

《普曜經》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
2009/04/01 丁玟伶

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 2042 的「䟽」字，Unicode 編碼是 47FD，由於 XP 並不支援，仍需使用構字式「𠩺𠩺𠩺」。

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：

《普曜經》使用舊版造字 56 個，字頻 372 次，這 56 個造字中，50 個可轉成 Windows XP 能顯示的字，字頻 362 次；另外 6 個字必須轉成構字式，字頻 10 次。

轉碼完成製作轉碼分析表，請參考附件一《普曜經》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 4134 的「衛」字，是「衛」的異體字，以標準字「衛」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. 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 5341 的「𠄎」。
 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 4495 的「𠄎」字，Unicode 字型呈現為「𠄎」，實際上仍為同一字。

附件一、《普曜經》轉碼分析表

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
1091	灑	4	8FF6	58CD	灑		
1400	窓	4	91F1	7A93	窓		
1642	脉	2	93A9	8109	脉		

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
1981	讐	1	95C2	8B90	讐		讎，異體字問題。
2042	䟽	3	9640	47FD		足厶流	
2253	齧	1	9776	4DA7		齒厶堯	
2506	飭	3	98F8	991D	飭		
2524	饒	2	994B	994D	饒		
3197		1	9D78			𠄎𠄎 首寸	𠄎，錯字，以程式全部取代。
3805	勅	23	8164	52C5	勅		
3814	却	12	816D	5374	却		
3824	叙	1	8177	53D9	叙		
3850	姪	1	81B3	59D9	姪		
3860	𡗗	8	81BD	5C20	𡗗		
3940	𡗗	2	824E	6B1D	𡗗		
3966	烟	3	8268	70DF	烟		
3975	𡗗	4	8271	7255	𡗗		Unicode 字型呈現差異。
3980	𡗗	2	8276	7328	𡗗		

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
3984	瑠	1	827A	7460	瑠		
4001	瘥	3	82AD	7602	瘥		
4016	着	1	82BC	7740	着		
4038	竈	1	82D2	7AC8	竈		
4041	豎	16	82D5	7AEA	豎		
4085	耻	1	8342	803B	耻		
4089	脇	4	8346	8107	脇		
4090	脚	4	8347	811A	脚		
4105	菓	1	8356	83D3	菓		
4134	衛	43	8373	885E	衛		衛，異體字問題。
4146	覩	112	83A1	89A9	覩		睹，異體字問題。
4176	輒	14	83BF	8F19	輒		
4184	迹	2	83C7	8FF9	迹		
4195	鉢	29	83D2	9262	鉢		
4203	隣	13	83DA	96A3	隣		
4206	榆	2	83DD	3BD3		木ㄩ俞	
4237	髡	1	83FC	9AE0	髡		

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
4278	亘	1	8466	4E98	亘		
4447	捏	1	8572	63D1	捏		
4495	麩	4	85C4	9EA8	麩		Unicode 字型 呈現差異。
4496	段	1	85C5	53DA	段		
4515	呪	7	85D8	546A	呪		
4561		3	8647			𠄎"厂 曲比𠄎	麤，錯字，以 程式全部取 代。
4723	塶	1	874C	58C3	塶		
4768	妬	5	8779	59AC	妬		妒，異體字問 題。
4856	氈	2	87F3	3CB2		疊厶毛	
4904	尅	1	8864	5C05	尅		
5098	寤	1	89AB	7AB9	寤		
5133	憇	1	89CE	6199	憇		
5290	携	2	8ACE	643A	携		
5300	鬪	8	8AD8	9B2A	鬪		
5341		1	8B42			彡厶壁	待造字。

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
5459	缺	1	8BDA	21647		垂厶夬	
5500	滯	1	8C44	21A7B		形彳 滯	恒，錯字，以程式全部取代。
5542	鑠	1	8C6E	93C1	鑠		
5588	恡	2	8CBE	6061	恡		
5595	塚	2	8CC5	579C	塚		
5809	踰	1	8DFE	8E79	踰		

二、手動取代表說明：

《普曜經》檔案中出現「●」字形，共計33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《普曜經》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、《普曜經》「●」字手動取代表

頁碼	檔案原文摘錄	原書字形	備註
p. 486	顏貌●怡	熙	

頁碼	檔案原文摘錄	原書字形	備註
p. 488	鷹鷓赤●	口 𠂔 隹	待造字
p. 491	無思議億●	亥	
p. 491	寂定·●首	至 𠂔 頁	
p. 491	時晏然●	𠂔 穴 月 未 𠂔	
p. 491	●拭衣服及	扌 𠂔 支	
p. 492	鬼神乾沓●	𠂔	
p. 494	天玉女●嗟	口 𠂔 咨	待造字
p. 496	日光·目●	目 𠂔 妾	
p. 496	鹿●腸	𠂔	
p. 500	慈愍·不●	持	
p. 505	若●縛頭	犛 𠂔 苟	
p. 505	遇此嬰兒●	持	
p. 505	羅刹健沓●	𠂔	
p. 507	樂諸天●嗟	口 𠂔 咨	待造字
p. 507	●突畜生	𠂔 角 豕 𠂔 𠂔	
p. 511	亦無涕●	氵 𠂔 垂	待造字
p. 511	亦不●	口 𠂔 血	待造字
p. 511	●千頭牛	𠂔 吉 爻 牛 𠂔	待造字
p. 515	數億●載劫	亥	
p. 518	溝坑布●●	金 𠂔 疾 鑠	
p. 519	三曰●●	言 𠂔 口	

頁碼	檔案原文摘錄	原書字形	備註
		言𠄎口	
p. 521	設復億●	𠄎	
p. 521	受●為佛釋	𠄎	
p. 521	●天見佛	𠄎	
p. 521	錠光佛時●	𠄎	
p. 523	震動億●國	𠄎	
p. 524	神捷沓●	𠄎	
p. 526	世億●諸	𠄎	
p. 526	億千●天	𠄎	
p. 533	未●·殖福	𠄎𠄎朋	待造字