

《如淨和尚語錄》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
2010/03/02 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 1772 的「藜」字，Unicode 編碼是 4527，由於 XP 並不支援，仍需使用構字式「梨」。

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：

《如淨和尚語錄》使用舊版造字 58 個，字頻 178 次，這 58 個造字中，53 個可轉成 Windows XP 能顯示的字，字頻 168 次；另外 5 個字必須轉成構字式，字頻 10 次。

轉碼完成製作轉碼分析表，請參考附件一《如淨和尚語錄》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 3819 的「廡」字，是「廡」的異體字，以標準字「廡」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 3978 的「猪」字，Unicode 字型呈現為「猪」，實際上仍為同一字。

附件一、《如淨和尚語錄》轉碼分析表

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
64	颺	1	FAA1	98B7	颺		
934	焮	1	8EF6	7200	焮		
1310	蠊	2	9175	45EB		虫 𧈧 麻	
1354	禊	1	91C3	79A9	禊		
1400	窓	1	91F1	7A93	窓		
1559		1	92F3			𠃉 广 目 无 口	廡，異體字問題。
1772	藜	1	946C	4527		艹 厶 梨	

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
1999	贊	6	95D4	8CDB	贊		
2534	埒	2	9955	213BA		祝  土	
2565	霏	1	9974	9714	霏		
2814	灑	1	9AF2	704B	灑		
3792	涼	2	8157	51C9	涼		涼，異體字問題。
3801	剗	2	8160	5257	剗		
3803	餽	1	8162	4B54		食  追	
3805	勅	7	8164	52C5	勅		
3814	却	24	816D	5374	却		
3817	廁	1	8170	53A0	廁		廁，異體字問題。
3819	廝	2	8172	53AE	廝		廝，異體字問題。
3826	叶	2	8179	53F6	叶		
3858	寃	6	81BB	5BC3	寃		寃，異體字問題。

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
3864	峯	10	81C1	5CEF	峯		
3900	挖	1	81E5	62D5	挖		
3931	椀	4	8245	6900	椀		
3940	鬱	3	824E	6B1D	鬱		
3944	冰	4	8252	6C37	冰		
3966	烟	9	8268	70DF	烟		
3978	猪	4	8274	732A	猪		Unicode 字型 呈現差異。
3984	瑠	2	827A	7460	瑠		
3992	嘗	2	82A4	751E	嘗		
4028	稟	1	82C8	7980	稟		
4035	穽	1	82CF	7A7D	穽		
4038	竈	1	82D2	7AC8	竈		
4041	豎	5	82D5	7AEA	豎		
4042	笋	2	82D6	7B0B	笋		
4044	筴	2	82D8	7B6F	筴		

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
4067	綉	1	82EF	7D89	綉		
4090	脚	19	8347	811A	脚		
4107	葱	2	8358	8471	葱		
4139	袴	1	8378	88B4	袴		
4184	迹	1	83C7	8FF9	迹		
4195	鉢	4	83D2	9262	鉢		
4262	麪	1	8456	9EAA	麪		Unicode 字型 呈現差異。
4278	亘	2	8466	4E98	亘		
4447	捏	6	8572	63D1	捏		
4448		1	8573			𠃉 纟 厂 墨 𠃉	纏，異體字問 題。
4558	廐	2	8644	53A9	廐		廐，異體字問 題。
4778	鞞	2	87A5	97B4	鞞		
4903	𠂔	2	8863	5C00	𠂔		
4957	崑	2	88BB	5D53	崑		

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
4981	皂	1	88D3	9CEC	皂		
5206	顛	4	8A58	29573		寧厶頁	
5290	携	2	8ACE	643A	携		
5453	涅	3	8BD4	23D40		形日 工	涅，異體字問題。
5463	肯	1	8BDE	65F9	肯		
5555	迥	2	8C7B	9008	迥		
5702	厨	1	8D71	53A8	厨		
5714	瓮	1	8D7D	74EB	瓮		
5797	灑	1	8DF2	7054	灑		

二、手動取代表說明：

《如淨和尚語錄》檔案中出現「●」字形，共計29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《如淨和尚語錄》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、《如淨和尚語錄》「●」字手動取代表

頁碼	檔案原文摘錄	原書字形	備註
p. 121	關●	換	有 2 處
p. 122	●訛	誦	有 2 處
p. 122	大海波心輕●	扌△聿	
p. 124	●跳	足△字	有 7 處
p. 124	大家贏得●	瞳	
p. 125	無柄木●	木△夕	
p. 125	海底泥牛●	瞳	
p. 127	晒●諸人	日△良	
p. 128	●火爐動也	口△力	
p. 128	●九旬垂釣	口△力	
p. 129	●沸	尸△豕	有 4 處
p. 129	箇團●果	鬮	
p. 129	剔團●也得	鬮	
p. 131	●·太平無象	口△力	
p. 131	●酒	瞳	
p. 131	●肉破	瞳	
p. 132	真快樂·●	口△力	
p. 132	喫飯●眠	瞳	