

《儒林外史》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
2010/02/23 丁玟伶

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 1691 的「鬚」字，Unicode 編碼是 4BFC，由於 XP 並不支援，仍需使用構字式「鬚𠂇狄」。

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：

《儒林外史》使用舊版造字 111 個，字頻 3981 次，這 111 個造字中，94 個可轉成 Windows XP 能顯示的字，字頻 3929 次；另外 17 個字必須轉成構字式，字頻 52 次。

轉碼完成製作轉碼分析表，請參考附件一《儒林外史》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 4750 的「卅」字，而原書使用之字為「𠄎」，查詢教育部異體字字典確認「𠄎」並非「卅」之異體字，因此歸為錯字，以程式取代為正確字形「𠄎」。
 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 3097 的「彳𠄎幸」。

附件一、《儒林外史》轉碼分析表

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
296	𠄎	5	FBEC			𠄎	
297	𠄎	26	FBED			𠄎	
299	𠄎	1	FBEF			𠄎	

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
300	◻	1	FBF0			◻	
304	⊘	1	FBF4			⊘	
755	...	1	FEE0			...	心，錯字，以程式全部取代。
770	扌	2	FEEF			扌	
1153	幻	2	9075	362D		土△幻	
1213	睽	1	90D3	7743	睽		
1278	碍	3	9155	788D	碍		
1332		8	91AD			扌△嵩	攜，異體字問題。
1359	襪	1	91C8	25725		衤△冀	
1513	糧	1	92C5	7CAE	糧		糧，異體字問題。
1691	鬚	2	93DA	4BFC		髟△狄	
1981	讐	2	95C2	8B90	讐		
2046	踪	13	9644	8E2A	踪		
2157	酌	1	96D5	9167	酌		
2346	鐳	1	97F5	9427	鐳		

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
2592	臟	1	99B1	8D1C	臟		
2749	枱	1	9AB1	67B1	枱		
3016	闐	1	9C60	28D91		門△臭	
3021	忸	1	9C65			忸	
3097		1	9CD3			彳△幸	待造字。
3168	瘡	1	9D5B	24EA5		疒△答	
3177	蠃	14	9D64	7F4E	蠃		
3789	冲	6	8154	51B2	冲		沖，異體字問題。
3797	凭	2	815C	51ED	凭		
3799	刦	3	815E	5226	刦		
3804	効	13	8163	52B9	効		
3805	勅	2	8164	52C5	勅		
3814	却	316	816D	5374	却		
3815	𡗗	1	816E	537A	𡗗		
3819	廡	1	8172	53AE	廡		廡，異體字問題。
3820	厦	4	8173	53A6	厦		
3825	叠	1	8178	53E0	叠		

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
3829	咏	2	817C	548F	咏		
3843	堦	4	81AC	5826	堦		
3846	堦	39	81AF	58FB	堦		
3849	獎	3	81B2	596C	獎		獎，異體字問題。
3858	冤	23	81BB	5BC3	冤		冤，異體字問題。
3864	峯	36	81C1	5CEF	峯		
3876	廻	4	81CD	5EFB	廻		
3883	徧	3	81D4	5FA7	徧		
3884	忽	8	81D5	6031	忽		
3909	擡	71	81EE	64E1	擡		
3934	榻	6	8248	69C5	榻		
3942	毡	2	8250	6BE1	毡		
3966	烟	6	8268	70DF	烟		
3974	牀	78	8270	7240	牀		
3976	犁	1	8272	7282	犁		
3978	猪	53	8274	732A	猪		Unicode 字型呈現差異。

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
3996	疎	9	82A8	758E	疎		
4008	癯	1	82B4	766F	癯		
4016	着	244 1	82BC	7740	着		
4036	窰	2	82D0	7AB0	窰		
4038	竈	1	82D2	7AC8	竈		
4040	竝	1	82D4	7ADD	竝		
4041	豎	1	82D5	7AEA	豎		
4042	筍	5	82D6	7B0B	筍		
4043	筭	1	82D7	7B53	筭		
4044	筍	6	82D8	7B6F	筍		
4074	繚	6	82F6	7E27	繚		
4077	罇	3	82F9	7F47	罇		
4080	羣	12	82FC	7FA3	羣		
4090	脚	155	8347	811A	脚		
4093	舐	4	834A	445B		舌 𠂔 氏	
4105	菓	8	8356	83D3	菓		
4107	葱	3	8358	8471	葱		
4112	蔴	7	835D	8534	蔴		

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
4132		266	8371			血△𠂇	眾，異體字問題。
4134	衛	25	8373	885E	衛		衛，異體字問題。
4139	袴	3	8378	88B4	袴		
4173	𨔵	18	83BC	8EAD	𨔵		
4184	迹	1	83C7	8FF9	迹		
4193	鈎	2	83D0	920E	鈎		
4195	鉢	3	83D2	9262	鉢		
4203	隣	1	83DA	96A3	隣		鄰，異體字問題。
4232	隳	1	83F7	96B2	隳		
4242	鬪	1	8442	9B2D	鬪		
4256	鷄	17	8450	9DC4	鷄		
4267	鼈	2	845B	9F08	鼈		
4268	叢	3	845C	9F17	叢		
4272		1	8460			厂△龍	龐，異體字問題。
4307	罵	53	84A5	99E1	罵		罵，異體字問題。

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
							題。
4308	刼	4	84A6	5227	刼		
4428	躡	3	855F	8E70	躡		
4432		1	8563			𠄎 纒 厂 墨 𠄎	纏，異體字問題。
4447	捏	14	8572	63D1	捏		
4518	响	3	85DB	54CD	响		
4684		1	86E4			禛 𠄎 受	禛，錯字，以程式全部取代。
4750	卅	1	8767	5344	卅		卅，錯字，以程式全部取代。
4805	瀆	1	87C0	51DF	瀆		
4816	亻	1	87CB			亻	
4904	尅	2	8864	5C05	尅		
4996	斲	1	88E2	65B5	斲		
5008	幪	1	88EE	5E5E	幪		
5015	寗	7	88F5	5BD7	寗		Unicode 字型

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
							呈現差異。
5075	彳	1	8972	72AD		彳	
5176		20	89F9			汜 厶 冗	沉，異體字問題。
5195		2	8A4D			形木 畝 且 口	檀，異體字問題。
5229	撐	1	8A6F	6490	撐		
5230		1	8A70			形土 畝 且 口	壇，異體字問題。
5290	携	9	8ACE	643A	携		
5300	鬪	1	8AD8	9B2A	鬪		
5559	欸	1	8CA1	6B35	欸		
5590	朶	8	8CC0	6736	朶		
5595	朶	2	8CC5	579C	朶		
5666	污	2	8D4D	6C5A	污		污，異體字問題。
5702	厨	8	8D71	53A8	厨		
5706	敍	41	8D75	654D	敍		敍，異體字問題。

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
5742	澁	1	8DBB	6F81	澁		