

《法演禪師語錄》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
2010/03/22 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 厶羅」。

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：

《法演禪師語錄》使用舊版造字 72 個，字頻 332 次，這 72 個造字中，66 個可轉成 Windows XP 能顯示的字，字頻 325 次；另外 6 個字必須轉成構字式，字頻 7 次。

轉碼完成製作轉碼分析表，請參考附件一《法演禪師語錄》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 4262 的「麪」字，Unicode 字型呈現為「麪」，實際上仍為同一字。
3. 待造字：Unicode 及漢字構形資料庫皆未收錄的舊漢籍造字，正在等待補造字中，所以「造字」欄空白無法看到字形。如編號 2704 的「𠂔𠂔了」。

附件一、《法演禪師語錄》轉碼分析表

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
814	𦵑	1	8E5C	87C7	𦵑		
842	𦵑	1	8E78	3AB7		𦵑	
887	𦵑	3	8EC7	9E0E	𦵑		
1133	𦵑	3	9061	74CC	𦵑		
1135	𦵑	1	9063	44B5		𦵑 𦵑好	
1400	窓	4	91F1	7A93	窓		

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
1536	麤	5	92DC	9E81	麤		
1591	羗	1	9354	7F97	羗		
1981	讐	3	95C2	8B90	讐		
2046	踪	1	9644	8E2A	踪		
2162	醜	1	96DA	918E	醜		
2278	鎗	1	97B1	936E	鎗		
2325	潤	1	97E0	77A4	潤		
2386	瑯	2	985E	7458	瑯		
2505	翻	3	98F7	98DC	翻		
2704		2	9A62			𠄎𠄎了	待造字。
2906	竽	1	9BB1	659A	竽		
3797	凭	3	815C	51ED	凭		
3805	勅	1	8164	52C5	勅		
3814	却	73	816D	5374	却		
3826	叶	1	8179	53F6	叶		
3843	堦	1	81AC	5826	堦		
3851	姹	1	81B4	59F9	姹		
3858	寃	7	81BB	5BC3	寃		冤，異體字問

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
							題。
3864	峯	39	81C1	5CEF	峯		
3900	挖	1	81E5	62D5	挖		
3909	擡	3	81EE	64E1	擡		
3931	椀	1	8245	6900	椀		
3940	鬱	4	824E	6B1D	鬱		
3944	氷	4	8252	6C37	氷		
3955	澗	4	825D	3D4E		汜 𠂔 間	澗，異體字問題。
3966	烟	2	8268	70DF	烟		
3979	猫	2	8275	732B	猫		
3984	瑠	1	827A	7460	瑠		
3996	疎	6	82A8	758E	疎		
4013	盃	1	82B9	76CC	盃		
4042	笋	1	82D6	7B0B	笋		
4045	箒	1	82D9	7B92	箒		
4090	脚	21	8347	811A	脚		
4107	葱	1	8358	8471	葱		

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
4118	藥	2	8363	8602	藥		
4121	懼	1	8366	39AC		忄 厶 羅	
4146	覩	1	83A1	89A9	覩		
4195	鉢	10	83D2	9262	鉢		
4207	霧	1	83DE	9736	霧		
4216	鞮	1	83E7	97C8	鞮		
4220	穎	4	83EB	9834	穎		
4240	髯	1	8440	9AF4	髯		
4256	鷄	3	8450	9DC4	鷄		
4262	麪	1	8456	9EAA	麪		Unicode 字型 呈現差異。
4267	鼈	4	845B	9F08	鼈		
4278	亘	1	8466	4E98	亘		
4381	倆	57	84EF	511E	倆		
4440	峇	1	856B	37E7		山 厶 号	
4447	捏	3	8572	63D1	捏		
4448		1	8573			形 纟 厂 墨 口	纏，異體字問 題。

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
4515	呪	1	85D8	546A	呪		
4546	鴟	1	85F7	9D76	鴟		
4738	睺	1	875B	777A	睺		
4739	鞞	1	875C	97DB	鞞		
4856	疊	1	87F3	3CB2		疊 厶 毛	
4957	崑	1	88BB	5D53	崑		
5111	旃	1	89B8	65BE	旃		
5119	博	1	89C0	613D	博		
5179	駟	1	89FC	99C8	駟		
5290	携	4	8ACE	643A	携		
5300	鬪	4	8AD8	9B2A	鬪		
5439	栢	1	8BC6	6822	栢		
5453	涅	2	8BD4	23D40		形 日 工 口	涅，異體字問題。
5555	迥	3	8C7B	9008	迥		
5590	朶	2	8CC0	6736	朶		
5753	潜	2	8DC6	6F5C	潜		

二、手動取代表說明：

《法演禪師語錄》檔案中出現「●」字形，共計15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《法演禪師語錄》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、《法演禪師語錄》「●」字手動取代表

頁碼	檔案原文摘錄	原書字形	備註
p. 650	●地鳥語	目𠃉舌	
p. 651	老●	口𠃉爾	待造字 有2處
p. 652	湍●	汜𠃉盾	待造字 有2處
p. 656	一場●	卜𠃉麼	
p. 657	隈●老古錐	𠃉𠃉崔	
p. 657	眾人●嗽	口𠃉𠃉	
p. 657	一擔矇●	目𠃉董	
p. 658	●八囉札	王𠃉朋	
p. 660	阿●佛	閱	
p. 662	拏雲●霧	爪𠃉隻	
p. 666	●●微風起	風𠃉弗 風𠃉弗	

頁碼	檔案原文摘錄	原書字形	備註
p. 669	●序	禘 𡗗寸	