

《典論一卷》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
2008/12/04 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 4455 的「劔」字，Unicode 編碼未收錄，仍需使用構字式「𠄎刀」。

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：

《典論一卷》使用舊版造字 19 個，字頻 41 次，這 19 個造字中，17 個可轉成 Windows XP 能顯示的字，字頻 29 次；另外 2 個字必須轉成構字式，字頻 12 次。

轉碼完成製作轉碼分析表，請參考附件一《典論一卷》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 982 的「𪛗」字，是「胸」的異體字，以標準字「胸」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 2449 的「名𪛗佳」字，而原書使用之字為「雒」，查詢教育部異體字字典確認「名𪛗佳」並非「雒」之異體字，因此歸為錯字，以程式取代為正確字形「雒」。

附件一、《典論一卷》轉碼分析表

| 編號 | 造字 | 頻次 | Big5 | Unicode | WinXP | 構字式 | 備註 |
|------|----|----|------|---------|-------|-----|---------------|
| 982 | 𪛗 | 1 | 8F67 | 80F7 | 𪛗 | | 胸，異體字問題。 |
| 1067 | 碁 | 1 | 8FDE | 68CA | 碁 | | |
| 2449 | | 1 | 98BF | | | 名𪛗佳 | 雒，錯字，以程式全部取代。 |
| 3868 | 卮 | 1 | 81C5 | 5DF5 | 卮 | | |
| 3966 | 烟 | 1 | 8268 | 70DF | 烟 | | |

| 編號 | 造字 | 頻次 | Big5 | Unicode | WinXP | 構字式 | 備註 |
|------|----|----|------|---------|-------|-------|----------|
| 4040 | 竝 | 8 | 82D4 | 7ADD | 竝 | | |
| 4080 | 羣 | 2 | 82FC | 7FA3 | 羣 | | |
| 4118 | 藥 | 1 | 8363 | 8602 | 藥 | | |
| 4132 | | 2 | 8371 | | | 血 ㄩ 丞 | 眾，異體字問題。 |
| 4134 | 衛 | 1 | 8373 | 885E | 衛 | | 衛，異體字問題。 |
| 4146 | 覩 | 1 | 83A1 | 89A9 | 覩 | | 睹，異體字問題。 |
| 4237 | 髡 | 2 | 83FC | 9AE0 | 髡 | | |
| 4260 | 麤 | 1 | 8454 | 9E9E | 麤 | | |
| 4455 | 劒 | 8 | 857A | | | 僉 ㄩ 刀 | |
| 4768 | 妬 | 2 | 8779 | 59AC | 妬 | | 妒，異體字問題。 |
| 5019 | 畧 | 4 | 88F9 | | | 畧 | |
| 5176 | | 1 | 89F9 | | | 冫 ㄩ 冗 | 沉，異體字問題。 |
| 5310 | 烏 | 1 | 8AE2 | 8204 | 烏 | | |
| 5706 | 敍 | 2 | 8D75 | 654D | 敍 | | 敘，異體字問 |

| 編號 | 造字 | 頻次 | Big5 | Unicode | WinXP | 構字式 | 備註 |
|----|----|----|------|---------|-------|-----|----|
| | | | | | | | 題。 |