

《穀山筆塵》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
2008/8/13 葉冠孜 製作

一、《穀山筆塵》一書(穀山筆塵.xml)使用舊版造字 141 個，字頻 539 次，詳如附件一。這 141 個造字中，112 個可轉成 Windows XP 能顯示的字，字頻 421 次；另外 29 個字必須轉成構字式，字頻 118 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 1332 的「扌𠂔𠂔」字，原字應為「扌𠂔崔乃」。
- 2、舊版漢籍錯字，必須人工取代：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，然而在所有頻次中，只有一部分是錯字，另一部分是正確的，因此無法全部取代，必須人工一一比對，將錯誤的部分更正，如編號 5710「𠂔」，共出現 5 次，但原書字形有 4 次為「禾𠂔」字，1 次為「𠂔」字，必須手動取代。
- 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「血𠂔」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 5749 的「片𠄎庸」。

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 3803 的「𩺰」字，Unicode 編碼為 4B54，由於 Windows XP 並不支援，仍須使用構字式「食𠄎追」。

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 檔頻次不同或頻次相同而需手動取代、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、取代：實際使用字形與舊漢籍造字不同時，於此欄列出。
- 丁、檔案原文摘錄：摘錄檔案(穀山筆塵.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。
- 戊、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、《穀山筆塵》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	10	10	FBEC			𠄎	
297	𠄏	5	5	FBED			𠄏	
298	𠄐	6	6	FBEE			𠄐	
299	𠄑	2	2	FBEF			𠄑	
300	𠄒	2	2	FBF0			𠄒	
429	𠄓	3	3	FCD4			𠄓	
639	𠄔	1	1	FE4A			𠄔	
739	𠄕	2	2	FED0			𠄕	
746	𠄖	1	1	FED7			𠄖	
759	𠄗	1	1	FEE4			𠄗	
765	𠄘	2	2	FEEA			𠄘	
784	𠄙	1	1	FEFD			𠄙	
994	𠄚	1	1	8F73	8E75	𠄚		
1133	𠄛	1	1	9061	74CC	𠄛		
1167	𠄜	3	3	90A5	7D97	𠄜		

編號	造字	字頻 (txt)	字頻 (xml)	Big5	Unicode	WinXP	構字式	備註
1226	罰	5	5	90E0	7F78	罰		
1285	礪	3	3	915C	40CB		石厶俞	
1332	儻	1	1	91AD			扌厶嵩	𠄎扌崔乃𠄎，舊漢籍 錯字，可用程式全部 取代。
1462	驗	1	1	9270	299B1		馬厶會	
1490	羸	1	1	92AE	7C5D	羸		
1555	緡	1	1	92EF	7DCD	緡		
1591	羗	12	12	9354	7F97	羗		
1601	袞	3	3	935E	4646		衤厶癸	
1646	脛	1	1	93AD	8117	脛		
1893	袞	5	5	9548	886E	袞		
1913	褱	2	2	955C	8948	褱		
1981	讐	1	1	95C2	8B90	讐		
1999	贊	1	1	95D4	8CDB	贊		
2044	踈	10	10	9642	8E08	踈		
2050	踈	1	1	9648	8E4F	踈		

編號	造字	字頻 (txt)	字頻 (xml)	Big5	Unicode	WinXP	構字式	備註
2107	胞	3	3	96A3	9AB2	胞		Unicode 呈現差異。
2197	鉄	2	2	96FD	9244	鉄		
2305	鎖	1	1	97CC	93BB	鎖		
2441	箸	1	1	98B7	3E19		父△者	
2524	饊	1	1	994B	994D	饊		
2829	月	1	1	9B42			月	
3789	冲	10	10	8154	51B2	冲		
3803	餽	2	2	8162	4B54		食△追	
3804	効	11	11	8163	52B9	効		
3805	勅	23	23	8164	52C5	勅		
3814	却	13	13	816D	5374	却		
3819	廝	1	1	8172	53AE	廝		
3824	叙	10	10	8177	53D9	叙		
3826	叶	2	2	8179	53F6	叶		
3829	咏	3	3	817C	548F	咏		
3836	喻	2	2	81A5	564F	喻		

編號	造字	字頻 (txt)	字頻 (xml)	Big5	Unicode	WinXP	構字式	備註
3840	坂	4	4	81A9	5742	坂		
3843	堦	1	1	81AC	5826	堦		
3852	媼	1	1	81B5	5A63	媼		
3858	寃	1	1	81BB	5BC3	寃		
3860	尠	1	1	81BD	5C20	尠		
3876	廻	2	2	81CD	5EFB	廻		
3877	迺	2	2	81CE	5EFC	迺		
3883	徧	2	2	81D4	5FA7	徧		
3891	慙	1	1	81DC	6159	慙		
3895	淑	1	1	81E0	6E86	淑		
3940	薈	2	2	824E	6B1D	薈		
3946	汹	1	1	8254	6C79	汹		
3966	烟	4	4	8268	70DF	烟		
3974	牀	7	7	8270	7240	牀		
3976	犁	1	1	8272	7282	犁		
3977	狗	4	4	8273	72E5	狗		

