

《青箱雜記》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
2008/5/6 葉冠孜 製作

一、《青箱雜記》一書(青箱雜記.xml)使用舊版造字 98 個，字頻 341 次，詳如附件一。這 98 個造字中，83 個可轉成 Windows XP 能顯示的字，字頻 277 次；另外 15 個字必須轉成構字式，字頻 64 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 1572 的「革口王」係「革呈」字之異體，故以「革呈」字取代。又編號 3966 的「烟」字，係「煙」字之簡體字，亦以標準字的「煙」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。

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 4856 的「鬣」字，Unicode 編碼為 3CB2，由於 Windows XP 並不支援，仍須使用構字式「疊毛」。

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、對於在缺字之外發現的錯字，詳細取代內容可參考附件二的校出錯字表。附件二校出錯字表的欄位說明如下：

- 甲、檔案原文摘錄：摘錄檔案(青箱雜記.xml)中包含錯字的文句段落，其中紅字底線的部分，為需要手動取代的字形。
- 乙、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、《青箱雜記》造字分析表

編號	造字	頻次 (txt)	頻次 (xml)	Big5	Unicode	WinXP	構字式	備註
296	𠄎	1	1	FBEC			𠄎	
297	𠄎	24	24	FBED			𠄎	
358	𠄎	3	3	FC6B			𠄎	
429	𠄎	1	1	FCD4			𠄎	
675	𠄎	1	1	FE6E			𠄎	
1067	𠄎	1	1	8FDE	68CA	𠄎		
1331	𠄎	7	7	91AC	3813		山 𠄎 蒙	
1520	𠄎	1	1	92CC	3B20		日 𠄎 煞	
1530	𠄎	1	1	92D6	7CFA	𠄎		
1572		1	1	9341			𠄎革口王 𠄎	革 𠄎 呈，異體字問題

編號	造字	頻次 (txt)	頻次 (xml)	Big5	Unicode	WinXP	構字式	備註
1885	蠱	1	1	9540	882D	蠱		
1912	襍	1	1	955B	894D	襍		
2046	踪	1	1	9644	8E2A	踪		
2157	酌	2	2	96D5	9167	酌		
2834	攵	3	3	9B47	6535	攵		
3789	冲	4	4	8154	51B2	冲		沖，異體字問題
3791	況	2	2	8156	51B5	況		況，異體字問題
3792	涼	2	2	8157	51C9	涼		涼，異體字問題
3797	凭	4	4	815C	51ED	凭		憑，異體字問題
3801	剗	1	1	8160	5257	剗		
3804	効	6	6	8163	52B9	効		
3805	勅	4	4	8164	52C5	勅		
3814	却	8	8	816D	5374	却		
3835	噍	1	1	81A4	564D	噍		
3843	堦	3	3	81AC	5826	堦		
3846	墀	7	7	81AF	58FB	墀		Unicode 呈現差異
3858	寃	3	3	81BB	5BC3	寃		冤，異體字問題

編號	造字	頻次 (txt)	頻次 (xml)	Big5	Unicode	WinXP	構字式	備註
3883	徧	3	3	81D4	5FA7	徧		
3886	悞	2	2	81D7	609E	悞		
3891	慙	1	1	81DC	6159	慙		慙，異體字問題
3909	擡	1	1	81EE	64E1	擡		
3935	槩	5	5	8249	69E9	槩		概，異體字問題
3942	毡	1	1	8250	6BE1	毡		
3966	烟	3	3	8268	70DF	烟		煙，異體字問題
3974	牀	4	4	8270	7240	牀		
3977	狗	1	1	8273	72E5	狗		
3978	猪	2	2	8274	732A	猪		Unicode 呈現差異
3980	獾	2	2	8276	7328	獾		
3981	珉	1	1	8277	73C9	珉		
3991	攜	2	2	82A3	64D5	攜		
3996	疎	14	14	82A8	758E	疎		
4016	着	3	3	82BC	7740	着		著，異體字問題
4038	竈	4	4	82D2	7AC8	竈		
4045	箒	1	1	82D9	7B92	箒		

