

# 《撫州曹山本寂禪師語錄》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  
2010/03/12 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 77 的「哪」字，Unicode 編碼是 35BF，由於 XP 並不支援，仍需使用構字式「口𠂔耶」。

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：

《撫州曹山本寂禪師語錄》使用舊版造字 19 個，字頻 87 次，這 19 個造字中，18 個可轉成 Windows XP 能顯示的字，字頻 85 次；另外 1 個字必須轉成構字式，字頻 2 次。

轉碼完成製作轉碼分析表，請參考附件一《撫州曹山本寂禪師語錄》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 5453 的「涅」字，是「涅」的異體字，以標準字「涅」取代。

附件一、《撫州曹山本寂禪師語錄》轉碼分析表

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
77	哪	2	FAAE	35BF		口𠂔耶	
938	燮	1	8EFA	7215	燮		
1151	剗	1	9073	5273	剗		
1536	麤	5	92DC	9E81	麤		
2935	穈	1	9BCE	7A6A	穈		
3805	勅	1	8164	52C5	勅		
3814	却	41	816D	5374	却		
3826	叶	1	8179	53F6	叶		
3864	峯	8	81C1	5CEF	峯		
3996	疎	3	82A8	758E	疎		
4001	瘧	1	82AD	7602	瘧		

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
4041	豎	2	82D5	7AEA	豎		
4089	脇	1	8346	8107	脇		
4090	脚	4	8347	811A	脚		
4134	衛	1	8373	885E	衛		衛，異體字問題。
4174	𦓐	2	83BD	8EB6	𦓐		
4180	鼻	2	83C3	8FA0	鼻		
4278	亘	3	8466	4E98	亘		
5453	涅	7	8BD4	23D40		形: 日 工: 工	涅，異體字問題。

## 二、手動取代表說明：

《撫州曹山本寂禪師語錄》檔案中出現「●」字形，共計3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《撫州曹山本寂禪師語錄》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、《撫州曹山本寂禪師語錄》「●」字手動取代表

頁碼	檔案原文摘錄	原書字形	備註
p. 530	彌勒阿●	閱	
p. 536	每日●什麼	瞳	
p. 542	相●脈脈	拄	