# 《蒲松齡集》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
2010/02/01　　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 1691 的「鬚」字，Unicode 編碼是 4BFC，由於 XP 並不支援，仍需使用構字式「髟㠯狄」。

**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：

《蒲松齡集》使用舊版造字 187 個，字頻 7689 次，這 187 個造字中，148 個可轉成 Windows XP 能顯示的字，字頻 6806 次；另外 39 個字必須轉成構字式，字頻 883 次。

轉碼完成製作轉碼分析表，請參考附件一《蒲松齡集》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. 部份使用不同字形，手動修改：檔案中為同一舊版造字，校對時卻發現原書使用兩個以上不同的字形，可能部份舊版造字為錯字或原書使用兩種異體字形，導致同一舊版造字轉碼時對應到兩個以上的字形；將一部份手動修改，剩餘的頻次則用程式取代成對應的 Unicode 或構字式。例：造字編號 2853 的「式」字頻次為 149，依書將 1 字手動修改為「弋⟁三」，剩餘 148 字轉為「式」。手動修改部份請參考附件二《蒲松齡集》手動取代表。
3. 錯字，以程式全部取代：與原書字形不符，並查詢教育部異體字字典確認非異體字後，皆歸類為錯字，以程式取代為正確字形。例：檔案中使用造字編號 4750 的「廾」字，而原書使用之字為「艹」，查詢教育部異體字字典確認「艹」並非「廾」之異體字，因此歸為錯字，以程式取代為正確字形「艹」。
4. 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。
5. 依原書修改未定義字形：編號 309 舊版造字「」代表未定義字型，依原書修改為書中字形。例：「」在書中的字形為「臭」，依原書修改為「臭」。

## 附件一、《蒲松齡集》轉碼分析表

| 編號 | 造字 | 頻次 | Big5 | Unicode | WinXP | 構字式 | 備註 |
|---|---|---|---|---|---|---|---|
| 33 | 蔔 | 4 | FA60 | 85A5 | 蔔 | | |
| 59 | 吠 | 15 | FA7A | 356D | | 口从天 | |
| 77 | 哪 | 3 | FAAE | 35BF | | 口从耶 | |
| 296 | 𠆢 | 31 | FBEC | | | 𠆢 | |
| 297 | 从 | 520 | FBED | | | 从 | |
| 298 | 𠔼 | 22 | FBEE | | | 𠔼 | |
| 299 | 形 | 30 | FBEF | | | 形 | |
| 300 | ⊡ | 30 | FBF0 | | | ⊡ | |
| 309 | ❓ | 4 | FBF9 | | | ❓ | 臭，依原書修改未定義字形。 |
| 510 | 㽘 | 1 | FD66 | | | 㽘 | |
| 520 | 囚 | 4 | FD70 | | | 囚 | 罒，異體字問題。 |
| 658 | 疒 | 2 | FE5D | | | 疒 | |
| 675 | 飠 | 22 | FE6E | | | 飠 | |
| 717 | 足 | 16 | FEBA | | | 足 | |
| 742 | 〆 | 7 | FED3 | | | 〆 | |
| 750 | 刂 | 25 | FEDB | | | 刂 | |

| 編號 | 造字 | 頻次 | Big5 | Unicode | WinXP | 構字式 | 備註 |
|---|---|---|---|---|---|---|---|
| 768 | 艹 | 10 | FEED | | | 艹 | |
| 770 | 扌 | 54 | FEEF | | | 扌 | |
| 772 | 辶 | 9 | FEF1 | | | 辶 | |
| 773 | 厶 | 4 | FEF2 | | | 厶 | |
| 777 | 宀 | 1 | FEF6 | | | 宀 | |
| 780 | 氵 | 7 | FEF9 | | | 氵 | |
| 784 | 亻 | 14 | FEFD | | | 亻 | |
| 814 | 蠆 | 1 | 8E5C | 87C7 | 蠆 | | |
| 1002 | 鳩 | 1 | 8F7B | 9D02 | 鳩 | | |
| 1137 | 塾 | 1 | 9065 | 58AA | 塾 | | |
| 1170 | 曈 | 2 | 90A8 | 7583 | 曈 | | |
| 1213 | 睃 | 3 | 90D3 | 7743 | 睃 | | |
| 1226 | 罸 | 2 | 90E0 | 7F78 | 罸 | | |
| 1321 | 礭 | 2 | 91A2 | 7936 | 礭 | | |
| 1394 | 啎 | 1 | 91EB | 35D6 | | 口厽舀 | |
| 1409 | 寵 | 4 | 91FA | 7AC9 | 寵 | | |
| 1464 | 鋷 | 1 | 9272 | 28A20 | | 金厽匊 | |
| 1491 | 攩 | 3 | 92AF | 6527 | 攩 | | |

