

《祖堂集》舊版造字轉碼說明

中研院資訊所文獻處理實驗室
中研院史語所漢籍電子文獻工作小組
2008/4/8 黃榮順 製作

一、《祖堂集》一書(祖堂集.xml)使用舊版造字 161 個，字頻 1013 次，詳如附件一。這 161 個造字中，115 個可轉成 Windows XP 可顯示的字，字頻 722 次；另外 46 個字必須轉成構字式，字頻 291 次。

二、附件一的造字分析表說明如下：

甲、編號：Big5 造字空間為 6217 個，編號由 1 到 6217。

乙、造字：舊版造字

丙、字頻(txt)：造字在「.txt」文件的出現次數

丁、字頻(xml)：造字在「.xml」文件的出現次數

戊、Big5：造字的 Big5 碼

己、Unicode：造字所對應的 Unicode 碼

庚、WinXP：造字在 Windows XP 的對應字形

辛、構字式：Windows XP 若無對應字形，則改採用構字式

壬、備註凡例：

- 1、校對問題，舊版漢籍錯字，可用程式全部取代：在舊版漢籍電子文獻中即存在的錯字，因校對時的疏漏而未更正，持續留存在新版漢籍電子文獻中；若該造字的所有頻次，皆屬於錯誤使用的錯字情形，可以用程式全部取代為正確字形。如編號 5121 的「𠂔」字，原字應為「𠂔」字。又如編號 374 的「𠂔」，係誤增字，原書所無，故手動刪去。
- 2、異體字問題：新版漢籍考量到使用者檢索及使用時的便利性，將用字原則改為除專詞等特殊情形之外，一律改用標準字呈現。如編號 982 的「𠂔」係「胸」字之異體，故以「胸」字取代。又如編號 1187 的「𠂔」係「蓋」字之簡體，故以「蓋」字取代。
- 3、Unicode 字型呈現差異：Unicode 字型與舊漢籍造字有些微差異，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 2439 的「𠂔」字，Unicode 字型呈現為「𠂔」，實際上仍為同一字。

- 三、 Unicode 目前收錄的漢字總數為 70194，分屬於三個不同區段，詳如表一。目前 Windows XP 只支援 CJK 認同表意文字區的 20902 個字，內碼為 4E00-9FFF。所以如造字編號 1032 的「憊」字，Unicode 編碼為 8FBB，由於 Windows XP 並不支援，仍須改用構字式「 𢇛 心 」。

表一、Unicode 的字數及編碼區段

Unicode	字集子集合	新增字數	新增編碼區段	總字數	WinXP
1.1 版	CJK 認同表意文字區	20902	4E00-9FFF	20902	支援
3.0 版	CJK 認同表意文字擴充 A 區	6582	3400-4DFF	27484	不支援
3.1 版	CJK 認同表意文字擴充 B 區	42710	20000-2A6D6	70194	不支援

- 四、 造字分析表「備註」欄標示「手動取代」者，詳細取代內容可參考附件二的手動取代表。附件二的手動取代表說明如下：

- 甲、 編號：Big5 造字空間為 6217 個，編號由 1 到 6217。
乙、 造字：舊漢籍造字。
丙、 檔案原文摘錄：摘錄檔案(祖堂集.xml)中包含該造字的文句段落，其中紅字底線的部分，為需要手動取代的原文。
丁、 手動取代結果：變更過後的檔案內容，其中藍字底線的部分，為修改後的結果。

- 五、 「祖堂集.xml」缺字外發現之錯字共 2 處，修正情形詳列附件三，附件三「祖堂集.xml」缺字外錯字修改表體例說明如下：

- 甲、 檔案原文摘錄：摘錄檔案(祖堂集.xml)中包含該錯字的文句段落，其中紅字底線的部分，為需要手動取代的原文。
乙、 手動取代結果：變更過後的檔案內容，其中藍字底線的部分，為修改後的結果。

