

# 《孫子》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  
2010/04/09 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 1845 的「諭」字，Unicode 編碼是 27B08，由於 XP 並不支援，仍需使用構字式「言𠄎翁」。

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：

《孫子》使用舊版造字 10 個，字頻 15 次，這 10 個造字中，9 個可轉成 Windows XP 能顯示的字，字頻 13 次；另外 1 個字必須轉成構字式，字頻 2 次。

轉碼完成製作轉碼分析表，請參考附件一《孫子》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 5299 的「𢆶」字，是「𢆶」的異體字，以標準字「𢆶」取代。

附件一、《孫子》轉碼分析表

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
847	𧈧	1	8E7D	7F3B	𧈧		
1845	諭	2	94D7	27B08		言𠂔翁	
2415	鬪	1	987B	95D8	鬪		鬪，異體字問題。
4030	秆	1	82CA	79C6	秆		
4080	羣	1	82FC	7FA3	羣		
4146	覩	1	83A1	89A9	覩		
4242	鬪	5	8442	9B2D	鬪		
4720		1	8749			土𠂔已	圮，異體字問題。
5043	𧈧	1	8952	5F4D	𧈧		

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
5299	戟	1	8AD7	39B8		卓𠄎戈	戟，異體字問題。

## 二、手動取代表說明：

《孫子》檔案中出現「☆」字形，共計 1 字，比對原書後以 Unicod 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. 檔案原文摘錄：摘錄《孫子》檔案有「☆」字形的文句。
3. 原書字形：「☆」在原書中之字形，以此字取代檔案原文的「☆」。
4. 備註：記錄修改相同詞句的次數或其他事項。

### 附件二、《孫子》「☆」字手動取代表

頁碼	檔案原文摘錄	原書字形	備註
p. 81	☆秆一石	𠄎𠄎忌	