

《遊仙窟》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
2008/5/02 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：

《遊仙窟》使用舊版造字 42 個，字頻 83 次，這 42 個造字中，38 個可轉成 Windows XP 能顯示的字，字頻 79 次；另外 4 個字必須轉成構字式，字頻 4 次。

轉碼完成製作轉碼分析表，請參考附件一《遊仙窟》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 2252 的「乱」字，是「亂」的異體字，以標準字「亂」取代。

附件一、《遊仙窟》轉碼分析表

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
888	榛	1	8EC8	3B88		木 ㄩ 奈	
1226	罰	1	90E0	7F78	罰		
1642	脉	2	93A9	8109	脉		
2044	踈	4	9642	8E08	踈		
2252	乱	8	9775	4E71	乱		亂，異體字問題。
2325	睏	1	97E0	77A4	睏		
2378	颯	1	9856	4AFB		風 ㄩ 日	
2415	鬪	1	987B	95D8	鬪		鬪，異體字問題。
2893	屬	4	9BA4	5C5E	属		屬，異體字問題。
2997	搯	1	9C4D	25081		形 土 ㄥ	

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
						口皿 	
3814	却	6	816D	5374	却		
3840	坂	1	81A9	5742	坂		
3846	壻	1	81AF	58FB	壻		
3851	姁	1	81B4	59F9	姁		
3909	擡	1	81EE	64E1	擡		
3984	瑠	1	827A	7460	瑠		
4012	鼓	1	82B8	76B7	鼓		鼓，異體字問題。
4016	着	4	82BC	7740	着		
4090	脚	4	8347	811A	脚		腳，異體字問題。
4105	菓	5	8356	83D3	菓		
4139	袴	1	8378	88B4	袴		
4146	覩	1	83A1	89A9	覩		睹，異體字問題。
4176	輒	1	83BF	8F19	輒		
4184	迹	3	83C7	8FF9	迹		
4195	鉢	2	83D2	9262	鉢		
4256	鷄	3	8450	9DC4	鷄		雞，異體字問題。

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
4260	麀	1	8454	9E9E	麀		
4267	鼈	1	845B	9F08	鼈		
4307	罵	1	84A5	99E1	罵		罵，異體字問題。
4460	勅	1	85A1	52D1	勅		
4768	妬	1	8779	59AC	妬		妒，異體字問題。
4844		1	87E7			𦉳𠂇凡	𦉳，異體字問題。
4904	尅	1	8864	5C05	尅		
4996	斲	1	88E2	65B5	斲		
5071	忝	2	896E	5FE9	忝		
5074	𠂇	2	8971	5C2B	𠂇		
5076	爇	1	8973	2435C		火𠂇參	
5104	礼	5	89B1	793C	礼		禮，異體字問題。
5290	携	1	8ACE	643A	携		
5588	恠	1	8CBE	6061	恠		
5595	塚	2	8CC5	579C	塚		
5709	蟀	1	8D78	872F	蟀		

二、手動取代表說明：

《遊仙窟》檔案中需手動修改的情況分為以下幾種：

(一) 《遊仙窟》檔案中出現「●」字形，共計 18 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《遊仙窟》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、《遊仙窟》「●」字手動取代表

頁碼	檔案原文摘錄	原書字形	備註
p. 3	腰支細細許，●	目厶兼	
p. 3	●眼子長長馨	目厶舌	
p. 4	率●乾煩	尔	
p. 5	丹青暉●	日厶美	
p. 5	行步絕娃●	女厶犀	
p. 5	香兒取酒。俄●	尔	
p. 6	●無文書	彡厶文	
p. 7	鶉●桂糝	月厶鐵	
p. 7	雉●狸脣	睪	
p. 17	雞●雉臃	月厶載	
p. 17	●棗石榴	木厶栗	
p. 18	率●則天邊雪落	尔	
p. 18	異種●姑	女厶音	
p. 18	万看万種●妍	女厶便	
p. 22	不因針而●	彡厶意	

頁碼	檔案原文摘錄	原書字形	備註
p. 22	閣●頭	幞	
p. 22	鼻裏痠●	疒△帝	待造字
p. 22	嬌鶯弄●人	彡△文	