

《水滸全傳》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
2008/6/24 黃榮順 製作

一、《水滸全傳》一書(水滸全傳.xml)使用舊版造字 114 個，字頻 407 次，詳如附件一。這 114 個造字中，85 個可轉成 Windows XP 可顯示的字，字頻 370 次；另外 29 個字必須轉成構字式，字頻 37 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、重碼問題，採用部件外字，不轉碼：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」。然而漢籍在舊轉新的過程中，若已將原舊漢籍造字的部分全數改為 Unicode 或構字式，則在 xml 出現的造字頻次皆採用部件外字，此時便不需轉碼。如編號 5797，字碼為 8DF2，對應的舊漢籍造字為「灩」，部件外字為「𪛗」，對原書發現書中皆取部件外字「𪛗」，故不需轉碼。
- 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 2881 的「僂」係「僂」字之異體，故以「僂」字取代。又如編號 2758 的「猷」係「猷」字之簡體，故以「猷」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 4145 的「羈」字，Unicode 字型呈現為「羈」，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 3014 的「木△扇」。

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。所以如造字編號 1647 的「沒」字，Unicode 編碼為 3CC7，由於 Windows XP 並不支援，仍須改用構字式「彳△父」。

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、「水滸全傳.xml」造字分析表

編號	造字	字頻 (txt)	字頻 (xml)	Big5	Unicode	WinXP	構字式	備註
314	𠃉	1	1	FBFE			𠃉	
658	疒	1	1	FE5D			疒	
1066	顛	1	1	8FDD	4AC7		須△丹	須△冉，異體字問題。
1067	碁	5	5	8FDE	68CA	碁		碁，異體字問題。
1146	甕	1	1	906E	7506	甕		

編號	造字	字頻 (txt)	字頻 (xml)	Big5	Unicode	WinXP	構字式	備註
1321	確	1	1	91A2	7936	確		
1559		2	2	92F3			廌广目无	廌，異體字問題。
1572		1	1	9341			廌革口王	革呈，異體字問題。
1647	汶	5	5	93AE	3CC7		氵 父	
1660	賺	1	1	93BB	8181	賺		
1772		1	1	946C			艹 梨	藜，異體字問題。
1781	蓋	4	4	9475	854B	蓋		
1926	覓	4	4	9569	8994	覓		
1995	通	1	1	95D0	23EE2		氵 通	
2021	趁	5	5	95EA	8D82	趁		
2132	癩	1	1	96BC	3FC0		疒 班	
2206	祆	2	2	9747	8884	祆		
2345	鑽	1	1	97F4	9404	鑽		
2454	鍬	2	2	98C4	28AFC		金 教	
2758	獻	3	3	9ABA	732E	獻		獻，異體字問題。
2770	絳	1	1	9AC6	7D98	絳		

編號	造字	字頻 (txt)	字頻 (xml)	Big5	Unicode	WinXP	構字式	備註
2864		1	1	9B65			艹 厶 音	菖，異體字問題。
2881	僂	7	7	9B76	5127	僂		僂，異體字問題。
2899		1	1	9BAA			亠 厶 粉 厶 大	彛，異體字問題。
2901		1	1	9BAC			忄 厶 左	忄 厶 左，異體字問題。
2938	簪	1	1	9BD1	7C2E	簪		簪，異體字問題。
2993	刃	1	1	9C49	5205	刃		刃，異體字問題。
3004	籊	3	3	9C54	7BD0	籊		
3014		1	1	9C5E			木 厶 扇	待造字。
3035	賓	1	1	9C73	8CD4	賓		賓，異體字問題。
3264	梦	1	1	9DDD	68A6	梦		夢，異體字問題。
3777	倅	1	1	8148	5008	倅		
3799	刼	2	2	815E	5226	刼		
3800		2	2	815F			杀 厶 司	剎，異體字問題。
3801	剗	2	2	8160	5257	剗		
3825	疊	5	5	8178	53E0	疊		疊，異體字問題。
3829	咏	1	1	817C	548F	咏		

編號	造字	字頻 (txt)	字頻 (xml)	Big5	Unicode	WinXP	構字式	備註
3830	咤	2	2	817D	54A4	咤		
3831	詰	1	1	817E	5586	詰		
3832	嘅	1	1	81A1	5605	嘅		Unicode 字型差異。
3849	獎	3	3	81B2	596C	獎		獎，異體字問題。
3870	昏	1	1	81C7	5E0B	昏		
3877	迺	4	4	81CE	5EFC	迺		迺，異體字問題。
3891	慙	2	2	81DC	6159	慙		慙，異體字問題。
3931	椀	1	1	8245	6900	椀		
3935	槩	3	3	8249	69E9	槩		概，異體字問題。
3940	藪	4	4	824E	6B1D	藪		
3946	汹	4	4	8254	6C79	汹		
3964	威	1	1	8266	70D5	威		
3976	犁	1	1	8272	7282	犁		
3984	瑠	1	1	827A	7460	瑠		
4014	蓋	11	11	82BA	8462	蓋		蓋，異體字問題。
4032	秣	1	1	82CC	2578A		禾𪔐未	秣，異體字問題。

