

《燕丹子》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
2010/06/11 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 797 的「掙」字，Unicode 編碼是 39D6，由於 XP 並不支援，仍需使用構字式「扌𠂔尾」。

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：

《燕丹子》使用舊版造字 5 個，字頻 6 次，這 5 個造字中，4 個可轉成 Windows XP 能顯示的字，字頻 5 次；另外 1 個字必須轉成構字式，字頻 1 次。

轉碼完成製作轉碼分析表，請參考附件一《燕丹子》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。

附件一、《燕丹子》轉碼分析表

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
797	挽	1	8E4B	39D6		扌厶尾	
1981	讐	2	95C2	8B90	讐		
3974	牀	1	8270	7240	牀		
4146	覩	1	83A1	89A9	覩		
4184	迹	1	83C7	8FF9	迹		