附件一、「祖堂集.xml」造字分析表

編號	造字	字頻 (txt)	字頻 (xml)	Big5	Unicode	WinXP	構字式	備註
296	𠂔	36	36	FBEC			𠂔	
297	𠂕	90	90	FBED			𠂕	
298	𠂖	52	52	FBEE			𠂖	
299	𠂗	2	2	FBEF			𠂗	
300	𠂘	2	2	FBF0			𠂘	
309	𠂙	2	2	FBF9			𠂙	○，校對問題，舊版漢籍錯字，可用程式全部取代。
374	𠂚	1	1	FC7B			𠂚	原書無字，手動刪除。
568	𠂛	2	2	FDC2			𠂛	
624	𠂜	1	1	FDFA			𠂜	
639	𠂝	1	1	FE4A			𠂝	
658	𠂞	11	11	FE5D			𠂞	
717	𠂟	2	2	FEBA			𠂟	
737	𠂠	1	1	FECE			𠂠	
750	𠂡	3	3	FEDB			𠂡	

編號	造字	字頻 (txt)	字頻 (xml)	Big5	Unicode	WinXP	構字式	備註
752	𠂇	1	1	FEDD			𠂇	
758	𠂈	1	1	FEE3			𠂈	
770	𠂉	3	3	FEEF			𠂉	
773	𠂊	1	1	FEF2			𠂊	
777	𠂋	1	1	FEF6			𠂋	
842	𠂌	4	4	8E78	3AB7		𠂌	
932	𠂍	2	2	8EF4	7683	兒		
972	𠂎	1	1	8F5D	7287	𠂎		
982	𠂏	7	7	8F67	80F7	𠂏		胸，異體字問題。
1024	𠂐	8	8	8FB3	8675	𠂐		
1032	𠂑	2	2	8FBB	2039D		𠂑 扶心 𠂑	
1044	𠂒	1	1	8FC7	511B	𠂒		
1056	𠂓	50	50	8FD3	9641	𠂓		
1081	𠂔	1	1	8FEC	9E16	𠂔		
1187	𠂕	10	10	90B9	76D6	𠂕		蓋，異體字問題。
1310	𠂖	5	5	9175	45EB		虫 𠂖 麻	

編號	造字	字頻 (txt)	字頻 (xml)	Big5	Unicode	WinXP	構字式	備註
1328	鈔	2	2	91A9	492C		金𠂔沙	
1396	罕	1	1	91ED	7A82	罕		
1400	窓	4	4	91F1	7A93	窓		
1513	糧	3	3	92C5	7CAE	糧		糧，異體字問題。
1536	麤	1	1	92DC	9E81	麤		Unicode 字型差異。
1547	𠂔	1	1	92E7	4E31	𠂔		
1642	脉	8	8	93A9	8109	脉		
1668	臈	4	4	93C3	81C8	臈		
1755	𠂔	1	1	945B			𠂔 𠂔修	
1905	𠂔	3	3	9554	88E9	𠂔		
1994	達	2	2	95CF	6FBE	達		
2044	踈	20	20	9642	8E08	踈		
2147	蟪	2	2	96CB	87EE	蟪		
2151	𠂔	7	7	96CF	961D	𠂔		
2242	𠂔	1	1	976B	46D2		巧𠂔言	
2269	𠂔	1	1	97A8	5215	𠂔		

編號	造字	字頻 (txt)	字頻 (xml)	Big5	Unicode	WinXP	構字式	備註
2439	磁	7	7	98B5	7920	磁		Unicode 字型差異。
2505	翻	1	1	98F7	98DC	翻		
2524	饒	1	1	994B	994D	饒		
2564	剌	7	7	9973	34E8		夾𠂔 𠂔	
2598	鄒	2	2	99B7	9381	鄒		
2671	畐	1	1	9A41	7550	畐		
2801	鑼	1	1	9AE5	947A	鑼		
2930	畚	1	1	9BC9	7568	畚		
3021	𠂔	6	6	9C65			𠂔	
3035	賓	3	3	9C73	8CD4	賓		賓，異體字問題。
3143	潑	1	1	9D42	6FFD	潑		
3215	炉	1	1	9DAC	7089	炉		
3770	个	8	8	8141	4E2A	个		個，異體字問題。
3780	儁	1	1	814B	5101	儁		
3796	決	1	1	815B	51B3	決		決，異體字問題。
3801	剗	10	10	8160	5257	剗		

