

《楊岐方會和尚語錄》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
2010/03/18 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 3803 的「鮓」字，Unicode 編碼是 4B54，由於 XP 並不支援，仍需使用構字式「食 𪛗 追」。

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：

《楊岐方會和尚語錄》使用舊版造字 17 個，字頻 65 次，這 17 個造字中，16 個可轉成 Windows XP 能顯示的字，字頻 64 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《楊岐方會和尚語錄》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、《楊岐方會和尚語錄》轉碼分析表

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
1400	窓	1	91F1	7A93	窓		
1536	龕	1	92DC	9E81	龕		
2157	酌	1	96D5	9167	酌		
2366	鎔	1	984A	93F4	鎔		
3803	飽	1	8162	4B54		食 𠂇 追	
3814	却	17	816D	5374	却		
3843	堦	1	81AC	5826	堦		
3864	峯	13	81C1	5CEF	峯		
3900	挖	2	81E5	62D5	挖		
3931	椀	1	8245	6900	椀		

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
3996	疎	3	82A8	758E	疎		
4090	脚	14	8347	811A	脚		
4195	鉢	4	83D2	9262	鉢		
4267	鼈	1	845B	9F08	鼈		
5300	鬪	1	8AD8	9B2A	鬪		
5453	涅	1	8BD4	23D40			涅，異體字問題。
5753	潛	2	8DC6	6F5C	潛		

二、手動取代表說明：

《楊岐方會和尚語錄》檔案中出現「●」字形，共計2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《楊岐方會和尚語錄》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、《楊岐方會和尚語錄》「●」字手動取代表

頁碼	檔案原文摘錄	原書字形	備註
p. 643	一時●瞎	瞿𠃉支	
p. 643	龍馬駒●跳	足𠃉字	