



《貪欣誤》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
2008/7/17 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 1347 的「𠄎」字，Unicode 編碼是 21E05，由於 XP 並不支援，仍需使用構字式「山弓女」。

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：

《貪欣誤》使用舊版造字 26 個，字頻 210 次，這 26 個造字中，24 個可轉成 Windows XP 能顯示的字，字頻 205 次；另外 2 個字必須轉成構字式，字頻 5 次。

轉碼完成製作轉碼分析表，請參考附件一《貪欣誤》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。

附件一、《貪欣誤》轉碼分析表

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
298	△	3	FBEE			△	
1347	焱	2	91BC	21E05		形山弓 女	
2046	踪	2	9644	8E2A	踪		
3789	冲	4	8154	51B2	冲		沖，異體字問題。
3791	况	11	8156	51B5	况		況，異體字問題。
3824	叙	1	8177	53D9	叙		
3826	叶	1	8179	53F6	叶		
3829	咏	1	817C	548F	咏		

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
3830	咤	1	817D	54A4	咤		
3966	烟	7	8268	70DF	烟		
4016	着	119	82BC	7740	着		
4036	窰	10	82D0	7AB0	窰		
4038	竈	1	82D2	7AC8	竈		
4065	紮	2	82ED	7D25	紮		
4068	綫	5	82F0	7DAB	綫		
4077	罇	2	82F9	7F47	罇		
4085	耻	1	8342	803B	耻		
4105	菓	3	8356	83D3	菓		
4107	葱	2	8358	8471	葱		
4132		16	8371			血𠂇𠂇	眾，異體字問題。
4193	鈎	2	83D0	920E	鈎		鈎，異體字問題。
4256	鷄	4	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	5	84A5	99E1	罵		罵，異體字問題。
4768	妬	3	8779	59AC	妬		妒，異體字問題。
5447	躡	1	8BCE	8E83	躡		
5753	潛	1	8DC6	6F5C	潛		