| 編號 | 造字 | 頻次 | Big5 | Unicode | WinXP | 構字式 | 備註 |
|---|---|---|---|---|---|---|---|
| 1513 | 粮 | 1 | 92C5 | 7CAE | 粮 | | 糧，異體字問題。 |
| 1565 | 縣 | 1 | 92F9 | 7DDC | 縣 | | |
| 1591 | 羌 | 1 | 9354 | | | 羌 | |
| 1598 | 翰 | 3 | 935B | 4A7A | | 革翁 | |
| 1641 | 胆 | 2 | 93A8 | 80C6 | 胆 | | 膽，異體字問題。 |
| 1653 | 腍 | 1 | 93B4 | 8142 | 腍 | | |
| 1691 | 髮 | 4 | 93DA | 4BFC | | 髟狄 | |
| 1791 | 盜 | 2 | 94A1 | | | 水皿 | |
| 1854 | 嗮 | 4 | 94E0 | 20E95 | | 口留 | |
| 1856 | 舐 | 11 | 94E2 | 74F3 | 舐 | | |
| 1926 | 覓 | 1 | 9569 | 8994 | 覓 | | |
| 1981 | 讋 | 1 | 95C2 | 8B90 | 讋 | | |
| 1994 | 達 | 1 | 95CF | 6FBE | 達 | | |
| 2021 | 趁 | 4 | 95EA | 8D82 | 趁 | | |
| 2063 | 餚 | 1 | 9655 | 29735 | | 飠崙 | |
| 2107 | 骲 | 1 | 96A3 | 9AB2 | 骲 | | |
| 2147 | 蟮 | 2 | 96CB | 87EE | 蟮 | | |

| 編號 | 造字 | 頻次 | Big5 | Unicode | WinXP | 構字式 | 備註 |
|---|---|---|---|---|---|---|---|
| 2305 | 鐫 | 1 | 97CC | 93BB | 鐫 | | |
| 2408 | 鬧 | 25 | 9874 | 9599 | 鬧 | | 鬧，異體字問題。 |
| 2679 | 哆 | 19 | 9A49 | 3597 | | 口⿰夕岁 | |
| 2826 | 廵 | 1 | 9AFE | 5EF5 | 廵 | | |
| 2853 | 弍 | 149 | 9B5A | 5F0D | 弍 | | 部份使用不同字形「弋⿱一⿱一三」，手動修改。 |
| 3004 | 箍 | 2 | 9C54 | 7BD0 | 箍 | | |
| 3029 | 鏬 | 5 | 9C6D | 941D | 鏬 | | |
| 3098 | 揸 | 17 | 9CD4 | 63F8 | 揸 | | Unicode 字型呈現差異。 |
| 3104 | 薟 | 1 | 9CDA | 861D | 薟 | | |
| 3169 | 砘 | 2 | 9D5C | 7818 | 砘 | | |
| 3177 | 鑭 | 4 | 9D64 | 7F4E | 鑭 | | |
| 3770 | 个 | 3 | 8141 | 4E2A | 个 | | |
| 3789 | 冲 | 108 | 8154 | 51B2 | 冲 | | 沖，異體字問題。 |

| 編號 | 造字 | 頻次 | Big5 | Unicode | WinXP | 構字式 | 備註 |
|---|---|---|---|---|---|---|---|
| 3791 | 况 | 86 | 8156 | 51B5 | 况 | | 況，異體字問題。 |
| 3793 | 减 | 2 | 8158 | 51CF | 减 | | 減，異體字問題。 |
| 3794 | 凑 | 28 | 8159 | 51D1 | 凑 | | 湊，異體字問題。 |
| 3796 | 决 | 38 | 815B | 51B3 | 决 | | 決，異體字問題。 |
| 3800 | | 1 | 815F | | | 杀 | 剎，異體字問題。 |
| 3804 | 効 | 5 | 8163 | 52B9 | 効 | | |
| 3805 | 勅 | 8 | 8164 | 52C5 | 勅 | | |
| 3814 | 却 | 316 | 816D | 5374 | 却 | | |
| 3815 | 卺 | 2 | 816E | 537A | 卺 | | |
| 3819 | 厮 | 1 | 8172 | 53AE | 厮 | | 廝，異體字問題。 |
| 3822 | 叁 | 11 | 8175 | 53C1 | 叁 | | |
| 3824 | 叙 | 25 | 8177 | 53D9 | 叙 | | |
| 3825 | 叠 | 36 | 8178 | 53E0 | 叠 | | |

