

《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
中央研究院語言所文獻語料小組
2010/02/06 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 77 的「哪」字，Unicode 編碼是 35BF，由於 XP 並不支援，仍需使用構字式「口𠂔耶」。

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：

《兒女英雄傳》使用舊版造字 178 個，字頻 7745 次，這 178 個造字中，147 個可轉成 Windows XP 能顯示的字，字頻 7394 次；另外 31 個字必須轉成構字式，字頻 351 次。

轉碼完成製作轉碼分析表，請參考附件一《兒女英雄傳》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. 錯字，以程式全部取代：與原書字形不符，並查詢教育部異體字字典確認非異體字後，皆歸類為錯字，以程式取代為正確字形。例：檔案中使用造字編號 5767 的「濶」字，而原書使用之字為「濶」，查詢教育部異體字字典確認「濶」並非「濶」之異體字，因此歸為錯字，以程式取代為正確字形「濶」。
 3. 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 3978 的「猪」字，Unicode 字型呈現為「猪」，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 2704 的「𠂔𠂔了」。

附件一、《兒女英雄傳》轉碼分析表

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
8	賦	1	FA47	27E10		貝𠂔貳	
20	籟	1	FA53			竹𠂔𠂔救	
77	哪	2	FAAE	35BF		口𠂔耶	

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
296	厶	16	FBEC			厶	
297	厶	212	FBED			厶	
298	厶	2	FBEE			厶	
299	形	3	FBEF			形	
300	口	3	FBF0			口	
301	88	5	FBF1			88	
432	業	9	FCD7			業	
493	冑	4	FD55	5190	冑		
628	繼	1	FDFE			繼	
639	彳	4	FE4A			彳	
648	彳	3	FE53			彳	
675	食	2	FE6E			食	
717	足	7	FEBA			足	
770	扌	12	FEEF			扌	
820	烟	2	8E62	70F1	烟		
1100	身	1	9040	8EAB	身		
1151	剗	4	9073	5273	剗		
1153	坳	3	9075	362D		土厶幻	



編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
1201	昵	2	90C7	7724	昵		
1213	睽	1	90D3	7743	睽		
1491	擷	2	92AF	6527	擷		
1565	絲	8	92F9	7DDC	絲		
1600	猷	7	935D	3E84		犴 ㄩ	
1641	胆	2	93A8	80C6	胆		膽，異體字問題。
1660	賺	1	93BB	8181	賺		
1854	啣	5	94E0	20E95		口 ㄩ 留	
1885	蠶	1	9540	882D	蠶		
1950	牝	19	95A3	7264	牝		
2046	踪	7	9644	8E2A	踪		
2053	躑	1	964B	8EA7	躑		
2107	髀	2	96A3	9AB2	髀		
2138	岳	2	96C2	7F36	岳		
2225	裓	2	975A	88D3	裓		
2228	腓	2	975D	818E	腓		
2343	鞆	1	97F2	4A5E		革 ㄩ 占	
2592	贓	6	99B1	8D1C	贓		

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
2667	歛	1	99FC	9FA1	歛		
2704		1	9A62			𠄎𠄎了	待造字。
2759	畱	1	9ABB	24CA2		亞𠄎田	畱，異體字問題。
2811		1	9AEF			亞𠄎田	畱，異體字問題。
3004	籊	3	9C54	7BD0	籊		
3019	𦏧	2	9C63	817C	𦏧		
3020	𦏨	1	9C64			𦏨	
3029	鋤	9	9C6D	941D	鋤		
3048	𦏩	3	9CA2	53C5	𦏩		
3099	𦏪	1	9CD5	68B9	𦏪		
3168	瘡	1	9D5B	24EA5		疒𠄎答	
3188	篋	1	9D6F	7BB2	篋		
3784	兪	2	814F	516A	兪		
3789	冲	23	8154	51B2	冲		冲，異體字問題。
3792	凉	13	8157	51C9	凉		凉，異體字問題。

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
3794	湊	2	8159	51D1	湊		湊，異體字問題。
3796	決	1	815B	51B3	決		決，異體字問題。
3797	凭	1	815C	51ED	凭		
3799	刳	1	815E	5226	刳		
3801	剗	6	8160	5257	剗		
3804	効	2	8163	52B9	効		
3814	却	922	816D	5374	却		
3815	𡗗	3	816E	537A	𡗗		
3817	厠	6	8170	53A0	厠		
3819	廡	12	8172	53AE	廡		廡，異體字問題。
3820	厦	4	8173	53A6	厦		
3826	叶	1	8179	53F6	叶		
3843	堦	17	81AC	5826	堦		
3846	堦	1	81AF	58FB	堦		
3849	獎	1	81B2	596C	獎		獎，異體字問題。

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
3858	寃	38	81BB	5BC3	寃		寃，異體字問題。
3864	峯	14	81C1	5CEF	峯		
3868	卮	2	81C5	5DF5	卮		
3876	廻	17	81CD	5EFB	廻		
3883	徧	5	81D4	5FA7	徧		
3884	忽	8	81D5	6031	忽		
3886	悞	15	81D7	609E	悞		
3899	担	1	81E4	62C5	担		
3909	擡	87	81EE	64E1	擡		
3921	暫	25	81FA	6673	暫		
3932	椽	1	8246	6936	椽		
3934	楫	22	8248	69C5	楫		
3936	棹	1	824A	6A70	棹		
3938	橈	32	824C	6AC8	橈		
3955	澗	4	825D	3D4E		澗 ㄇ 澗	
3966	烟	57	8268	70DF	烟		
3974	牀	58	8270	7240	牀		
3975	牕	75	8271	7255	牕		

