

《寶女所問經》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
2009/05/06 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 2042 的「䟽」字，Unicode 編碼是 47FD，由於 XP 並不支援，仍需使用構字式「𠃉𠃉𠃉」。

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：

《寶女所問經》使用舊版造字 19 個，字頻 100 次，這 19 個造字中，18 個可轉成 Windows XP 能顯示的字，字頻 99 次；另外 1 個字必須轉成構字式，字頻 1 次。

轉碼完成製作轉碼分析表，請參考附件一《寶女所問經》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 1978 的「讚」字，是「讚」的異體字，以標準字「讚」取代。

附件一、《寶女所問經》轉碼分析表

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
1111	瓚	2	904B	749D	瓚		
1536	麤	2	92DC	9E81	麤		
1978	讚	5	95BF	8B83	讚		讚，異體字問題。
1981	讐	1	95C2	8B90	讐		讐，異體字問題。
2042	䟽	1	9640	47FD		足厶充	
2305	鑽	1	97CC	93BB	鑽		
3805	勅	1	8164	52C5	勅		
3824	叙	1	8177	53D9	叙		
3860	𡗗	2	81BD	5C20	𡗗		
3992	嘗	1	82A4	751E	嘗		

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
4012	鼓	1	82B8	76B7	鼓		鼓，異體字問題。
4059	糝	1	82E7	7CC7	糝		
4132		1	8371			血𠄎𠄎	眾，異體字問題。
4146	覩	66	83A1	89A9	覩		睹，異體字問題。
5133	憇	1	89CE	6199	憇		
5300	鬪	5	8AD8	9B2A	鬪		
5316	晉	4	8AE8	664B	晉		晉，異體字問題。
5437	并	3	8BC4	5E77	并		
5666	汚	1	8D4D	6C5A	汚		汚，異體字問題。

二、手動取代表說明：

《寶女所問經》檔案中出現「●」字形，共計18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《寶女所問經》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、《寶女所問經》「●」字手動取代表

頁碼	檔案原文摘錄	原書字形	備註
p. 453	百千億●	亥	
p. 454	七無●	革厶亢	待造字
p. 454	●說。八口演柔軟	革厶印	
p. 455	無慳●塵勞	恪	
p. 456	七財而為貪●	恪	
p. 457	所●惜亦無	恪	
p. 459	●。侍從圍繞	亥	
p. 459	●數如恒沙	亥	
p. 459	炤濟諸群●	利厶人厶水	改為黎
p. 460	九十二●民	亥	
p. 462	施而不●	恪	
p. 464	無所●惜	恪	
p. 465	說●無之	謔	
p. 466	捷沓●阿須	愬	
p. 469	無節●腸如鹿	足厶專	
p. 469	胸前自然●	卍	
p. 472	慳貪愛●	恪	

頁碼	檔案原文摘錄	原書字形	備註
p. 473	及捷沓●	想	