

《西溪叢語》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
2008/6/25 葉冠孜 製作

一、《西溪叢語》一書(西溪叢語.xml)使用舊版造字 115 個，字頻 453 次，詳如附件一。這 115 個造字中，83 個可轉成 Windows XP 能顯示的字，字頻 304 次；另外 32 個字必須轉成構字式，字頻 149 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 3255 的「善」字，原字應為「嘉」。
- 2、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 3789 的「冲」係「沖」字之異體，故以「沖」字取代。
- 3、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。
- 4、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 4910 的「𠄎 𠄎 𠄎」。

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 4340 的「撮」字，Unicode 編碼為 3A67，由於 Windows XP 並不支援，仍須使用構字式「扌𠂔暴」。

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、《西溪叢語》造字分析表

編號	造字	字頻 (txt)	字頻 (xml)	Big5	Unicode	WinXP	構字式	備註
296	𠂔	9	9	FBEC			𠂔	
297	𠂔	30	30	FBED			𠂔	
299	𠂔	11	11	FBEF			𠂔	
300	𠂔	11	11	FBF0			𠂔	
479	𠂔	2	2	FD47			𠂔	
650	𠂔	2	2	FE55			𠂔	
675	𠂔	3	3	FE6E			𠂔	
693	𠂔	3	3	FEA2			𠂔	
738	𠂔	1	1	FECF			𠂔	




編號	造字	字頻 (txt)	字頻 (xml)	Big5	Unicode	WinXP	構字式	備註
739	勹	2	2	FED0			勹	
750	勹	4	4	FEDB			勹	
758	𠂇	1	1	FEE3			𠂇	
762	人	9	9	FEE7			人	
763	𠂇	2	2	FEE8			𠂇	
766	𠂇	1	1	FEEB			𠂇	
768	𠂇	1	1	FEED			𠂇	
770	扌	2	2	FEEF			扌	
814	蟻	4	4	8E5C	87C7	蟻		蟻，異體字問題。
820	烟	1	1	8E62	70F1	烟		
899	馳	7	7	8ED3	99DE	馳		
936	鷓	1	1	8EF8	4D0A		義𪗇鳥	
1133	瓌	3	3	9061	74CC	瓌		
1308	蕙	1	1	9173	85FC	蕙		
1572		1	1	9341			𪗇革口王𪗇	靺，異體字問題。
1759	澆	7	7	945F	6EF0	澆		

編號	造字	字頻 (txt)	字頻 (xml)	Big5	Unicode	WinXP	構字式	備註
1981	讐	1	1	95C2	8B90	讐		
2044	踈	1	1	9642	8E08	踈		
2152	蕻	1	1	96D0	979F	蕻		
2322	鏘	2	2	97DD	4948		金𠂔莽	
2638	鷓	4	4	99DF	9D55	鷓		
2793	鋤	1	1	9ADD	9378	鋤		
3015	毆	1	1	9C5F	6BB9	毆		
3238		1	1	9DC3			𠂔西广艹系𠂔	醪，異體字問題。
3255	善	1	1	9DD4	5584	善		嘉，校對問題，舊版漢籍錯字，可用程式全部取代。
3787	冢	2	2	8152	51A2	冢		
3789	冲	1	1	8154	51B2	冲		沖，異體字問題。
3801	剗	1	1	8160	5257	剗		
3805	勅	6	6	8164	52C5	勅		
3806	勑	1	1	8165	52E0	勑		
3807	驗	3	3	8166	9A10	驗		

編號	造字	字頻 (txt)	字頻 (xml)	Big5	Unicode	WinXP	構字式	備註
3814	却	10	10	816D	5374	却		卻，異體字問題。
3829	咏	1	1	817C	548F	咏		
3851	姁	2	2	81B4	59F9	姁		
3858	寃	1	1	81BB	5BC3	寃		
3864	峯	17	17	81C1	5CEF	峯		峰，異體字問題。
3883	徧	3	3	81D4	5FA7	徧		
3884	忽	2	2	81D5	6031	忽		
3886	悞	1	1	81D7	609E	悞		
3921	皙	2	2	81FA	6673	皙		
3935	槩	4	4	8249	69E9	槩		概，異體字問題。
3964	威	1	1	8266	70D5	威		
3966	烟	5	5	8268	70DF	烟		煙，異體字問題。
3968	煨	1	1	826A	7145	煨		
3974	牀	7	7	8270	7240	牀		
3981	珉	2	2	8277	73C9	珉		
3985	璫	1	1	827B	7447	璫		Unicode 呈現差異。

