

《汾陽無德禪師語錄》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
2010/03/17 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 77 的「哪」字，Unicode 編碼是 35BF，由於 XP 並不支援，仍需使用構字式「口𠂔耶」。

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：

《汾陽無德禪師語錄》使用舊版造字 74 個，字頻 357 次，這 74 個造字中，68 個可轉成 Windows XP 能顯示的字，字頻 341 次；另外 6 個字必須轉成構字式，字頻 16 次。

轉碼完成製作轉碼分析表，請參考附件一《汾陽無德禪師語錄》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. 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 2752 的「木𣎵栗」。

附件一、《汾陽無德禪師語錄》轉碼分析表

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
77	哪	6	FAAE	35BF		口𣎵耶	
814	墓	2	8E5C	87C7	墓		
887	鷲	3	8EC7	9E0E	鷲		
934	焮	3	8EF6	7200	焮		
943	禡	1	8F40	8975	禡		
1400	窓	1	91F1	7A93	窓		
1536	麤	7	92DC	9E81	麤		
1565	繇	1	92F9	7DDC	繇		
2524	饒	1	994B	994D	饒		

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
2563	搨	1	9972	6502	搨		
2564	刺	2	9973	34E8		夾厶司	
2565	霆	2	9974	9714	霆		
2572	銜	2	997B	8858	銜		
2671	畱	1	9A41	7550	畱		
2731	霍	4	9A7D			宀会隹	
2752		2	9AB4	235D6		木厶栗	待造字。
2786	躬	3	9AD6	8EB3	躬		
3804	効	3	8163	52B9	効		
3805	勅	2	8164	52C5	勅		
3814	却	91	816D	5374	却		
3826	叶	1	8179	53F6	叶		
3864	峯	41	81C1	5CEF	峯		
3891	慙	1	81DC	6159	慙		
3900	挖	1	81E5	62D5	挖		
3909	擡	8	81EE	64E1	擡		
3937	檝	2	824B	6A9D	檝		
3940	薈	1	824E	6B1D	薈		

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
3944	冰	10	8252	6C37	冰		
3955	澗	4	825D	3D4E		灬 冫 間	澗，異體字問題。
3980	獾	1	8276	7328	獾		
3984	瑠	4	827A	7460	瑠		
3989	璿	3	82A1	74C8	璿		
3990	甄	5	82A2	750E	甄		
3992	嘗	1	82A4	751E	嘗		
3996	疎	11	82A8	758E	疎		
4001	瘧	2	82AD	7602	瘧		
4041	豎	1	82D5	7AEA	豎		
4042	笋	1	82D6	7B0B	笋		
4045	箒	1	82D9	7B92	箒		
4089	脇	2	8346	8107	脇		
4090	脚	38	8347	811A	脚		
4118	藁	3	8363	8602	藁		
4139	袴	1	8378	88B4	袴		
4146	覩	5	83A1	89A9	覩		

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
4155	訶	1	83AA	8B0C	訶		
4184	迹	2	83C7	8FF9	迹		
4193	鈎	1	83D0	920E	鈎		
4195	鉢	7	83D2	9262	鉢		
4223	顛	1	83EE	984B	顛		
4240	髯	1	8440	9AF4	髯		
4256	鷄	2	8450	9DC4	鷄		
4267	鼈	4	845B	9F08	鼈		
4271	齧	5	845F	9F69	齧		
4447	捏	2	8572	63D1	捏		
4484	鼎	1	85B9	25109		鼎	
4546	鴟	3	85F7	9D76	鴟		
4738	睽	1	875B	777A	睽		
4768	妬	1	8779	59AC	妬		
4903	吋	1	8863	5C00	吋		
4957	崑	4	88BB	5D53	崑		
5133	憇	1	89CE	6199	憇		
5163	憂	1	89EC	621E	憂		

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
5206	顛	1	8A58	29573		寧𠄎頁	
5290	携	2	8ACE	643A	携		
5300	鬪	4	8AD8	9B2A	鬪		
5360	鄺	1	8B55	913D	鄺		
5396	鎔	1	8B79	938B	鎔		
5453	涅	11	8BD4	23D40		形: 日 工: 工	涅，異體字問題。
5463	岿	1	8BDE	65F9	岿		
5555	迥	4	8C7B	9008	迥		
5559	欸	1	8CA1	6B35	欸		
5621	泪	1	8CDF	6CEA	泪		
5742	澁	1	8DBB	6F81	澁		
5753	潜	4	8DC6	6F5C	潜		

二、手動取代表說明：

《汾陽無德禪師語錄》檔案中出現「●」字形，共計26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《汾陽無德禪師語錄》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、《汾陽無德禪師語錄》「●」字手動取代表

頁碼	檔案原文摘錄	原書字形	備註
p. 596	禪是骨●	骨厶侖	
p. 596	不●慈悲	恪	有 2 處
p. 597	不是性●衲	𠂔𠂔 桑心𠂔	
p. 598	●破頑癡石	翟厶支	
p. 604	方曲●木頭	彖厶見	
p. 604	我早是●跳	足厶亭	
p. 608	截●閉關	金厶戡	
p. 609	前●跪云	足厶胡	
p. 611	出云崔禪●	𠂔	
p. 611	瞽瘕●痲	𠂔𠂔 君巾𠂔	
p. 613	鵝湖●步	足厶聚	
p. 616	優波●多	𠂔	
p. 616	知師不●	恪	
p. 618	代云豈●	恪	
p. 621	此經皆●跪	足厶胡	
p. 621	●明有記	口厶覽	改為口厶覽
p. 622	一條曲●杖	彖厶見	

頁碼	檔案原文摘錄	原書字形	備註
p. 622	鬣●眼睛	耳𠂔聶	
p. 622	●開懵鈍	瞿𠂔支	
p. 623	輪●山疊	車𠂔困	
p. 623	黠●兩晴	兼𠂔頁	
p. 623	●地撐天	瞿𠂔支	
p. 626	●閔宗師	門△元	
p. 626	●鬟髻髻	髻△般	
p. 627	開口●甥前	外𠂔男	待造字