

# 《中論》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  
2008/12/02 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 4827 的「𩇛」字，Unicode 編碼是 3DC0，由於 XP 並不支援，仍需使用構字式「〇〇火厶孔」。

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：

《中論》使用舊版造字 58 個，字頻 145 次，這 58 個造字中，55 個可轉成 Windows XP 能顯示的字，字頻 142 次；另外 3 個字必須轉成構字式，字頻 3 次。

轉碼完成製作轉碼分析表，請參考附件一《中論》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 3796 的「決」字，是「決」的異體字，以標準字「決」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 5220 的「毋」字，而原書使用之字為「毋」，查詢教育部異體字字典確認「毋」並非「毋」之異體字，因此歸為錯字，以程式取代為正確字形「毋」。
  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。
  4. 待造字：Unicode 及漢字構形資料庫皆未收錄的舊漢籍造字，正在等待補造字中，所以「造字」欄空白無法看到字形。如編號 3097 的「彳𠂔幸」。

附件一、《中論》轉碼分析表

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
915	遲	2	8EE3	905F	遲		遲，異體字問題。
1067	棊	1	8FDE	68CA	棊		

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
1981	讐	7	95C2	8B90	讐		讎，異體字問題。
1999	贇	1	95D4	8CDB	贇		
2044	踈	2	9642	8E08	踈		
2191	鉞	2	96F7			金 ㄩ 戍	鉞，異體字問題。
2434	随	1	98B0	968F	随		
2516	銑	2	9943	9919	銑		
2947	聰	1	9BDA	8066	聰		
3035	賔	7	9C73	8CD4	賔		
3096	歲	4	9CD2	5D57	歲		
3097		1	9CD3			彳 ㄩ 幸	待造字。
3791	况	14	8156	51B5	况		况，異體字問題。
3796	决	4	815B	51B3	决		決，異體字問題。
3804	効	1	8163	52B9	効		
3852	媼	1	81B5	5A63	媼		
3883	徧	1	81D4	5FA7	徧		

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
3889	慤	1	81DA	6164	慤		
3935	槩	1	8249	3BA3	槩		Unicode 字型 呈現差異。
3944	氷	1	8252	6C37	氷		
3961	灾	1	8263	707E	灾		
3977	狗	2	8273	72E5	狗		
3992	嘗	2	82A4	751E	嘗		
3996	疎	1	82A8	758E	疎		
4011	皞	1	82B7	76A1	皞		皞，異體字問 題。
4040	竝	1	82D4	7ADD	竝		
4080	羣	1	82FC	7FA3	羣		
4081	翱	1	82FD	7FFA	翱		
4085	耻	6	8342	803B	耻		
4132		28	8371			血𠂇𠂇	眾，異體字問 題。
4136	衰	1	8375	887A	衰		
4140	衽	1	8379	88B5	衽		
4144	霸	1	837D	8987	霸		Unicode 字型

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
							呈現差異。
4145	羈	2	837E	898A	羈		Unicode 字型 呈現差異。
4155	訶	1	83AA	8B0C	訶		
4156	諂	1	83AB	8B1F	諂		
4173	耽	1	83BC	8EAD	耽		
4184	迹	1	83C7	8FF9	迹		
4193	鈎	1	83D0	920E	鈎		鈎，異體字問題。
4203	隣	1	83DA	96A3	隣		鄰，異體字問題。
4220	穎	1	83EB	9834	穎		
4250	鱸	1	844A	9C0C	鱸		
4400	堯	1	8543	5160	堯		
4502	俎	1	85CB	723C	俎		
4643	徃	8	86BB	5F83	徃		
4723	塲	1	874C	58C3	塲		
4827	𦉳	1	87D6	3DC0		𠄎火厶 𦉳	

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
4910		4	886A			汙 𠄎 𠄎	淫，錯字，以程式全部取代。
4996	斲	1	88E2	65B5	斲		
5015	寗	2	88F5	5BD7	寗		
5034	弃	1	8949	5F03	弃		
5046	彝	1	8955	5F5C	彝		
5119	博	3	89C0	613D	博		
5220	毋	1	8A66	6BCC	毋		毋，錯字，以程式全部取代。
5371	冑	3	8B60	808E	冑		
5395	規	1	8B78	4093		矢 𠄎 見	
5768	濶	3	8DD5	6FF6	濶		闊，異體字問題。
5792	籀	1	8DED	7C55	籀		

## 二、手動取代表說明：

《中論》檔案中出現「●」字形，共計6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《中論》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、《中論》「●」字手動取代表

頁碼	檔案原文摘錄	原書字形	備註
p. 9	●而成之者無幾	ノ厶地	
p. 17	孺子之所以●膝	足厶困	
p. 24	而無名貪●之	胃	
p. 25	相●是非易位	胃	
p. 31	其夭者惟顏回●	扌厶處	
p. 33	虞也大道●矣	寢	