

《論語》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
2010/04/23 丁致伶

一、舊版造字轉碼分析：

《論語》使用舊版造字 17 個，字頻 48 次，這 17 個造字中，皆可轉成 Windows XP 能顯示的字，字頻 48 次。

轉碼完成製作轉碼分析表，請參考附件一《論語》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 4134 的「衛」字，是「衛」的異體字，以標準字「衛」取代。
 2. 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 5705 的「廐」字，Unicode 字型呈現為「廐」，實際上仍為同一字。

附件一、《論語》轉碼分析表

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
2152	葶	2	96D0	979F	葶		
3787	冢	1	8152	51A2	冢		
3810	匱	1	8169	5335	匱		
3921	皙	3	81FA	6673	皙		
3924	菴	1	81FD	671E	菴		
3930	惇	3	8244	6901	惇		
3949	涖	2	8257	6D96	涖		
3976	犁	1	8272	7282	犁		
4038	竈	1	82D2	7AC8	竈		
4080	羣	4	82FC	7FA3	羣		
4132		11	8371			血会𠂇	眾，異體字問題。
4134	衛	12	8373	885E	衛		衛，異體字問題。
4184	迹	1	83C7	8FF9	迹		
4242	鬪	1	8442	9B2D	鬪		
4268	叢	1	845C	9F17	叢		
4507	吳	2	85D0	5434	吳		吳，異體字問

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
							題。
5705	廐	1	8D74	5ED0	廐		Unicode 字型 呈現差異。

二、手動取代表說明：

《論語》檔案中出現「●」字形，共計1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《論語》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、《論語》「●」字手動取代表

頁碼	檔案原文摘錄	原書字形	備註
p. 137	子路無●諾	宿	