

《簷曝雜記》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
2008/6/2 陳維君 製作

一、《簷曝雜記》一書(簷曝雜記.xml)使用舊版造字 121 個，字頻 382 次，詳如附件一。這 121 個造字中，102 個可轉成 Windows XP 能顯示的字，字頻 284 次；另外 19 個字必須轉成構字式，字頻 98 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 3858 的「寃」係「冤」字之異體，故以「冤」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 1616 的「運」字，Unicode 編碼為 284E6，由於 Windows XP 並不支援，仍須使用構字式「里」。

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

附件一、《簷曝雜記》造字分析表

編號	造字	字頻 (txt)	字頻 (xml)	Big5	Unicode	WinXp	構字式	備註
74	髻	1	1	FAAB	29B06		髻彡𠂇	髻，異體字問題。
296	𠂇	2	2	FBEC			𠂇	
297	𠂈	5	5	FBED			𠂈	
298	𠂉	1	1	FBEE			𠂉	
717	𠂊	2	2	FEBA			𠂊	
752	𠂋	1	1	FEDD			𠂋	
784	𠂌	1	1	FEFD			𠂌	
814	𠂍	2	2	8E5C	87C7	𠂍		𠂍，異體字問題。
821	𠂎	1	1	8E63	7718	𠂎		
841	𠂏	1	1	8E77	7932	𠂏		𠂏，異體字問題。
853	𠂐	1	1	8EA5	7134	𠂐		
1006	𠂑	1	1	8FA1	4C50		魚𠂒孚	
1111	𠂓	1	1	904B	749D	𠂓		
1146	𠂔	2	2	906E	7506	𠂔		
1278	𠂕	1	1	9155	788D	𠂕		𠂕，異體字問題。

編號	造字	字頻 (txt)	字頻 (xml)	Big5	Unicode	WinXp	構字式	備註
1547	𠂇	1	1	92E7	4E31	𠂇		
1591	羗	1	1	9354	7F97	羗		羗，異體字問題。
1616	逕	1	1	936D	284E6		之△里	
1809	困	1	1	94B3	56E6	困		
1981	讐	1	1	95C2	8B90	讐		讐，異體字問題。
2087	霽	3	3	966D	291A9		雨△澍	
2162	醜	1	1	96DA	918E	醜		
2408	鬧	1	1	9874	9599	鬧		鬧，異體字問題。
2624		1	1	99D1			舟△義	舨，異體字問題。
3770	个	9	9	8141	4E2A	个		個，異體字問題。
3789	冲	1	1	8154	51B2	冲		沖，異體字問題。
3792	凉	1	1	8157	51C9	凉		涼，異體字問題。
3799	刼	7	7	815E	5226	刼		
3802	劓	1	1	8161	529A	劓		
3807	驗	5	5	8166	9A10	驗		
3814	却	2	2	816D	5374	却		卻，異體字問題。
3817	厠	2	2	8170	53A0	厠		廁，異體字問題。
3820	厦	1	1	8173	53A6	厦		廈，異體字問題。

編號	造字	字頻 (txt)	字頻 (xml)	Big5	Unicode	WinXp	構字式	備註
3825	疊	2	2	8178	53E0	疊		疊，異體字問題。
3826	叶	1	1	8179	53F6	叶		
3829	咏	3	3	817C	548F	咏		
3830	咤	1	1	817D	54A4	咤		
3831	喆	3	3	817E	5586	喆		
3840	坂	1	1	81A9	5742	坂		
3846	壻	2	2	81AF	58FB	壻		Unicode 呈現差異。
3849	獎	2	2	81B2	596C	獎		獎，異體字問題。
3858	冤	2	2	81BB	5BC3	冤		冤，異體字問題。
3864	峯	1	1	81C1	5CEF	峯		峰，異體字問題。
3868	卮	2	2	81C5	5DF5	卮		
3883	徧	2	2	81D4	5FA7	徧		
3886	悞	5	5	81D7	609E	悞		
3909	擡	2	2	81EE	64E1	擡		
3929	梨	1	1	8243	68C3	梨		梨，異體字問題。
3931	椀	1	1	8245	6900	椀		
3935	槩	1	1	8249	69E9	槩		概，異體字問題。
3936	棹	1	1	824A	6A70	棹		

