

《大明英烈傳》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
2010/01/08 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 5105 的「馱」字，Unicode 編碼是 4B7E，由於 XP 並不支援，仍需使用構字式「馬𠂔犬」。

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：

《大明英烈傳》使用舊版造字 89 個，字頻 1234 次，這 89 個造字中，77 個可轉成 Windows XP 能顯示的字，字頻 1192 次；另外 12 個字必須轉成構字式，字頻 42 次。

轉碼完成製作轉碼分析表，請參考附件一《大明英烈傳》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 4145 的「羈」字，Unicode 字型呈現為「羈」，實際上仍為同一字。

附件一、《大明英烈傳》轉碼分析表

| 編號 | 造字 | 頻次 | Big5 | Unicode | WinXP | 構字式 | 備註 |
|------|----|----|------|---------|-------|-------|----------|
| 20 | 𠄎 | 2 | FA53 | | | 𠄎 𠄎 𠄎 | |
| 296 | 𠄎 | 1 | FBEC | | | 𠄎 | |
| 297 | 𠄎 | 15 | FBED | | | 𠄎 | |
| 299 | 𠄎 | 1 | FBEF | | | 𠄎 | |
| 300 | 𠄎 | 1 | FBF0 | | | 𠄎 | |
| 639 | 𠄎 | 1 | FE4A | | | 𠄎 | |
| 750 | 𠄎 | 1 | FEDB | | | 𠄎 | |
| 1073 | | 1 | 8FE4 | | | 𠄎 𠄎 𠄎 | 𠄎，異體字問題。 |

| 編號 | 造字 | 頻次 | Big5 | Unicode | WinXP | 構字式 | 備註 |
|------|----|-----|------|---------|-------|-----|----------|
| 2346 | 鐳 | 2 | 97F5 | 9427 | 鐳 | | |
| 2993 | 办 | 1 | 9C49 | 5205 | 办 | | |
| 3010 | 攷 | 1 | 9C5A | 5746 | 攷 | | |
| 3227 | 艚 | 1 | 9DB8 | 826A | 艚 | | |
| 3255 | 善 | 1 | 9DD4 | 5584 | 善 | | |
| 3784 | 兪 | 119 | 814F | 516A | 兪 | | |
| 3789 | 冲 | 19 | 8154 | 51B2 | 冲 | | |
| 3799 | 刼 | 1 | 815E | 5226 | 刼 | | |
| 3804 | 効 | 4 | 8163 | 52B9 | 効 | | |
| 3805 | 勅 | 4 | 8164 | 52C5 | 勅 | | |
| 3819 | 廝 | 3 | 8172 | 53AE | 廝 | | 廝，異體字問題。 |
| 3830 | 咤 | 4 | 817D | 54A4 | 咤 | | |
| 3843 | 堦 | 16 | 81AC | 5826 | 堦 | | |
| 3846 | 壻 | 3 | 81AF | 58FB | 壻 | | |
| 3849 | 獎 | 4 | 81B2 | 596C | 獎 | | 獎，異體字問題。 |
| 3858 | 寃 | 11 | 81BB | 5BC3 | 寃 | | 寃，異體字問題。 |

| 編號 | 造字 | 頻次 | Big5 | Unicode | WinXP | 構字式 | 備註 |
|------|----|-----|------|---------|-------|-------|----------|
| 3864 | 峯 | 17 | 81C1 | 5CEF | 峯 | | |
| 3868 | 卮 | 2 | 81C5 | 5DF5 | 卮 | | |
| 3876 | 廻 | 10 | 81CD | 5EFB | 廻 | | |
| 3945 | | 1 | 8253 | | | 彳 厶 义 | 汙，異體字問題。 |
| 3966 | 烟 | 18 | 8268 | 70DF | 烟 | | |
| 3974 | 牀 | 19 | 8270 | 7240 | 牀 | | |
| 3976 | 犁 | 2 | 8272 | 7282 | 犁 | | |
| 4011 | 皞 | 1 | 82B7 | 76A1 | 皞 | | |
| 4016 | 着 | 538 | 82BC | 7740 | 着 | | |
| 4028 | 稟 | 6 | 82C8 | 7980 | 稟 | | |
| 4032 | 秣 | 1 | 82CC | 2578A | | 禾 厶 未 | 秣，異體字問題。 |
| 4034 | 穉 | 1 | 82CE | 7A49 | 穉 | | |
| 4035 | 穽 | 1 | 82CF | 7A7D | 穽 | | |
| 4039 | 竝 | 1 | 82D3 | 7ADA | 竝 | | |
| 4043 | 筭 | 1 | 82D7 | 7B53 | 筭 | | |
| 4044 | 筓 | 4 | 82D8 | 7B6F | 筓 | | |
| 4045 | 筓 | 1 | 82D9 | 7B92 | 筓 | | |

