

《顏元集》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

2008/4/17 陳建安 製作

一、《顏元集》一書(顏元集.xml)使用舊版造字 212 個，字頻 1398 次，詳如附件一。這 212 個造字中，174 個可轉成 Windows XP 能顯示的字，字頻 1188 次；另外 38 個字必須轉成構字式，字頻 210 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 5516 的「○○束」字，原字為「棘」。
- 2、校對問題，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，有一部分是錯字，一部分又是另外一個字，又由於頻次相同，故無法讓電腦自動取代，必須人工一一比對更正。如編號 4750「廿」，共出現 3 次，但原書字形有 1 次為「廿」字，2 次為「𠂇」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 1000 的「王𠂇已」係「玕」字之異體，故以「玕」字取代。又編號 3825 的「疊」字，係「疊」字之簡體字，亦以標準字的「疊」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 4158 的「譌」字，Unicode 字型呈現為「譌」，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「片𠂔庸」。

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 797 的「挽」字，Unicode 編碼為 39D6，由於 Windows XP 並不支援，仍須使用構字式「扌𠂔」。

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的欄位說明如下：

甲、編號：Big5 造字空間為 6217 個，編號由 1 到 6217。

乙、取代：當實際使用字形與造字不同時，在此標出。

丙、檔案原文摘錄：摘錄檔案(顏元集.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。

附件一、《顏元集》造字分析表

編號	造字	頻次 (txt)	頻次 (xml)	Big5	Unicode	WinXP	構字式	備註
33	葛	4	4	FA60	85A5	葛		
296	𠂔	18	18	FBEC			𠂔	

編號	造字	頻次 (txt)	頻次 (xml)	Big5	Unicode	WinXP	構字式	備註
297	𠂔	42	42	FBED			𠂔	
299	𠂔	4	4	FBEF			𠂔	
300	𠂔	4	4	FBF0			𠂔	
301	𠂔	1	1	FBF1			𠂔	
450		1	1	FCE9				丿，校對問題，舊版漢籍錯字，可用程式全部取代。
658	𠂔	2	2	FE5D			𠂔	
675	食	1	1	FE6E			食	
715	𠂔	1	1	FEB8			𠂔	
717	𠂔	1	1	FEBA			𠂔	
752	𠂔	3	3	FEDD			𠂔	
770	𠂔	3	3	FEEF			𠂔	
785	ノ	1	1	FEFE			ノ	
797	𠂔	1	1	8E4B	39D6		𠂔 𠂔 𠂔	
814	𠂔	1	1	8E5C	87C7	𠂔		𠂔，異體字問題。
820	𠂔	5	5	8E62	70F1	𠂔		
821	𠂔	2	2	8E63	7718	𠂔		
841	𠂔	1	1	8E77	7932	𠂔		

編號	造字	頻次 (txt)	頻次 (xml)	Big5	Unicode	WinXP	構字式	備註
893	𪗇	1	1	8ECD	5CE9	𪗇		
1000		1	1	8F79			王𪗇已	玘，異體字問題。
1201	昵	1	1	90C7	7724	昵		
1205	眎	2	2	90CB	770E	眎		
1226	罰	12	12	90E0	7F78	罰		
1267	𪗇	1	1	914A	6B1B	𪗇		
1278	碍	1	1	9155	788D	碍		
1326	𪗇	5	5	91A7	2213C		須𪗇巾	
1337	慌	1	1	91B2	385B		巾𪗇荒	
1354	褱	1	1	91C3	79A9	褱		
1383	糠	1	1	91E0	7A45	糠		
1513	糧	1	1	92C5	7CAE	糧		糧，異體字問題。
1547	𪗇	1	1	92E7	4E31	𪗇		
1565	𪗇	12	12	92F9	7DDC	𪗇		
1616	𪗇	1	1	936D	284E6		𪗇𪗇里	
1641	胆	1	1	93A8	80C6	胆		膽，異體字問題。
1646	𪗇	1	1	93AD	8117	𪗇		
1981	𪗇	13	13	95C2	8B90	𪗇		

