

《莊子》(標記)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
2008/04/03 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 4228 的「滫」字，Unicode 編碼是 24162，由於 XP 並不支援，仍需使用構字式「灬 𠂔 齋」。

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：

《莊子》(標記)使用舊版造字 85 個，字頻 1101 次，這 85 個造字中，73 個可轉成 Windows XP 能顯示的字，字頻 1085 次；另外 12 個字必須轉成構字式，字頻 16 次。

轉碼完成製作轉碼分析表，請參考附件一《莊子》(標記)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. 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 2768 的「𦰩𦰪千」字，而原書使用之字為「竿」，查詢教育部異體字字典確認「𦰩𦰪千」並非「竿」之異體字，因此歸為錯字，以程式取代為正確字形「竿」。
3. 待造字：Unicode 及漢字構形資料庫皆未收錄的舊漢籍造字，正在等待補造字中，所以「造字」欄空白無法看到字形。如編號 2811 的「𦰪𦰫田」。

附件一、《莊子》(標記)轉碼分析表

| 編號 | 造字 | 頻次 | Big5 | Unicode | WinXP | 構字式 | 備註 |
|------|----|----|------|---------|-------|-----|----------|
| 826 | 癱 | 2 | 8E68 | 7655 | 癱 | | |
| 949 | 柎 | 1 | 8F46 | 678F | 柎 | | |
| 982 | 𦰪 | 4 | 8F67 | 80F7 | 𦰪 | | 胸，異體字問題。 |
| 994 | 楚 | 8 | 8F73 | 8E75 | 楚 | | |
| 1024 | 虵 | 1 | 8FB3 | 8675 | 虵 | | |

| 編號 | 造字 | 頻次 | Big5 | Unicode | WinXP | 構字式 | 備註 |
|------|----|----|------|---------|-------|-----|---------------|
| 1040 | 聃 | 45 | 8FC3 | 803C | 聃 | | |
| 1043 | 頤 | 1 | 8FC6 | 2945E | | 𠃉𠃉頁 | |
| 1044 | 儻 | 2 | 8FC7 | 511B | 儻 | | |
| 1077 | 稟 | 1 | 8FE8 | 7A3E | 稟 | | |
| 1099 | 睥 | 1 | 8FFE | 66A4 | 睥 | | |
| 1133 | 瓌 | 1 | 9061 | 74CC | 瓌 | | |
| 1383 | 穰 | 2 | 91E0 | 7A45 | 穰 | | |
| 1646 | 脗 | 1 | 93AD | 8117 | 脗 | | |
| 2050 | 蹠 | 1 | 9648 | 8E4F | 蹠 | | |
| 2191 | 鉞 | 1 | 96F7 | | | 金𠃉戍 | 鉞，異體字問題。 |
| 2722 | 吡 | 1 | 9A74 | 5421 | 吡 | | |
| 2728 | 妲 | 2 | 9A7A | 59B8 | 妲 | | |
| 2768 | | 1 | 9AC4 | | | 𠃉𠃉千 | 竿，錯字，以程式全部取代。 |
| 2783 | 蚤 | 3 | 9AD3 | 8689 | 蚤 | | 蚊，異體字問題。 |
| 2785 | 𠃉 | 1 | 9AD5 | 8C40 | 𠃉 | | |
| 2811 | | 2 | 9AEF | | | 𠃉𠃉田 | 待造字。 |

| 編號 | 造字 | 頻次 | Big5 | Unicode | WinXP | 構字式 | 備註 |
|------|----|----|------|---------|-------|-----|---------------|
| 2817 | 𠂔 | 4 | 9AF5 | 3AC4 | | 𠂔 | 備，錯字，以程式全部取代。 |
| 2822 | 愜 | 1 | 9AFA | 6051 | 愜 | | |
| 3787 | 冢 | 1 | 8152 | 51A2 | 冢 | | |
| 3799 | 刳 | 1 | 815E | 5226 | 刳 | | |
| 3866 | 崐 | 1 | 81C3 | 5D10 | 崐 | | |
| 3868 | 卮 | 4 | 81C5 | 5DF5 | 卮 | | |
| 3869 | 帀 | 1 | 81C6 | 5E00 | 帀 | | |
| 3883 | 徧 | 6 | 81D4 | 5FA7 | 徧 | | |
| 3891 | 慙 | 2 | 81DC | 6159 | 慙 | | |
| 3929 | 犁 | 2 | 8243 | 68C3 | 犁 | | |
| 3974 | 牀 | 1 | 8270 | 7240 | 牀 | | |
| 3976 | 犁 | 1 | 8272 | 7282 | 犁 | | |
| 3980 | 獾 | 5 | 8276 | 7328 | 獾 | | |
| 4014 | 蓋 | 1 | 82BA | 8462 | 蓋 | | 蓋，異體字問題。 |
| 4025 | 礪 | 2 | 82C5 | 7934 | 礪 | | |
| 4034 | 穉 | 1 | 82CE | 7A49 | 穉 | | |
| 4038 | 竈 | 3 | 82D2 | 7AC8 | 竈 | | |

