

《太子瑞應本起經》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
2009/03/12 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：

《太子瑞應本起經》使用舊版造字 30 個，字頻 76 次，這 30 個造字中，29 個可轉成 Windows XP 能顯示的字，字頻 72 次；另外 1 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《太子瑞應本起經》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. 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「𪗇」字，Unicode 字型呈現為「𪗇」，實際上仍為同一字。

附件一、《太子瑞應本起經》轉碼分析表

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
1039	稟	1	8FC2	69C0	稟		
1642	脉	1	93A9	8109	脉		
3039	覩	3	9C77	89A9	覩		睹，異體字問題。
3805	勅	5	8164	52C5	勅		
3814	却	6	816D	5374	却		
3850	姪	1	81B3	59D9	姪		
3884	忽	2	81D5	6031	忽		
3940	爵	2	824E	6B1D	爵		

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
3966	烟	2	8268	70DF	烟		
3980	獾	2	8276	7328	獾		
4001	瘥	1	82AD	7602	瘥		
4041	豎	1	82D5	7AEA	豎		
4085	耻	1	8342	803B	耻		
4089	脇	2	8346	8107	脇		
4090	脚	1	8347	811A	脚		
4176	輒	2	83BF	8F19	輒		
4184	迹	1	83C7	8FF9	迹		
4195	鉢	23	83D2	9262	鉢		
4203	隣	3	83DA	96A3	隣		
4455	劒	1	857A			僉  刃	劒，異體字問題。
4495	麇	3	85C4	9EA8	麇		Unicode 字型呈現差異。
4515	呪	1	85D8	546A	呪		
4558	廐	1	8644	53A9	廐		廐，異體字問題。
4723	墻	1	874C	58C3	墻		

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
4775	妣	1	87A2	59B7	妣		
4856	疊	4	87F3	3CB2		疊厶毛	
5132	憇	1	89CD	6198	憇		
5133	憇	1	89CE	6199	憇		
5442	栴	1	8BC9	67BF	栴		
5753	潛	1	8DC6	6F5C	潛		

二、手動取代表說明：

《太子瑞應本起經》檔案中出現「●」字形，共計10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《太子瑞應本起經》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、《太子瑞應本起經》「●」字手動取代表

頁碼	檔案原文摘錄	原書字形	備註
p. 473	菩薩已得●言	言厶巳	改為記
p. 473	馬寶朱髦●	長厶葛	待造字

頁碼	檔案原文摘錄	原書字形	備註
p. 473	可稱●	言𠄎巳	改為記
p. 474	●紺色	目𠄎妾	
p. 474	鹿●腸・釣鎖	膾	
p. 475	天於●中	窓	
p. 475	肝膽腸●	月𠄎胃	
p. 475	大王何宜●來	扌	
p. 477	設復億●神武備	孩	
p. 477	●天見佛擒魔眾	孩	