編號	造字	字頻 (txt)	字頻 (xml)	Big5	Unicode	WinXP	構字式	備註
3805	勅	3	3	8164	52C5	勅		
3814	却	6	6	816D	5374	却		
3825	疊	1	1	8178	53E0	疊		疊，異體字問題。
3826	叶	1	1	8179	53F6	叶		
3829	咏	3	3	817C	548F	咏		
3834	嗶	1	1	81A3	5637	嗶		
3840	坂	2	2	81A9	5742	坂		
3843	堦	15	15	81AC	5826	堦		
3859	尔	2	2	81BC	5C14	尔		爾，異體字問題。
3864	峯	1	1	81C1	5CEF	峯		峰，異體字問題。
3866	崐	1	1	81C3	5D10	崐		
3869	币	5	5	81C6	5E00	币		
3891	慙	6	6	81DC	6159	慙		
3892	喉	7	7	81DD	3B0B		日 𠂔 侯	
3897	慙	2	2	81E2	615C	慙		
3903	簞	1	1	81E8	25BE4		𠂔 𠂔 韋	

編號	造字	字頻 (txt)	字頻 (xml)	Big5	Unicode	WinXP	構字式	備註
3924	朥	1	1	81FD	671E	朥		
3937	檝	1	1	824B	6A9D	檝		
3940	蓳	11	11	824E	6B1D	蓳		
3974	牀	1	1	8270	7240	牀		
3984	瑠	3	3	827A	7460	瑠		
3996	疎	19	19	82A8	758E	疎		
4038	竈	10	10	82D2	7AC8	竈		
4039	竝	1	1	82D3	7ADA	竝		
4040	竝	2	2	82D4	7ADD	竝		
4041	竝	10	10	82D5	7AEA	竝		
4042	笋	1	1	82D6	7B0B	笋		
4044	筴	4	4	82D8	7B6F	筴		
4045	筴	3	3	82D9	7B92	筴		
4078		2	2	82FA			已𠂇升	异，異體字問題。
4089	脇	5	5	8346	8107	脇		
4095	舩	1	1	834C	8229	舩		

編號	造字	字頻 (txt)	字頻 (xml)	Big5	Unicode	WinXP	構字式	備註
4105	菓	9	9	8356	83D3	菓		
4109	葛	1	1	835A	84AD	葛		
4121	懼	1	1	8366	39AC		忄 厶 羅	
4139	袴	1	1	8378	88B4	袴		
4146	覩	69	69	83A1	89A9	覩		
4156	諂	2	2	83AB	8B1F	諂		
4173	耽	13	13	83BC	8EAD	耽		
4184	迹	2	2	83C7	8FF9	迹		跡，異體字問題。
4193	鈎	10	10	83D0	920E	鈎		
4195	鉢	50	50	83D2	9262	鉢		
4198	鑛	1	1	83D5	945B	鑛		
4203	隣	1	1	83DA	96A3	隣		鄰，異體字問題。
4207	霧	1	1	83DE	9736	霧		
4221	頰	1	1	83EC	9833	頰		
4226	飡	3	3	83F1	98E1	飡		
4267	鼈	1	1	845B	9F08	鼈		