| 編號 | 造字 | 頻次 | Big5 | Unicode | WinXP | 構字式 | 備註 |
|------|----|----|------|---------|-------|--------------------------------------|--------------|
| 4055 | 粃 | 1 | 82E3 | 7C83 | 粃 | | |
| 4080 | 羣 | 15 | 82FC | 7FA3 | 羣 | | |
| 4081 | 翱 | 1 | 82FD | 7FFA | 翱 | | |
| 4089 | 脇 | 1 | 8346 | 8107 | 脇 | | |
| 4104 | 莅 | 1 | 8355 | 8385 | 莅 | | |
| 4107 | 葱 | 1 | 8358 | 8471 | 葱 | | |
| 4119 | | 2 | 8364 | | | 形 ^艹 正頁 已久 ^口 | 夔，異體字 問題。 |
| 4132 | | 56 | 8371 | | | 血 ^厶 𠂇 | 眾，異體字 問題。 |
| 4134 | 衛 | 18 | 8373 | 885E | 衛 | | 衛，異體字 問題。 |
| 4140 | 衽 | 1 | 8379 | 88B5 | 衽 | | 衽，異體字 問題。 |
| 4146 | 覩 | 13 | 83A1 | 89A9 | 覩 | | |
| 4160 | 適 | 1 | 83AF | 8B81 | 適 | | |
| 4184 | 迹 | 20 | 83C7 | 8FF9 | 迹 | | |
| 4226 | 飡 | 1 | 83F1 | 98E1 | 飡 | | |
| 4227 | 飡 | 1 | 83F2 | 98E6 | 飡 | | |
| 4228 | 瀟 | 1 | 83F3 | 24162 | | 灬 ^灬 蕭 | |
| 4238 | 髭 | 1 | 83FD | 9AE2 | 髭 | | |

| 編號 | 造字 | 頻次 | Big5 | Unicode | WinXP | 構字式 | 備註 |
|------|----|----|------|---------|-------|--------------|----------------------------------|
| 4239 | 髯 | 1 | 83FE | 9AE5 | 髯 | | |
| 4242 | 鬪 | 8 | 8442 | 9B2D | 鬪 | | |
| 4250 | 鱈 | 6 | 844A | 9C0C | 鱈 | | |
| 4265 | 鼃 | 5 | 8459 | 9F03 | 鼃 | | |
| 4267 | 鼈 | 4 | 845B | 9F08 | 鼈 | | |
| 4455 | 劔 | 1 | 857A | | | 僉 ㄩ 刀 | |
| 4458 | 勸 | 1 | 857D | 52CC | 勸 | | |
| 4562 | 逕 | 1 | 8648 | 28563 | | 形 辵 四 午 □ | |
| 4578 | | 2 | 8658 | | | 形 巾 欠 非 □ | 形 巾 欠 非 □，錯字， 以程式全部 取代。 |
| 4588 | 鮒 | 1 | 8662 | 4C54 | | 魚 ㄩ 攸 | |
| 4607 | 狫 | 1 | 8675 | 3E60 | | 豸 ㄩ 屯 | |
| 4817 | 寶 | 2 | 87CC | 5BF3 | 寶 | | 寶，異體字 問題。 |
| 4924 | | 1 | 8878 | | | 冗 ㄩ 曾 | 曾，異體字 問題。 |
| 4987 | 躡 | 1 | 88D9 | 4831 | | 足 ㄩ 屬 | |
| 5091 | 鄒 | 1 | 89A4 | 92E3 | 鄒 | | |

| 編號 | 造字 | 頻次 | Big5 | Unicode | WinXP | 構字式 | 備註 |
|------|----|-----|------|---------|-------|-----|----------|
| 5108 | 惛 | 2 | 89B5 | 60FD | 惛 | | |
| 5112 | 园 | 1 | 89B9 | 56ED | 园 | | |
| 5244 | 夔 | 1 | 8A7E | 3738 | | 薛△女 | |
| 5261 | 挽 | 1 | 8AB1 | 639C | 挽 | | |
| 5302 | 昏 | 1 | 8ADA | 662C | 昏 | | |
| 5316 | 晋 | 1 | 8AE8 | 664B | 晋 | | 晉，異體字問題。 |
| 5335 | 蚤 | 2 | 8AFB | 459F | | 亡△虫 | |
| 5371 | 冃 | 1 | 8B60 | 808E | 冃 | | |
| 5437 | 并 | 1 | 8BC4 | 5E77 | 并 | | |
| 5470 | 棟 | 1 | 8BE5 | 68C5 | 棟 | | |
| 5508 | 髻 | 1 | 8C4C | 9B10 | 髻 | | |
| 5585 | 寓 | 2 | 8CBB | 3762 | | 宀△禹 | |
| 5739 | | 3 | 8DB8 | | | 冗△古 | 𠂔，異體字問題。 |
| 5753 | 潜 | 2 | 8DC6 | 6F5C | 潜 | | |
| 5860 | 无 | 794 | C6D3 | 65E0 | 无 | | |

