

《孔叢子》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
2010/08/06 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 74 的「髻」字，Unicode 編碼是 29B06，由於 XP 並不支援，仍需使用構字式「髻𠄎𠄎」。

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：

《孔叢子》使用舊版造字 54 個，字頻 154 次，這 54 個造字中，51 個可轉成 Windows XP 能顯示的字，字頻 151 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《孔叢子》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，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 4132 的「血𠂇𠂇」字，是「眾」的異體字，以標準字「眾」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 4091 的「鼻」字，Unicode 字型呈現為「鼻」，實際上仍為同一字。

附件一、《孔叢子》轉碼分析表

| 編號 | 造字 | 頻次 | Big5 | Unicode | WinXP | 構字式 | 備註 |
|------|----|----|------|---------|-------|-----|-----------------|
| 74 | 髻 | 1 | FAAB | 29B06 | | 髻𠂇𠂇 | |
| 980 | 𠂇 | 1 | 8F65 | 72AB | 𠂇 | | |
| 1014 | 玠 | 1 | 8FA9 | 73CE | 玠 | | |
| 1238 | 疾 | 14 | 90EC | 77E6 | 疾 | | |
| 1925 | 羈 | 1 | 9568 | 8989 | 羈 | | Unicode 字型呈現差異。 |
| 1981 | 𠂇 | 2 | 95C2 | 8B90 | 𠂇 | | |
| 2044 | 踈 | 1 | 9642 | 8E08 | 踈 | | |

| 編號 | 造字 | 頻次 | Big5 | Unicode | WinXP | 構字式 | 備註 |
|------|----|----|------|---------|-------|-------|------------|
| 2415 | 鬪 | 1 | 987B | 95D8 | 鬪 | | |
| 2505 | 翻 | 1 | 98F7 | 98DC | 翻 | | |
| 2516 | 飾 | 1 | 9943 | 9919 | 飾 | | |
| 2811 | | 2 | 9AEF | | | 𠂇 𠂈 田 | 雷，異體字問題。 |
| 2935 | 穈 | 1 | 9BCE | 7A6A | 穈 | | |
| 2977 | 却 | 1 | 9BF8 | | | 去 𠂉 卩 | |
| 3035 | 賓 | 10 | 9C73 | 8CD4 | 賓 | | |
| 3048 | 叅 | 1 | 9CA2 | 53C5 | 叅 | | |
| 3791 | 况 | 3 | 8156 | 51B5 | 况 | | 况，異體字問題。 |
| 3796 | 決 | 1 | 815B | 51B3 | 決 | | 決，異體字問題。 |
| 3804 | 効 | 2 | 8163 | 52B9 | 効 | | |
| 3817 | 廁 | 1 | 8170 | 53A0 | 廁 | | 廁，異體字問題。 |
| 3876 | 廻 | 2 | 81CD | 5EFB | 廻 | | |
| 3883 | 徧 | 3 | 81D4 | 5FA7 | 徧 | | |
| 3935 | 槩 | 1 | 8249 | 3BA3 | 槩 | | Unicode 字型 |

| 編號 | 造字 | 頻次 | Big5 | Unicode | WinXP | 構字式 | 備註 |
|------|----|----|------|---------|-------|-----|-----------------|
| | | | | | | | 呈現差異。 |
| 3940 | 爵 | 1 | 824E | 6B1D | 爵 | | |
| 3974 | 牀 | 2 | 8270 | 7240 | 牀 | | |
| 3979 | 猫 | 1 | 8275 | 732B | 猫 | | |
| 3996 | 疎 | 1 | 82A8 | 758E | 疎 | | |
| 4010 | 皐 | 1 | 82B6 | 7690 | 皐 | | 皐，異體字問題。 |
| 4014 | 蓋 | 4 | 82BA | 8462 | 蓋 | | |
| 4028 | 稟 | 1 | 82C8 | 7980 | 稟 | | |
| 4038 | 竈 | 1 | 82D2 | 7AC8 | 竈 | | |
| 4057 | 莖 | 8 | 82E5 | 585F | 莖 | | |
| 4080 | 羣 | 9 | 82FC | 7FA3 | 羣 | | |
| 4085 | 耻 | 4 | 8342 | 803B | 耻 | | |
| 4091 | 皐 | 4 | 8348 | 81EF | 皐 | | Unicode 字型呈現差異。 |
| 4132 | | 30 | 8371 | | | 血会𠂇 | 眾，異體字問題。 |
| 4145 | 羈 | 1 | 837E | 898A | 羈 | | Unicode 字型呈現差異。 |

