現「●」字形，共計57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《悲華經》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、《悲華經》「●」字手動取代表

頁碼	檔案原文摘錄	原書字形	備註
p. 168	馬瑙為●	革𦰇茸	有2處
p. 168	菩薩得授記●	荊	

頁碼	檔案原文摘錄	原書字形	備註
p. 170	●帝目帝郁	羊厶面厶毛	待造字
p. 170	●帝 賒蜜致	面厶毛	
p. 170	研●阿婆婆	閱	
p. 170	婆波●婆叉	吡	
p. 170	路伽●達舍	少厶兔	
p. 170	斯提●頻提	昔	
p. 170	崩伽摩●尼	少厶兔	
p. 172	慳●見	恪	
p. 172	慳●者	恪	
p. 172	不慳●法	恪	
p. 173	帝叉●浮彌	商厶衣	待造字
p. 173	●毘地遮迦	少厶兔	
p. 173	阿●摩都	少厶兔	
p. 177	行水●舉鉢	收	
p. 177	為●蓮華臺	革厶茸	
p. 183	菩薩受佛記●	荊	
p. 184	●阜大小	培	
p. 187	頗梨琉璃車●	𠃉王巨木𠃉	
p. 194	憍慢慳●	恪	有 2 處
p. 194	汝名為阿●	閱	

頁碼	檔案原文摘錄	原書字形	備註
p. 194	阿●白佛言	閱	
p. 194	●阜土沙礫	培	
p. 194	便利之患痰●	瘥	
p. 194	慳●嫉妒	恪	有 7 處
p. 195	如來讚阿●言	閱	
p. 195	●如來	閱	有 4 處
p. 195	阿●菩薩	閱	有 5 處
p. 195	●菩薩	閱	
p. 196	第十一王子●	菅	
p. 203	●·令十力	荊	
p. 204	授其記●	荊	
p. 206	●在四流	漂△寸	
p. 207	●阜谿谷溝壑	培	
p. 209	相得受記●	荊	
p. 211	貪著慳●	恪	
p. 211	乃至慳●	恪	
p. 216	譬如●	甘	
p. 220	●而退轉者	荊	
p. 223	香積以為●	**△積	
p. 227	痰●風寒	瘥	