

# 《元好問全集》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  
2008/5/21 陳建安 製作

一、《元好問全集》一書(元好問全集.xml)使用舊版造字 277 個，字頻 1318 次，詳如附件一。這 277 個造字中，224 個可轉成 Windows XP 能顯示的字，字頻 1141 次；另外 53 個字必須轉成構字式，字頻 177 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 4433 的「其 $\Delta$ 少」字，原字為「𪛗」。

2、重碼問題，部分採用舊漢籍造字，部分採用部件外字，人工取代：舊漢籍造字與構字式部件外字，皆是建立在 Big5 的使用者造字區中，兩者所使用的碼區是重疊的，在同時使用構字式和舊漢籍造字的新版漢籍中，這些重疊的字碼可能被用以表達兩個不同字形，我們稱之為「重碼問題」。在使用構字式的 xml 檔中，舊漢籍造字也會呈現為重碼的部件外字，因而必須將採用舊漢籍造字部分以對應的 Unicode 或構字式取代，才能呈現正確字形；至於採用部件外字者則不需轉碼。如編號 5571，字碼 8CAD，對應的舊漢籍造字為「樞」，部件外字為「棊」，一一比對原書，確認是「樞」者，再以手動取代。

- 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 837 的「炎△毛」係「毯」字之異體，故以「毯」字取代。又編號 5316 的「晋」字，係「晉」字之簡體字，亦以標準字的「晉」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 3846 的「壻」字，Unicode 字型呈現為「壻」，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 1101 的「王△堂」。

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「掙」字，Unicode 編碼為 39D6，由於 Windows XP 並不支援，仍須使用構字式「扌△扌」。

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、造字：舊版造字字形。
  - 丙、檔案原文摘錄：摘錄檔案(元好問全集.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、《元好問全集》造字分析表

編號	造字	字頻 (txt)	字頻 (xml)	Big5	Unicode	WinXP	構字式	備註
44	颺	1	1	FA6B	295D7		風△貝	颺，異體字問題。
298	△	3	3	FBEE			△	
319	肩	3	3	FC44			肩	
362	考	1	1	FC6F			考	
489	隶	1	1	FD51			隶	
658	疒	2	2	FE5D			疒	
715	彡	1	1	FEB8			彡	
757	匕	1	1	FEE2			匕	
778	、	1	1	FEF7			、	
780	彡	1	1	FEF9			彡	
793	鞮	3	3	8E47	97E4	鞮		
797	扞	3	3	8E4B	39D6		扌△尾	
826	癰	1	1	8E68	7655	癰		
837	毳	1	1	8E73	3C9C		炎△毛	毳，異體字問題。
870	躡	2	2	8EB6	2A60F		足△齒	躡，異體字問題。
887	鷲	1	1	8EC7	9E0E	鷲		
912	鷓	1	1	8EE0	9DC0	鷓		

編號	造字	字頻 (txt)	字頻 (xml)	Big5	Unicode	WinXP	構字式	備註
972	犇	1	1	8F5D	7287	犇		
1010	珏	6	6	8FA5	73CF	珏		
1017	肩	5	5	8FAC	3795		肩	肩，異體字問題。
1024	虵	2	2	8FB3	8675	虵		
1039	稟	15	15	8FC2	69C0	稟		
1040	聃	1	1	8FC3	803C	聃		
1056	阨	1	1	8FD3	9641	阨		
1077	稟	2	2	8FE8	7A3E	稟		稿，異體字問題。
1080	丹	1	1	8FEB	5184	丹		
1088	暘	1	1	8FF3	757C	暘		
1101		1	1	9041			王𠂔堂	待造字。
1103	𠂔	1	1	9043	4731		谷𠂔曼	
1104	𠂔	1	1	9044	472A		谷𠂔九	
1126	璿	13	13	905A	74B9	璿		
1133	瓌	1	1	9061	74CC	瓌		
1151	劓	1	1	9073	5273	劓		
1205	眎	7	7	90CB	770E	眎		
1274	碣	1	1	9151	7854	碣		

