

# 《京本通俗小說》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  
2008/9/01 丁玫伶

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 1616 的「運」字，Unicode 編碼是 284E6，由於 XP 並不支援，仍需使用構字式「辶△里」。

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：

《京本通俗小說》使用舊版造字 55 個，字頻 587 次，這 55 個造字中，52 個可轉成 Windows XP 能顯示的字，字頻 581 次；另外 3 個字必須轉成構字式，字頻 6 次。

轉碼完成製作轉碼分析表，請參考附件一《京本通俗小說》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。
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 4263 的「麩」字，Unicode 字型呈現為「麩」，實際上仍為同一字。

附件一、《京本通俗小說》轉碼分析表

| 編號   | 造字 | 頻次 | Big5 | Unicode | WinXP | 構字式 | 備註       |
|------|----|----|------|---------|-------|-----|----------|
| 297  | 厶  | 3  | FBED |         |       | 厶   |          |
| 298  | 厶  | 1  | FBEE |         |       | 厶   |          |
| 1039 | 稟  | 3  | 8FC2 | 69C0    | 稟     |     |          |
| 1491 | 擷  | 1  | 92AF | 6527    | 擷     |     |          |
| 1616 | 逕  | 2  | 936D | 284E6   |       | 之厶里 |          |
| 2999 | 咆  | 1  | 9C4F | 5523    | 咆     |     |          |
| 3791 | 况  | 6  | 8156 | 51B5    | 况     |     | 况，異體字問題。 |

| 編號   | 造字 | 頻次  | Big5 | Unicode | WinXP | 構字式 | 備註       |
|------|----|-----|------|---------|-------|-----|----------|
| 3794 | 湊  | 6   | 8159 | 51D1    | 湊     |     | 湊，異體字問題。 |
| 3796 | 決  | 2   | 815B | 51B3    | 決     |     | 決，異體字問題。 |
| 3805 | 勅  | 2   | 8164 | 52C5    | 勅     |     |          |
| 3814 | 却  | 205 | 816D | 5374    | 却     |     |          |
| 3819 | 廝  | 6   | 8172 | 53AE    | 廝     |     | 廝，異體字問題。 |
| 3824 | 叙  | 8   | 8177 | 53D9    | 叙     |     |          |
| 3825 | 疊  | 2   | 8178 | 53E0    | 疊     |     |          |
| 3830 | 咤  | 1   | 817D | 54A4    | 咤     |     |          |
| 3843 | 堦  | 1   | 81AC | 5826    | 堦     |     |          |
| 3846 | 壻  | 12  | 81AF | 58FB    | 壻     |     |          |
| 3858 | 寃  | 12  | 81BB | 5BC3    | 寃     |     | 寃，異體字問題。 |
| 3909 | 擡  | 10  | 81EE | 64E1    | 擡     |     |          |
| 3931 | 椀  | 1   | 8245 | 6900    | 椀     |     |          |

| 編號   | 造字 | 頻次  | Big5 | Unicode | WinXP | 構字式 | 備註                  |
|------|----|-----|------|---------|-------|-----|---------------------|
| 3966 | 烟  | 5   | 8268 | 70DF    | 烟     |     |                     |
| 3968 | 煨  | 1   | 826A | 7145    | 煨     |     |                     |
| 3974 | 牀  | 2   | 8270 | 7240    | 牀     |     |                     |
| 3978 | 猪  | 7   | 8274 | 732A    | 猪     |     | Unicode 字型<br>呈現差異。 |
| 3996 | 疎  | 3   | 82A8 | 758E    | 疎     |     |                     |
| 4016 | 着  | 123 | 82BC | 7740    | 着     |     |                     |
| 4028 | 稟  | 12  | 82C8 | 7980    | 稟     |     |                     |
| 4038 | 竈  | 1   | 82D2 | 7AC8    | 竈     |     |                     |
| 4055 | 粃  | 1   | 82E3 | 7C83    | 粃     |     |                     |
| 4080 | 羣  | 3   | 82FC | 7FA3    | 羣     |     |                     |
| 4090 | 脚  | 32  | 8347 | 811A    | 脚     |     |                     |
| 4114 | 藥  | 1   | 835F | 8616    | 藥     |     |                     |
| 4132 |    | 43  | 8371 |         |       | 血𠄎𠄎 | 眾，異體字問題。            |
| 4146 | 覩  | 1   | 83A1 | 89A9    | 覩     |     | 睹，異體字問題。            |

| 編號   | 造字 | 頻次 | Big5 | Unicode | WinXP | 構字式 | 備註              |
|------|----|----|------|---------|-------|-----|-----------------|
| 4167 | 賚  | 1  | 83B6 | 8CEB    | 賚     |     |                 |
| 4193 | 鈎  | 2  | 83D0 | 920E    | 鈎     |     | 鈎，異體字問題。        |
| 4203 | 隣  | 14 | 83DA | 96A3    | 隣     |     | 鄰，異體字問題。        |
| 4232 | 隲  | 1  | 83F7 | 96B2    | 隲     |     |                 |
| 4242 | 鬪  | 2  | 8442 | 9B2D    | 鬪     |     |                 |
| 4263 | 麩  | 1  | 8457 | 9EAF    | 麩     |     | Unicode 字型呈現差異。 |
| 4266 | 鼃  | 1  | 845A | 9F02    | 鼃     |     |                 |
| 4307 | 罵  | 14 | 84A5 | 99E1    | 罵     |     | 罵，異體字問題。        |
| 4447 | 捏  | 2  | 8572 | 63D1    | 捏     |     |                 |
| 4702 | 場  | 1  | 86F6 | 5872    | 場     |     |                 |
| 4768 | 妬  | 1  | 8779 | 59AC    | 妬     |     | 妒，異體字問題。        |
| 4817 | 寶  | 13 | 87CC | 5BF3    | 寶     |     | 寶，異體字問題。        |

| 編號   | 造字 | 頻次 | Big5 | Unicode | WinXP | 構字式   | 備註              |
|------|----|----|------|---------|-------|-------|-----------------|
|      |    |    |      |         |       |       | 題。              |
| 5176 |    | 6  | 89F9 |         |       | 汙 ㄨㄣˊ | 沉，異體字問題。        |
| 5240 | 擗  | 1  | 8A7A | 6481    | 擗     |       |                 |
| 5434 | 湟  | 1  | 8BC1 | 6E7C    | 湟     |       |                 |
| 5551 | 賈  | 1  | 8C77 | 8CF7    | 賈     |       |                 |
| 5555 | 迥  | 1  | 8C7B | 9008    | 迥     |       |                 |
| 5583 | 殮  | 1  | 8CB9 | 98F1    | 殮     |       |                 |
| 5666 | 污  | 1  | 8D4D | 6C5A    | 污     |       | 污，異體字問題。        |
| 5697 | 煑  | 1  | 8D6C | 7151    | 煑     |       | Unicode 字型呈現差異。 |
| 5706 | 敘  | 1  | 8D75 | 654D    | 敘     |       | 敘，異體字問題。        |