

《警世通言》舊版造字轉碼說明

中研院資訊所文獻處理實驗室
中央研究院語言所文獻語料小組
2008/8/07 丁致伶

轉碼主要工作是把檔案中的舊版造字轉換成 Windows XP 能支援的 Unicode 字形，Unicode 目前共收錄漢字 70194 個字，而 XP 只能支援 20902 個字（詳如表一），不支援之字將以構字式表達。例：造字編號 74 的「髻」字，Unicode 編碼是 29B06，由於 XP 並不支援，仍需使用構字式「髻𠄎𠄎」。

表一、Unicode 的字數及編碼區段

Unicode	新增字數	新增編碼區段	總字數	WinXP
1.1 版	20902	4E00-9FFF	20902	支援
3.0 版	6582	3400-4DFF	27484	不支援
3.1 版	42710	20000-2A6D6	70194	不支援

一、舊版造字轉碼分析：

《警世通言》使用舊版造字 136 個，字頻 2994 次，這 136 個造字中，128 個可轉成 Windows XP 能顯示的字，字頻 2963 次；另外 8 個字必須轉成構字式，字頻 31 次。

轉碼完成製作轉碼分析表，請參考附件一《警世通言》轉碼分析表，欄位說明如下：

- (一) 編號：Big5 造字空間為 6217 個，編號由 1 到 6217。
- (二) 造字：舊版造字。
- (三) 頻次：舊版造字在文件的出現次數。
- (四) Big5：造字的 Big5 碼。
- (五) Unicode：造字所對應的 Unicode 碼。

(六) WinXP：造字在 Windows XP 的對應字形。

(七) 構字式：Windows XP 無法對應字形改用構字式。

(八) 備註凡例：備註欄中記錄轉碼後字形及修改原因，凡例如下：

1. 異體字問題：為了使用者查詢和使用的方便，在處理異體字時最主要的方式是以標準字取代，除非是專有名詞或特殊情形，如：人名、地名等。例：造字編號 4134 的「衛」字，是「衛」的異體字，以標準字「衛」取代。
2. Unicode 字型呈現差異：Unicode 字型與舊版造字有些微差異，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 4263 的「麩」字，Unicode 字型呈現為「麩」，實際上仍為同一字。

附件一、《警世通言》轉碼分析表

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
63	啐	2	FA7E	54F0	啐		
74	髻	3	FAAB	29B06		髻𠄎𠄏	
899	馳	6	8ED3	99DE	馳		
1039	稟	3	8FC2	69C0	稟		
1278	碍	4	9155	788D	碍		
1409	寵	1	91FA	7AC9	寵		
1491	擷	4	92AF	6527	擷		
1591	羗	1	9354	7F97	羗		
1597	鞞	1	935A	4A6B		革𠄎兩	


編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
1616	逕	22	936D	284E6		辶△里	
1642	脉	9	93A9	8109	脉		
1781	葢	1	9475	854B	葢		
1784	萼	1	9478	855A	萼		
1905	棍	1	9554	88E9	棍		
1981	讐	1	95C2	8B90	讐		讎，異體字問題。
1995	通	1	95D0	23EE2		辶△通	
2046	踪	4	9644	8E2A	踪		
2053	躡	2	964B	8EA7	躡		
2147	鞞	1	96CB	87EE	蟪		
2190	稔	1	96F6	6BFA	稔		
2432	鞞	1	98AE	29350		革△登	
2711	禧	1	9A69	3D59		辶△喜	
2999	咆	3	9C4F	5523	咆		
3016	闕	1	9C60	28D91		門△臭	
3770	个	1	8141	4E2A	个		
3784	兪	99	814F	516A	兪		
3789	冲	16	8154	51B2	冲		沖，異體字問題。

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
3791	况	48	8156	51B5	况		
3794	湊	28	8159	51D1	湊		湊，異體字問題。
3797	凭	5	815C	51ED	凭		
3801	剗	2	8160	5257	剗		
3804	効	5	8163	52B9	効		
3805	勅	20	8164	52C5	勅		
3813	邛	2	816C	536D	邛		
3814	却	827	816D	5374	却		
3820	厦	1	8173	53A6	厦		
3825	疊	4	8178	53E0	疊		
3829	咏	1	817C	548F	咏		
3830	咤	1	817D	54A4	咤		
3835	噍	1	81A4	564D	噍		
3840	坂	1	81A9	5742	坂		
3843	堦	25	81AC	5826	堦		
3858	寃	65	81BB	5BC3	寃		寃，異體字問題。
3864	峯	35	81C1	5CEF	峯		
3868	卮	3	81C5	5DF5	卮		

