

《醉醒石》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
2008/8/25 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 3935 的「槩」字，Unicode 編碼是 3BA3，由於 XP 並不支援，仍需使用構字式「既 $\text{\textcircled{木}}$ 」。

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：

《醉醒石》使用舊版造字 113 個，字頻 1104 次，這 113 個造字中，96 個可轉成 Windows XP 能顯示的字，字頻 1052 次；另外 17 個字必須轉成構字式，字頻 52 次。

轉碼完成製作轉碼分析表，請參考附件一《醉醒石》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 3789 的「冲」字，是「沖」的異體字，以標準字「沖」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 1219 的「宀厶巢」字，而原書使用之字為「窠」，查詢教育部異體字字典確認「宀厶巢」並非「窠」之異體字，因此歸為錯字，以程式取代為正確字形「窠」。
3. 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。

附件一、《醉醒石》轉碼分析表

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
296	厶	1	FBEC			厶	
297	厶	18	FBED			厶	
298	厶	14	FBEE			厶	
299	形	1	FBEF			形	
300	形	1	FBF0			形	

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
639	丿	4	FE4A			丿	
763	卩	1	FEE8			卩	
768	卩	1	FEE8			卩	
768	卩	1	FEE8			卩	
770	扌	2	FEEF			扌	
1213	睽	1	90D3	7743	睽		
1219		1	90D9			宀會巢	窠，錯字，以程式全部取代。
1641	胆	1	93A8	80C6	胆		膽，異體字問題。
1781	蓋	1	9475	854B	蓋		
2046	踪	5	9644	8E2A	踪		
2157	酌	1	96D5	9167	酌		
2999	嗷	3	9C4F	5523	嗷		
3021	丩	1	9C65			丩	
3098	揼	1	9CD4	63F8	揼		Unicode 字型呈現差異。

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
3163	呈	1	9D56	573C		呈	
3164		1	9D57			土𠂔冊	土𠂔冊，異體字問題。
3255	善	25	9DD4	5584	善		
3787	冢	1	8152	51A2	冢		
3789	冲	3	8154	51B2	冲		沖，異體字問題。
3791	况	53	8156	51B5	况		況，異體字問題。
3792	凉	1	8157	51C9	凉		涼，異體字問題。
3793	减	1	8158	51CF	减		減，異體字問題。
3794	凑	17	8159	51D1	凑		湊，異體字問題。
3796	决	24	815B	51B3	决		決，異體字問題。

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
3799	刼	1	815E	5226	刼		
3804	効	2	8163	52B9	効		
3805	勅	4	8164	52C5	勅		
3814	却	220	816D	5374	却		
3819	廝	2	8172	53AE	廝		廝，異體字問題。
3824	叙	3	8177	53D9	叙		
3825	疊	2	8178	53E0	疊		
3829	咏	1	817C	548F	咏		
3830	咤	1	817D	54A4	咤		
3843	堦	2	81AC	5826	堦		
3849	獎	1	81B2	596C	獎		獎，異體字問題。
3858	寃	41	81BB	5BC3	寃		寃，異體字問題。
3864	峯	2	81C1	5CEF	峯		
3868	卮	1	81C5	5DF5	卮		

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
3883	徧	1	81D4	5FA7	徧		
3886	悞	12	81D7	609E	悞		
3891	慙	1	81DC	6159	慙		
3899	担	3	81E4	62C5	担		
3909	擡	17	81EE	64E1	擡		
3921	暫	1	81FA	6673	暫		
3935	槩	1	8249	3BA3		既 _木	
3938	橈	1	824C	6AC8	橈		
3966	烟	16	8268	70DF	烟		
3974	牀	2	8270	7240	牀		
3977	狗	1	8273	72E5	狗		
3978	猪	1	8274	732A	猪		
3996	疎	9	82A8	758E	疎		
4015	眦	1	82BB	7726	眦		
4016	着	270	82BC	7740	着		
4028	稟	5	82C8	7980	稟		
4034	穉	2	82CE	7A49	穉		

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
4035	窳	2	82CF	7A7D	窳		
4036	窳	1	82D0	7AB0	窳		
4038	窳	2	82D2	7AC8	窳		
4041	豎	3	82D5	7AEA	豎		
4042	笋	1	82D6	7B0B	笋		
4044	筴	1	82D8	7B6F	筴		
4068	綫	5	82F0	7DAB	綫		
4074	縹	2	82F6	7E27	縹		
4080	羣	6	82FC	7FA3	羣		
4089	脇	1	8346	8107	脇		
4090	脚	15	8347	811A	脚		
4094	館	27	834B	8218	館		
4107	葱	1	8358	8471	葱		
4112	蔴	1	835D	8534	蔴		
4114	藜	1	835F	8616	藜		
4115	藁	1	8360	85C1	藁		
4129	虬	3	836E	866C	虬		

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
4132		76	8371			血𠂇𠂇	眾，異體字問題。
4139	袴	2	8378	88B4	袴		
4142	裛	1	837B	8943	裛		裛，異體字問題。
4145	羈	2	837E	898A	羈		Unicode 字型呈現差異。
4167	賈	1	83B6	8CEB	賈		
4173	駝	6	83BC	8EAD	駝		
4176	輒	1	83BF	8F19	輒		
4184	迹	3	83C7	8FF9	迹		
4195	鉢	2	83D2	9262	鉢		
4203	隣	21	83DA	96A3	隣		鄰，異體字問題。
4232	隲	1	83F7	96B2	隲		
4256	鷄	8	8450	9DC4	鷄		雞，異體字問題。

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
4267	鼈	1	845B	9F08	鼈		
4275	义	1	8463	4E49	义		
4307	罵	25	84A5	99E1	罵		罵，異體字問題。
4409	冏	1	854C	518B	冏		冏，異體字問題。
4447	捏	4	8572	63D1	捏		
4588	鮒	1	8662	4C54		魚△攸	
4768	妬	5	8779	59AC	妬		妒，異體字問題。
4779	鞣	1	87A6	97B2	鞣		
4785	贗	1	87AC	8D0B	贗		
4805	瀆	2	87C0	51DF	瀆		瀆，異體字問題。
4808	健	1	87C3	3493		亻△達	
4817	寶	17	87CC	5BF3	寶		寶，異體字問題。

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
4904	尅	4	8864	5C05	尅		
5011	驚	1	88F1	9DD4	驚		
5013	汜	1	88F3			汜	
5034	弃	2	8949	5F03		弃	
5065	忼	2	8968	5FFC	忼		
5174		2	89F7			𠄎刀兂 、𠄎	兔，異體字問題。
5193	號	1	8A4B	8662	號		
5218		15	8A64			𠄎宀寬 、𠄎	寬，異體字問題。
5549	藁	1	8C75	85F3	藁		
5583	殮	2	8CB9	98F1	殮		
5666	污	2	8D4D	6C5A	污		
5697	蕘	1	8D6C	7151	蕘		
5742	澁	1	8DBB	6F81	澁		