編號	造字	字頻 (txt)	字頻 (xml)	Big5	Unicode	WinXp	構字式	備註
3938	橈	1	1	824C	6AC8	橈		
3966	烟	13	13	8268	70DF	烟		煙，異體字問題。
3968	煨	2	2	826A	7145	煨		
3974	牀	1	1	8270	7240	牀		
3975	牕	4	4	8271	7255	牕		Unicode 呈現差異。
3976	犁	1	1	8272	7282	犁		犁，異體字問題。
3978	猪	2	2	8274	732A	猪		豬，異體字問題。
3989	璿	3	3	82A1	74C8	璿		
3991	攜	6	6	82A3	64D5	攜		
3996	疎	2	2	82A8	758E	疎		
4016	着	5	5	82BC	7740	着		著，異體字問題。
4024	礮	5	5	82C4	792E	礮		
4028	稟	1	1	82C8	7980	稟		稟，異體字問題。
4041	豎	2	2	82D5	7AEA	豎		豎，異體字問題。
4044	筋	1	1	82D8	7B6F	筋		
4074	縑	1	1	82F6	7E27	縑		
4080	羣	7	7	82FC	7FA3	羣		群，異體字問題。
4090	脚	10	10	8347	811A	脚		腳，異體字問題。

編號	造字	字頻 (txt)	字頻 (xml)	Big5	Unicode	WinXp	構字式	備註
4091	臯	3	3	8348	81EF	臯		
4097	芑	1	1	834E			艹 厶 巳	芑，異體字問題。
4104	蒞	1	1	8355	8385	蒞		蒞，異體字問題。
4107	葱	3	3	8358	8471	葱		蔥，異體字問題。
4112	蔴	5	5	835D	8534	蔴		
4118	藜	2	2	8363	8602	藜		
4129	虬	2	2	836E	866C	虬		虬，異體字問題。
4131	𧈧	1	1	8370	8842	𧈧		𧈧，異體字問題。
4132		12	12	8371			血 厶 𠂇	眾，異體字問題。
4137	袜	1	1	8376	889C	袜		
4139	袴	2	2	8378	88B4	袴		
4146	覩	1	1	83A1	89A9	覩		
4167	賈	1	1	83B6	8CEB	賈		
4174	𧈧	1	1	83BD	8EB6	𧈧		
4176	輒	3	3	83BF	8F19	輒		
4184	迹	5	5	83C7	8FF9	迹		
4193	鈎	6	6	83D0	920E	鈎		鈎，異體字問題。
4242	鬪	8	8	8442	9B2D	鬪		Unicode 呈現差異。

編號	造字	字頻 (txt)	字頻 (xml)	Big5	Unicode	WinXp	構字式	備註
4262	麪	2	2	8456	9EAA	麪		
4278	亘	1	1	8466	4E98	亘		
4307	罵	5	5	84A5	99E1	罵		罵，異體字問題。
4309	刼	1	1	84A7	523C	刼		劫，異體字問題。
4342	倮	16	16	84C8	502E	倮		
4447	捏	1	1	8572	63D1	捏		捏，異體字問題。
4507	吳	22	22	85D0	5434	吳		吳，異體字問題。
4515	呪	4	4	85D8	546A	呪		咒，異體字問題。
4554	唬	1	1	8640	55C1	唬		
4603	憇	4	4	8671	6187	憇		
4769	患	1	1	877A	60E5	患		
4785	贗	3	3	87AC	8D0B	贗		贗，異體字問題。
4908	届	1	1	8868	5C4A	届		屆，異體字問題。
4933	崗	3	3	88A3	5CDD	崗		
5019	畧	27	27	88F9			畧	略，異體字問題。
5101	馱	2	2	89AE	4B7E		馬厶犬	馱，異體字問題。
5173	漱	1	1	89F6	6F44	漱		漱，異體字問題。
5176		1	1	89F9			氵厶冗	沉，異體字問題。

編號	造字	字頻 (txt)	字頻 (xml)	Big5	Unicode	WinXp	構字式	備註
5194		1	1	8A4C			𠄎扌 畝且 𠄎	擅，異體字問題。
5200		2	2	8A52			𠄎ノ 冂米大 𠄎	奧，異體字問題。
5221		33	33	8A67			𠄎白米 𠄎	粵，異體字問題。
5241	鶩	1	1	8A7B	9D5E	鶩		鵝，異體字問題。
5289	慤	1	1	8ACD	6164	慤		慤，異體字問題。
5310	烏	1	1	8AE2	8204	烏		
5555	迥	1	1	8C7B	9008	迥		迥，異體字問題。
5558	麓	2	2	8C7E	9E84	麓		
5582	鞣	1	1	8CB8	97B5	鞣		
5693	𪗇	2	2	8D68	9E7B	𪗇		
5697	𪗇	4	4	8D6C	7151	𪗇		煮，異體字問題。
5702	厨	1	1	8D71	53A8	厨		廚，異體字問題。
5705	廐	1	1	8D74	5ED0	廐		Unicode 呈現差異。
5706	敘	7	7	8D75	654D	敘		敘，異體字問題。
5742	澁	1	1	8DBB	6F81	澁		
5768	潤	6	6	8DD5	6FF6	潤		潤，異體字問題。