編號	造字	頻次 (txt)	頻次 (xml)	Big5	Unicode	WinXP	構字式	備註
2046	踪	1	1	9644	8E2A	踪		
2053	躡	1	1	964B	8EA7	躡		
2151	𠂔	3	3	96CF	961D	𠂔		
2157	酎	1	1	96D5	9167	酎		
2168	醕	1	1	96E0	91BB	醕		
2408	鬧	1	1	9874	9599	鬧		鬧，異體字問題。
2501	颼	1	1	98F3	98C3	颼		
2651	鸛	2	2	99EC	9E1C	鸛		
2834	𠂔	1	1	9B47	6535	𠂔		
2885	𠂔	1	1	9B7A	544C	𠂔		
2989		3	3	9C45			鳥𠂔	𠂔，異體字問題。
3007	𠂔	1	1	9C57	70A7	𠂔		
3029	𠂔	1	1	9C6D	941D	𠂔		
3112		2	2	9CE2			●	繕，校對問題，舊版漢籍錯字，可用程式全部取代。
3245	𠂔	1	1	9DCA	7D24	𠂔		
3255	善	361	361	9DD4	5584	善		
3770	个	1	1	8141	4E2A	个		個，異體字問題。

編號	造字	頻次 (txt)	頻次 (xml)	Big5	Unicode	WinXP	構字式	備註
3787	冢	6	6	8152	51A2	冢		
3788	慤	5	5	8153	22871		殼𠂇心	慤，異體字問題。
3791	況	75	75	8156	51B5	況		況，異體字問題。
3792	凉	3	3	8157	51C9	凉		凉，異體字問題。
3794	湊	2	2	8159	51D1	湊		湊，異體字問題。
3801	剗	1	1	8160	5257	剗		
3804	効	2	2	8163	52B9	効		
3805	勅	2	2	8164	52C5	勅		
3814	却	13	13	816D	5374	却		
3815	𡗗	6	6	816E	537A	𡗗		
3817	廁	5	5	8170	53A0	廁		廁，異體字問題。
3819	廐	3	3	8172	53AE	廐		廐，異體字問題。
3820	厦	3	3	8173	53A6	厦		厦，異體字問題。
3824	叙	5	5	8177	53D9	叙		叙，異體字問題。
3825	疊	2	2	8178	53E0	疊		疊，異體字問題。
3826	叶	2	2	8179	53F6	叶		
3829	咏	5	5	817C	548F	咏		
3831	喆	3	3	817E	5586	喆		

編號	造字	頻次 (txt)	頻次 (xml)	Big5	Unicode	WinXP	構字式	備註
3852	嫵	11	11	81B5	5A63	嫵		
3860	尠	2	2	81BD	5C20	尠		
3864	峯	15	15	81C1	5CEF	峯		峰，異體字問題。
3868	卮	1	1	81C5	5DF5	卮		卮，異體字問題。
3869	币	1	1	81C6	5E00	币		
3874		1	1	81CB			𠂇 𠂇 稟	懍，異體字問題。
3883	徧	17	17	81D4	5FA7	徧		
3886	悞	2	2	81D7	609E	悞		
3889	慤	7	7	81DA	6164	慤		慤，異體字問題。
3891	慙	3	3	81DC	6159	慙		慙，異體字問題。
3924	朞	2	2	81FD	671E	朞		期，異體字問題。
3930	憵	15	15	8244	6901	憵		
3949	洩	1	1	8257	6D96	洩		
3965	裁	1	1	8267	70D6	裁		
3968	煅	3	3	826A	7145	煅		
3977	猢	2	2	8273	72E5	猢		
3978	猪	2	2	8274	732A	猪		Unicode 呈現差異。
3979	猫	2	2	8275	732B	猫		

編號	造字	頻次 (txt)	頻次 (xml)	Big5	Unicode	WinXP	構字式	備註
3990	甌	1	1	82A2	750E	甌		
3996	疎	4	4	82A8	758E	疎		
4000	疴	2	2	82AC	75B4	疴		
4008	癰	3	3	82B4	766F	癰		
4011	皞	1	1	82B7	76A1	皞		Unicode 呈現差異。
4013	𡗗	6	6	82B9	76CC	𡗗		
4016	着	1	1	82BC	7740	着		著，異體字問題。
4018	𦉳	1	1	82BE	77AF	𦉳		
4028	𦉳	11	11	82C8	7980	𦉳		𦉳，異體字問題。
4034	穉	8	8	82CE	7A49	穉		
4038	竈	27	27	82D2	7AC8	竈		
4041	豎	1	1	82D5	7AEA	豎		
4044	筋	3	3	82D8	7B6F	筋		
4045	箒	2	2	82D9	7B92	箒		
4055	粃	1	1	82E3	7C83	粃		
4068	綫	4	4	82F0	7DAB	綫		線，異體字問題。
4077	罇	1	1	82F9	7F47	罇		
4080	羣	67	67	82FC	7FA3	羣		群，異體字問題。

