

# 《新語》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  
2010/05/06 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 3894 的「筋」字，Unicode 編碼是 4225，由於 XP 並不支援，仍需使用構字式「𠂇𠂇𠂇𠂇」。

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：

《新語》使用舊版造字 30 個，字頻 80 次，這 30 個造字中，29 個可轉成 Windows XP 能顯示的字，字頻 79 次；另外 1 個字必須轉成構字式，字頻 1 次。

轉碼完成製作轉碼分析表，請參考附件一《新語》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 4132 的「血𠂇𠂇」字，是「眾」的異體字，以標準字「眾」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 4910 的「𠂇𠂇𠂇」字，而原書使用之字為「淫」，查詢教育部異體字字典確認「淫」並非「𠂇𠂇𠂇」之異體字，因此歸為錯字，以程式取代為正確字形「淫」。
  3. 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 3935 的「槩」字，Unicode 字型呈現為「槩」，實際上仍為同一字。

附件一、《新語》轉碼分析表

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
899	駝	1	8ED3	99DE	駝		
1044	儻	1	8FC7	511B	儻		
1226	罰	4	90E0	7F78	罰		
1565	絲	2	92F9	7DDC	絲		
1906	衰	1	9555	88E6	衰		

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
1981	讐	2	95C2	8B90	讐		
3791	况	1	8156	51B5	况		况，異體字問題。
3804	効	1	8163	52B9	効		
3810	匱	1	8169	5335	匱		
3894	筋	1	81DF	4225		𠂇 𠂇 筋	
3935	槩	1	8249	3BA3	槩		Unicode 字型呈現差異。
3937	檝	1	824B	6A9D	檝		
3976	犁	1	8272	7282	犁		
3985	璫	1	827B	7447	璫		Unicode 字型呈現差異。
3996	疎	2	82A8	758E	疎		
4040	竝	1	82D4	7ADD	竝		
4077	罇	1	82F9	7F47	罇		
4080	羣	9	82FC	7FA3	羣		
4132		33	8371			血 𠂇 𠂇	眾，異體字問題。
4146	覩	4	83A1	89A9	覩		

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
4178	輓	1	83C1	8F2D	輓		
4437	廩	1	8568	5EEA	廩		
4507	吳	1	85D0	5434	吳		吳，異體字問題。
4904	尅	1	8864	5C05	尅		
4910		1	886A			汙 Δ 畧	淫，錯字，以程式全部取代。
5031	閒	1	8946	95B4	閒		
5034	弃	2	8949	5F03	弃		棄，異體字問題。
5434	湟	1	8BC1	6E7C	湟		
5705	廡	1	8D74	5ED0	廡		Unicode 字型呈現差異。
5706	敘	1	8D75	654D	敘		敘，異體字問題。

## 二、手動取代表說明：

《新語》檔案中出現「●」字形，共計4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. 檔案原文摘錄：摘錄《新語》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、《新語》「●」字手動取代表

頁碼	檔案原文摘錄	原書字形	備註
p. 79	齊人使優●	从△冉	
p. 102	九●之隄	山△亢	
p. 102	傍之枯楊·●	山△累	
p. 102	●結屈	山△累	