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
3876	廻	3	81CD	5EFB	廻		
3899	担	1	81E4	62C5	担		
3909	擡	58	81EE	64E1	擡		
3924	菴	1	81FD	671E	菴		期，異體字問題。
3927	柏	1	8241	6855	柏		
3931	椀	1	8245	6900	椀		
3934	桶	6	8248	69C5	桶		
3938	橈	4	824C	6AC8	橈		
3963	炆	6	8265	7081	炆		
3966	烟	15	8268	70DF	烟		
3968	煨	1	826A	7145	煨		
3974	牀	4	8270	7240	牀		
3977	狗	1	8273	72E5	狗		
3978	猪	12	8274	732A	猪		猪，異體字問題。
3979	猫	3	8275	732B	猫		猫，異體字問題。
3989	璿	1	82A1	74C8	璿		
3996	疎	26	82A8	758E	疎		

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
4010	皐	2	82B6	7690	皐		皐，異體字問題。
4012	鼓	3	82B8	76B7	鼓		鼓，異體字問題。
4016	着	790	82BC	7740	着		
4024	礮	1	82C4	792E	礮		
4036	窰	2	82D0	7AB0	窰		
4038	竈	21	82D2	7AC8	竈		
4040	竝	1	82D4	7ADD	竝		
4042	笋	1	82D6	7B0B	笋		
4044	筍	9	82D8	7B6F	筍		
4045	箒	5	82D9	7B92	箒		
4055	粃	2	82E3	7C83	粃		
4069	綑	2	82F1	7DB3	綑		
4074	縹	6	82F6	7E27	縹		
4077	罇	1	82F9	7F47	罇		
4080	羣	30	82FC	7FA3	羣		
4090	脚	73	8347	811A	脚		
4105	菓	15	8356	83D3	菓		
4112	蔴	10	835D	8534	蔴		

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
4115	藁	1	8360	85C1	藁		
4129	虬	2	836E	866C	虬		
4132		345	8371			血𠄎𠄎	眾，異體字問題。
4134	衛	21	8373	885E	衛		衛，異體字問題。
4138	裘	1	8377	88A0	裘		
4139	袴	2	8378	88B4	袴		
4140	衽	1	8379	88B5	衽		
4142	褰	2	837B	8943	褰		褰，異體字問題。
4145	羈	2	837E	898A	羈		Unicode 字型呈現差異。
4146	覩	11	83A1	89A9	覩		睹，異體字問題。
4167	賫	2	83B6	8CEB	賫		
4171		1	83BA			𠄎𠄎𠄎 封𠄎	蹶，異體字問題。
4173	𨔵	1	83BC	8EAD	𨔵		
4174	𨔶	1	83BD	8EB6	𨔶		
4184	迹	7	83C7	8FF9	迹		

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
4193	鈎	3	83D0	920E	鈎		鈎，異體字問題。
4195	鉢	8	83D2	9262	鉢		
4242	鬪	20	8442	9B2D	鬪		
4247	鮎	1	8447	9B8E	鮎		
4251	鰕	6	844B	9C15	鰕		
4256	鷄	11	8450	9DC4	鷄		雞，異體字問題。
4263	麩	1	8457	9EAF	麩		Unicode 字型呈現差異。
4267	鼈	3	845B	9F08	鼈		
4278	亘	4	8466	4E98	亘		
4309	刼	2	84A7	523C	刼		
4428	躡	2	855F	8E70	躡		
4455	劒	1	857A			僉  刀	
4460	勅	1	85A1	52D1	勅		
4496	段	20	85C5	53DA	段		
4515	呪	1	85D8	546A	呪		
4702	場	3	86F6	5872	場		
4768	妬	6	8779	59AC	妬		妒，異體字問題。