| 編號 | 造字 | 頻次 | Big5 | Unicode | WinXP | 構字式 | 備註 |
|---|---|---|---|---|---|---|---|
| 3826 | 叶 | 2 | 8179 | 53F6 | 叶 | | |
| 3834 | 嗶 | 2 | 81A3 | 5637 | 嗶 | | |
| 3843 | 堦 | 1 | 81AC | 5826 | 堦 | | |
| 3846 | 堺 | 21 | 81AF | 58FB | 堺 | | |
| 3858 | 寃 | 162 | 81BB | 5BC3 | 寃 | | 冤，異體字問題。 |
| 3864 | 峯 | 2 | 81C1 | 5CEF | 峯 | | |
| 3866 | 崐 | 1 | 81C3 | 5D10 | 崐 | | |
| 3883 | 徧 | 17 | 81D4 | 5FA7 | 徧 | | |
| 3884 | 忽 | 1 | 81D5 | | | 忽 | |
| 3886 | 悞 | 2 | 81D7 | 609E | 悞 | | |
| 3900 | 拕 | 1 | 81E5 | 62D5 | 拕 | | |
| 3909 | 擡 | 191 | 81EE | 64E1 | 擡 | | |
| 3938 | 橈 | 8 | 824C | 6AC8 | 橈 | | |
| 3942 | 毡 | 32 | 8250 | 6BE1 | 毡 | | |
| 3946 | 汹 | 4 | 8254 | 6C79 | 汹 | | |
| 3966 | 烟 | 23 | 8268 | 70DF | 烟 | | |
| 3974 | 牀 | 266 | 8270 | 7240 | 牀 | | |
| 3976 | 犂 | 3 | 8272 | 7282 | 犂 | | |

| 編號 | 造字 | 頻次 | Big5 | Unicode | WinXP | 構字式 | 備註 |
|---|---|---|---|---|---|---|---|
| 3977 | 狗 | 1 | 8273 | 72E5 | 狗 | | |
| 3978 | 猪 | 48 | 8274 | 732A | 猪 | | Unicode 字型呈現差異。 |
| 3979 | 猫 | 1 | 8275 | 732B | 猫 | | |
| 3990 | 甎 | 4 | 82A2 | 750E | 甎 | | |
| 3996 | 疎 | 13 | 82A8 | 758E | 疎 | | |
| 4000 | 疔 | 3 | 82AC | 75B4 | 疔 | | |
| 4007 | 癙 | 2 | 82B3 | 7667 | 癙 | | |
| 4012 | 皷 | 1 | 82B8 | 76B7 | 皷 | | 鼓，異體字問題。 |
| 4013 | 盌 | 3 | 82B9 | 76CC | 盌 | | |
| 4016 | 着 | 3456 | 82BC | 7740 | 着 | | |
| 4021 | 磊 | 1 | 82C1 | 78AF | 磊 | | |
| 4024 | 礙 | 26 | 82C4 | 792E | 礙 | | |
| 4028 | 稟 | 12 | 82C8 | 7980 | 稟 | | |
| 4033 | 稭 | 9 | 82CD | 7A2D | 稭 | | |
| 4036 | 窯 | 4 | 82D0 | 7AB0 | 窯 | | |
| 4038 | 竈 | 6 | 82D2 | 7AC8 | 竈 | | |

