

# 《金陵清涼院文益禪師語錄》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  
2010/03/16 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 4121 的「懼」字，Unicode 編碼是 39AC，由於 XP 並不支援，仍需使用構字式「忄 𠂔羅」。

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：

《金陵清涼院文益禪師語錄》使用舊版造字 19 個，字頻 64 次，這 19 個造字中，18 個可轉成 Windows XP 能顯示的字，字頻 63 次；另外 1 個字必須轉成構字式，字頻 1 次。

轉碼完成製作轉碼分析表，請參考附件一《金陵清涼院文益禪師語錄》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。
2. 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 4450 的「翔」字，Unicode 字型呈現為「𨇖」，實際上仍為同一字。

附件一、《金陵清涼院文益禪師語錄》轉碼分析表

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
1400	窓	4	91F1	7A93	窓		
2598	鄒	2	99B7	9381	鄒		
2885	叫	1	9B7A	544C	叫		
3814	却	20	816D	5374	却		
3824	叙	1	8177	53D9	叙		
3864	峯	12	81C1	5CEF	峯		
3876	廻	1	81CD	5EFB	廻		
3931	椀	1	8245	6900	椀		

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
3990	甄	1	82A2	750E	甄		
3996	疎	2	82A8	758E	疎		
4090	脚	10	8347	811A	脚		
4121	懼	1	8366	39AC		忄 ㄩ 羅	
4134	衛	1	8373	885E	衛		衛，異體字問題。
4146	覩	2	83A1	89A9	覩		
4198	鑛	1	83D5	945B	鑛		
4450	狃	1	8575	524F	狃		Unicode 字型呈現差異。
4619	但	1	86A3	4F39	但		
5131	窻	1	89CC	7ABB	窻		Unicode 字型呈現差異。
5453	涅	1	8BD4	23D40		灑 日 工 口	涅，異體字問題。

## 二、手動取代表說明：

《金陵清涼院文益禪師語錄》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《金陵清涼院文益禪師語錄》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、《金陵清涼院文益禪師語錄》「●」字手動取代表

頁碼	檔案原文摘錄	原書字形	備註
p. 593	● · 師別云	口△力	
p. 594	● · 昭與一眾	灑	
p. 594	一眾 · ●	小△摩	