

《密菴和尚語錄》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
2010/03/29 丁玟伶

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 ㄩㄛ ㄌㄨㄛˊ」。

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：

《密菴和尚語錄》使用舊版造字 81 個，字頻 360 次，這 81 個造字中，77 個可轉成 Windows XP 能顯示的字，字頻 355 次；另外 4 個字必須轉成構字式，字頻 5 次。

轉碼完成製作轉碼分析表，請參考附件一《密菴和尚語錄》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。
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、《密菴和尚語錄》轉碼分析表

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
814	墓	1	8E5C	87C7	墓		
1400	窓	2	91F1	7A93	窓		
1466	櫛	1	9274	2383D		形木西 朝	
1491	擷	2	92AF	6527	擷		
1784	萼	1	9478	855A	萼		
1999	贊	8	95D4	8CDB	贊		
2162	醜	1	96DA	918E	醜		
2278	鎗	1	97B1	936E	鎗		

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
2366	鎔	1	984A	93F4	鎔		
2386	瑯	2	985E	7458	瑯		
2409	闕	3	9875	95A7	闕		
2534	堊	2	9955	213BA		祝会土	
2563	搨	2	9972	6502	搨		
3770	个	1	8141	4E2A	个		
3789	冲	1	8154	51B2	冲		
3801	剗	2	8160	5257	剗		
3804	効	1	8163	52B9	効		
3805	勅	4	8164	52C5	勅		
3806	勑	1	8165	52E0	勑		
3814	却	83	816D	5374	却		
3817	廁	2	8170	53A0	廁		廁，異體字問題。
3819	廡	1	8172	53AE	廡		廡，異體字問題。
3824	叙	4	8177	53D9	叙		
3826	叶	1	8179	53F6	叶		
3843	堦	1	81AC	5826	堦		

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
3858	寃	6	81BB	5BC3	寃		冤，異體字問題。
3864	峯	24	81C1	5CEF	峯		
3900	挖	2	81E5	62D5	挖		
3940	鬱	1	824E	6B1D	鬱		
3944	氷	8	8252	6C37	氷		
3955	澗	1	825D	3D4E		汜 澗	澗，異體字問題。
3966	烟	2	8268	70DF	烟		
3968	煨	8	826A	7145	煨		
3978	猪	4	8274	732A	猪		Unicode 字型呈現差異。
3981	珉	2	8277	73C9	珉		
3984	瑠	1	827A	7460	瑠		
3992	嘗	7	82A4	751E	嘗		
3996	疎	2	82A8	758E	疎		
4012	鼓	1	82B8	76B7	鼓		鼓，異體字問題。
4028	稟	1	82C8	7980	稟		

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
4041	豎	23	82D5	7AEA	豎		
4042	笋	2	82D6	7B0B	笋		
4045	箒	2	82D9	7B92	箒		
4067	綉	3	82EF	7D89	綉		
4089	脇	1	8346	8107	脇		
4090	脚	37	8347	811A	脚		
4121	懼	1	8366	39AC		忄 厶 羅	
4134	衛	1	8373	885E	衛		衛，異體字問題。
4139	袴	1	8378	88B4	袴		
4144	霸	7	837D	8987	霸		Unicode 字型呈現差異。
4146	覩	3	83A1	89A9	覩		
4184	迹	3	83C7	8FF9	迹		
4195	鉢	13	83D2	9262	鉢		
4207	霧	1	83DE	9736	霧		
4220	穎	1	83EB	9834	穎		
4226	凜	1	83F1	98E1	凜		
4232	隲	1	83F7	96B2	隲		

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
4240	鬚	1	8440	9AF4	鬚		
4256	鷄	1	8450	9DC4	鷄		
4262	麪	1	8456	9EAA	麪		Unicode 字型 呈現差異。
4272		2	8460			厂 △ 龍	龐，異體字問 題。
4278	亘	3	8466	4E98	亘		
4389	瑫	2	84F7	746B	瑫		
4447	捏	7	8572	63D1	捏		
4495	麪	3	85C4	9EA8	麪		Unicode 字型 呈現差異。
4515	呪	2	85D8	546A	呪		
4546	鷓	3	85F7	9D76	鷓		
4566	礪	1	864C	78F5	礪		
4778	鞞	6	87A5	97B4	鞞		
4870	壘	1	8842	97F2	壘		
4903	吋	1	8863	5C00	吋		
4950	嶮	1	88B4	5D18	嶮		
5206	顛	1	8A58	29573		寧 △ 頁	

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
5290	携	2	8ACE	643A	携		
5300	鬪	2	8AD8	9B2A	鬪		
5453	涅	4	8BD4	23D40		 	涅，異體字問題。
5555	迥	8	8C7B	9008	迥		
5559	欸	3	8CA1	6B35	欸		
5590	朶	2	8CC0	6736	朶		
5595	朶	2	8CC5	579C	朶		
5753	潜	1	8DC6	6F5C	潜		

二、手動取代表說明：

《密菴和尚語錄》檔案中出現「●」字形，共計41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《密菴和尚語錄》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、《密菴和尚語錄》「●」字手動取代表

頁碼	檔案原文摘錄	原書字形	備註
----	--------	------	----

頁碼	檔案原文摘錄	原書字形	備註
p. 958	●跳	足△字	有 3 處
p. 959	乾●匝匝	勾	
p. 961	秦時●轆鑽	車△度	
p. 961	●訛	諄	有 4 處
p. 961	●出來	目△董	
p. 962	行及坐臥●	口△尔	
p. 962	喫飯●	口△尔	
p. 963	澠●	彳△盾	待造字 有 3 處
p. 963	復雨●●	毛△參 毛△參	
p. 964	關●子	楨	有 2 處
p. 964	無處晒●	日△良	
p. 964	●忸△羅·華藏	忸△麼	
p. 965	與麼●苴	品若	
p. 965	重添搯●	扌△奎	
p. 966	一任●	足△字	
p. 967	磽磽●●	諄諄	有 2 處
p. 969	扇子●	足△字	
p. 971	良久云[二]●	斬△耳	
p. 972	●·明投暗合	扌△奎	

頁碼	檔案原文摘錄	原書字形	備註
p. 972	[二]●語	嚙	
p. 972	力●希	口△力	
p. 974	●匝·只將	勾	
p. 974	摸●	扌△索	有 2 處
p. 977	●瞎	瞿△欠	有 2 處
p. 978	頂門●	瞿△欠	
p. 978	墮在●盤中	米△离	
p. 981	是隨摟●	揪	