

《天湊巧》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
2008/7/08 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 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：

《天湊巧》使用舊版造字 30 個，字頻 189 次，這 30 個造字中，24 個可轉成 Windows XP 能顯示的字，字頻 178 次；另外 6 個字必須轉成構字式，字頻 11 次。

轉碼完成製作轉碼分析表，請參考附件一《天湊巧》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 4899 的「寒~~厶~~氷」字，而原書使用之字為「寨」，「寒~~厶~~氷」為錯字，以程式取代為正確字形「寨」。

附件一、《天湊巧》轉碼分析表

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
74	髻	1	FAAB	29B06		髻 厶 丐	
297	厶	1	FBED			厶	
298	厶	6	FBEE			厶	
299	形	1	FBEF			形	
300	口	1	FBF0			口	
1213	睽	1	90D3	7743	睽		
1888	苗	1	9543	3464		亻 厶 苗	
2046	踪	1	9644	8E2A	踪		
3789	冲	1	8154	51B2	冲		冲，異體字問

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
							題。
3791	况	10	8156	51B5	况		况，異體字問題。
3824	叙	3	8177	53D9	叙		
3826	叶	1	8179	53F6	叶		
3829	咏	1	817C	548F	咏		
3886	悞	5	81D7	609E	悞		
3966	烟	5	8268	70DF	烟		
4000	疴	1	82AC	75B4	疴		
4016	着	90	82BC	7740	着		
4038	竈	2	82D2	7AC8	竈		
4065	紮	1	82ED	7D25	紮		
4066	綺	1	82EE	7D5D	綺		
4068	綫	5	82F0	7DAB	綫		
4073	纒	1	82F5	7E6E	纒		
4107	葱	1	8358	8471	葱		
4132		29	8371			血𠂇𠂇	眾，異體字問題。
4256	鷄	5	8450	9DC4	鷄		雞，異體字問題。

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
4307	罵	7	84A5	99E1	罵		罵，異體字問題。
4372		2	84E6			𠄎广回 禾𠄎	廩，錯字，以程式全部取代。
4768	妬	3	8779	59AC	妬		妒，異體字問題。
4784	鞞	1	87AB	9793	鞞		
4899		1	885F			寒𠄎水	寨，錯字，以程式全部取代。