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
3977	狗	1	8273	72E5	狗		
3978	猪	7	8274	732A	猪		Unicode 字型 呈現差異。
3990	甄	12	82A2	750E	甄		
3996	疎	12	82A8	758E	疎		
4008	癯	1	82B4	766F	癯		
4010	臯	1	82B6	7690	臯		臯，異體字問 題。
4012	鼓	1	82B8	76B7	鼓		鼓，異體字問 題。
4013	盃	43	82B9	76CC	盃		
4015	眦	2	82BB	7726	眦		
4016	着	4556	82BC	7740	着		
4024	礮	2	82C4	792E	礮		
4028	稟	20	82C8	7980	稟		
4033	稽	4	82CD	7A2D	稽		
4038	竈	7	82D2	7AC8	竈		
4041	豎	1	82D5	7AEA	豎		
4043	筭	1	82D7	7B53	筭		

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
4044	筋	6	82D8	7B6F	筋		
4045	箒	2	82D9	7B92	箒		
4055	粃	1	82E3	7C83	粃		
4058	糲	3	82E6	7CC9	糲		
4067	綉	1	82EF	7D89	綉		
4069	綑	22	82F1	7DB3	綑		
4074	縑	7	82F6	7E27	縑		
4080	羣	54	82FC	7FA3	羣		
4090	脚	218	8347	811A	脚		
4093	舐	1	834A	445B		舌  氏	
4105	菓	25	8356	83D3	菓		
4107	葱	6	8358	8471	葱		
4112	蔴	31	835D	8534	蔴		
4130	衅	2	836F	8845	衅		
4132		210	8371			血  衞	眾，異體字問題。
4134	衞	29	8373	885E	衞		衞，異體字問題。
4138	裘	4	8377	88A0	裘		

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
4139	袴	7	8378	88B4	袴		
4140	裱	3	8379	88B5	裱		
4146	覩	1	83A1	89A9	覩		
4173	躄	47	83BC	8EAD	躄		
4178	輓	3	83C1	8F2D	輓		
4184	迹	2	83C7	8FF9	迹		
4188	邨	9	83CB	90A8	邨		
4193	鈎	14	83D0	920E	鈎		
4195	鉢	5	83D2	9262	鉢		
4218	葦	1	83E9	97EE	葦		
4223	顛	14	83EE	984B	顛		
4242	鬪	14	8442	9B2D	鬪		
4247	鮎	1	8447	9B8E	鮎		
4256	鷄	36	8450	9DC4	鷄		
4262	麪	39	8456	9EAA	麪		Unicode 字型 呈現差異。
4307	罵	28	84A5	99E1	罵		罵，異體字問 題。
4309	刼	4	84A7	523C	刼		

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
4318		2	84B0			𠄎畝且 毛◻	𦉳，異體字問題。
4376	儂	41	84EA	510D	儂		
4432		6	8563			𠄎糸厂 墨◻	纏，異體字問題。
4447	捏	10	8572	63D1	捏		
4448		1	8573			𠄎糸厂 墨◻	纏，異體字問題。
4477		1	85B2			𠄎艹勹 一◻	芍，異體字問題。
4515	呪	1	85D8	546A	呪		
4518	响	5	85DB	54CD	响		
4533	唵	1	85EA	552B	唵		
4616	壳	1	867E	58F3	壳		
4750	卅	1	8767	5344	卅		卌，錯字，以程式全部取代。
4768	妬	3	8779	59AC	妬		
4799	顛	2	87BA	9856	顛		

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
4800	瞪	10	87BB	5654	瞪		
4802	吼	14	87BD	20BBE		口厶兀	
4805	瀆	2	87C0	51DF	瀆		瀆，異體字問題。
4816	彳	21	87CB			彳	
4904	尅	3	8864	5C05	尅		
5013	彳	4	88F3			彳	
5045	飪	1	8954	9901	飪		
5075	彳	1	8972	72AD		彳	
5101	馱	3	89AE	4B7E		馬厶犬	
5176		33	89F9			彳厶兀	沉，異體字問題。
5180	拘	1	89FD	6285	拘		
5195		4	8A4D			形木畝 且	檀，異體字問題。
5229	撐	1	8A6F	6490	撐		
5240	摔	1	8A7A	6481	摔		
5290	携	14	8ACE	643A	携		
5300	鬪	21	8AD8	9B2A	鬪		

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
5398	滄	1	8B7B	6E7B	滄		
5434	湟	5	8BC1	6E7C	湟		
5555	迥	5	8C7B	9008	迥		
5590	朶	48	8CC0	6736	朶		
5591	杞	2	8CC1	233CC		木△巳	杞，異體字問題。
5595	朶	1	8CC5	579C	朶		
5666	污	6	8D4D	6C5A	污		污，異體字問題。
5697	煑	12	8D6C	7151	煑		
5702	厨	27	8D71	53A8	厨		
5706	敘	31	8D75	654D	敘		敘，異體字問題。
5767	濶	1	8DD4	24050		氵△閣	濶，錯字，以程式全部取代。
5768	濶	3	8DD5	6FF6	濶		

二、手動取代表說明：

《兒女英雄傳》檔案中出現「●」字形，共計2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《兒女英雄傳》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、《兒女英雄傳》「●」字手動取代表

頁碼	檔案原文摘錄	原書字形	備註
p. 510	這麼對●着	睺	
p. 638	公子的●膊	月𠂇合	