

《兒女英雄傳》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
2008/6/25 黃榮順 製作

一、《兒女英雄傳》一書(兒女英雄傳.xml)使用舊版造字 58 個，字頻 224 次，詳如附件一。這 58 個造字中，41 個可轉成 Windows XP 可顯示的字，字頻 185 次；另外 17 個字必須轉成構字式，字頻 39 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 5767 的「濶」字，原字應為「濶」，以標準字「闊」取代。

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 3815 的「𪗇」係「𪗇」字之異體，故以「𪗇」字取代。

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 2704 的「𪗇了」。

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 77 的「哪」字，Unicode 編碼為 35BF，由於 Windows XP 並不支援，仍須改用構字式「口

厶耶」。

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、「兒女英雄傳.xml」造字分析表

編號	造字	字頻 (txt)	字頻 (xml)	Big5	Unicode	WinXP	構字式	備註
8	賦	1	1	FA47	27E10		貝厶貳	膩，異體字問題。
20	藪	1	1	FA53			竹厶叻	
77	哪	2	2	FAAE	35BF		口厶耶	
493	胃	4	4	FD55	5190	胃		
1153	坳	3	3	9075	362D		土厶幻	坳，異體字問題。
1201	昵	2	2	90C7	7724	昵		
1600	猷	7	7	935D	3E84		犴厶俞	
1660	賺	1	1	93BB	8181	賺		
1854	啣	5	5	94E0	20E95		口厶留	
1950	牝	19	19	95A3	7264	牝		
2107	髀	2	2	96A3	9AB2	髀		

編號	造字	字頻 (txt)	字頻 (xml)	Big5	Unicode	WinXP	構字式	備註
2225	裊	2	2	975A	88D3	裊		
2592	贓	6	6	99B1	8D1C	贓		
2667	歛	1	1	99FC	9FA1	歛		
2704		1	1	9A62			𠄎𠄎了	待造字。
2759	畱	1	1	9ABB	24CA2		亞𠄎田	留，異體字問題。
2811		1	1	9AEF			亞𠄎田	留，異體字問題。
3004	籊	3	3	9C54	7BD0	籊		
3019	腩	2	2	9C63	817C	腩		
3029	鋤	9	9	9C6D	941D	鋤		
3099	栺	1	1	9CD5	68B9	栺		櫛，異體字問題。
3168	瘡	1	1	9D5B	24EA5		疒△答	
3188	箊	1	1	9D6F	7BB2	箊		
3784	兪	2	2	814F	516A	兪		兪，異體字問題。
3799	刼	1	1	815E	5226	刼		
3801	剗	6	6	8160	5257	剗		
3815	𣎵	3	3	816E	537A	𣎵		𣎵，異體字問題。

編號	造字	字頻 (txt)	字頻 (xml)	Big5	Unicode	WinXP	構字式	備註
3826	叶	1	1	8179	53F6	叶		
3849	獎	1	1	81B2	596C	獎		獎，異體字問題。
3868	卮	2	2	81C5	5DF5	卮		
3921	皙	25	25	81FA	6673	皙		
3932	椶	1	1	8246	6936	椶		
3955	澗	4	4	825D	3D4E		氵 ㄩ 間	
3977	狗	1	1	8273	72E5	狗		
4008	癯	1	1	82B4	766F	癯		
4015	眦	2	2	82BB	7726	眦		
4043	筭	1	1	82D7	7B53	筭		
4045	箒	2	2	82D9	7B92	箒		
4055	粃	1	1	82E3	7C83	粃		
4093	舐	1	1	834A	445B		舌 ㄩ 氏	舐，異體字問題。
4188	邨	9	9	83CB	90A8	邨		
4218	萑	1	1	83E9	97EE	萑		
4247	鮎	1	1	8447	9B8E	鮎		

編號	造字	字頻 (txt)	字頻 (xml)	Big5	Unicode	WinXP	構字式	備註
4318		2	2	84B0			𠄎 尙且毛 𠄎	𧇗，異體字問題。
4376	儂	41	41	84EA	510D	儂		
4448		1	1	8573			𠄎 纒 厂墨 𠄎	纏，異體字問題。
4616	壳	1	1	867E	58F3	壳		
4799	顛	2	2	87BA	9856	顛		
4800	瞪	10	10	87BB	5654	瞪		
5045	餛	1	1	8954	9901	餛		
5180	拘	1	1	89FD	6285	拘		
5229	撐	1	1	8A6F	6490	撐		
5240	摔	1	1	8A7A	6481	摔		
5398	涸	1	1	8B7B	6E7B	涸		
5434		5	5	8BC1			𠄎 汜 白工 𠄎	渥，異體字問題。
5591	杞	2	2	8CC1	233CC		木 𠄎 巳	杞，異體字問題。
5697	煮	12	12	8D6C	7151	煮		煮，異體字問題。
5767	濶	1	1	8DD4	24050		汜 𠄎 閣	闊，校對問題，舊版漢籍錯字，可用程式全部取代。