

《醒世姻緣》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
2008/8/13 葉冠孜 製作

一、《醒世姻緣》一書(醒世姻緣.xml)使用舊版造字 129 個，字頻 790 次，詳如附件一。這 129 個造字中，107 個可轉成 Windows XP 能顯示的字，字頻 409 次；另外 22 個字必須轉成構字式，字頻 381 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 2999 的「口𠂔𠂔」係「喙」字之異體，故以「喙」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 4242 的「鬪」字，Unicode 字型呈現為「鬪」，實際上仍為同一字。
- 3、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 3097 的「彳𠂔幸」。

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 1598 的「翰」字，Unicode 編碼為 4A7A，由於 Windows XP 並不支援，仍須使用構字式「革𠂔翁」。

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(醒世姻緣.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、《醒世姻緣》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△丐	
296	△	3	3	FBEC			△	
297	△	216	216	FBED			△	
298	△	2	2	FBEE			△	
299	形	3	3	FBEF			形	
300	□	3	3	FBF0			□	
658	疒	1	1	FE5D			疒	

編號	造字	字頻 (txt)	字頻 (xml)	Big5	Unicode	WinXP	構字式	備註
675	食	62	62	FE6E			食	
681	厂	1	1	FE74			厂	
717	足	3	3	FEBA			足	
770	扌	41	41	FEEF			扌	
1213	睽	5	5	90D3	7743	睽		
1267	孺	3	3	914A	6B1B	孺		
1383	糠	2	2	91E0	7A45	糠		
1491	擷	2	2	92AF	6527	擷		
1598	翰	5	5	935B	4A7A		革叺翁	
1691	髻	16	16	93DA	4BFC		髻叺狄	
1905	裊	2	2	9554	88E9	裊		
1994	達	3	3	95CF	6FBE	達		
2053	躡	12	12	964B	8EA7	躡		
2097	朒	5	5	9677	26782		月叺荅	
2157	酌	2	2	96D5	9167	酌		
2246	炆	1	1	976F	7079	炆		

編號	造字	字頻 (txt)	字頻 (xml)	Big5	Unicode	WinXP	構字式	備註
2520	饊	1	1	9947	993B	饊		
2881	僂	1	1	9B76	5127	僂		
2999		5	5	9C4F			口𠂇自	咍，異體字問題。
3010	攷	32	32	9C5A	5746	攷		
3021	𠂇	4	4	9C65			𠂇	
3029	鋤	1	1	9C6D	941D	鋤		
3054	蝮	1	1	9CA8	45B5		蝮	
3097		2	2	9CD3			彳𠂇幸	待造字。
3098	揸	2	2	9CD4	63F8	揸		Unicode 呈現差異。
3160	坟	4	4	9D53	575F	坟		
3177	𧈧	9	9	9D64	7F4E	𧈧		
3770	个	1	1	8141	4E2A	个		
3805	勅	10	10	8164	52C5	勅		
3830	咤	1	1	817D	54A4	咤		
3834	嗶	1	1	81A3	5637	嗶		
3836	喻	1	1	81A5	564F	喻		

編號	造字	字頻 (txt)	字頻 (xml)	Big5	Unicode	WinXP	構字式	備註
3868	卮	4	4	81C5	5DF5	卮		
3883	徧	4	4	81D4	5FA7	徧		
3884	忽	1	1	81D5	6031	忽		
3886	悞	1	1	81D7	609E	悞		
3890	憑	1	1	81DB	6142	憑		
3891	慙	3	3	81DC	6159	慙		
3909	擡	1	1	81EE	64E1	擡		
3928	椶	6	6	8242	685A	椶		
3931	椀	3	3	8245	6900	椀		
3934	楈	1	1	8248	69C5	楈		
3942	毡	28	28	8250	6BE1	毡		
3968	煨	3	3	826A	7145	煨		
3970	焗	1	1	826C	714F	焗		
3974	牀	17	17	8270	7240	牀		
3976	犁	1	1	8272	7282	犁		
3990	甄	1	1	82A2	750E	甄		