編號	造字	字頻 (txt)	字頻 (xml)	Big5	Unicode	WinXP	構字式	備註
3996	疎	3	3	82A8	758E	疎		
4016	着	7	7	82BC	7740	着		著，異體字問題。
4034	穉	5	5	82CE	7A49	穉		
4038	竈	7	7	82D2	7AC8	竈		
4042	笋	12	12	82D6	7B0B	笋		
4051	籜	1	1	82DF	7C35	籜		
4052	簾	1	1	82E0	7C34	簾		
4077	罇	1	1	82F9	7F47	罇		
4080	羣	5	5	82FC	7FA3	羣		群，異體字問題。
4090	脚	3	3	8347	811A	脚		腳，異體字問題。
4091	臯	7	7	8348	81EF	臯		
4114	藥	2	2	835F	8616	藥		
4118	藥	4	4	8363	8602	藥		
4129	虬	2	2	836E	866C	虬		
4132		11	11	8371			血𠂇𠂇	眾，異體字問題。
4134	衛	5	5	8373	885E	衛		衛，異體字問題。

編號	造字	字頻 (txt)	字頻 (xml)	Big5	Unicode	WinXP	構字式	備註
4146	覩	3	3	83A1	89A9	覩		
4158	譌	1	1	83AD	8B4C	譌		Unicode 呈現差異。
4174	躑	1	1	83BD	8EB6	躑		
4180	臯	1	1	83C3	8FA0	臯		
4184	迹	6	6	83C7	8FF9	迹		跡，異體字問題。
4193	鈎	2	2	83D0	920E	鈎		鈎，異體字問題。
4195	鉢	4	4	83D2	9262	鉢		
4226	凜	1	1	83F1	98E1	凜		
4232	隲	4	4	83F7	96B2	隲		
4240	髯	3	3	8440	9AF4	髯		
4242	鬪	3	3	8442	9B2D	鬪		Unicode 呈現差異。
4256	鷄	4	4	8450	9DC4	鷄		雞，異體字問題。
4265	鼃	3	3	8459	9F03	鼃		
4267	鼈	4	4	845B	9F08	鼈		
4278	亘	5	5	8466	4E98	亘		
4340	掬	1	1	84C6	3A67		扌 𠂔 暴	

編號	造字	字頻 (txt)	字頻 (xml)	Big5	Unicode	WinXP	構字式	備註
4373	鼯	3	3	84E7			鼠  晶	
4460	勅	1	1	85A1	52D1	勅		
4507	吳	40	40	85D0	5434	吳		吳，異體字問題。
4515	呪	1	1	85D8	546A	呪		
4531	攢	1	1	85E8	6505	攢		攢，異體字問題。
4756	呂	7	7	876D	382F		呂	
4768	妬	1	1	8779	59AC	妬		妒，異體字問題。
4784	鞞	1	1	87AB	9793	鞞		
4817	寶	1	1	87CC	5BF3	寶		寶，異體字問題。
4906		1	1	8866			二  儿	示，異體字問題。
4907	橐	2	2	8867	69D6	橐		橐，異體字問題。
4910		3	3	886A			彳  畷	待造字。
5019	畧	1	1	88F9			畧	略，異體字問題。
5028	昇	8	8	8943	662A	昇		
5040	鞫	1	1	894F	4A71		革  俞	
5176		19	19	89F9			彳  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
5225	卹	1	1	8A6B	460F		血𠂔卩	卹，異體字問題。
5278	苦	2	2	8AC2	44B7		艹𠂔舌	
5482	璽	1	1	8BF1	8812	璽		
5541	蠹	1	1	8C6D	8827	蠹		
5559	欸	4	4	8CA1	6B35	欸		
5590	朶	1	1	8CC0	6736	朶		朶，異體字問題。
5693	𪗇	1	1	8D68	9E7B	𪗇		
5705	廐	1	1	8D74	5ED0	廐		廐，異體字問題。
5706	敘	20	20	8D75	654D	敘		敘，異體字問題。
5768	潤	1	1	8DD5	6FF6	潤		潤，異體字問題。
5792	籓	1	1	8DED	7C55	籓		