編號	造字	字頻 (txt)	字頻 (xml)	Big5	Unicode	WinXP	構字式	備註
4036	窰	1	1	82D0	7AB0	窰		
4040	竝	17	17	82D4	7ADD	竝		
4045	筵	4	4	82D9	7B92	筵		
4089	脇	15	15	8346	8107	脇		
4109	葛	1	1	835A	84AD	葛		
4129	虬	4	4	836E	866C	虬		
4144	霸	1	1	837D	8987	霸		Unicode 字型差異。
4145	羈	2	2	837E	898A	羈		Unicode 字型差異。
4213	鞞	1	1	83E4	97BE	鞞		
4220	穎	2	2	83EB	9834	穎		穎，異體字問題。
4250	鮓	1	1	844A	9C0C	鮓		
4251	鰕	2	2	844B	9C15	鰕		
4267	鼈	3	3	845B	9F08	鼈		
4279	驄	2	2	8467	9A18	驄		
4283		6	6	846B	622F	戲		戲，異體字問題。
4287	織	7	7	846F	7E8E	織		織，異體字問題。

編號	造字	字頻 (txt)	字頻 (xml)	Big5	Unicode	WinXP	構字式	備註
4316	艷	3	3	84AE	8276	艷		艷，異體字問題。
4375		1	1	84E9			月 厶 郤	腳，異體字問題。
4410	巖	1	1	854D	21F71		山 厶 截	
4448		1	1	8573			形 纟 厂 墨 口	纏，異體字問題。
4450	𪗇	1	1	8575	524F	𪗇		𪗇，異體字問題。
4496	段	9	9	85C5	53DA	段		段，異體字問題。
4504	𪗈	2	2	85CD	541A	𪗈		
4604	擦	1	1	8672	22D32		扌 厶 絜	
4679	塏	1	1	86DF	5818	塏		
4779	鞣	2	2	87A6	97B2	鞣		
4784	鞣	9	9	87AB	9793	鞣		
4907	橐	1	1	8867	69D6	橐		橐，異體字問題。
4934	鬢	1	1	88A4	9B09	鬢		
4940	韵	5	5	88AA	97F5	韵		
4980		1	1	88D2			蓺 厶 火	熱，異體字問題。
5005	幫	1	1	88EB	5E2E	幫		幫，異體字問題。

編號	造字	字頻 (txt)	字頻 (xml)	Big5	Unicode	WinXP	構字式	備註
5027	窓	1	1	8942	724E	窓		
5066	瑣	1	1	8969	24A0F		王𠂔頁	瑣，異體字問題。
5074	尫	1	1	8971	5C2B	尫		尫，異體字問題。
5086		1	1	897D			冗𠂔卓	罩，異體字問題。
5119	博	1	1	89C0	613D	博		
5163	憂	1	1	89EC	621E	憂		憂，異體字問題。
5180	拘	11	11	89FD	6285	拘		
5225	邨	1	1	8A6B	460F		血𠂔卩	邨，異體字問題。
5229	撐	2	2	8A6F	6490	撐		
5240	捧	15	15	8A7A	6481	捧		
5277	逞	2	2	8AC1			形辶山王	
5281	餓	1	1	8AC5	994A	餓		
5285	睂	1	1	8AC9	7742	睂		眉，異體字問題。
5316	晉	1	1	8AE8	664B	晉		晉，異體字問題。
5324	裊	2	2	8AF0	88E0	裊		
5327	嶂	1	1	8AF3	5D2A	嶂		

編號	造字	字頻 (txt)	字頻 (xml)	Big5	Unicode	WinXP	構字式	備註
5328		1	1	8AF4			目 𠃉 𠃉	眇，異體字問題。
5339	𦍋	1	1	8B40	59E7	𦍋		
5434		1	1	8BC1			𠃉 𠃉 白 工 𠃉	涅，異體字問題。
5439	栢	8	8	8BC6	6822	栢		
5451	玳	1	1	8BD2	7398	玳		
5466	倏	1	1	8BE1	5010	倏		倏，異體字問題。
5541	蠹	1	1	8C6D	8827	蠹		蠹，異體字問題。
5621	泪	8	8	8CDF	6CEA	泪		淚，異體字問題。
5697	煑	1	1	8D6C	7151	煑		煮，異體字問題。
5756	粹	1	1	8DC9	22519		彳 𠃉 卒	
5778		1	1	8DDF			𠃉 𠃉 田 𠃉 𠃉	𠃉，異體字問題。
5779	揿	1	1	8DE0	645A	揿		
5797	灑	4	113	8DF2	7054	灑		𠃉，重碼問題，採用部件外字，不轉碼。