

# 《雲門匡真禪師廣錄》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  
2010/03/15 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 2534 的「壑」字，Unicode 編碼是 213BA，由於 XP 並不支援，仍需使用構字式「祝 $\text{厶}$ 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：

《雲門匡真禪師廣錄》使用舊版造字 46 個，字頻 503 次，這 46 個造字中，38 個可轉成 Windows XP 能顯示的字，字頻 487 次；另外 8 個字必須轉成構字式，字頻 16 次。

轉碼完成製作轉碼分析表，請參考附件一《雲門匡真禪師廣錄》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 5453 的「涅」字，是「涅」的異體字，以標準字「涅」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 4554 的「唬」字，而原書使用之字為「睇」，查詢教育部異體字字典確認「睇」並非「唬」之異體字，因此歸為錯字，以程式取代為正確字形「睇」。
  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「木𣎵栗」。
  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 4262 的「麪」字，Unicode 字型呈現為「麪」，實際上仍為同一字。

附件一、《雲門匡真禪師廣錄》轉碼分析表

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
15	畝	1	FA4E	7546	畝		
814	墓	18	8E5C	87C7	墓		
931	磻	3	8EF3	791A	磻		

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
1146	甕	1	906E	7506	甕		
1400	窓	1	91F1	7A93	窓		
1536	麤	1	92DC	9E81	麤		
1854	啣	2	94E0	20E95		口 厶 留	
2085	櫛	1	966B	69F5	櫛		
2298	殼	1	97C5	3C89		殼 厶 卵	
2505	翻	2	98F7	98DC	翻		
2534	埒	3	9955	213BA		祝 厶 土	
2742	抄	1	9AAA	6331	抄		
2752		1	9AB4	235D6		木 厶 栗	待造字。
3803	飽	4	8162	4B54		食 厶 追	
3805	勅	6	8164	52C5	勅		
3814	却	118	816D	5374	却		
3864	峯	82	81C1	5CEF	峯		
3909	擡	2	81EE	64E1	擡		
3931	椀	18	8245	6900	椀		
3940	鬱	3	824E	6B1D	鬱		
3977	狗	1	8273	72E5	狗		

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
3992	嘗	1	82A4	751E	嘗		
3996	疎	8	82A8	758E	疎		
4001	瘥	2	82AD	7602	瘥		
4041	豎	1	82D5	7AEA	豎		
4045	箒	2	82D9	7B92	箒		
4090	脚	50	8347	811A	脚		
4134	衛	1	8373	885E	衛		衛，異體字問題。
4195	鉢	34	83D2	9262	鉢		
4251	緞	4	844B	9C15	緞		
4256	鷄	2	8450	9DC4	鷄		
4262	麪	6	8456	9EAA	麪		Unicode 字型呈現差異。
4278	亘	2	8466	4E98	亘		
4381	儻	81	84EF	511E	儻		
4389	瑫	2	84F7	746B	瑫		
4515	呪	2	85D8	546A	呪		
4554	唬	1	8640	55C1	唬		睇，錯字，以程式全部取

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
							代。
4856	疊	1	87F3	3CB2		疊厶毛	
5046	彝	1	8955	5F5C	彝		
5074	尫	1	8971	5C2B	尫		
5300	鬪	3	8AD8	9B2A	鬪		
5393		1	8B76	3E6E		豸厶百	馬厶白，錯字，以程式全部取代。
5453	涅	20	8BD4	23D40		形; 日 工	涅，異體字問題。
5555	迥	2	8C7B	9008	迥		
5703	齒	2	8D72	9F53	齒		
5783	教	3	8DE4	3A8D		孛厶攵	

## 二、手動取代表說明：

《雲門匡真禪師廣錄》檔案中出現「●」字形，共計46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《雲門匡真禪師廣錄》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、《雲門匡真禪師廣錄》「●」字手動取代表

頁碼	檔案原文摘錄	原書字形	備註
p. 545	曬●	日厶良	有 2 處
p. 547	一擔搯●	扌厶垂	
p. 548	野●漢	裡	有 5 處
p. 548	這掠虛漢●	言厶宰	
p. 548	問●嬾之子	岭	
p. 550	關●	榦	有 2 處
p. 552	●却飯了	噶	
p. 553	●臀後	月厶屈	
p. 554	盂●飯	噶	有 3 處
p. 558	以空●空	噶	
p. 559	任打野●	裡	
p. 561	無●啄處	鷓	
p. 563	摩●顯正	少厶兔	
p. 564	口似木●	木厶突	
p. 565	代云●●	口厶部啉	
p. 566	師云●	口厶部	
p. 566	作麼生是●	口厶部	
p. 566	還將得●	饌	

頁碼	檔案原文摘錄	原書字形	備註
p. 567	完●道將	𪛗	
p. 568	蒸餅●轉云	捌	
p. 568	某甲也●麼	溜	有 2 處
p. 568	爾打野●	裡	
p. 568	●飯漢	噎	
p. 568	完●餅角子	𪛗	
p. 568	●底把將來	𪛗	
p. 569	將●錢來	貝𪛗親	
p. 570	又云吽●	口𪛗發	
p. 570	又云句●	口𪛗爾	待造字
p. 572	即●●·師云	亻𪛗荅 瓜𪛗至	
p. 572	便●麼去	溜	
p. 572	脫却●	允𪛗骨	改為𪛗
p. 573	秦時●轆	車𪛗度	有 2 處
p. 575	及長落●	髟𪛗采	
p. 576	髟●猶生	髟𪛗采	