

《圓悟佛果禪師語錄》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
2010/03/25 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 2534 的「堊」字，Unicode 編碼是 213BA，由於 XP 並不支援，仍需使用構字式「祝 厶 土」。

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：

《圓悟佛果禪師語錄》使用舊版造字 109 個，字頻 1169 次，這 109 個造字中，100 個可轉成 Windows XP 能顯示的字，字頻 1126 次；另外 9 個字必須轉成構字式，字頻 43 次。

轉碼完成製作轉碼分析表，請參考附件一《圓悟佛果禪師語錄》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 2417 的「檄」字，而原書使用之字為「木𠂇𠂇」，查詢教育部異體字字典確認「木𠂇𠂇」並非「檄」之異體字，因此歸為錯字，以程式取代為正確字形「木𠂇𠂇」。
 3. 待造字：Unicode 及漢字構形資料庫皆未收錄的舊漢籍造字，正在等待補造字中，所以「造字」欄空白無法看到字形。如編號 2752 的「木𠂇栗」。
 4. 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、《圓悟佛果禪師語錄》轉碼分析表

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
814	墓	1	8E5C	87C7	墓		
887	鷲	1	8EC7	9E0E	鷲		
893	峩	1	8ECD	5CE9	峩		

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
1151	剗	3	9073	5273	剗		
1400	窓	2	91F1	7A93	窓		
1530	糺	1	92D6	7CFA	糺		
1536	麤	8	92DC	9E81	麤		
2044	踈	1	9642	8E08	踈		
2225	穢	1	975A	88D3	穢		
2232		1	9761			𠄎厂墨 卩𠄎	𠄎，異體字問題。
2278	鋤	1	97B1	936E	鋤		
2386	瑯	2	985E	7458	瑯		
2417	檄	2	987D	2362E		木𠄎敝	木𠄎敝，錯字，以程式全部取代。
2502	颺	1	98F4	98C8	颺		
2505	飜	2	98F7	98DC	飜		
2524	饜	1	994B	994D	饜		
2534	埜	3	9955	213BA		祝会土	
2598	鄒	4	99B7	9381	鄒		
2669	獯	1	99FE	7379	獯		

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
2752		1	9AB4	235D6		木𠂇栗	待造字。
3002	韜	1	9C52	97B1	韜		
3789	冲	1	8154	51B2	冲		沖，異體字問題。
3792	凉	2	8157	51C9	凉		涼，異體字問題。
3801	剗	15	8160	5257	剗		
3804	効	4	8163	52B9	効		
3805	勅	20	8164	52C5	勅		
3814	却	333	816D	5374	却		
3819	廡	2	8172	53AE	廡		廡，異體字問題。
3820	厦	1	8173	53A6	厦		
3824	叙	1	8177	53D9	叙		
3831	喆	2	817E	5586	喆		
3858	寃	7	81BB	5BC3	寃		冤，異體字問題。
3864	峯	152	81C1	5CEF	峯		
3884	忽	1	81D5	6031	忽		

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
3900	挖	18	81E5	62D5	挖		
3903	筆	1	81E8	25BE4		竹 聿 聿	
3909	擡	7	81EE	64E1	擡		
3931	椀	3	8245	6900	椀		
3932	椶	1	8246	6936	椶		
3937	檝	2	824B	6A9D	檝		
3940	鬱	7	824E	6B1D	鬱		
3944	氷	22	8252	6C37	氷		
3946	汹	1	8254	6C79	汹		
3955	澗	5	825D	3D4E		氵 間	澗，異體字問題。
3977	狗	1	8273	72E5	狗		
3978	猪	2	8274	732A	猪		Unicode 字型呈現差異。
3979	猫	3	8275	732B	猫		
3984	溜	3	827A	7460	溜		
3992	嘗	3	82A4	751E	嘗		
3996	疎	17	82A8	758E	疎		
4001	瘥	1	82AD	7602	瘥		

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
4013	盃	2	82B9	76CC	盃		
4014	蓋	1	82BA	8462	蓋		
4028	稟	11	82C8	7980	稟		
4038	竈	15	82D2	7AC8	竈		
4041	豎	30	82D5	7AEA	豎		
4042	笋	3	82D6	7B0B	笋		
4044	筋	1	82D8	7B6F	筋		
4067	綉	2	82EF	7D89	綉		
4089	脇	2	8346	8107	脇		
4090	脚	108	8347	811A	脚		
4091	阜	3	8348	81EF	阜		Unicode 字型 呈現差異。
4100	苳	1	8351	82FD	苳		
4107	葱	2	8358	8471	葱		
4134	衛	2	8373	885E	衛		衛，異體字問 題。
4139	袴	2	8378	88B4	袴		
4146	覩	3	83A1	89A9	覩		
4174	躑	12	83BD	8EB6	躑		