編號	造字	字頻 (txt)	字頻 (xml)	Big5	Unicode	WinXP	構字式	備註
3994	畀	1	1	82A6	754A	畀		
3996	疎	7	7	82A8	758E	疎		
4000	疴	1	1	82AC	75B4	疴		
4004	瘿	1	1	82B0	763B	瘿		
4012	鼓	1	1	82B8	76B7	鼓		
4013	盃	2	2	82B9	76CC	盃		
4024	礮	1	1	82C4	792E	礮		
4033	稽	4	4	82CD	7A2D	稽		
4034	穉	2	2	82CE	7A49	穉		
4035	穿	1	1	82CF	7A7D	穿		
4038	竈	1	1	82D2	7AC8	竈		
4041	豎	3	3	82D5	7AEA	豎		
4042	笋	5	5	82D6	7B0B	笋		
4044	筋	20	20	82D8	7B6F	筋		
4045	箒	1	1	82D9	7B92	箒		
4055	粃	12	12	82E3	7C83	粃		

編號	造字	字頻 (txt)	字頻 (xml)	Big5	Unicode	WinXP	構字式	備註
4057	莖	2	2	82E5	585F	莖		
4067	綉	3	3	82EF	7D89	綉		
4069	綑	6	6	82F1	7DB3	綑		
4074	繚	7	7	82F6	7E27	繚		
4077	罇	1	1	82F9	7F47	罇		
4089	脇	1	1	8346	8107	脇		
4091	臯	11	11	8348	81EF	臯		
4098	芪	1	1	834F	82AA	芪		
4105	菓	3	3	8356	83D3	菓		
4107	葱	2	2	8358	8471	葱		
4112	蔴	4	4	835D	8534	蔴		
4115	藁	4	4	8360	85C1	藁		
4129	虬	1	1	836E	866C	虬		
4131	劬	2	2	8370	8842	劬		
4139	袴	1	1	8378	88B4	袴		
4146	覩	1	1	83A1	89A9	覩		

編號	造字	字頻 (txt)	字頻 (xml)	Big5	Unicode	WinXP	構字式	備註
4155	訶	1	1	83AA	8B0C	訶		
4169	趨	2	2	83B8	8DA6	趨		
4173	耽	7	7	83BC	8EAD	耽		
4178	輓	2	2	83C1	8F2D	輓		
4184	迹	3	3	83C7	8FF9	迹		
4195	鉢	4	4	83D2	9262	鉢		
4218	菲	5	5	83E9	97EE	菲		
4232	隲	2	2	83F7	96B2	隲		
4240	髯	1	1	8440	9AF4	髯		
4242	鬪	7	7	8442	9B2D	鬪		Unicode 呈現差異。
4247	鮎	1	1	8447	9B8E	鮎		
4252	鰓	1	1	844C			魚𩺰眾	
4259	麇	1	1	8453	9E95	麇		Unicode 呈現差異。
4260	麇	1	1	8454	9E9E	麇		
4262	麪	1	1	8456	9EAA	麪		
4263	麪	2	2	8457	9EAF	麪		

編號	造字	字頻 (txt)	字頻 (xml)	Big5	Unicode	WinXP	構字式	備註
4267	鼈	13	13	845B	9F08	鼈		
4270	鼾	3	3	845E	9F45	鼾		
4450	翔	4	4	8575	524F	翔		Unicode 呈現差異。
4493	犇	1	1	85C2	3E40		犇 𠂇字	
4515	呪	3	3	85D8	546A	呪		
4518	响	4	4	85DB	54CD	响		
4616	壳	5	5	867E	58F3	壳		
4619	但	1	1	86A3	4F39	但		
4735	壠	2	2	8758	58E0	壠		
4768	妬	3	3	8779	59AC	妬		
4771	麈	1	1	877C	9E85	麈		
4779	鞫	1	1	87A6	97B2	鞫		
4784	鞞	5	5	87AB	9793	鞞		
4790	賺	2	2	87B1	81C1	賺		
4816	彳	4	4	87CB			彳	
5008	僕	15	15	88EE	5E5E	僕		

編號	造字	字頻 (txt)	字頻 (xml)	Big5	Unicode	WinXP	構字式	備註
5013	汙	1	1	88F3			汙	
5015	甯	1	1	88F5	5BD7	甯		
5255	涇	1	1	8AAB	6DB6	涇		
5281	餓	3	3	8AC5	994A	餓		
5408	絳	1	1	8BA7	7D4D	絳		
5433	爰	1	1	8BC0	590B	爰		
5447	躡	1	1	8BCE	8E83	躡		
5543	養	5	5	8C6F	9B9D	養		
5621	泪	5	5	8CDF	6CEA	泪		
5779	揿	4	4	8DE0	645A	揿		

附件二、《醒世姻緣》校出錯字表

檔案原文摘錄	手動取代結果
蒸食 ㄟ 衲食 ㄟ 磨	蒸 釐 食 ㄟ 磨