編號	造字	字頻 (txt)	字頻 (xml)	Big5	Unicode	WinXP	構字式	備註
3978	猪	2	2	8274	732A	猪		Unicode 呈現差異。
3989	璿	1	1	82A1	74C8	璿		
3990	甄	1	1	82A2	750E	甄		
3996	疎	4	4	82A8	758E	疎		
4000	疴	1	1	82AC	75B4	疴		
4004	瘠	1	1	82B0	763B	瘠		
4010	阜	1	1	82B6	7690	阜		
4012	鼓	3	3	82B8	76B7	鼓		
4015	眦	4	4	82BB	7726	眦		
4016	着	11	11	82BC	7740	着		Unicode 呈現差異。
4028	稟	2	2	82C8	7980	稟		
4041	豎	1	1	82D5	7AEA	豎		
4044	筋	1	1	82D8	7B6F	筋		
4066	綺	1	1	82EE	7D5D	綺		
4068	綫	1	1	82F0	7DAB	綫		
4080	羣	35	35	82FC	7FA3	羣		

編號	造字	字頻 (txt)	字頻 (xml)	Big5	Unicode	WinXP	構字式	備註
4082	耇	1	1	82FE	8008	耇		耇，舊漢籍錯字，可用程式全部取代
4085	耻	5	5	8342	803B	耻		
4089	脇	2	2	8346	8107	脇		
4090	脚	6	6	8347	811A	脚		
4091	臯	7	7	8348	81EF	臯		
4107	葱	1	1	8358	8471	葱		
4114	藜	1	1	835F	8616	藜		
4115	藁	1	1	8360	85C1	藁		
4124	蝨	1	1	8369	8771	蝨		
4129	虬	1	1	836E	866C	虬		
4131	𠂔	2	2	8370	8842	𠂔		
4132		46	46	8371			血𠂔𠂔	眾，異體字問題。
4134	衛	6	6	8373	885E	衛		
4145	羈	1	1	837E	898A	羈		Unicode 呈現差異。
4146	覩	2	2	83A1	89A9	覩		
4173	駝	4	4	83BC	8EAD	駝		

編號	造字	字頻 (txt)	字頻 (xml)	Big5	Unicode	WinXP	構字式	備註
4174	𦍋	1	1	83BD	8EB6	𦍋		
4176	輒	1	1	83BF	8F19	輒		
4184	迹	8	8	83C7	8FF9	迹		
4193	鈎	9	9	83D0	920E	鈎		
4203	隣	1	1	83DA	96A3	隣		
4219	類	1	1	83EA	982E	類		
4232	隲	2	2	83F7	96B2	隲		
4242	鬪	1	1	8442	9B2D	鬪		Unicode 呈現差異。
4256	鷄	2	2	8450	9DC4	鷄		
4266	鼃	1	1	845A	9F02	鼃		
4309	刼	2	2	84A7	523C	刼		
4320	臭	1	1	84B2	369F		𦍋𠂇大	
4450	𦍋	7	7	8575	524F	𦍋		Unicode 呈現差異。
4461	勛	2	2	85A2	52D4	勛		
4469	勺	2	2	85AA	5304	勺		
4507	吳	50	50	85D0	5434	吳		

編號	造字	字頻 (txt)	字頻 (xml)	Big5	Unicode	WinXP	構字式	備註
4515	呪	1	1	85D8	546A	呪		
4584	嘉	5	5	865E	569E	嘉		
4603	憇	1	1	8671	6187	憇		
4607	狫	1	1	8675	3E60		犭 厶 屯	
4715	嘸	1	1	8744	5688	嘸		
4735	墮	1	1	8758	58E0	墮		
4768	妬	7	7	8779	59AC	妬		
4904	尅	1	1	8864	5C05	尅		
4957	崑	5	5	88BB	5D53	崑		
5008	幞	1	1	88EE	5E5E	幞		
5176		12	12	89F9			冫 厶 冗	沉，異體字問題。
5290	携	6	6	8ACE	643A	携		
5299	戟	1	1	8AD7	39B8		卓 厶 戈	
5358	麓	4	4	8B53	23BDB		阝 牙 女 厂 毛 阝	
5422	梯	1	1	8BB5	67F9	梯		
5439	栢	2	2	8BC6	6822	栢		

編號	造字	字頻 (txt)	字頻 (xml)	Big5	Unicode	WinXP	構字式	備註
5467	𦉳	1	1	8BE2	7439	𦉳		
5515	軌	2	2	8C53			車𠂔丸	
5555	迥	2	2	8C7B	9008	迥		
5559	欸	2	2	8CA1	6B35	欸		
5578	弛	1	1	8CB4	38AE		弓𠂔屯	
5588	恠	1	1	8CBE	6061	恠		
5704	𧯓	1	1	8D73	6C0A	𧯓		
5706	敍	3	3	8D75	654D	敍		
5710	𧯓	5	5	8D79	7A06	𧯓		禾𠂔吕，舊版漢籍錯字，必須人工取代。Unicode 呈現差異。
5713	躓	1	1	8D7C	481E		足𠂔戚	
5720	垧	4	4	8DA5	578D	垧		
5721	頓	3	3	8DA6	9814	頓		
5742	澁	1	1	8DBB	6F81	澁		
5749		1	1	8DC2			片𠂔庸	待造字。
5753	潜	3	3	8DC6	6F5C	潜		

附件二、《穀山筆塵》手動取代表

編號	造字	取代	檔案原文摘錄	手動取代結果
5710	稻	禾  呂	禾不因種而生曰 <u>猥</u>	禾不因種而生曰 <u>禾呂</u>
			貴家大族皆自出禾 <u>猥</u>	貴家大族皆自出禾 <u>禾呂</u>
			今所謂 <u>猥</u> 生是也	今所謂 <u>禾呂</u> 生是也
			以上數「 <u>猥</u> 」字	以上數「 <u>禾呂</u> 」字
	稻	天啟本皆作「 <u>猥</u> 」	天啟本皆作「 <u>稻</u> 」	