編號	造字	頻次 (txt)	頻次 (xml)	Big5	Unicode	WinXP	構字式	備註
4089	脇	1	1	8346	8107	脇		
4090	脚	18	18	8347	811A	脚		
4091	臯	2	2	8348	81EF	臯		
4104	莅	2	2	8355	8385	莅		
4107	葱	2	2	8358	8471	葱		
4114	蘖	2	2	835F	8616	蘖		
4115	藁	6	6	8360	85C1	藁		
4134	衛	27	27	8373	885E	衛		衛，異體字問題。
4138	裘	1	1	8377	88A0	裘		
4139	袴	2	2	8378	88B4	袴		
4140	衽	7	7	8379	88B5	衽		
4144	霸	1	1	837D	8987	霸		Unicode 呈現差異。
4145	羈	6	6	837E	898A	羈		Unicode 呈現差異。
4146	覩	18	18	83A1	89A9	覩		
4158	譌	1	1	83AD	8B4C	譌		Unicode 呈現差異。
4173	𨔵	2	2	83BC	8EAD	𨔵		
4176	輒	4	4	83BF	8F19	輒		
4178	輓	3	3	83C1	8F2D	輓		

編號	造字	頻次 (txt)	頻次 (xml)	Big5	Unicode	WinXP	構字式	備註
4184	迹	43	43	83C7	8FF9	迹		
4186	遡	2	2	83C9	9061	遡		
4188	邨	1	1	83CB	90A8	邨		
4190	邨	1	1	83CD	90C4	邨		
4193	鈎	1	1	83D0	920E	鈎		
4203	隣	4	4	83DA	96A3	隣		鄰，異體字問題。
4208	磧	2	2	83DF	9767	磧		
4213	韡	3	3	83E4	97BE	韡		
4217	韡	2	2	83E8	97E0	韡		
4218	萑	1	1	83E9	97EE	萑		
4220	穎	6	6	83EB	9834	穎		穎，異體字問題。
4234	羸	1	1	83F9	9A58	羸		
4237	髡	1	1	83FC	9AE0	髡		髡，異體字問題。
4240	髡	3	3	8440	9AF4	髡		
4242	鬪	7	7	8442	9B2D	鬪		Unicode 呈現差異。
4250	鰭	2	2	844A	9C0C	鰭		
4256	鷄	7	7	8450	9DC4	鷄		
4262	麪	1	1	8456	9EAA	麪		

編號	造字	頻次 (txt)	頻次 (xml)	Big5	Unicode	WinXP	構字式	備註
4267	鼈	1	1	845B	9F08	鼈		
4268	𧈧	1	1	845C	9F17	𧈧		
4278	亘	1	1	8466	4E98	亘		
4287	纖	1	1	846F	7E8E	纖		纖，異體字問題。
4307	罵	14	14	84A5	99E1	罵		罵，異體字問題。
4308	刼	2	2	84A6	5227	刼		劫，異體字問題。
4309	刼	2	2	84A7	523C	刼		劫，異體字問題。
4447	捏	1	1	8572	63D1	捏		捏，異體字問題。
4460	勅	5	5	85A1	52D1	勅		
4476		47	47	85B1			良 Δ P	即，異體字問題。
4482	厯	1	1	85B7	53AA	厯		
4496	段	1	1	85C5	53DA	段		段，異體字問題。
4504	𠂔	4	4	85CD	541A	𠂔		
4507	𠂔	33	33	85D0	5434	𠂔		𠂔，異體字問題。
4515	呪	1	1	85D8	546A	呪		
4659	垲	4	4	86CB	57DB	垲		
4698	𡵓	1	1	86F2	5850	𡵓		
4702	場	1	1	86F6	5872	場		

