

《法句譬喻經》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 888 的「捺」字，Unicode 編碼是 3B88，由於 XP 並不支援，仍需使用構字式「木𠂔奈」。

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：

《法句譬喻經》使用舊版造字 54 個，字頻 291 次，這 54 個造字中，46 個可轉成 Windows XP 能顯示的字，字頻 274 次；另外 8 個字必須轉成構字式，字頻 17 次。

轉碼完成製作轉碼分析表，請參考附件一《法句譬喻經》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 5498 的「𠄎來女厂來[◻]」字，而原書使用之字為「墻」，「𠄎來女厂來[◻]」為錯字，以程式取代為正確字形「墻」。
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、《法句譬喻經》轉碼分析表

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
888	捺	4	8EC8	3B88		木𠄎奈	
1032	憊	1	8FBB	2039D		𠄎𠄎 𠄎 心[◻]	
1978	讚	8	95BF	8B83	讚		讚，異體字問題。

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
2044	踈	3	9642	8E08	踈		
2877	倭	2	9B72			亻 ㄩ 妾	
2969	譖	3	9BF0	8B5B	譖		
3805	勅	17	8164	52C5	勅		
3808	勾	3	8167	5303	勾		
3814	却	28	816D	5374	却		
3824	叙	3	8177	53D9	叙		
3834	嗶	1	81A3	5637	嗶		
3836	喻	1	81A5	564F	喻		
3860	尠	1	81BD	5C20	尠		
3891	慙	2	81DC	6159	慙		
3940	藪	2	824E	6B1D	藪		
3966	烟	1	8268	70DF	烟		
3978	猪	2	8274	732A	猪		Unicode 字型 呈現差異。
3984	瑠	9	827A	7460	瑠		
4028	稟	2	82C8	7980	稟		
4036	窰	4	82D0	7AB0	窰		
4041	豎	4	82D5	7AEA	豎		

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
4080	羣	3	82FC	7FA3	羣		
4089	脇	1	8346	8107	脇		
4090	脚	4	8347	811A	脚		
4105	菓	3	8356	83D3	菓		
4134	衛	64	8373	885E	衛		衛，異體字問題。
4146	覩	9	83A1	89A9	覩		睹，異體字問題。
4173	駝	1	83BC	8EAD	駝		
4184	迹	20	83C7	8FF9	迹		
4193	鈎	6	83D0	920E	鈎		鈎，異體字問題。
4195	鉢	17	83D2	9262	鉢		
4206	榆	1	83DD	3BD3		木ㄩ翁	
4216	鞮	1	83E7	97C8	鞮		
4226	凜	4	83F1	98E1	凜		
4240	髻	1	8440	9AF4	髻		
4278	亘	1	8466	4E98	亘		
4336	阜	1	84C2	278BB		角厶牛	

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
4342	倮	3	84C8	502E	倮		
4515	呪	11	85D8	546A	呪		
4531	攢	1	85E8	6505	攢		
4696	燿	6	86F0	3E0C		火△霍	
4768	妬	1	8779	59AC	妬		妒，異體字問題。
4775	姝	1	87A2	59B7	姝		
4825	闢	1	87D4	28D66		門△俞	
4904	尅	1	8864	5C05	尅		
5002	鞫	1	88E8	4A6D		鞫	
5034	弃	1	8949	5F03	弃		棄，異體字問題。
5300	鬪	12	8AD8	9B2A	鬪		
5453	涅	7	8BD4	23D40		氵日 工□	涅，異體字問題。
5496	羈	2	8C40	7F82	羈		
5498		1	8C42			形來女 厂來□	墻，錯字，以程式全部取代。

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
5588	恠	1	8CBE	6061	恠		
5709	蟀	2	8D78	872F	蟀		
5753	潛	1	8DC6	6F5C	潛		

二、手動取代表說明：

《法句譬喻經》檔案中出現「●」字形，共計21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《法句譬喻經》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、《法句譬喻經》「●」字手動取代表

頁碼	檔案原文摘錄	原書字形	備註
p. 576	如●●水	𠄎𠄎巾 雨𠄎井	
p. 576	●脹臭爛	𠄎	
p. 578	身體●脹	𠄎	
p. 579	安吉終無●橫	𠄎	
p. 579	明月珠●絡	垂	

頁碼	檔案原文摘錄	原書字形	備註
p. 582	●意顧斯	垂	
p. 583	豈使東宮●	門△視	
p. 584	法服●面	疒△土	
p. 584	●而畫眉	疒△土	
p. 590	●泣受報	口△罩	
p. 591	為人所●	扌	
p. 593	●惜經義	恪	
p. 596	塘●墻壁樹	揆	
p. 596	慳貪塘●	揆	
p. 602	何為塘●	揆	
p. 602	長者●然	慴	
p. 602	●爛其臭	縫	
p. 604	如冥●明	門△視	
p. 607	不理臣●民	薛△女	
p. 609	●明·唯願	門△視	