

《萬松老人評唱天童覺和尚頌古從容庵錄》 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
2010/03/05 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 1310 的「蠟」字，Unicode 編碼是 45EB，由於 XP 並不支援，仍需使用構字式「虫𠩺麻」。

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：

《萬松老人評唱天童覺和尚頌古從容庵錄》使用舊版造字 137 個，字頻 1129 次，這 137 個造字中，122 個可轉成 Windows XP 能顯示的字，字頻 1088 次；另外 15 個字必須轉成構字式，字頻 41 次。

轉碼完成製作轉碼分析表，請參考附件一《萬松老人評唱天童覺和尚頌古從容庵錄》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 4203 的「隣」字，是「鄰」的異體字，以標準字「鄰」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 1403 的「穴厶兔」字，而原書使用之字為「冤」，查詢教育部異體字字典確認「冤」並非「穴厶兔」之異體字，因此歸為錯字，以程式取代為正確字形「冤」。
 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 3978 的「猪」字，Unicode 字型呈現為「猪」，實際上仍為同一字。

附件一、《萬松老人評唱天童覺和尚頌古從容庵錄》轉碼分析表

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
77	哪	2	FAAE	35BF		口厶耶	
786	蕩	1	8E40	26FC6		++厶碣	
814	墓	1	8E5C	87C7	墓		
893	峩	8	8ECD	5CE9	峩		

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
1067	碁	8	8FDE	68CA	碁		
1151	劄	1	9073	5273	劄		
1310	蠊	3	9175	45EB		虫厶麻	
1387	腎	1	91E4	7796	腎		
1400	窓	9	91F1	7A93	窓		
1403		1	91F4			穴厶兔	兔，錯字，以程式全部取代。
1463	擢	1	9271	651E	擢		
1489	馥	1	92AD	24CD6		番厶去	
1520	嚟	1	92CC	3B20		日厶煞	
1536	麤	8	92DC	9E81	麤		
1559		1	92F3			形广目 无口	廐，異體字問題。
1597	鞞	2	935A	4A6B		革厶兩	
1642	脉	1	93A9	8109	脉		
1772	藜	1	946C	4527		艹厶梨	
1981	讐	5	95C2	8B90	讐		
1999	贊	7	95D4	8CDB	贊		

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
2084	啟	2	966A	6550	啟		
2085	穗	2	966B	69F5	穗		
2148	磨	1	96CC	8B4D	磨		
2232		1	9761			形厂墨 卩口	鄺，異體字問題。
2251	反	1	9774	3906		忄厶反	
2278	銜	1	97B1	936E	銜		
2415	鬪	1	987B	95D8	鬪		鬪，異體字問題。
2454	鍬	1	98C4	28AFC		金厶教	
2501	颯	1	98F3	98C3	颯		
2502	颯	1	98F4	98C8	颯		
2505	翻	2	98F7	98DC	翻		
2520	糕	1	9947	993B	糕		
2534	塋	1	9955	213BA		祝会土	
2592	贓	1	99B1	8D1C	贓		
2671	畱	3	9A41	7550	畱		
2758	献	1	9ABA	732E	献		
2831	啖	1	9B44	55DE	啖		

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
2928		1	9BC7			𠄎王虍 𠄎八目 𠄎	璿，錯字，以程式全部取代。
3770	个	1	8141	4E2A	个		
3801	剗	4	8160	5257	剗		
3804	効	2	8163	52B9	効		
3805	勅	6	8164	52C5	勅		
3814	却	241	816D	5374	却		
3824	叙	4	8177	53D9	叙		
3826	叶	1	8179	53F6	叶		
3831	喆	6	817E	5586	喆		
3834	嗶	1	81A3	5637	嗶		Unicode 字型呈現差異。
3840	坂	1	81A9	5742	坂		
3858	冤	13	81BB	5BC3	冤		冤，異體字問題。
3864	峯	194	81C1	5CEF	峯		
3866	崐	2	81C3	5D10	崐		
3876	廻	6	81CD	5EFB	廻		

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
3884	忽	8	81D5	6031	忽		
3891	慙	1	81DC	6159	慙		
3900	挖	2	81E5	62D5	挖		
3909	擡	12	81EE	64E1	擡		
3924	菴	1	81FD	671E	菴		
3931	椀	4	8245	6900	椀		
3932	椶	1	8246	6936	椶		
3940	爵	3	824E	6B1D	爵		
3944	冰	13	8252	6C37	冰		
3955	澗	1	825D	3D4E		澗 𠄎 澗	澗，異體字問題。
3966	烟	2	8268	70DF	烟		
3977	狗	2	8273	72E5	狗		
3978	猪	3	8274	732A	猪		Unicode 字型呈現差異。
3979	猫	24	8275	732B	猫		
3981	珉	3	8277	73C9	珉		
3984	瑠	5	827A	7460	瑠		
3992	嘗	30	82A4	751E	嘗		

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
3996	疎	53	82A8	758E	疎		
4016	着	1	82BC	7740	着		
4019	矧	2	82BF	77F4	矧		
4028	稟	2	82C8	7980	稟		
4035	穽	3	82CF	7A7D	穽		
4036	窰	1	82D0	7AB0	窰		
4038	竈	9	82D2	7AC8	竈		
4041	豎	24	82D5	7AEA	豎		
4042	笋	6	82D6	7B0B	笋		
4044	筴	1	82D8	7B6F	筴		
4045	箒	6	82D9	7B92	箒		
4081	翹	1	82FD	7FFA	翹		Unicode 字型 呈現差異。
4089	脇	3	8346	8107	脇		
4090	脚	77	8347	811A	脚		
4115	藁	2	8360	85C1	藁		
4121	懼	15	8366	39AC		忄 ㄇ 羅	
4134	衛	2	8373	885E	衛		衛，異體字問 題。