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
4184	迹	9	83C7	8FF9	迹		
4195	鉢	19	83D2	9262	鉢		
4203	隣	6	83DA	96A3	隣		鄰，異體字問題。
4207	霧	1	83DE	9736	霧		
4213	鞞	1	83E4	97BE	鞞		
4220	穎	6	83EB	9834	穎		
4226	凜	3	83F1	98E1	凜		
4232	隲	1	83F7	96B2	隲		
4240	髯	3	8440	9AF4	髯		
4256	鷄	5	8450	9DC4	鷄		
4267	鼈	3	845B	9F08	鼈		
4278	亘	55	8466	4E98	亘		
4337	睽	1	84C3	25210		目 𠃉 柴	
4381	備	11	84EF	511E	備		
4428	蹶	6	855F	8E70	蹶		
4447	捏	10	8572	63D1	捏		
4448		4	8573			形 纟 厂 墨 口	纏，異體字問題。

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
4450	狃	1	8575	524F	狃		Unicode 字型呈現差異。
4515	呪	1	85D8	546A	呪		
4714	廛	1	8743	222E8		广 厶 墨	
4768	妬	3	8779	59AC	妬		
4778	鞞	10	87A5	97B4	鞞		
5008	幞	2	88EE	5E5E	幞		
5163	憂	3	89EC	621E	憂		
5206	顛	30	8A58	29573		寧 厶 頁	
5253		3	8AA9			卜 厶 冂 厶 冂	冂 厶 冂，異體字問題。
5290	携	9	8ACE	643A	携		
5300	鬪	4	8AD8	9B2A	鬪		
5316	晉	1	8AE8	664B	晉		晉，異體字問題。
5360	鄜	4	8B55	913D	鄜		
5361	鴟	1	8B56	969D	鴟		子 厶 鳥，錯字，以程式全部取代。

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
5442	栞	1	8BC9	67BF	栞		
5453	涅	21	8BD4	23D40		 	涅，異體字問題。
5466	倏	1	8BE1	5010	倏		
5555	迥	33	8C7B	9008	迥		
5559	欸	2	8CA1	6B35	欸		
5590	朶	1	8CC0	6736	朶		
5595	朶	4	8CC5	579C	朶		
5702	厨	4	8D71	53A8	厨		
5753	潜	7	8DC6	6F5C	潜		
5797	灑	1	8DF2	7054	灑		

二、手動取代表說明：

《圓悟佛果禪師語錄》檔案中出現「●」字形，共計54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《圓悟佛果禪師語錄》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、《圓悟佛果禪師語錄》「●」字手動取代表

頁碼	檔案原文摘錄	原書字形	備註
p. 713	●跳	足△字	有 7 處
p. 723	秦時●轆鑽	車△度	
p. 725	●不著	扌△索	
p. 731	千里外絕●	誦	
p. 733	眾眼●●	目△荅 目△荅	
p. 736	力●叱	口△力	
p. 741	●●論實事	石△暴 石△暴	
p. 745	霧●注	雨△汙	
p. 746	團●	團	有 2 處
p. 747	兩輪天氣●	璣	
p. 747	破衲太●毳	毛△監	
p. 748	●多尊者	璣	
p. 748	鐵樹●鬆石	髟△龍	
p. 748	●訛	誦	有 13 處
p. 754	空裏●啄	啣△鳥	改為鵠
p. 755	朦朦●●	目△董 目△董	
p. 758	瞢瞢●●	目△董	

頁碼	檔案原文摘錄	原書字形	備註
		目𠂔董	
p. 758	不勞●	𠂔	
p. 758	密密處有●	諄	
p. 766	教爾●不得	𠂔	
p. 781	教筌●爾	𠂔	
p. 789	無●啄處	𠂔	
p. 789	蛙步●泥沙	馬𠂔展	
p. 792	首座●	口𠂔尔	有 2 處
p. 792	寺主●	口𠂔尔	有 2 處
p. 792	維那●	口𠂔尔	有 2 處
p. 795	若不是●公	𠂔	
p. 804	道崇寧●	口𠂔爾	待造字