| 編號 | 造字 | 頻次 | Big5 | Unicode | WinXP | 構字式 | 備註 |
|------|----|----|------|---------|-------|-------|----------|
| 4146 | 覩 | 5 | 83A1 | 89A9 | 覩 | | |
| 4156 | 諂 | 2 | 83AB | 8B1F | 諂 | | |
| 4184 | 迹 | 1 | 83C7 | 8FF9 | 迹 | | |
| 4203 | 隣 | 4 | 83DA | 96A3 | 隣 | | 鄰，異體字問題。 |
| 4256 | 鷄 | 1 | 8450 | 9DC4 | 鷄 | | |
| 4259 | 麇 | 1 | 8453 | 9E95 | 麇 | | |
| 4275 | 义 | 1 | 8463 | 4E49 | 义 | | |
| 4437 | 廩 | 1 | 8568 | 5EEA | 廩 | | |
| 4460 | 勅 | 1 | 85A1 | 52D1 | 勅 | | |
| 4493 | 特 | 1 | 85C2 | 3E40 | | 牛 厶 寺 | |
| 5074 | 尫 | 3 | 8971 | 5C2B | 尫 | | |
| 5098 | 寤 | 1 | 89AB | 7AB9 | 寤 | | |
| 5147 | 恠 | 1 | 89DC | 6060 | 恠 | | |
| 5290 | 携 | 2 | 8ACE | 643A | 携 | | |
| 5300 | 鬪 | 6 | 8AD8 | 9B2A | 鬪 | | |
| 5408 | 紕 | 1 | 8BA7 | 7D4D | 紕 | | |
| 5439 | 栢 | 1 | 8BC6 | 6822 | 栢 | | |
| 5558 | 龕 | 1 | 8C7E | 9E84 | 龕 | | |

二、手動取代表說明：

《孔叢子》檔案中需手動修改的情況分為以下幾種：

(一)《孔叢子》檔案中出現「●」字形，共計 8 字，比對原書後以 Unicod 字形或構字式手動取代，並製作「●」字手動取代表，請參考附件二。欄位說明如下：

1. 頁碼：「●」字形所在位置之原書頁碼。
2. 檔案原文摘錄：摘錄《孔叢子》檔案有「●」字形的文句。
3. 原書字形：「●」在原書中之字形，以此字取代檔案原文的「●」。
4. 備註：記錄修改相同詞句的次數或其他事項。

附件二、《孔叢子》「●」字手動取代表

| 頁碼 | 檔案原文摘錄 | 原書字形 | 備註 |
|---------|--------|------|-------|
| p. 1588 | 乃為●災 | 生△月 | 改為眚 |
| p. 1588 | ●者不可生 | 夕△人 | |
| p. 1589 | ●枳棘充路 | 之△市 | |
| p. 1590 | 麟出而● | 夕△人 | |
| p. 1600 | 柔順終以不● | 敝△大 | |
| p. 1601 | ●牛元武 | 牛△麗 | |
| p. 1608 | ●嬖寵之官 | 口△文 | |
| p. 1618 | 手搏●獸 | 足△呂免 | 改為足△鼯 |

(二)《孔叢子》檔案中，對於舊版造字以外發現的錯字，以手動修改，並製作錯字修改表，請參考附件三。欄位說明如下：

1. 檔案原文摘錄：摘錄檔案《孔叢子》中的文句段落，其中紅字底線的部分，為需要手動取代的原文。
2. 手動取代結果：變更過後的檔案內容，其中藍字底線的部分，為

修改後的結果。

3. 備註：記錄修改相同詞句的次數或其他事項。

附件三、《孔叢子》缺字外錯字修改

| 檔案原文摘錄 | 手動取代結果 | 備註 |
|--------------|--------------|----|
| <u>第</u> 之私房 | <u>第</u> 之私房 | |