編號	造字	字頻 (txt)	字頻 (xml)	Big5	Unicode	WinXP	構字式	備註
1332		27	27	91AD			扌厶嵩	攜，異體字問題。
1383	糠	1	1	91E0	7A45	糠		
1416	埃	1	1	9242	7AE2	埃		
1433	窆	1	1	9253	53DC	窆		
1513	糧	1	1	92C5	7CAE	糧		糧，異體字問題。
1523	砧	1	1	92CF	7872	砧		
1530	紕	1	1	92D6	7CFA	紕		
1532	搢	2	2	92D8	636A	搢		扌厶昏，校對問題，舊版漢籍錯字，可用程式全部取代。
1557	羴	1	1	92F1			羊厶歷	
1600	猷	1	1	935D	3E84		犛厶俞	
1642	脉	2	2	93A9	8109	脉		
1646	脰	2	2	93AD	8117	脰		
1658	囟	1	1	93B9	5705	囟		
1664	腸	1	1	93BF	8193	腸		腸，異體字問題。
1704	擘	1	1	93E7	3BAE		木厶挈	
1808	棚	3	3	94B2	3BB6		木厶朔	
1885	逢	2	2	9540	882D	逢		

編號	造字	字頻 (txt)	字頻 (xml)	Big5	Unicode	WinXP	構字式	備註
1893	袞	18	18	9548	886E	袞		袞，異體字問題。
1912	襍	1	1	955B	894D	襍		
1926	覓	2	2	9569	8994	覓		覓，異體字問題。
1981	讐	6	6	95C2	8B90	讐		
1999	贊	1	1	95D4	8CDB	贊		贊，異體字問題。
2043	蹕	1	1	9641	8E01	蹕		
2046	踪	2	2	9644	8E2A	踪		
2105	菱	7	7	96A1	84E1	菱		
2128	眈	1	1	96B8	6623	眈		
2157	酌	1	1	96D5	9167	酌		
2162	醜	1	1	96DA	918E	醜		
2168	醜	3	3	96E0	91BB	醜		
2188	羅	9	9	96F4	4366		囗 㠩 離	
2371	玠	2	2	984F	3EB9		王 㠩 弁	
2427	嶼	1	1	98A9			山 㠩 郭	
2452	蠹	3	3	98C2	45F6		虫 㠩 葛	
2472	靄	1	1	98D6	9741	靄		
2524	饒	2	2	994B	994D	饒		


編號	造字	字頻 (txt)	字頻 (xml)	Big5	Unicode	WinXP	構字式	備註
2569	磻	1	1	9978	40C9		石厶昏	磻，異體字問題。
2592	贓	2	2	99B1	8D1C	贓		贓，異體字問題。
2603	輦	1	1	99BC	4874		車厶童	
2648	灑	1	1	99E9	2A19F		灑厶勅厶鳥	
2650	鸛	3	3	99EB	9E18	鸛		
2651	鸚	2	2	99EC	9E1C	鸚		
2753	棧	1	1	9AB5	69F0	棧		
2811		24	24	9AEF			亞厶田	留，異體字問題。
2812	舉	1	1	9AF0	64E7	舉		舉，異體字問題。
2882	富	1	1	9B77	51A8	富		富，異體字問題。
3097		1	1	9CD3			彳厶幸	待造字。
3113	彥	1	1	9CE3	6245	彥		
3269	憑	1	1	9DE2	51F4	憑		
3780	雋	1	1	814B	5101	雋		
3788	慤	2	2	8153	22871		殼厶心	慤，異體字問題。
3789	冲	5	5	8154	51B2	冲		冲，異體字問題。
3791	况	79	79	8156	51B5	况		况，異體字問題。
3792	凉	5	5	8157	51C9	凉		凉，異體字問題。

編號	造字	字頻 (txt)	字頻 (xml)	Big5	Unicode	WinXP	構字式	備註
3796	決	76	76	815B	51B3	決		決，異體字問題。
3797	凭	4	4	815C	51ED	凭		憑，異體字問題。
3799	刼	1	1	815E	5226	刼		劫，異體字問題。
3801	剗	11	11	8160	5257	剗		
3802	劓	1	1	8161	529A	劓		
3805	勅	1	1	8164	52C5	勅		
3808	勾	1	1	8167	5303	勾		
3809	匱	12	12	8168	5333	匱		
3816	厓	34	34	816F	5393	厓		
3826	叶	15	15	8179	53F6	叶		
3829	咏	10	10	817C	548F	咏		
3831	喆	2	2	817E	5586	喆		
3833	嘑	1	1	81A2	5611	嘑		
3835	噍	4	4	81A4	564D	噍		
3843	堦	2	2	81AC	5826	堦		
3846	堦	34	34	81AF	58FB	堦		Unicode 呈現差異。
3849	獎	9	9	81B2	596C	獎		獎，異體字問題。
3850	姪	1	1	81B3	59D9	姪		