編號	造字	頻次 (txt)	頻次 (xml)	Big5	Unicode	WinXP	構字式	備註
4067	綉	2	2	82EF	7D89	綉		繡，異體字問題
4074	繚	1	1	82F6	7E27	繚		
4080	羣	14	14	82FC	7FA3	羣		群，異體字問題
4081	翱	1	1	82FD	7FFA	翱		
4090	脚	1	1	8347	811A	脚		
4091	臯	2	2	8348	81EF	臯		
4107	葱	2	2	8358	8471	葱		
4114	藥	1	1	835F	8616	藥		
4129	虬	1	1	836E	866C	虬		
4132		4	4	8371			血𠂇𠂇	眾，異體字問題
4134	衛	7	7	8373	885E	衛		衛，異體字問題
4139	袴	2	2	8378	88B4	袴		
4146	覩	9	9	83A1	89A9	覩		
4155	訶	2	2	83AA	8B0C	訶		
4158	譌	1	1	83AD	8B4C	譌		Unicode 呈現差異
4176	輒	1	1	83BF	8F19	輒		
4184	迹	4	4	83C7	8FF9	迹		

編號	造字	頻次 (txt)	頻次 (xml)	Big5	Unicode	WinXP	構字式	備註
4203	隣	11	11	83DA	96A3	隣		鄰，異體字問題
4220	穎	4	4	83EB	9834	穎		穎，異體字問題
4236	瞞	1	1	83FB	77D2	瞞		
4237	髡	1	1	83FC	9AE0	髡		髡，異體字問題
4242	鬪	1	1	8442	9B2D	鬪		Unicode 呈現差異
4254	鷓	1	1	844E	9D7E	鷓		
4271	齧	2	2	845F	9F69	齧		
4309	劫	1	1	84A7	523C	劫		劫，異體字問題
4387	啗	3	3	84F5	20E57		口 𠂔 哲	
4450	翔	2	2	8575	524F	翔		創，異體字問題
4460	勅	2	2	85A1	52D1	勅		敕，異體字問題
4507	吳	60	60	85D0	5434	吳		吳，異體字問題
4515	容	2	2	85D8	546A	呪		咒，異體字問題
4768	妬	2	2	8779	59AC	妬		妒，異體字問題
4791	羴	3	3	87B2	7F93	羴		
4856	毳	1	1	87F3	3CB2		疊 𠂔 毛	
4940	韻	1	1	88AA	97F5	韻		韻，異體字問題

編號	造字	頻次 (txt)	頻次 (xml)	Big5	Unicode	WinXP	構字式	備註
4957	崑	1	1	88BB	5D53	崑		
4961	盼	1	1	88BF	80B9	盼		盼，異體字問題
5025		1	1	8940			千△干	干，異體字問題
5028	昇	6	6	8943	662A	昇		
5034	弃	1	1	8949	5F03	弃		
5133	憇	1	1	89CE	6199	憇		
5147	恠	1	1	89DC	6060	恠		
5173	漱	1	1	89F6	6F44	漱		漱，異體字問題
5176		12	12	89F9			氵△冗	沉，異體字問題
5439	栢	1	1	8BC6	6822	栢		
5463	岢	2	2	8BDE	65F9	岢		
5483	眈	2	2	8BF2	7721	眈		
5542	鑠	3	3	8C6E	93C1	鑠		
5555	迥	1	1	8C7B	9008	迥		
5558	龕	3	3	8C7E	9E84	龕		
5590	朶	1	1	8CC0	6736	朶		朶，異體字問題
5592	桌	1	1	8CC2	3B9A		桌	

編號	造字	頻次 (txt)	頻次 (xml)	Big5	Unicode	WinXP	構字式	備註
5621	泪	1	1	8CDF	6CEA	泪		淚，異體字問題
5625	滙	3	3	8CE3	3D8F		彳 厶 隱	
5706	敘	1	1	8D75	654D	敘		敘，異體字問題

附件二、《青箱雜記》校出錯字表

檔案原文摘錄	手動取代結果
木 厶 數 厶 攴 字形。	木 厶 數 字形。