

《錢塘湖隱濟顛禪師語錄》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
2008/7/21 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 1616 的「運」字，Unicode 編碼是 284E6，由於 XP 並不支援，仍需使用構字式「辶△里」。

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：

《錢塘湖隱濟顛禪師語錄》使用舊版造字 17 個，字頻 270 次，這 17 個造字中，13 個可轉成 Windows XP 能顯示的字，字頻 247 次；另外 4 個字必須轉成構字式，字頻 23 次。

轉碼完成製作轉碼分析表，請參考附件一《錢塘湖隱濟顛禪師語錄》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 4256 的「鷄」字，是「雞」的異體字，以標準字「雞」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 4262 的「麪」字，Unicode 字型呈現為「麪」，實際上仍為同一字。

附件一、《錢塘湖隱濟顛禪師語錄》轉碼分析表

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
297	厶	14	FBED			厶	
1616	運	5	936D	284E6		辶厶里	
1854	啣	1	94E0	20E95		口厶留	
3164		3	9D57			土厶冊	土厶冊，異體字問題。
3825	疊	1	8178	53E0	疊		疊，異體字問題。
3829	咏	1	817C	548F	咏		

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
3966	烟	6	8268	70DF	烟		
4016	着	64	82BC	7740	着		
4038	竈	3	82D2	7AC8	竈		
4042	笋	13	82D6	7B0B	笋		
4112	蔴	6	835D	8534	蔴		
4132		125	8371			血𠄎𠄎	眾，異體字問題。
4139	袴	1	8378	88B4	袴		
4195	鉢	12	83D2	9262	鉢		
4256	鷄	6	8450	9DC4	鷄		雞，異體字問題。
4262	麪	3	8456	9EAA	麪		Unicode 字型呈現差異。
4307	罵	6	84A5	99E1	罵		罵，異體字問題。