編號	造字	字頻 (txt)	字頻 (xml)	Big5	Unicode	WinXP	構字式	備註
3851	姹	2	2	81B4	59F9	姹		
3852	媼	4	4	81B5	5A63	媼		
3855	孽	2	2	81B8	5B7C	孽		孽，異體字問題。
3860	尠	1	1	81BD	5C20	尠		
3866	崑	1	1	81C3	5D10	崑		崑，異體字問題。
3868	卮	5	5	81C5	5DF5	卮		卮，異體字問題。
3869	币	3	3	81C6	5E00	币		
3874		1	1	81CB			忄 厶 稟	慄，異體字問題。
3884	忽	6	6	81D5	6031	忽		
3909	擡	1	1	81EE	64E1	擡		
3911	攬	2	2	81F0	651F	攬		
3916	敷	2	2	81F5	656D	敷		
3929	黎	2	2	8243	68C3	黎		
3931	椀	3	3	8245	6900	椀		
3935	槩	13	13	8249	69E9	槩		概，異體字問題。
3949	涖	10	10	8257	6D96	涖		
3955	澗	2	2	825D	3D4E		氵 厶 間	澗，異體字問題。
3961	灾	6	6	8263	707E	灾		災，異體字問題。

編號	造字	字頻 (txt)	字頻 (xml)	Big5	Unicode	WinXP	構字式	備註
3965	裁	3	3	8267	70D6	裁		
3974	牀	59	59	8270	7240	牀		
3978	猪	3	3	8274	732A	猪		Unicode 呈現差異。
3980	獾	1	1	8276	7328	獾		
3981	珉	2	2	8277	73C9	珉		
3985	璫	1	1	827B	7447	璫		Unicode 呈現差異。
3987	穀	2	2	827D	7474	穀		
3990	甄	1	1	82A2	750E	甄		
3993	畀	1	1	82A5	2233F		畀	畀，異體字問題。
3994	畀	2	2	82A6	754A	畀		
3996	疎	12	12	82A8	758E	疎		
4008	癯	2	2	82B4	766F	癯		
4011	皤	1	1	82B7	76A1	皤		
4013	盃	7	7	82B9	76CC	盃		
4015	眦	2	2	82BB	7726	眦		
4024	礮	1	1	82C4	792E	礮		
4025	礮	5	5	82C5	7934	礮		
4028	稟	9	9	82C8	7980	稟		稟，異體字問題。

編號	造字	字頻 (txt)	字頻 (xml)	Big5	Unicode	WinXP	構字式	備註
4033	稽	3	3	82CD	7A2D	稽		
4034	穉	2	2	82CE	7A49	穉		
4035	穽	4	4	82CF	7A7D	穽		
4038	竈	12	12	82D2	7AC8	竈		
4040	竝	25	25	82D4	7ADD	竝		
4042	笋	5	5	82D6	7B0B	笋		
4044	筍	5	5	82D8	7B6F	筍		
4055	粃	1	1	82E3	7C83	粃		
4069	綑	1	1	82F1	7DB3	綑		
4081	駟	1	1	82FD	7FFA	駟		
4085	耻	8	8	8342	803B	耻		
4089	脇	7	7	8346	8107	脇		
4093	舐	2	2	834A	445B		舌  氏	舐，異體字問題。
4104	莅	9	9	8355	8385	莅		
4107	葱	1	1	8358	8471	葱		
4115	藁	9	9	8360	85C1	藁		
4129	虬	6	6	836E	866C	虬		
4130	衅	4	4	836F	8845	衅		

