

《弁而釵》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
2008/7/09 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 1616 的「運」字，Unicode 編碼是 284E6，由於 XP 並不支援，仍需使用構字式「辶△里」。

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：

《弁而釵》使用舊版造字 45 個，字頻 429 次，這 45 個造字中，35 個可轉成 Windows XP 能顯示的字，字頻 384 次；另外 10 個字必須轉成構字式，字頻 45 次。

轉碼完成製作轉碼分析表，請參考附件一《弁而釵》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 3791 的「況」字，是「況」的異體字，以標準字「況」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「卅」，「卅」為錯字，以程式取代為正確字形「卅」。

附件一、《弁而釵》轉碼分析表

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
296	厶	4	FBEC			厶	
297	厶	15	FBED			厶	
298	厶	1	FBEE			厶	
309	?	10	FBF9			?	龙，錯字，以程式全部取代。
687	彡	1	FE7A			彡	
750	丩	1	FEDB			丩	
759	广	1	FEE4			广	
1616	迺	1	936D	284E6		迺厶里	

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
1905	棍	2	9554	88E9	棍		
2046	踪	15	9644	8E2A	踪		
2152	葶	1	96D0	979F	葶		
2346	鐳	16	97F5	9427	鐳		
3789	冲	1	8154	51B2	冲		沖，異體字問題。
3791	况	10	8156	51B5	况		況，異體字問題。
3805	勅	11	8164	52C5	勅		
3824	叙	5	8177	53D9	叙		
3825	疊	5	8178	53E0	疊		疊，異體字問題。
3829	咏	1	817C	548F	咏		
3830	咤	1	817D	54A4	咤		
3836	喻	2	81A5	564F	喻		
3868	卮	1	81C5	5DF5	卮		
3886	悞	12	81D7	609E	悞		
3900	挖	2	81E5	62D5	挖		
3909	擡	10	81EE	64E1	擡		
3938	橈	2	824C	6AC8	橈		

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
3966	烟	14	8268	70DF	烟		
4000	疴	1	82AC	75B4	疴		
4016	着	163	82BC	7740	着		
4039	疇	1	82D3	7ADA	疇		
4055	粃	1	82E3	7C83	粃		
4065	紮	6	82ED	7D25	紮		
4068	綫	2	82F0	7DAB	綫		
4085	耻	12	8342	803B	耻		
4091	臯	1	8348	81EF	臯		
4105	菓	2	8356	83D3	菓		
4129	虬	3	836E	866C	虬		
4132		54	8371			血𠂇𠂇	眾，異體字問題。
4139	袴	4	8378	88B4	袴		
4167	賈	2	83B6	8CEB	賈		
4256	鷄	9	8450	9DC4	鷄		雞，異體字問題。
4307	罵	5	84A5	99E1	罵		罵，異體字問題。
4750	卅	1	8767			卅	卅，錯字，以程式全部取

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
							代。
4768	妬	6	8779	59AC	妬		妒，異體字問題。
4801	斌	1	87BC	5A2C	斌		
4816	彳	10	87CB			彳	