

《百喻經》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
2009/04/20 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：

《百喻經》使用舊版造字 27 個，字頻 85 次，這 27 個造字中，26 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 13 次。

轉碼完成製作轉碼分析表，請參考附件一《百喻經》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 5497 的「衿」字，而原書使用之字為「裏」，「衿」為錯字，以程式取代為正確字形「裏」。

附件一、《百喻經》轉碼分析表

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
1591	羗	8	9354	7F97	羗		
1978	讚	2	95BF	8B83	讚		讚，異體字問題。
2044	踈	1	9642	8E08	踈		
2325	睭	7	97E0	77A4	睭		
2524	饑	1	994B	994D	饑		
3801	剗	1	8160	5257	剗		
3804	効	10	8163	52B9	効		
3805	勅	7	8164	52C5	勅		

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
3814	却	3	816D	5374	却		
3860	尠	1	81BD	5C20	尠		
3940	爵	1	824E	6B1D	爵		
3992	嘗	1	82A4	751E	嘗		
4041	豎	1	82D5	7AEA	豎		
4090	脚	8	8347	811A	脚		
4134	衛	1	8373	885E	衛		衛，異體字問題。
4146	覩	2	83A1	89A9	覩		睹，異體字問題。
4173	躄	3	83BC	8EAD	躄		
4184	迹	1	83C7	8FF9	迹		
4195	鉢	3	83D2	9262	鉢		
4496	段	1	85C5	53DA	段		
4515	呪	2	85D8	546A	呪		
4856	毳	13	87F3	3CB2		疊厶毛	
5300	鬪	2	8AD8	9B2A	鬪		
5453	涅	2	8BD4	23D40		形: 日 工	涅，異體字問題。

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
5497	𡗗	1	8C41	8873	𡗗		裏，錯字，以程式全部取代。
5542	𡗗	1	8C6E	93C1	𡗗		
5714	𡗗	1	8D7D	74EB	𡗗		

二、手動取代表說明：

《百喻經》檔案中出現「●」字形，共計 10 字，比對原書後以 Unicod 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《百喻經》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、《百喻經》「●」字手動取代表

頁碼	檔案原文摘錄	原書字形	備註
p. 543	日中●取牛	𡗗𡗗牛	待造字
p. 543	當頓●取	𡗗𡗗牛	待造字
p. 543	牽牛來欲●取	𡗗𡗗牛	待造字
p. 543	大●薪油	𡗗𡗗積	
p. 549	用稻穀●水	麥𡗗戈	

頁碼	檔案原文摘錄	原書字形	備註
p. 549	純以稻●不如	麥△戈	
p. 555	身毛孰●來	毛△參	
p. 555	而●加神仙	枉	
p. 556	●橫打狗	枉	
p. 556	●螫喪身	𧈧虫扌斤口	