| 編號 | 造字 | 頻次 | Big5 | Unicode | WinXP | 構字式 | 備註 |
|------|----|----|------|---------|-------|-----|-----------------|
| 4067 | 綉 | 1 | 82EF | 7D89 | 綉 | | |
| 4074 | 縹 | 3 | 82F6 | 7E27 | 縹 | | |
| 4080 | 羣 | 57 | 82FC | 7FA3 | 羣 | | |
| 4107 | 葱 | 1 | 8358 | 8471 | 葱 | | |
| 4112 | 蔴 | 15 | 835D | 8534 | 蔴 | | |
| 4129 | 虬 | 51 | 836E | 866C | 虬 | | |
| 4132 | | 2 | 8371 | | | 血𠄎𠄎 | 眾，異體字問題。 |
| 4134 | 衛 | 19 | 8373 | 885E | 衛 | | 衛，異體字問題。 |
| 4139 | 袴 | 1 | 8378 | 88B4 | 袴 | | |
| 4144 | 霸 | 1 | 837D | 8987 | 霸 | | Unicode 字型呈現差異。 |
| 4145 | 羈 | 1 | 837E | 898A | 羈 | | Unicode 字型呈現差異。 |
| 4146 | 覩 | 4 | 83A1 | 89A9 | 覩 | | |
| 4173 | 𧈧 | 4 | 83BC | 8EAD | 𧈧 | | |
| 4180 | 臯 | 1 | 83C3 | 8FA0 | 臯 | | |
| 4186 | 遡 | 1 | 83C9 | 9061 | 遡 | | |

| 編號 | 造字 | 頻次 | Big5 | Unicode | WinXP | 構字式 | 備註 |
|------|----|----|------|---------|-------|-----|----------|
| 4193 | 鈎 | 6 | 83D0 | 920E | 鈎 | | |
| 4218 | 葦 | 1 | 83E9 | 97EE | 葦 | | |
| 4221 | 頰 | 1 | 83EC | 9833 | 頰 | | |
| 4256 | 鷄 | 3 | 8450 | 9DC4 | 鷄 | | |
| 4267 | 鼈 | 2 | 845B | 9F08 | 鼈 | | |
| 4307 | 罵 | 22 | 84A5 | 99E1 | 罵 | | 罵，異體字問題。 |
| 4308 | 刼 | 3 | 84A6 | 5227 | 刼 | | |
| 4309 | 刼 | 15 | 84A7 | 523C | 刼 | | |
| 4447 | 捏 | 1 | 8572 | 63D1 | 捏 | | |
| 4738 | 睺 | 4 | 875B | 777A | 睺 | | |
| 4742 | | 3 | 875F | | | 衣△執 | 褻，異體字問題。 |
| 4798 | | 1 | 87B9 | | | 白△夕 | 囟，異體字問題。 |
| 4805 | 瀆 | 3 | 87C0 | 51DF | 瀆 | | 瀆，異體字問題。 |
| 4816 | 亻 | 4 | 87CB | | | 亻 | |
| 4904 | 尅 | 6 | 8864 | 5C05 | 尅 | | |

| 編號 | 造字 | 頻次 | Big5 | Unicode | WinXP | 構字式 | 備註 |
|------|----|----|------|---------|-------|--|----------|
| 5008 | 幪 | 1 | 88EE | 5E5E | 幪 | | |
| 5019 | 畧 | 1 | 88F9 | | | 畧 | |
| 5101 | 馱 | 3 | 89AE | 4B7E | | 馬厶犬 | |
| 5174 | | 3 | 89F7 | | | 形刀兕 、  | 兔，異體字問題。 |
| 5176 | | 30 | 89F9 | | | 彳厶冗 | 沉，異體字問題。 |
| 5218 | | 31 | 8A64 | | | 形宀寬 、  | 寬，異體字問題。 |
| 5225 | 邨 | 8 | 8A6B | 460F | | 血厶卩 | |
| 5246 | | 5 | 8AA2 | | | 形正人 月<<  | 踰，異體字問題。 |
| 5290 | 携 | 5 | 8ACE | 643A | 携 | | |
| 5299 | 戟 | 25 | 8AD7 | 39B8 | | 卓厶戈 | 戟，異體字問題。 |
| 5300 | 鬪 | 2 | 8AD8 | 9B2A | 鬪 | | |
| 5583 | 殮 | 1 | 8CB9 | 98F1 | 殮 | | |
| 5590 | 朶 | 20 | 8CC0 | 6736 | 朶 | | |
| 5595 | 朶 | 1 | 8CC5 | 579C | 朶 | | |

| 編號 | 造字 | 頻次 | Big5 | Unicode | WinXP | 構字式 | 備註 |
|------|----|----|------|---------|-------|-----|----------|
| 5632 | 替 | 4 | 8CEA | 3B31 | | 替 | |
| 5702 | 厨 | 2 | 8D71 | 53A8 | 厨 | | |
| 5706 | 敍 | 13 | 8D75 | 654D | 敍 | | 敍，異體字問題。 |
| 5768 | 濶 | 1 | 8DD5 | 6FF6 | 濶 | | |