| 編號 | 造字 | 頻次 | Big5 | Unicode | WinXP | 構字式 | 備註 |
|---|---|---|---|---|---|---|---|
| 4041 | 竪 | 2 | 82D5 | 7AEA | 竪 | | |
| 4042 | 笋 | 6 | 82D6 | 7B0B | 笋 | | |
| 4045 | 篬 | 2 | 82D9 | 7B92 | 篬 | | |
| 4055 | 粃 | 1 | 82E3 | 7C83 | 粃 | | |
| 4067 | 綉 | 2 | 82EF | 7D89 | 綉 | | |
| 4068 | 綫 | 3 | 82F0 | 7DAB | 綫 | | |
| 4080 | 羣 | 87 | 82FC | 7FA3 | 羣 | | |
| 4089 | 脇 | 2 | 8346 | 8107 | 脇 | | |
| 4090 | 脚 | 217 | 8347 | 811A | 脚 | | |
| 4094 | 舘 | 2 | 834B | 8218 | 舘 | | |
| 4098 | 芺 | 6 | 834F | 82AA | 芺 | | |
| 4105 | 菓 | 17 | 8356 | 83D3 | 菓 | | |
| 4107 | 葱 | 12 | 8358 | 8471 | 葱 | | |
| 4112 | 蘢 | 11 | 835D | 8534 | 蘢 | | |
| 4115 | 藁 | 2 | 8360 | 85C1 | 藁 | | |
| 4124 | 蝨 | 2 | 8369 | 8771 | 蝨 | | |
| 4131 | 劒 | 4 | 8370 | 8842 | 劒 | | |
| 4132 | | 348 | 8371 | | | 血𠙵丞 | 眾，異體字問題。 |

| 編號 | 造字 | 頻次 | Big5 | Unicode | WinXP | 構字式 | 備註 |
|---|---|---|---|---|---|---|---|
| 4134 | 衞 | 6 | 8373 | 885E | 衞 | | 衛，異體字問題。 |
| 4155 | 訶 | 2 | 83AA | 8B0C | 訶 | | |
| 4173 | 躬 | 19 | 83BC | 8EAD | 躬 | | |
| 4178 | 輿 | 2 | 83C1 | 8F2D | 輿 | | |
| 4193 | 鈎 | 10 | 83D0 | 920E | 鈎 | | |
| 4195 | 鉢 | 1 | 83D2 | 9262 | 鉢 | | |
| 4203 | 隣 | 5 | 83DA | 96A3 | 隣 | | 鄰，異體字問題。 |
| 4218 | 韮 | 3 | 83E9 | 97EE | 韮 | | |
| 4242 | 鬬 | 9 | 8442 | 9B2D | 鬬 | | |
| 4256 | 鶏 | 57 | 8450 | 9DC4 | 鶏 | | |
| 4262 | 麪 | 6 | 8456 | 9EAA | 麪 | | Unicode 字型呈現差異。 |
| 4267 | 鼈 | 5 | 845B | 9F08 | 鼈 | | |
| 4277 | | 1 | 8465 | | | 金⿰乂 | 釵，異體字問題。 |
| 4307 | 罵 | 494 | 84A5 | 99E1 | 罵 | | 罵，異體字問題。 |

| 編號 | 造字 | 頻次 | Big5 | Unicode | WinXP | 構字式 | 備註 |
|---|---|---|---|---|---|---|---|
| 4308 | 劫 | 1 | 84A6 | 5227 | 劫 | | |
| 4309 | 刼 | 5 | 84A7 | 523C | 刼 | | |
| 4350 | 侷 | 3 | 84D0 | 5058 | 侷 | | |
| 4376 | 僂 | 19 | 84EA | 510D | 僂 | | |
| 4428 | 躕 | 5 | 855F | 8E70 | 躕 | | |
| 4447 | 揑 | 29 | 8572 | 63D1 | 揑 | | |
| 4510 | 吒 | 5 | 85D3 | 5493 | 吒 | | |
| 4531 | 攢 | 1 | 85E8 | 6505 | 攢 | | |
| 4539 | 崢 | 1 | 85F0 | 20C9C | | 口厽爭 | |
| 4616 | 壳 | 5 | 867E | 58F3 | 壳 | | 壳，異體字問題。 |
| 4735 | 壠 | 7 | 8758 | 58E0 | 壠 | | |
| 4750 | 廾 | 1 | 8767 | 5344 | 廾 | | 艹，錯字，以程式全部取代。 |
| 4768 | 妬 | 10 | 8779 | 59AC | 妬 | | |
| 4778 | 鞴 | 2 | 87A5 | 97B4 | 鞴 | | |
| 4784 | 鞮 | 3 | 87AB | 9793 | 鞮 | | |
| 4800 | 噔 | 1 | 87BB | 5654 | 噔 | | |