二、手動取代表說明：

《莊子》檔案中出現「●」字形，共計 34 字，比對原書後以 Unicode 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. 標記行碼：語言所標記檔裡設定的行碼。
3. 檔案原文摘錄：摘錄《莊子》(標記)檔案有「●」字形的文句。
4. 原書字形：「●」在原書中之字形，以此字取代檔案原文的「●」。
5. 備註：記錄修改相同詞句的次數或其他事項。

附件二、《莊子》(標記)「●」字手動取代表

| 頁碼 | 標記行碼 | 檔案原文摘錄 | 原書字形 | 備註 |
|--------|------|---|----------|----|
| p. 51 | 352 | 慮(NI) 嘆(NI) 變 (NI) ● | 執△心 | |
| p. 56 | 391 | ●然(VI)[+poly] | 𠄎△尔 | |
| P. 216 | 1789 | 甕●犬瘦 (NB1)[+prop] | 央△瓦 | |
| p. 242 | 1981 | 相(DH) ●(VPY) 以 (P) | 形; 宙 尙 □ | |
| p. 259 | 2135 | 𠄎●(VI)[+rd] 而 (C) 鑑 | 足△鮮 | |
| p. 268 | 2250 | 以(P) 死(NI) 為 (VG) 決● (NA2)[+attr] | 疒△丸 | |
| p. 330 | 2828 | 連(VJ) 之(NH) 以 (P) 羈●(NA2)[+co] | 鬲 | |
| p. 522 | 4960 | ●(VC1) 膚(NA2) | 形口𠄎日 □ | |
| p. 527 | 5050 | ●蠆(NA1)[+co] 之 (T) 尾 | 厂△蠆 | |
| p. 552 | 5302 | ●(VP) 淳(NI) 散 | ; △泉 | |

| 頁碼 | 標記行碼 | 檔案原文摘錄 | 原書字形 | 備註 |
|--------|------|----------------------------------|---------|-------|
| | | (VP) | | |
| p. 592 | 5690 | 一(S) 足(NA1) ● 蹕(VA)[+rd] | 跲 | |
| p. 624 | 6144 | 得(VC2) 水(NA2) 則(C) 為(VG) ● | 漚 | |
| p. 625 | 6161 | 比(VH1) 乎(P) 不 ●(NA1)[+others] | 尊 | |
| p. 642 | 6310 | 注(VA) 者(NH) ● (VH1) | 𠂇 𠂇 昏 | |
| p. 648 | 6367 | 三(S) 月(NA5) ● (VC1) 汝(NH) | 𠂇 𠂇 𠂇 | |
| p. 722 | 7361 | ●斛(NA2)[+co] 不 (DC) | 缺 | 有 2 處 |
| p. 746 | 7711 | 墮(VP) 其(NH) 天 ● | 失 𠂇 衣 | |
| p. 765 | 7961 | 是非(NI)[+co] 相 (DH) ●(VC1Y) | 𠂇 𠂇 𠂇 𠂇 | |
| p. 775 | 8062 | 日中(NA5)[+others] 穴(VP)[+nv] ● | 𠂇 | |
| p. 795 | 8269 | ●(VH1N) 內(NG) 者 | 券 | |
| p. 795 | 8271 | ●(VH1N) 外(NG) 者 | 券 | |
| p. 795 | 8287 | 鎔(NB2)[+prop] ● (NB2)[+prop] | 鄒 | |
| p. 805 | 8355 | ●(NI) 也(T) | 黷 | |

| 頁碼 | 標記行碼 | 檔案原文摘錄 | 原書字形 | 備註 |
|---------|-------|---|-------|----|
| p. 846 | 8804 | 委蛇(VI)[+rd] 攫● | 𠄎扌又爪𠄎 | |
| p. 868 | 9042 | 雞●(NA1)[+others] 也 | 广△雍 | |
| p. 943 | 9862 | 皆●(NI)[+co] 可以 | 女△威 | |
| p. 1037 | 11669 | 於(P) 十(S) ● (NA4) | 將△食 | |
| p. 1037 | 11670 | 而(C) 五(S) ● (NA4) 先(NG) 饋 | 將△食 | |
| p. 1037 | 11679 | 而(C) ●(VC1) 其 (NH) | 𠄎巾女非𠄎 | |
| p. 1037 | 11680 | 夫(NH) ●人 (NA1)[+attr] | 將△食 | |
| p. 1062 | 11925 | 子(NH) 為(VG) ● 粉 | 𠄎巾女非𠄎 | |
| p. 1063 | 11941 | 吾葬具豈不●邪 | 備 | |
| p. 1091 | 12266 | 不(DC) 免(VH1) 於 (P) ●斷 (NI)[+others] | 𩺰 | |
| P. 216 | 1789 | 甕●犬瘦 (NB1)[+prop] | 央△瓦 | |