

《法句經》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
2009/04/21 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 5364 的「𠄎」字，Unicode 編碼是 239D1，由於 XP 並不支援，仍需使用構字式「歹𠄎凶」。

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：

《法句經》使用舊版造字 25 個，字頻 53 次，這 25 個造字中，23 個可轉成 Windows XP 能顯示的字，字頻 51 次；另外 2 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《法句經》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。

附件一、《法句經》轉碼分析表

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
1032	憊	1	8FBB	2039D		𠄎 𠄎 心 𠄎	
1978	讚	1	95BF	8B83	讚		讚，異體字問題。
3791	況	1	8156	51B5	況		況，異體字問題。
3805	勅	2	8164	52C5	勅		
3814	却	6	816D	5374	却		
3824	叙	2	8177	53D9	叙		
3897	慙	1	81E2	615C	慙		
3940	鬱	2	824E	6B1D	鬱		
3980	獮	1	8276	7328	獮		

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
3996	疎	1	82A8	758E	疎		
4105	菓	1	8356	83D3	菓		
4134	衛	1	8373	885E	衛		衛，異體字問題。
4146	覩	7	83A1	89A9	覩		睹，異體字問題。
4176	輒	1	83BF	8F19	輒		
4184	迹	8	83C7	8FF9	迹		
4193	鈎	1	83D0	920E	鈎		鈎，異體字問題。
4195	鉢	2	83D2	9262	鉢		
4342	倮	1	84C8	502E	倮		
4700	埴	1	86F4	586A	埴		
4904	尅	3	8864	5C05	尅		
5300	鬪	4	8AD8	9B2A	鬪		
5364	殂	1	8B59	239D1		歹 厶 凶	
5588	恠	1	8CBE	6061	恠		
5709	蟀	1	8D78	872F	蟀		
5753	潜	2	8DC6	6F5C	潜		

二、手動取代表說明：

《法句經》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《法句經》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、《法句經》「●」字手動取代表

頁碼	檔案原文摘錄	原書字形	備註
p. 559	如●●水	𣵀𣵀巾 雨𣵀井	
p. 559	人為此●擾	𣵀𣵀參	待造字
p. 563	意使作●	𣵀𣵀苟	
p. 563	退見悔●	悒	
p. 564	妖●見福	薛𣵀女	
p. 570	怨●良人	譖	
p. 570	以●治士	枉	
p. 574	修正不阿●	枉	