

《前漢書平話》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
2009/12/14 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 4093 的「舐」字，Unicode 編碼是 445B，由於 XP 並不支援，仍需使用構字式「舌𠂔氏」。

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：

《前漢書平話》使用舊版造字 18 個，字頻 40 次，這 18 個造字中，15 個可轉成 Windows XP 能顯示的字，字頻 37 次；另外 3 個字必須轉成構字式，字頻 3 次。

轉碼完成製作轉碼分析表，請參考附件一《前漢書平話》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. 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。

附件一、《前漢書平話》轉碼分析表

| 編號 | 造字 | 頻次 | Big5 | Unicode | WinXP | 構字式 | 備註 |
|------|----|----|------|---------|-------|-----|-----------------|
| 297 | 𠄎 | 1 | FBED | | | 𠄎 | |
| 780 | 彳 | 1 | FEF9 | | | 彳 | |
| 2833 | 刃 | 1 | 9B46 | 5204 | 刃 | | |
| 3826 | 叶 | 1 | 8179 | 53F6 | 叶 | | |
| 3840 | 坂 | 1 | 81A9 | 5742 | 坂 | | |
| 3843 | 堦 | 6 | 81AC | 5826 | 堦 | | |
| 3909 | 擡 | 1 | 81EE | 64E1 | 擡 | | |
| 3978 | 猪 | 1 | 8274 | 732A | 猪 | | Unicode 字型呈現差異。 |

| 編號 | 造字 | 頻次 | Big5 | Unicode | WinXP | 構字式 | 備註 |
|------|----|----|------|---------|-------|-------|----------|
| 4042 | 笋 | 1 | 82D6 | 7B0B | 笋 | | |
| 4080 | 羣 | 11 | 82FC | 7FA3 | 羣 | | |
| 4093 | 舐 | 1 | 834A | 445B | | 舌 𠂔 舐 | |
| 4134 | 衛 | 1 | 8373 | 885E | 衛 | | 衛，異體字問題。 |
| 4146 | 覩 | 3 | 83A1 | 89A9 | 覩 | | |
| 4173 | 躄 | 1 | 83BC | 8EAD | 躄 | | |
| 4309 | 刼 | 1 | 84A7 | 523C | 刼 | | |
| 4460 | 勅 | 3 | 85A1 | 52D1 | 勅 | | |
| 4768 | 妬 | 2 | 8779 | 59AC | 妬 | | |
| 5551 | 賈 | 3 | 8C77 | 8CF7 | 賈 | | |