編號	造字	頻次 (txt)	頻次 (xml)	Big5	Unicode	WinXP	構字式	備註
4750	𠂔	3	3	8767			𠂔	𠂔、𠂔，校對問題，舊版漢籍錯字，必須人工取代。
4761	𣎵	1	1	8772	6AA9	𣎵		𣎵，異體字問題。
4767	懷	3	3	8778	61B9	懷		
4768	妬	1	1	8779	59AC	妬		妒，異體字問題。
4805	瀆	1	1	87C0	51DF	瀆		瀆，異體字問題。
4823	窑	1	1	87D2	7A91	窑		
4861	嫫	1	1	87F8	5B3A	嫫		
4878	冗	3	3	884A	5B82	冗		冗，異體字問題。
4953	𢇛	1	1	88B7	22835		𢇛 𢇛𢇛𢇛	
4996	斲	3	3	88E2	65B5	斲		
5008	幞	4	4	88EE	5E5E	幞		
5015	甯	3	3	88F5	5BD7	甯		
5019	畧	9	9	88F9			畧	略，異體字問題。
5094	𦍋	1	1	89A7	26407		𦍋	
5095	来	4	4	89A8	6765	来		來，異體字問題。
5101	馱	1	1	89AE	4B7E		馬 𠂔 犬	馱，異體字問題。
5111	旆	1	1	89B8	65BE	旆		旆，異體字問題。

編號	造字	頻次 (txt)	頻次 (xml)	Big5	Unicode	WinXP	構字式	備註
5131	窻	1	1	89CC	7ABB	窻		Unicode 呈現差異。
5147	恠	1	1	89DC	6060	恠		
5163	憂	1	1	89EC	621E	憂		
5176		3	3	89F9			氵厶冗	沉，異體字問題。
5200		2	2	8A52			𠂔ノ冂米大	奧，異體字問題。
5218		28	28	8A64			𠂔ㄣ覓、	寬，異體字問題。
5220	𠂔	1	1	8A66	6BCC	𠂔		𠂔，校對問題，舊版漢籍錯字，可用程式全部取代。
5225	邨	6	6	8A6B	460F		血厶冂	
5229	撐	5	5	8A6F	6490	撐		
5289	慤	11	11	8ACD	6164	慤		慤，異體字問題。
5290	携	2	2	8ACE	643A	携		
5292	𠂔	3	3	8AD0	65C9	𠂔		
5300	鬪	1	1	8AD8	9B2A	鬪		
5310	烏	1	1	8AE2	8204	烏		
5381	𦍋	3	3	8B6A	8103	𦍋		Unicode 呈現差異。
5392	𦍋	3	3	8B75	60A4	𦍋		
5439	栢	1	1	8BC6	6822	栢		

編號	造字	頻次 (txt)	頻次 (xml)	Big5	Unicode	WinXP	構字式	備註
5466	倏	2	2	8BE1	5010	倏		Unicode 呈現差異。
5473	攷	1	1	8BE8	22BF1		扌𠂔𠂔	
5505	耑	1	1	8C49	264BE		老𠂔占	耑，校對問題，舊版漢籍錯字，可用程式全部取代。
5516		1	1	8C54			𠂔束	棘，校對問題，舊版漢籍錯字，可用程式全部取代。
5552	弃	1	1	8C78	5F06	弃		
5559	欸	1	1	8CA1	6B35	欸		
5579	窰	2	2	8CB5	5D08	窰		
5583	殮	1	1	8CB9	98F1	殮		
5590	朶	2	2	8CC0	6736	朶		
5591	杞	2	2	8CC1	233CC		木𠂔巳	杞，異體字問題。
5621	汨	5	5	8CDF	6CEA	汨		淚，異體字問題。
5679	簞	1	1	8D5A	7C51	簞		
5689	𡵓	1	1	8D64	20087		𡵓	
5702	厨	4	4	8D71	53A8	厨		廚，異體字問題。
5706	敘	33	33	8D75	654D	敘		敘，異體字問題。
5749		1	1	8DC2			片𠂔庸	待造字。

附件二、《顏元集》手動取代表

編號	取代	檔案原文摘錄	手動取代結果
4750	卅	庸 <u>卅</u> 卩之習者	庸 <u>卅</u> 卩之習者
		鋤 <u>卅</u> 賢	鋤 <u>卅</u> 賢
	廿	<u>卅</u> 史作金 欣經作	<u>卅</u> 史作金 欣經作