編號	造字	字頻 (txt)	字頻 (xml)	Big5	Unicode	WinXP	構字式	備註
4271	𪗇	1	1	845F	9F69	𪗇		
4278	𪗈	13	13	8466	4E98	𪗈		
4307	罵	1	1	84A5	99E1	罵		罵，異體字問題。
4336	𪗉	6	6	84C2	278BB		角𪗉牛	
4350	侶	1	1	84D0	5058	侶		
4389	𪗊	4	4	84F7	746B	𪗊		
4395	𪗋	1	1	84FD	9721	𪗋		
4460	勅	60	60	85A1	52D1	勅		
4463	攏	1	1	85A4	3A72		扌𪗌隹	
4489		1	1	85BE			口𪗍𠂔	号，異體字問題。
4495	𪗎	2	2	85C4	9EA8	𪗎		
4508	攖	1	1	85D1	3A1E		扌𪗏素	
4546	𪗐	1	1	85F7	9D76	𪗐		
4566	𪗑	5	5	864C	78F5	𪗑		
4581	𪗒	1	1	865B	847C	𪗒		
4711	𪗓	1	1	8740	2145E		土𪗔遂	

編號	造字	字頻 (txt)	字頻 (xml)	Big5	Unicode	WinXP	構字式	備註
4738	睺	3	3	875B	777A	睺		
4750	卅	13	13	8767			卅	卅，異體字問題。
4816	亻	1	1	87CB			亻	
4908	屆	3	3	8868	5C4A	屆		屆，異體字問題。
4909	弥	5	5	8869	5F25	弥		彌，異體字問題。
5008	幪	3	3	88EE	5E5E	幪		
5013	彳	4	4	88F3			彳	
5027	牕	1	1	8942	724E	牕		
5034	弃	7	7	8949	5F03	弃		棄，異體字問題。
5051	輦	1	1	895A	4870		輦△車	
5071	忝	6	6	896E	5FE9	忝		
5099	惣	40	40	89AC	60E3	惣		
5115	颶	2	2	89BC	4B03		風△立	
5121		1	1	89C2			𠂇𠂇 丰方攴	𠂇 △教，校對問題，舊版漢籍錯字，可用程式全部取代。
5147	恠	12	12	89DC	6060	恠		

編號	造字	字頻 (txt)	字頻 (xml)	Big5	Unicode	WinXP	構字式	備註
5149	𩇛	2	2	89DE	2676B		要厶月	腰，異體字問題。
5179	𩇛	6	6	89FC	99C8	𩇛		
5186	𩇛	1	1	8A44	5C17	𩇛		
5206	𩇛	1	1	8A58	29573		寧厶頁	
5300	𩇛	3	3	8AD8	9B2A	𩇛		
5315	𩇛	3	3	8AE7	3CC2		𩇛 厶公	
5327	𩇛	1	1	8AF3	5D2A	𩇛		
5338	𩇛	1	1	8AFE	668E	𩇛		
5360	𩇛	2	2	8B55	913D	𩇛		
5396	𩇛	1	1	8B79	938B	𩇛		
5439	𩇛	3	3	8BC6	6822	𩇛		
5478	𩇛	24	24	8BED	98F0	𩇛		
5493	𩇛	1	1	8BFC			𩇛	
5542	𩇛	6	6	8C6E	93C1	𩇛		
5555	𩇛	7	7	8C7B	9008	𩇛		
5588	𩇛	3	3	8CBE	6061	𩇛		

編號	造字	字頻 (txt)	字頻 (xml)	Big5	Unicode	WinXP	構字式	備註
5621	泪	2	2	8CDF	6CEA	泪		淚，異體字問題。
5753	潜	9	9	8DC6	6F5C	潜		潛，異體字問題。
5769	邠	11	11	8DD6	90CD	邠		
5860	无	9	9	C6D3	65E0	无		

附件二、「祖堂集.xml」手動取代表

編號	造字	檔案原文摘錄	手動取代結果
374	𠂔	遶轉兩三 <u>𠂔</u>	遶轉兩三

附件三、「祖堂集.xml」缺字外錯字取代表

檔案原文摘錄	手動取代結果
是第幾月 <u>個</u>	是第幾月 <u>？</u>
天 _ナ <u>𠂔</u> 罵什摩債負	天 _ナ <u>𠂔</u> 罵什摩債負