

《西遊記》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
2008/6/24 黃榮順 製作

一、《西遊記》一書(西遊記.xml)使用舊版造字 80 個，字頻 183 次，詳如附件一。這 80 個造字中，69 個可轉成 Windows XP 可顯示的字，字頻 167 次；另外 11 個字必須轉成構字式，字頻 16 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 5528，字碼為 8C60，對應的舊漢籍造字為「糠」，部件外字為「𩺰」，對原書發現書中皆取部件外字「𩺰」，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 3794 的「湊」係「湊」字之異體，故以「湊」字取代。又如編號 3899 的「担」係「擔」字之簡體，故以「擔」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。

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 1600 的「猷」字，Unicode 編碼為 3E84，由於 Windows XP 並不支援，仍須改用構字式「犛 𠂔 俞」。

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、「西遊記.xml」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	2	2	FAAB	29B06		髻 厶 丂	髻，異體字問題。
912	鷓	1	1	8EE0	9DC0	鷓		
1219		1	1	90D9			宀 厶 巢	窠，異體字問題。
1278	碍	1	1	9155	788D	碍		
1321	確	11	11	91A2	7936	確		
1591	羗	1	1	9354	7F97	羗		
1600	猷	1	1	935D	3E84		犛 𠂔 俞	

編號	造字	字頻 (txt)	字頻 (xml)	Big5	Unicode	WinXP	構字式	備註
1691	鬚	1	1	93DA	4BFC		髟彡狄	
1905	棍	2	2	9554	88E9	棍		
1925	羈	1	1	9568	8989	羈		Unicode 字型差異。
1926	覓	3	3	9569	8994	覓		
2001		1	1	95D6			皿彡貝	賈，異體字問題。
2147	蟪	1	1	96CB	87EE	蟪		
2345	鑽	3	3	97F4	9404	鑽		
2524	饊	1	1	994B	994D	饊		
2589	筴	1	1	99AE	7B21	筴		
2592	臟	2	2	99B1	8D1C	臟		
2616	鱈	1	1	99C9	9C6F	鱈		
2658	鞞	4	4	99F3	97C2	鞞		
3016	闕	1	1	9C60	28D91		門△臭	闕，異體字問題。
3161	壘	1	1	9D54	58DC	壘		
3177	蠱	1	1	9D64	7F4E	蠱		
3777	俵	4	4	8148	5008	俵		

編號	造字	字頻 (txt)	字頻 (xml)	Big5	Unicode	WinXP	構字式	備註
3794	湊	2	2	8159	51D1	湊		湊，異體字問題。
3801	剗	3	3	8160	5257	剗		
3804	効	1	1	8163	52B9	効		效，異體字問題。
3815	𪗇	1	1	816E	537A	𪗇		
3820	厦	2	2	8173	53A6	厦		廈，異體字問題。
3829	咏	2	2	817C	548F	咏		
3840	坂	4	4	81A9	5742	坂		
3843	堦	1	1	81AC	5826	堦		
3849	獎	1	1	81B2	596C	獎		獎，異體字問題。
3868	卮	3	3	81C5	5DF5	卮		
3876	廻	2	2	81CD	5EFB	廻		迴，異體字問題。
3891	慙	1	1	81DC	6159	慙		慚，異體字問題。
3899	担	1	1	81E4	62C5	担		擔，異體字問題。
3902	拊	1	1	81E7	636C	拊		
3932	櫓	1	1	8246	6936	櫓		
3938	橈	1	1	824C	6AC8	橈		

編號	造字	字頻 (txt)	字頻 (xml)	Big5	Unicode	WinXP	構字式	備註
3976	犁	2	2	8272	7282	犁		
3981	珉	1	1	8277	73C9	珉		
3984	瑠	1	1	827A	7460	瑠		
3989	璿	1	1	82A1	74C8	璿		
4000	疴	6	6	82AC	75B4	疴		
4002	瘡	3	3	82AE	7607	瘡		
4028	稟	1	1	82C8	7980	稟		稟，異體字問題。
4129	虬	3	3	836E	866C	虬		
4140	衽	1	1	8379	88B5	衽		
4145	羈	4	4	837E	898A	羈		Unicode 字型差異。
4176	輒	1	1	83BF	8F19	輒		輒，異體字問題。
4184	迹	4	4	83C7	8FF9	迹		跡，異體字問題。
4220	穎	1	1	83EB	9834	穎		穎，異體字問題。
4231	駝	1	1	83F6	9A0C	駝		
4266	鼃	1	1	845A	9F02	鼃		
4267	鼈	11	11	845B	9F08	鼈		

編號	造字	字頻 (txt)	字頻 (xml)	Big5	Unicode	WinXP	構字式	備註
4309	劫	23	23	84A7	523C	劫		劫，異體字問題。
4318		1	1	84B0			𠄎畝且毛	𠄎，異體字問題。
4345		3	3	84CB			口𠄎救	嗽，異體字問題。
4376	傻	1	1	84EA	510D	傻		
4504	呬	2	2	85CD	541A	呬		
4531	攢	1	1	85E8	6505	攢		攢，異體字問題。
4575	繼	1	1	8655	7E7F	繼		
4768	妬	2	2	8779	59AC	妬		妒，異體字問題。
4790	賺	1	1	87B1	81C1	賺		
4803	腭	2	2	87BE	816D	腭		
4805	瀆	1	1	87C0	51DF	瀆		瀆，異體字問題。
5180	拘	1	1	89FD	6285	拘		
5209	摑	3	3	8A5B	6386	摑		
5229	撐	3	3	8A6F	6490	撐		
5241	鶯	1	1	8A7B	9D5E	鶯		鶯，異體字問題。
5299	戟	2	2	8AD7	39B8		卓𠄎戈	戟，異體字問題。

編號	造字	字頻 (txt)	字頻 (xml)	Big5	Unicode	WinXP	構字式	備註
5347	減	1	1	8B48	617D	減		
5528	糠	1	2	8C60	69FA	糠		𧈧，重碼問題，採用部件外字，不轉碼。
5555		3	3	8C7B	9008	迥		迥，異體字問題。
5589	廚	2	2	8CBF	2228D		𠂇广豆寸	廚，異體字問題。
5595	塚	5	5	8CC5	579C	塚		塚，異體字問題。
5697	煮	4	4	8D6C	7151	煮		煮，異體字問題。
5705	廐	1	1	8D74	5ED0	廐		廐，異體字問題。
5742	澁	1	1	8DBB	6F81	澁		澁，異體字問題。
5778		1	1	8DDF			𠂇田	鼉，異體字問題。