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
4146	覩	6	83A1	89A9	覩		
4173	躡	1	83BC	8EAD	躡		
4184	迹	6	83C7	8FF9	迹		
4195	鉢	30	83D2	9262	鉢		
4203	隣	9	83DA	96A3	隣		鄰，異體字問題。
4216	鞞	1	83E7	97C8	鞞		
4220	穎	2	83EB	9834	穎		
4223	顛	1	83EE	984B	顛		
4250	鰭	1	844A	9C0C	鰭		
4251	鰕	2	844B	9C15	鰕		
4254	鴟	1	844E	9D7E	鴟		
4256	鷄	11	8450	9DC4	鷄		
4267	鼈	6	845B	9F08	鼈		
4271	齧	1	845F	9F69	齧		
4278	亘	7	8466	4E98	亘		
4381	僮	4	84EF	511E	僮		
4389	瑤	1	84F7	746B	瑤		
4428	躡	2	855F	8E70	躡		

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
4432		5	8563			𠄎糸厂 墨𠄎	纏，異體字問題。
4447	捏	9	8572	63D1	捏		
4515	呪	5	85D8	546A	呪		
4519	咲	2	85DC	54B2	咲		
4716	塗	1	8745	57FF	塗		
4738	睺	1	875B	777A	睺		
4739	鞞	1	875C	97DB	鞞		
4763	斃	1	8774	6586	斃		
4778	鞞	2	87A5	97B4	鞞		
4903	吋	1	8863	5C00	吋		
4907	橐	2	8867	69D6	橐		
4950	嶮	1	88B4	5D18	嶮		
4981	鳧	3	88D3	9CEC	鳧		
4996	斲	6	88E2	65B5	斲		
5008	僕	4	88EE	5E5E	僕		
5119	博	1	89C0	613D	博		
5120	虛	2	89C1	8657	虛		
5206	顛	5	8A58	29573		寧𠄎頁	

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
5225	卹	3	8A6B	460F		血厶卩	
5241	鷺	8	8A7B	9D5E	鷺		
5290	携	6	8ACE	643A	携		
5300	鬪	9	8AD8	9B2A	鬪		
5315	汧	3	8AE7	3CC2		氵厶公	
5316	晉	7	8AE8	664B	晉		晉，異體字問題。
5386	塙	1	8B6F	365B		土厶專	
5453	涅	16	8BD4	23D40		形氵日 工	涅，異體字問題。
5472	瓦	10	8BE7		瓦		
5555	迥	8	8C7B	9008	迥		
5559	欸	2	8CA1	6B35	欸		
5621	泪	6	8CDF	6CEA	泪		
5706	敘	1	8D75	654D	敘		敘，異體字問題。
5742	澁	1	8DBB	6F81	澁		
5753	潜	13	8DC6	6F5C	潜		

二、手動取代表說明：

《萬松老人評唱天童覺和尚頌古從容庵錄》檔案中出現「●」字形，共計74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《萬松老人評唱天童覺和尚頌古從容庵錄》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、《萬松老人評唱天童覺和尚頌古從容庵錄》「●」字手動取代表

頁碼	檔案原文摘錄	原書字形	備註
p. 227	護國三●	𠂇 厶麼	有 4 處
p. 227	黃檗●糟	𠂇	有 3 處
p. 228	那堪上曲●木	𠂇 厶見	
p. 231	場●	𠂇 厶麼	有 10 處
p. 232	●和習語	𠂇	
p. 234	捧著楫●	𠂇	
p. 238	●訛	𠂇	有 5 處
p. 239	摸●	𠂇 厶索	有 3 處
p. 239	非慳●	𠂇	
p. 239	如須彌山●	山 厶突	
p. 246	頌三●	𠂇 厶麼	
p. 246	●泥沙	馬 厶展	有 3 處

頁碼	檔案原文摘錄	原書字形	備註
p. 247	未免●	炆△目	
p. 252	背後底●	灑	
p. 252	孔●鞋時	扌△兜	
p. 260	關●	椽	
p. 260	●酒糟	嚏	有 2 處
p. 260	愛罵人·●	嚏	
p. 261	●子·有此	椽	
p. 265	扇子●	足△字	
p. 265	秦時●輅鑽	車△度	有 2 處
p. 266	撒手捨筌●	罍	
p. 266	●者所以在兔	罍	
p. 266	而忘●	罍	
p. 266	者象之●也	罍	
p. 266	忘●捨筌	罍	
p. 267	獨露身●	灑	有 2 處
p. 267	昭與眾●	卜△麼	
p. 271	●●毳毳	監△毛 監△毛	
p. 271	●毳一半	監△毛	
p. 275	●跳	足△字	有 3 處
p. 276	相●·僧曰	灑	

頁碼	檔案原文摘錄	原書字形	備註
p. 277	●𠂇 𠂇羅·還有	𠂇 𠂇麼	
p. 277	相嚙●底	嚙	
p. 277	裏擣●餅	食 𠂇茲	
p. 278	掙犁拽●	金 𠂇罷	
p. 278	優波●多	攴	
p. 278	順手骨●	揅	
p. 279	怨他河曲●	𠂇 𠂇見	
p. 281	牽巨鉤●	金 𠂇𠂇	
p. 282	如一●墨	形 金 之 手 形	
p. 286	乾坤底●	擣	
p. 286	宇宙底●	擣	
p. 288	伏手骨●	揅	
p. 291	箇●頭底	誦	