

《眾經撰雜譬喻》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 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：

《眾經撰雜譬喻》使用舊版造字 20 個，字頻 77 次，這 20 個造字中，19 個可轉成 Windows XP 能顯示的字，字頻 75 次；另外 1 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《眾經撰雜譬喻》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. 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、《眾經撰雜譬喻》轉碼分析表

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
1978	讚	5	95BF	8B83	讚		讚，異體字問題。
1981	讐	1	95C2	8B90	讐		讐，異體字問題。
2044	踈	1	9642	8E08	踈		
2505	翻	2	98F7	98DC	翻		
2524	饜	4	994B	994D	饜		
3805	勅	5	8164	52C5	勅		
3814	却	7	816D	5374	却		

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
3955	澗	1	825D	3D4E		氵 厶 間	澗，異體字問題。
3978	猪	3	8274	732A	猪		Unicode 字型呈現差異。
4016	着	1	82BC	7740	着		
4105	菓	2	8356	83D3	菓		
4134	衛	7	8373	885E	衛		衛，異體字問題。
4176	輒	1	83BF	8F19	輒		
4195	鉢	16	83D2	9262	鉢		
4515	呪	9	85D8	546A	呪		
4799	顛	1	87BA	9856	顛		
4856	疊	2	87F3	3CB2		疊 厶 毛	
4904	尅	1	8864	5C05	尅		
5300	鬪	2	8AD8	9B2A	鬪		
5453	涅	6	8BD4	23D40		形 日 工 口	涅，異體字問題。

二、手動取代表說明：

《眾經撰雜譬喻》檔案中出現「●」字形，共計 8 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《眾經撰雜譬喻》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、《眾經撰雜譬喻》「●」字手動取代表

頁碼	檔案原文摘錄	原書字形	備註
p. 532	●乳日得一斗	殼△牛	待造字
p. 532	不復*●牛	殼△牛	待造字
p. 532	入*●牛乳	殼△牛	待造字
p. 532	久不*●	殼△牛	待造字
p. 532	日不*●	殼△牛	待造字
p. 535	是●車	研	
p. 541	人●能却之	虫△轟	
p. 541	●爛消索	虫△轟	