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
4778	鞞	2	87A5	97B4	鞞		
4805	瀆	4	87C0	51DF	瀆		瀆，異體字問題。
4903	吋	8	8863	5C00	吋		
4904	尅	3	8864	5C05	尅		
4996	斲	2	88E2	65B5	斲		
5008	幪	4	88EE	5E5E	幪		
5074	𠂔	2	8971	5C2B	𠂔		
5123	悞	1	89C4	6130	悞		
5163	憂	2	89EC	621E	憂		
5189	宀	1	8A47	5B82	宀		
5240	擗	2	8A7A	6481	擗		
5281	餼	1	8AC5	994A	餼		
5290	携	26	8ACE	643A	携		
5299	戟	1	8AD7	39B8		卓△戈	戟，異體字問題。
5437	并	1	8BC4	5E77	并		
5551	賈	18	8C77	8CF7	賈		
5555	迥	3	8C7B	9008	迥		
5587	歛	1	8CBD	6B58	歛		

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
5590	朶	1	8CC0	6736	朶		
5621	泪	4	8CDF	6CEA	泪		淚，異體字問題。
5704	瓊	3	8D73	6C0A	瓊		氈，異體字問題。
5706	敘	24	8D75	654D	敘		敘，異體字問題。
5768	濶	3	8DD5	6FF6	濶		闊，異體字問題。
5779	揿	1	8DE0	645A	揿		

二、手動取代表說明：

《警世通言》檔案中需手動修改的情況分為以下幾種：

(一)《警世通言》檔案中出現「●」字形，共計 104 字，比對原書後以 Unicode 字形或構字式手動取代，並製作「●」字手動取代表，請參考附件二欄位說明如下：

1. 頁碼：「●」字形所在位置之原書頁碼。
2. 檔案原文摘錄：摘錄《警世通言》檔案有「●」字形的文句。
3. 原書字形：「●」在原書中之字形，以此字取代檔案原文的「●」。
4. 備註：記錄修改相同詞句的次數或其他事項。

附件二、《警世通言》「●」字手動取代表

頁碼	檔案原文摘錄	原書字形	備註
p. 52	敲檯打●	檯	

頁碼	檔案原文摘錄	原書字形	備註
p. 54	被誣●罪	贓	
p. 55	李直●污狼籍	贓	
p. 62	只得掙●起來	門△坐	
p. 95	●出江	撻	
p. 95	酒是燒身●焰	焯	
p. 99	一個個●鬆	蓬	
p. 101	必是貪●所致	贓	
P102	楊●嘴	羊△束	有 3 處
p. 102	正劈着●子	孛△頁	
p. 106	直●到嚮水閘邊	汜△吞	
p. 112	緝拿真●真盜	贓	
p. 138	又是●●大官府	嗶嚙	
p. 148	●送些東西	扌△弃	
p. 150	將去●賭花費	闕	
p. 150	拿獲●盜時	贓	
p. 153	拷得真●出來	贓	
p. 154	李大家去起●	贓	
p. 154	姐夫床下●物	贓	
p. 156	要●露	贓	
p. 156	追出●來	贓	
p. 158	●燭	熾	
p. 158	真●真賊	贓	
p. 159	且●物未見破散	贓	
p. 159	姑照給主●例	贓	
p. 166	頭上●鬆	蓬	

頁碼	檔案原文摘錄	原書字形	備註
p. 170	把親妹六●	嫫	
p. 171	坐●萬兩	脏	
p. 171	萬兩●銀交納	脏	
p. 175	妹子六●改聘	嫫	
p. 175	六●以死自誓	嫫	
p. 175	六●年長未嫁	嫫	
p. 175	六●堅執不從	嫫	
p. 175	●出一身廣瘡	闕	
p. 175	全賴六●一身	嫫	
p. 175	六●自家	嫫	
p. 175	六●是個女中丈夫	嫫	
p. 176	六●小姐	嫫	有 2 處
p. 215	有現成●●	饜饜	
p. 216	個肉●	食 𠂇 巴	
p. 223	支分●匠鋸木	金 𠂇 解	
p. 243	●出水面	彳 𠂇 吞	
p. 246	紅光●●	烟烟	
p. 259	等他●滿肉	臙	
p. 265	王匠●了金哥一眼	睇	
p. 272	追●不恕衙門	脏	
p. 281	桂遷把眼一●	睇	
p. 283	再透一個窟●	窰	
p. 285	丘●已成	壠	
p. 300	焉能磨●吾行止	湟	
p. 310	請燒●子	𠂇 𠂇 奄	

