

# 《古夫于亭雜錄》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  
2008/8/11 葉冠孜 製作

一、《古夫于亭雜錄》一書(古夫于亭雜錄.xml)使用舊版造字 89 個，字頻 300 次，詳如附件一。這 89 個造字中，72 個可轉成 Windows XP 能顯示的字，字頻 226 次；另外 17 個字必須轉成構字式，字頻 74 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、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 4132 的「血~~𠄎~~𠄎」係「眾」字之異體，故以「眾」字取代。
- 2、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 3846 的「壻」字，Unicode 字型呈現為「壻」，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 2811 的「𠄎~~𠄎~~田」。

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。所以造字編號 3892 的「睽」字，Unicode 編碼為 3B0B，由於 Windows XP 並不支援，仍須使用構字式「日~~𠄎~~侯」。

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

附件一、《古夫于亭雜錄》造字分析表

編號	造字	字頻 (txt)	字頻 (xml)	Big5	Unicode	WinXP	構字式	備註
296	𠄎	4	4	FBEC			𠄎	
297	𠄎	17	17	FBED			𠄎	
298	𠄎	5	5	FBEE			𠄎	
543	段	2	2	FDA9			段	
717	𠄎	1	1	FEBA			𠄎	
770	𠄎	1	1	FEEF			𠄎	
772	𠄎	1	1	FEF1			𠄎	
814	𠄎	1	1	8E5C	87C7	𠄎		
821	𠄎	1	1	8E63	7718	𠄎		
1109	𠄎	1	1	9049	7496	𠄎		
1212	𠄎	8	8	90D2	39B7		𠄎 𠄎 戈	
1544	𠄎	1	1	92E4	7D59	𠄎		

編號	造字	字頻 (txt)	字頻 (xml)	Big5	Unicode	WinXP	構字式	備註
1642	脉	4	4	93A9	8109	脉		
1784	萼	3	3	9478	855A	萼		
1926	覓	1	1	9569	8994	覓		
1944	簪	1	1	957B	8A67	簪		
1981	讐	3	3	95C2	8B90	讐		
2811		3	3	9AEF			𠄎𠄎田	待造字。
3259	載	1	1	9DD8	97EF	載		
3787	冢	2	2	8152	51A2	冢		
3789	冲	5	5	8154	51B2	冲		
3791	况	6	6	8156	51B5	况		
3797	凭	2	2	815C	51ED	凭		
3814	却	6	6	816D	5374	却		
3816	厓	1	1	816F	5393	厓		
3817	厠	1	1	8170	53A0	厠		
3819	廡	3	3	8172	53AE	廡		
3826	叶	1	1	8179	53F6	叶		

編號	造字	字頻 (txt)	字頻 (xml)	Big5	Unicode	WinXP	構字式	備註
3829	咏	3	3	817C	548F	咏		
3831	喆	1	1	817E	5586	喆		
3846	壻	1	1	81AF	58FB	壻		Unicode 呈現差異。
3849	獎	1	1	81B2	596C	獎		
3858	寃	1	1	81BB	5BC3	寃		
3868	卮	1	1	81C5	5DF5	卮		
3892	睺	5	5	81DD	3B0B		日 厶 侯	
3932	稷	1	1	8246	6936	稷		
3968	煨	1	1	826A	7145	煨		
3974	牀	4	4	8270	7240	牀		
3979	猫	2	2	8275	732B	猫		
3990	甄	2	2	82A2	750E	甄		
4013	盃	2	2	82B9	76CC	盃		
4028	稟	1	1	82C8	7980	稟		
4036	窰	5	5	82D0	7AB0	窰		
4041	豎	1	1	82D5	7AEA	豎		

編號	造字	字頻 (txt)	字頻 (xml)	Big5	Unicode	WinXP	構字式	備註
4055	粍	1	1	82E3	7C83	粍		
4068	綫	1	1	82F0	7DAB	綫		
4080	羣	19	19	82FC	7FA3	羣		
4081	翱	1	1	82FD	7FFA	翱		
4107	葱	1	1	8358	8471	葱		
4129	虬	3	3	836E	866C	虬		
4132		15	15	8371			血𠂇𠂇	眾，異體字問題。
4134	衛	21	21	8373	885E	衛		
4146	覩	2	2	83A1	89A9	覩		
4153	識	2	2	83A8	8A90	識		
4173	駝	2	2	83BC	8EAD	駝		
4176	輒	2	2	83BF	8F19	輒		
4184	迹	5	5	83C7	8FF9	迹		
4188	邨	11	11	83CB	90A8	邨		
4193	鈎	4	4	83D0	920E	鈎		
4195	鉢	4	4	83D2	9262	鉢		

編號	造字	字頻 (txt)	字頻 (xml)	Big5	Unicode	WinXP	構字式	備註
4234	贏	3	3	83F9	9A58	贏		
4242	鬪	7	7	8442	9B2D	鬪		Unicode 呈現差異。
4251	鰕	5	5	844B	9C15	鰕		
4263	麩	1	1	8457	9EAF	麩		Unicode 呈現差異。
4265	鼃	3	3	8459	9F03	鼃		
4287	織	1	1	846F	7E8E	織		
4399	塼	1	1	8542	213D6		土 厶 盍	
4504	呬	1	1	85CD	541A	呬		
4507	吳	31	31	85D0	5434	吳		
4702	場	1	1	86F6	5872	場		
4735	壠	1	1	8758	58E0	壠		
4763	斆	1	1	8774	6586	斆		
4768	妬	2	2	8779	59AC	妬		
4785	贗	3	3	87AC	8D0B	贗		
4817	寶	2	2	87CC	5BF3	寶		
4902	充	1	1	8862	5156	充		

編號	造字	字頻 (txt)	字頻 (xml)	Big5	Unicode	WinXP	構字式	備註
5111	旆	1	1	89B8	65BE	旆		
5174		1	1	89F7			𠄎刀兂、◻	待造字。
5200		3	3	8A52			𠄎ノ冂米大◻	奧，異體字問題。
5218	寬	4	4	8A64	5BDB	寬		
5221		3	3	8A67			𠄎白米𠄎◻	粵，異體字問題。
5225	邨	3	3	8A6B	460F		血𠄎冂	
5310	烏	1	1	8AE2	8204	烏		
5434	湟	2	2	8BC1	6E7C	湟		
5512	𠄎	1	1	8C50	8C3A	𠄎		
5515	軌	1	1	8C53			車𠄎丸	
5552	𠄎	1	1	8C78	5F06	𠄎		
5707	棧	1	1	8D76	6830	棧		
5721	頓	2	2	8DA6	9814	頓		