

# 《寓圃雜記》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  
2008/8/12 葉冠孜 製作

一、《寓圃雜記》一書(寓圃雜記.xml)使用舊版造字 71 個，字頻 234 次，詳如附件一。這 71 個造字中，61 個可轉成 Windows XP 能顯示的字，字頻 218 次；另外 10 個字必須轉成構字式，字頻 16 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 4750 的「卅」字，原字應為「卅」。
- 2、標記問題：標記問題主要是漢籍電子文獻在舊轉新的過程中，採用了新的標記語言，而在修改標記時，對於某些標題句做了增刪的動作，因為這些標題句的內容包含了缺字，所以這些更改標記的動作造成 txt 檔與 xml 檔缺字頻次不符的情形，處理方式為將未修改的頻次以手動方式取代之。如編號 4507 的「吳」字，xml 檔在 93 頁多了一個標題，因而造成頻次不同，處理方式為將未修改的頻次手動取代之為「吳」。
- 3、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「血𠂇」係「眾」字之異體，故以「眾」字取代。
- 4、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 4145 的「羈」字，Unicode 字型呈現為「羈」，實際上仍為同一字。

5、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「𠂔田」。

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 1707 的「𠂔」字，Unicode 編碼為 349C，由於 Windows XP 並不支援，仍須使用構字式「𠂔與」。

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、txt 檔與 xml 檔頻次不同且造字分析表「備註」欄有備註問題情況者，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(寓圃雜記.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。文句後方標示(全部取代)者，表示執行全部取代的動作，不一一列舉所有原文。
- 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。文句後方標示(全部取代)者，表示執行全部取代的動作，不一一列舉所有修改後結果。

附件一、《寓圃雜記》造字分析表

編號	造字	字頻 (txt)	字頻 (xml)	Big5	Unicode	WinXP	構字式	備註
296	𡗗	1	1	FBEC			𡗗	
297	𡗘	2	2	FBED			𡗘	
1010	珏	2	2	8FA5	73CF	珏		
1226	罰	3	3	90E0	7F78	罰		
1642	脉	1	1	93A9	8109	脉		
1707	𡗙	1	1	93EA	349C		彳 𡗘 與	
1926	覓	4	4	9569	8994	覓		
1981	讐	1	1	95C2	8B90	讐		
2044	踈	1	1	9642	8E08	踈		
2046	踪	1	1	9644	8E2A	踪		
2138	缶	1	1	96C2	7F36	缶		
2502	颺	3	3	98F4	98C8	颺		
2811		1	1	9AEF			亞 𡗗 田	待造字。
3096	歲	1	1	9CD2	5D57	歲		
3789	冲	1	1	8154	51B2	冲		

編號	造字	字頻 (txt)	字頻 (xml)	Big5	Unicode	WinXP	構字式	備註
3791	况	6	6	8156	51B5	况		
3796	决	1	1	815B	51B3	决		
3814	却	4	4	816D	5374	却		
3822	叁	1	1	8175	53C1	叁		
3824	叙	2	2	8177	53D9	叙		
3858	寃	8	8	81BB	5BC3	寃		
3868	卮	1	1	81C5	5DF5	卮		
3883	徧	1	1	81D4	5FA7	徧		
3884	忽	1	1	81D5	6031	忽		
3899	担	2	2	81E4	62C5	担		
3935	槩	7	7	8249	69E9	槩		Unicode 呈現差異。
3940	爵	5	5	824E	6B1D	爵		
3961	灾	1	1	8263	707E	灾		
3966	烟	3	3	8268	70DF	烟		
3968	煨	1	1	826A	7145	煨		
3978	猪	2	2	8274	732A	猪		Unicode 呈現差異。

編號	造字	字頻 (txt)	字頻 (xml)	Big5	Unicode	WinXP	構字式	備註
3979	猫	2	2	8275	732B	猫		
4012	鼓	2	2	82B8	76B7	鼓		
4016	着	1	1	82BC	7740	着		Unicode 呈現差異。
4036	窰	1	1	82D0	7AB0	窰		
4038	竈	1	1	82D2	7AC8	竈		
4067	綉	6	6	82EF	7D89	綉		
4080	羣	9	9	82FC	7FA3	羣		
4085	耻	3	3	8342	803B	耻		
4089	脇	2	2	8346	8107	脇		
4107	葱	1	1	8358	8471	葱		
4114	藥	1	1	835F	8616	藥		
4115	藁	1	1	8360	85C1	藁		
4129	虬	1	1	836E	866C	虬		
4132		5	5	8371			血𠂇𠂇	眾，異體字問題。
4134	衛	8	8	8373	885E	衛		
4145	羈	1	1	837E	898A	羈		Unicode 呈現差異。

編號	造字	字頻 (txt)	字頻 (xml)	Big5	Unicode	WinXP	構字式	備註
4167	賚	1	1	83B6	8CEB	賚		
4174	躑	1	1	83BD	8EB6	躑		
4176	輒	9	9	83BF	8F19	輒		
4193	鈎	1	1	83D0	920E	鈎		
4203	隣	3	3	83DA	96A3	隣		
4232	隲	2	2	83F7	96B2	隲		
4256	鷄	1	1	8450	9DC4	鷄		
4307	罵	1	1	84A5	99E1	罵		
4428	躑	1	1	855F	8E70	躑		
4460	勅	5	5	85A1	52D1	勅		
4478	淙	1	1	85B3	3CFF		汙 ㄇ 奈	
4507	吳	71	72	85D0	5434	吳		標記問題。
4750	卅	1	1	8767			卅	卅，校對問題，舊版漢籍錯字，可用程式全部取代。
4768	妬	1	1	8779	59AC	妬		
4785	贗	1	1	87AC	8D0B	贗		

編號	造字	字頻 (txt)	字頻 (xml)	Big5	Unicode	WinXP	構字式	備註
5019	畧	1	1	88F9			畧	
5176		2	2	89F9			彳 厶 冂	沉，異體字問題。
5218	寬	1	2	8A64	5BDB	寬		標記問題。
5225	邨	1	1	8A6B	460F		血 厶 卩	
5408	絀	1	1	8BA7	7D4D	絀		
5551	賈	1	1	8C77	8CF7	賈		
5555	迥	1	1	8C7B	9008	迥		
5559	欸	3	3	8CA1	6B35	欸		
5772	澆	6	6	8DD9	6FD9	澆		

附件二、《寓圃雜記》手動取代表

編號	造字	檔案原文摘錄	手動取代結果
4507	吳	<u>垂</u> (全部取代)	<u>吳</u> (全部取代)
5218	寬	<u>半</u> (全部取代)	<u>寬</u> (全部取代)