頁碼	檔案原文摘錄	原書字形	備註
p. 310	把●子錢馬	𠂇 𠂇 奄	
p. 310	保叔塔燒●子	𠂇 𠂇 奄	
p. 310	燒了●子	𠂇 𠂇 奄	
p. 311	叔塔燒●子	𠂇 𠂇 奄	
p. 314	真●正賊	贓	有 2 處
p. 336	解佩江●浪得聲	皋	
p. 349	公子●他	闕	
p. 360	聞知兒子●院	闕	
p. 365	●得衣衫藍縷	闕	
p. 376	將去小二●項下	𠂇 𠂇 頁	
p. 408	兼追所詐之●	贓	
p. 410	琥珀●	王 𠂇 車	
p. 410	●玳瑁塔	璪	
p. 412	看的人●肩疊背	扌 𠂇 牙	
p. 418	領單色舊●先捲	衤 𠂇 奚	
p. 419	裏面掣玷抽●	扌 戶 冨 冫	
p. 435	况本婦甫能●	門 𠂇 爭	
p. 435	●得病好	門 𠂇 坐	
p. 437	鳴●者鬧	扌 𠂇 郎	待造字
p. 450	陽詹●先生	日 𠂇 危	待造字
p. 455	長姊事南昌●君	目 𠂇 亏	改為目 𠂇 亏
p. 455	遺下一子●烈	目 𠂇 亏	改為目 𠂇 亏
p. 455	汝子●烈	目 𠂇 亏	改為目 𠂇 亏
p. 455	●母曰	目 𠂇 亏	改為目 𠂇 亏
p. 455	吾與●烈哥哥	目 𠂇 亏	改為目 𠂇 亏

頁碼	檔案原文摘錄	原書字形	備註
p. 455	何獨與●兄同行	目 𠃉 𠃉	改為目 𠃉 𠃉
p. 459	一任他●●磅磅	石 𠃉 兵 石 𠃉 兵	待造字
p. 462	細細輕飄如麥●	夾 𠃉 𠃉	改為麪
P. 462	虺虺●	口 𠃉 𠃉 𠃉 虎	
P. 462	●可畏	口 𠃉 𠃉 𠃉 虎	
p. 463	●烈字道微	目 𠃉 𠃉	改為目 𠃉 𠃉
p. 464	施岑●烈曰	目 𠃉 𠃉	改為目 𠃉 𠃉
p. 464	施●欣然領命	目 𠃉 𠃉	改為目 𠃉 𠃉
p. 464	此蓋施●眺	目 𠃉 𠃉	改為目 𠃉 𠃉
p. 464	施岑謂●烈曰	目 𠃉 𠃉	改為目 𠃉 𠃉
p. 466	施岑與●烈	目 𠃉 𠃉	改為目 𠃉 𠃉
p. 466	施●二人	目 𠃉 𠃉	改為目 𠃉 𠃉
p. 467	●烈	目 𠃉 𠃉	改為目 𠃉 𠃉
p. 472	地名●洋九里潭	； 𠃉 𠃉 茶	
p. 476	●●璫璫	王 𠃉 吉 王 𠃉 吉	
p. 476	●響	石 𠃉 𠃉 邦	
p. 486	●烈與其母	目 𠃉 𠃉	改為目 𠃉 𠃉
p. 488	●烈六人	目 𠃉 𠃉	改為目 𠃉 𠃉
p. 488	雲陽詹●二人	日 𠃉 𠃉 危	待造字

(二)《警世通言》檔案中，對於舊版造字以外發現的錯字，以手動修改，並製作錯字修改表，請參考附件三欄位說明如下：

1. 檔案原文摘錄：摘錄檔案《警世通言》中的文句段落，其中紅字
底線的部分，為需要手動取代的原文。
2. 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。

3. 備註：記錄修改相同詞句的次數或其他事項。

附件三、《警世通言》缺字外錯字修改

檔案原文摘錄	手動取代結果	備註
料應 <u>褻</u> 姒逞嬌容	料應 <u>褻</u> 姒逞嬌容	