

《五色石》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
2008/8/21 丁政伶

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 74 的「髻」字，Unicode 編碼是 29B06，由於 XP 並不支援，仍需使用構字式「髻𠄎𠄎」。

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：

《五色石》使用舊版造字 94 個，字頻 2113 次，這 94 個造字中，72 個可轉成 Windows XP 能顯示的字，字頻 1901 次；另外 22 個字必須轉成構字式，字頻 212 次。

轉碼完成製作轉碼分析表，請參考附件一《五色石》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 1513 的「粮」字，是「糧」的異體字，以標準字「糧」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 4145 的「羈」字，Unicode 字型呈現為「羈」，實際上仍為同一字。
3. 依原書修改未定義字形：編號 309 舊版造字「?

附件一、《五色石》轉碼分析表

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
74	髻	1	FAAB	29B06		髻𠄎𠄎	
296	𠄎	4	FBEC			𠄎	
297	𠄎	21	FBED			𠄎	
298	𠄎	5	FBEE			𠄎	
299	𠄎	43	FBEF			𠄎	
300	𠄎	41	FBF0			𠄎	
309	?	19	FBF9			?	两，依原書修

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
							改未定義字形。
639	丿	1	FE4A			丿	
676	𠂇	3	FE6F			𠂇	
690	𠂈	13	FE7D			𠂈	
738	𠂉	2	FECF			𠂉	
742	𠂊	13	FED3			𠂊	
763	𠂋	1	FEE8			𠂋	
765	𠂌	2	FEEA			𠂌	
768	𠂍	20	FEED			𠂍	
770	𠂎	7	FEEF			𠂎	
772	𠂏	1	FEF1			𠂏	
796	邨	1	8E4A	2868D		比𠂐𠂋	
1213	睽	1	90D3	7743	睽		
1226	罰	2	90E0	7F78	罰		
1513	糧	17	92C5	7CAE	糧		糧，異體字問題。
1616	逕	4	936D	284E6		逕△里	
1641	胆	7	93A8	80C6	胆		膽，異體字問題。

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
1642	脉	3	93A9	8109	脉		
1999	贄	1	95D4	8CDB	贄		
2046	踪	10	9644	8E2A	踪		
2409	闕	3	9875	95A7	闕		
2415	鬪	6	987B	95D8	鬪		
2448	隼	1	98BE	96BD	隼		
2524	饒	2	994B	994D	饒		
2826	巡	4	9AFE	5EF5	巡		
2834	攵	3	9B47	6535		攵	
2999	咆	2	9C4F	5523	咆		
3770	个	544	8141	4E2A	个		
3789	冲	4	8154	51B2	冲		冲，異體字問題。
3791	况	29	8156	51B5	况		况，異體字問題。
3792	凉	1	8157	51C9	凉		凉，異體字問題。
3793	减	6	8158	51CF	减		减，異體字問題。
3794	凑	5	8159	51D1	凑		凑，異體字問題。

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
3796	決	13	815B	51B3	決		決，異體字問題。
3814	却	285	816D	5374	却		
3815	𪗇	2	816E	537A	𪗇		
3819	廝	17	8172	53AE	廝		廝，異體字問題。
3824	叙	34	8177	53D9	叙		
3829	咏	40	817C	548F	咏		
3858	寃	40	81BB	5BC3	寃		寃，異體字問題。
3864	峯	2	81C1	5CEF	峯		
3909	擡	15	81EE	64E1	擡		
3938	橈	1	824C	6AC8	橈		
3966	烟	5	8268	70DF	烟		
3974	牀	1	8270	7240	牀		
3979	猫	1	8275	732B	猫		猫，異體字問題。
4016	着	347	82BC	7740	着		
4028	稟	25	82C8	7980	稟		
4041	豎	1	82D5	7AEA	豎		
4042	笋	1	82D6	7B0B	笋		

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
4067	綉	8	82EF	7D89	綉		
4085	耻	4	8342	803B	耻		
4090	脚	26	8347	811A	脚		
4094	館	25	834B	8218	館		
4105	菓	1	8356	83D3	菓		
4132		78	8371			血会𠂇	眾，異體字問題。
4138	裘	1	8377	88A0	裘		
4145	羈	1	837E	898A	羈		Unicode 字型呈現差異。
4146	覩	6	83A1	89A9	覩		睹，異體字問題。
4173	𪗇	1	83BC	8EAD	𪗇		
4184	迹	19	83C7	8FF9	迹		
4193	鈎	15	83D0	920E	鈎		
4203	隣	17	83DA	96A3	隣		鄰，異體字問題。
4242	鬪	2	8442	9B2D	鬪		
4256	鷄	2	8450	9DC4	鷄		雞，異體字問題。
4307	罵	38	84A5	99E1	罵		罵，異體字問題。

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
							題。
4348	惚	1	84CE	506C	惚		
4504	呬	1	85CD	541A	呬		
4702	場	76	86F6	5872	場		
4768	妬	1	8779	59AC	妬		妒，異體字問題。
4785	贗	1	87AC	8D0B	贗		
4798		3	87B9			白△夕	窗，異體字問題。
4817	寶	31	87CC	5BF3	寶		寶，異體字問題。
4940	韵	9	88AA	97F5	韵		
5019	畧	1	88F9			畧	
5034	弃	6	8949	5F03		弃	
5104	礼	2	89B1	793C	礼		禮，異體字問題。
5176		13	89F9			氵△冗	沉，異體字問題。
5290	携	22	8ACE	643A	携		
5300	鬪	1	8AD8	9B2A	鬪		
5434	湟	1	8BC1	6E7C	湟		

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
5549	藁	1	8C75	85F3	藁		
5583	殮	1	8CB9	98F1	殮		
5666	污	6	8D4D	6C5A	污		污，異體字問題。
5697	蕘	1	8D6C	7151	蕘		
5702	厨	1	8D71	53A8	厨		厨，異體字問題。
5753	潜	2	8DC6	6F5C	潜		
5768	潤	5	8DD5	6FF6	潤		