編號	造字	字頻 (txt)	字頻 (xml)	Big5	Unicode	WinXP	構字式	備註
4131	𠂔	2	2	8370	8842	𠂔		
4136	衰	2	2	8375	887A	衰		
4138	衰	2	2	8377	88A0	衰		
4142	衰	4	4	837B	8943	衰		
4146	覩	4	4	83A1	89A9	覩		
4155	訶	1	1	83AA	8B0C	訶		
4172	𠂔	3	3	83BB	8E77	𠂔		𠂔，異體字問題。
4178	輓	7	7	83C1	8F2D	輓		
4180	臯	1	1	83C3	8FA0	臯		
4181	薛	5	5	83C4	8FA5	薛		薛，異體字問題。
4184	迹	89	89	83C7	8FF9	迹		
4186	遡	1	1	83C9	9061	遡		
4193	鈎	6	6	83D0	920E	鈎		
4195	鉢	10	10	83D2	9262	鉢		
4198	鑛	4	4	83D5	945B	鑛		
4216	韃	5	5	83E7	97C8	韃		
4218	菲	1	1	83E9	97EE	菲		
4221	頰	1	1	83EC	9833	頰		

編號	造字	字頻 (txt)	字頻 (xml)	Big5	Unicode	WinXP	構字式	備註
4222	頰	3	3	83ED	9847	頰		
4225	顛	3	3	83F0	9866	顛		
4226	凜	1	1	83F1	98E1	凜		
4232	隲	1	1	83F7	96B2	隲		
4249	鱗	4	4	8449	9BFF	鱗		
4254	鷓	2	2	844E	9D7E	鷓		
4256	鷄	9	9	8450	9DC4	鷄		
4259	麇	1	1	8453	9E95	麇		Unicode 呈現差異。
4262	麪	8	8	8456	9EAA	麪		
4263	麴	2	2	8457	9EAF	麴		
4267	鼈	8	8	845B	9F08	鼈		
4268	鼈	1	1	845C	9F17	鼈		
4278	亘	4	4	8466	4E98	亘		
4301		1	1	847D			𠃉广夾土𠃉	瘞，異體字問題。
4308	劫	2	2	84A6	5227	劫		劫，異體字問題。
4345		1	1	84CB			口𠃉救	嗽，異體字問題。
4350	侶	2	2	84D0	5058	侶		
4399	壻	1	1	8542	213D6		土𠃉盍	

編號	造字	字頻 (txt)	字頻 (xml)	Big5	Unicode	WinXP	構字式	備註
4428	躡	1	1	855F	8E70	躡		躡，異體字問題。
4433		1	1	8564			其△少	𠂔，校對問題，舊版漢籍錯字，可用程式全部取代。
4437	廩	1	1	8568	5EEA	廩		廩，異體字問題。
4469	勺	8	8	85AA	5304	勺		
4480	岫	2	2	85B5	21D84		山△包	
4505	碯	1	1	85CE	25500		石△旬	
4531	攢	1	1	85E8	6505	攢		攢，異體字問題。
4533	唵	1	1	85EA	552B	唵		
4554	唬	1	1	8640	55C1	唬		
4566	礪	16	16	864C	78F5	礪		
4603	憇	2	2	8671	6187	憇		
4629	坳	4	4	86AD	577F	坳		
4637	时	1	1	86B5	65F6	时		時，異體字問題。
4698	瘞	2	2	86F2	5850	瘞		
4702	場	1	1	86F6	5872	場		場，異體字問題。
4735	壠	9	9	8758	58E0	壠		
4739	鞞	1	1	875C	97DB	鞞		

編號	造字	字頻 (txt)	字頻 (xml)	Big5	Unicode	WinXP	構字式	備註
4754	辰	5	5	876B	200A2		辰	
4766	揜	1	1	8777	6366	揜		
4768	妬	6	6	8779	59AC	妬		妒，異體字問題。
4769	憊	1	1	877A	60E5	憊		
4775	媿	10	10	87A2	59B7	媿		
4778	鞞	1	1	87A5	97B4	鞞		
4785	贗	7	7	87AC	8D0B	贗		贗，異體字問題。
4822	尪	2	2	87D1	5C29	尪		尪，異體字問題。
4878	宄	7	7	884A	5B82	宄		宄，異體字問題。
4902	兗	8	8	8862	5156	兗		兗，異體字問題。
4904	尅	1	1	8864	5C05	尅		
4907	橐	2	2	8867	69D6	橐		橐，異體字問題。
4927	岝	1	1	887B	5C9E	岝		
4940	韵	7	7	88AA	97F5	韵		
4957	崑	4	4	88BB	5D53	崑		
4996	斲	1	1	88E2	65B5	斲		
5008	幞	1	1	88EE	5E5E	幞		
5019	畧	1	1	88F9			畧	略，異體字問題。

