

《魏忠賢小說斥奸書》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
2008/7/16 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 3788 的「慤」字，Unicode 編碼是 22871，由於 XP 並不支援，仍需使用構字式「殼厶心」。

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：

《魏忠賢小說斥奸書》使用舊版造字 54 個，字頻 455 次，這 54 個造字中，50 個可轉成 Windows XP 能顯示的字，字頻 446 次；另外 4 個字必須轉成構字式，字頻 9 次。

轉碼完成製作轉碼分析表，請參考附件一《魏忠賢小說斥奸書》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、《魏忠賢小說斥奸書》轉碼分析表

| 編號 | 造字 | 頻次 | Big5 | Unicode | WinXP | 構字式 | 備註 |
|------|----|----|------|---------|-------|-----|----------|
| 297 | 厶 | 6 | FBED | | | 厶 | |
| 675 | 食 | 1 | FE6E | | | 食 | |
| 908 | 燂 | 7 | 8EDC | 71DD | 燂 | | |
| 1642 | 脉 | 1 | 93A9 | 8109 | 脉 | | 脈，異體字問題。 |
| 2046 | 踪 | 4 | 9644 | 8E2A | 踪 | | |
| 2157 | 酌 | 1 | 96D5 | 9167 | 酌 | | |
| 2638 | 鷓 | 1 | 99DF | 9D55 | 鷓 | | |
| 3787 | 豕 | 9 | 8152 | 51A2 | 豕 | | |
| 3788 | 慙 | 1 | 8153 | 22871 | | 殼厶心 | |
| 3789 | 冲 | 6 | 8154 | 51B2 | 冲 | | 沖，異體字問題。 |
| 3791 | 况 | 22 | 8156 | 51B5 | 况 | | 況，異體字問 |

| 編號 | 造字 | 頻次 | Big5 | Unicode | WinXP | 構字式 | 備註 |
|------|----|-----|------|---------|-------|-----|----------|
| | | | | | | | 題。 |
| 3805 | 勅 | 1 | 8164 | 52C5 | 勅 | | |
| 3815 | 𪗇 | 1 | 816E | 537A | 𪗇 | | |
| 3824 | 叙 | 6 | 8177 | 53D9 | 叙 | | |
| 3825 | 疊 | 4 | 8178 | 53E0 | 疊 | | 疊，異體字問題。 |
| 3829 | 咏 | 6 | 817C | 548F | 咏 | | |
| 3849 | 獎 | 2 | 81B2 | 596C | 獎 | | 獎，異體字問題。 |
| 3851 | 姍 | 1 | 81B4 | 59F9 | 姍 | | |
| 3868 | 卮 | 1 | 81C5 | 5DF5 | 卮 | | |
| 3891 | 慙 | 1 | 81DC | 6159 | 慙 | | |
| 3909 | 擡 | 2 | 81EE | 64E1 | 擡 | | |
| 3966 | 烟 | 12 | 8268 | 70DF | 烟 | | |
| 3974 | 牀 | 1 | 8270 | 7240 | 牀 | | |
| 4008 | 癯 | 2 | 82B4 | 766F | 癯 | | |
| 4015 | 眦 | 1 | 82BB | 7726 | 眦 | | |
| 4016 | 着 | 236 | 82BC | 7740 | 着 | | |
| 4035 | 窳 | 1 | 82CF | 7A7D | 窳 | | |
| 4036 | 窳 | 3 | 82D0 | 7AB0 | 窳 | | |

| 編號 | 造字 | 頻次 | Big5 | Unicode | WinXP | 構字式 | 備註 |
|------|----|----|------|---------|-------|-----|----------|
| 4038 | 竈 | 1 | 82D2 | 7AC8 | 竈 | | |
| 4044 | 筋 | 1 | 82D8 | 7B6F | 筋 | | |
| 4055 | 粃 | 1 | 82E3 | 7C83 | 粃 | | |
| 4067 | 綉 | 8 | 82EF | 7D89 | 綉 | | |
| 4068 | 綫 | 8 | 82F0 | 7DAB | 綫 | | |
| 4073 | 縑 | 1 | 82F5 | 7E6E | 縑 | | |
| 4074 | 縑 | 2 | 82F6 | 7E27 | 縑 | | |
| 4080 | 羣 | 32 | 82FC | 7FA3 | 羣 | | |
| 4085 | 耻 | 5 | 8342 | 803B | 耻 | | |
| 4091 | 臯 | 1 | 8348 | 81EF | 臯 | | |
| 4132 | | 2 | 8371 | | | 血𠂇𠂇 | 眾，異體字問題。 |
| 4167 | 賈 | 4 | 83B6 | 8CEB | 賈 | | |
| 4169 | 趁 | 3 | 83B8 | 8DA6 | 趁 | | |
| 4184 | 迹 | 2 | 83C7 | 8FF9 | 迹 | | |
| 4186 | 遡 | 1 | 83C9 | 9061 | 遡 | | |
| 4193 | 鈎 | 2 | 83D0 | 920E | 鈎 | | 鈎，異體字問題。 |
| 4195 | 鉢 | 2 | 83D2 | 9262 | 鉢 | | |
| 4242 | 鬪 | 6 | 8442 | 9B2D | 鬪 | | |

| 編號 | 造字 | 頻次 | Big5 | Unicode | WinXP | 構字式 | 備註 |
|------|----|----|------|---------|-------|-----|----------|
| 4256 | 鷄 | 9 | 8450 | 9DC4 | 鷄 | | 雞，異體字問題。 |
| 4307 | 罵 | 9 | 84A5 | 99E1 | 罵 | | 罵，異體字問題。 |
| 4460 | 勅 | 1 | 85A1 | 52D1 | 勅 | | |
| 4768 | 妬 | 1 | 8779 | 59AC | 妬 | | 妒，異體字問題。 |
| 4817 | 寶 | 2 | 87CC | 5BF3 | 寶 | | 寶，異體字問題。 |
| 4904 | 尅 | 6 | 8864 | 5C05 | 尅 | | |
| 5075 | 彳 | 1 | 8972 | | | 彳 | |
| 5338 | 暎 | 4 | 8AFE | 668E | 暎 | | |