

《慎子》舊版造字轉碼說明

中研院資訊所文獻處理實驗室
中研院史語所漢籍電子文獻工作小組
2008/3/7 陳建安 製作

一、《慎子》一書(慎子.xml)使用舊版造字 9 個，字頻 13 次，詳如附件一。這 9 個造字中，7 個可轉成 Windows XP 能顯示的字，字頻 9 次；另外 2 個字必須轉成構字式，字頻 4 次。

二、附件一的造字分析表說明如下：

甲、編號：Big5 造字空間為 6217 個，編號由 1 到 6217。

乙、造字：舊版造字

丙、字頻(txt)：造字在「.txt」文件的出現次數

丁、字頻(xml)：造字在「.xml」文件的出現次數

戊、Big5：造字的 Big5 碼

己、Unicode：造字所對應的 Unicode 碼

庚、WinXP：造字在 Windows XP 的對應字形

辛、備註凡例：

1、異體字問題：新版漢籍考量到使用者檢索及使用時的便利性，將用字原則改為除專詞等特殊情形之外，一律改用標準字呈現。如編號 4132 的「血𩇑𩇑」係「眾」字之異體，故以「眾」字取代。

2、Unicode 字型呈現差異：Unicode 字型與舊漢籍造字有些微差異，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 4242 的「鬪」字，Unicode 字型呈現為「鬪」，實際上仍為同一字。

三、《慎子》中未確定字形「●」共 1 處，詳細處理內容可參考附件二的未確定字形「●」手動取代表。附件二《慎子》未確定字形「●」手動取代表欄位說明如下：

甲、檔案原文摘錄：摘錄檔案(慎子.xml)中包含未確定字形「●」的文句段落，其中紅字底線的部分，為需要手動取代的字形。

乙、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。

附件一、《慎子》造字分析表

編號	造字	頻次 (txt)	頻次 (xml)	Big5	Unicode	WinXP	構字式	備註
3855	孽	2	2	81B8	5B7C	孽		孽，異體字問題。
3996	疎	2	2	82A8	758E	疎		
4080	羣	1	1	82FC	7FA3	羣		群，異體字問題。
4132		3	3	8371			血𠂇𠂇	眾，異體字問題。
4146	覩	1	1	83A1	89A9	覩		
4212	鞞	1	1	83E3	97B8	鞞		
4242	鬪	1	1	8442	9B2D	鬪		Unicode 呈現差異。
5034	弃	1	1	8949	5F03	弃		棄，異體字問題。
5306		1	1	8ADE			𠂇𠂇𠂇	簡，異體字問題。

附件一、《慎子》未確定字形「●」手動取代表

檔案原文摘錄	手動取代結果
●窮谷野走十里	<u>足</u> <u>𠂇</u> <u>𠂇</u> 窮谷野走十里