編號	造字	字頻 (txt)	字頻 (xml)	Big5	Unicode	WinXP	構字式	備註
5048	婚	5	5	8957	36F0		女𠂔昏	婚，異體字問題。
5101	馱	2	2	89AE	4B7E		馬𠂔犬	馱，異體字問題。
5109	榘	1	1	89B6	6A10	榘		
5117	惚	1	1	89BE	6121	惚		
5126	暮	1	1	89C7	8B29	暮		
5127	澥	1	1	89C8	6EB5	澥		
5133	憇	2	2	89CE	6199	憇		
5169	戇	3	3	89F2	6239	戇		
5173	漱	1	1	89F6	6F44	漱		漱，異體字問題。
5176		6	6	89F9			氵𠂔冗	沉，異體字問題。
5225	邨	17	17	8A6B	460F		血𠂔卩	邨，異體字問題。
5229	撐	1	1	8A6F	6490	撐		
5239	擎	1	1	8A79	6486	擎		
5290	携	9	9	8ACE	643A	携		
5302	昏	5	5	8ADA	662C	昏		昏，異體字問題。
5306		10	10	8ADE			𠂔𠂔閒	簡，異體字問題。
5316	晋	2	2	8AE8	664B	晋		晉，異體字問題。
5338	暎	2	2	8AFE	668E	暎		



編號	造字	字頻 (txt)	字頻 (xml)	Big5	Unicode	WinXP	構字式	備註
5371	冃	4	4	8B60	808E	冃		
5381	脛	2	2	8B6A	8103	脛		Unicode 呈現差異。
5392	息	12	12	8B75	60A4	息		
5398	涓	1	1	8B7B	6E7B	涓		
5400	酌	3	3	8B7D	48F1		酉ㄩ句	
5409	盞	1	1	8BA8	76CB	盞		
5442	枿	3	3	8BC9	67BF	枿		
5451	玳	8	8	8BD2	7398	玳		
5467	琴	1	1	8BE2	7439	琴		
5484	楫	1	1	8BF3	6973	楫		
5515	軌	2	2	8C53			車ㄩ丸	
5541	蠹	1	1	8C6D	8827	蠹		
5549	藁	8	8	8C75	85F3	藁		
5552	奔	1	1	8C78	5F06	奔		
5571	樛	1	2	8CAD	6A45	樛		與棊重碼，重碼問題，部分採用舊漢籍造字，部分採用部件外字，人工取代。
5584	悞	1	1	8CBA	611E	悞		
5586	獻	1	1	8CBC	6B57	獻		

編號	造字	字頻 (txt)	字頻 (xml)	Big5	Unicode	WinXP	構字式	備註
5587	歛	1	1	8CBD	6B58	歛		
5590	朶	1	1	8CC0	6736	朶		朶，異體字問題。
5591	杞	4	4	8CC1	233CC		木𠂇巳	杞，異體字問題。
5654	籊	1	1	8D41	7C4B	籊		
5685	懣	2	2	8D60	22960		忄𠂇雙	
5703	𪗇	3	3	8D72	9F53	𪗇		
5704	𪗈	2	2	8D73	6C0A	𪗈		
5706	𪗉	13	13	8D75	654D	𪗉		𪗉，異體字問題。
5720	𪗊	3	3	8DA5	578D	𪗊		
5730	𪗋	4	4	8DAF	9638	𪗋		
5752	𪗌	6	6	8DC5	63D4	𪗌		

附件二、《元好問全集》手動取代表

編號	造字	檔案原文摘錄	手動取代結果
5571	樛	蓋規 <del>棊</del> 仰山而	蓋規 <del>樛</del> 仰山而