| 編號 | 造字 | 頻次 | Big5 | Unicode | WinXP | 構字式 | 備註 |
|---|---|---|---|---|---|---|---|
| 4805 | 瀆 | 6 | 87C0 | 51DF | 瀆 | | 瀆，異體字問題。 |
| 4816 | 亻 | 2 | 87CB | | | 亻 | |
| 4904 | 尅 | 1 | 8864 | 5C05 | 尅 | | |
| 4966 | 癇 | 8 | 88C4 | 764E | 癇 | | |
| 5008 | 幙 | 3 | 88EE | 5E5E | 幙 | | |
| 5074 | 厊 | 1 | 8971 | 5C2B | 厊 | | |
| 5075 | 犭 | 3 | 8972 | 72AD | | 犭 | |
| 5094 | 翉 | 2 | 89A7 | 26407 | | 翉 | |
| 5095 | 来 | 1 | 89A8 | 6765 | 来 | | |
| 5147 | 恮 | 2 | 89DC | 6060 | 恮 | | |
| 5174 | | 2 | 89F7 | | | ⿰刀兎、⊙ | 兔，異體字問題。 |
| 5176 | | 3 | 89F9 | | | 氵⿱冗 | 沉，異體字問題。 |
| 5180 | 拘 | 1 | 89FD | 6285 | 拘 | | |
| 5218 | | 9 | 8A64 | | | ⿰宀莧、⊙ | 寬，異體字問題。 |
| 5229 | 撐 | 3 | 8A6F | 6490 | 撐 | | |

| 編號 | 造字 | 頻次 | Big5 | Unicode | WinXP | 構字式 | 備註 |
|---|---|---|---|---|---|---|---|
| 5240 | 摔 | 2 | 8A7A | 6481 | 摔 | | |
| 5281 | 鐵 | 2 | 8AC5 | 994A | 鐵 | | |
| 5402 | 朮 | 3 | 8BA1 | 672F | 朮 | | 朮，異體字問題。 |
| 5433 | 夋 | 1 | 8BC0 | | | 夋 | |
| 5437 | 幷 | 3 | 8BC4 | 5E77 | 幷 | | |
| 5559 | 欶 | 2 | 8CA1 | 6B35 | 欶 | | |
| 5590 | 朵 | 36 | 8CC0 | 6736 | 朵 | | 部份使用不同字形「朵」，手動修改。 |
| 5591 | 杞 | 3 | 8CC1 | 233CC | | 木巳 | 杞，異體字問題。 |
| 5595 | 垛 | 5 | 8CC5 | 579C | 垛 | | |
| 5621 | 泪 | 2 | 8CDF | 6CEA | 泪 | | |
| 5653 | 暎 | 1 | 8D40 | 3B09 | | 日央 | |
| 5697 | 烫 | 10 | 8D6C | 7151 | 烫 | | |
| 5742 | 澔 | 6 | 8DBB | 6F81 | 澔 | | |

## 二、手動取代表說明：

《蒲松齡集》檔案中需手動修改的情況分為以下幾種：

（一）《蒲松齡集》檔案中之舊版造字，因字形部份不同而無法全部以程式取代，皆手動修改成Windows XP能顯示的字或構字式，並製作手動取代表，請參考附件二。欄位說明如下：

1. 編號：Big5 造字空間為 6217 個，編號由 1 到 6217。
2. 造字：舊版造字字形。
3. 取代：實際使用字形與舊版造字不同時，於此欄列出。
4. 檔案原文摘錄：摘錄檔案《蒲松齡集》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。
5. 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。
6. 備註：記錄修改相同辭句的次數或其他事項。

**附件二、《蒲松齡集》手動取代表**

| 編號 | 造字 | 取代 | 檔案原文摘錄 | 手動取代結果 | 備註 |
|---|---|---|---|---|---|
| 2853 | 弍 | 弋𠃌三 | 丟了參　__，撂了 | 丟了參　弋𠃌三，撂了 | |
| 5590 | 朵 | 朵 | 尋思起 朵 | 尋思起朵 | |
| | | | 貪慌擺劃這 朵 | 貪慌擺劃這朵 | |

（二）《蒲松齡集》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《蒲松齡集》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、《蒲松齡集》缺字外錯字修改

| 檔案原文摘錄 | 手動取代結果 | 備註 |
| --- | --- | --- |
| 朿骨！你央及 | 朿骨！你央及 | |