

《法言》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
2010/06/02 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 5354 的「𠄎」字，Unicode 編碼是 23173，由於 XP 並不支援，仍需使用構字式「日𠄎𠄎」。

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：

《法言》使用舊版造字 29 個，字頻 102 次，這 29 個造字中，27 個可轉成 Windows XP 能顯示的字，字頻 95 次；另外 2 個字必須轉成構字式，字頻 7 次。

轉碼完成製作轉碼分析表，請參考附件一《法言》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 5034 的「弃」字，Unicode 字型呈現為「弃」，實際上仍為同一字。

附件一、《法言》轉碼分析表

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
841	礪	3	8E77	7932	礪		
1067	碁	2	8FDE	68CA	碁		
1091	漚	1	8FF6	58CD	漚		
1565	繇	1	92F9	7DDC	繇		
2044	踈	1	9642	8E08	踈		
2409	闕	2	9875	95A7	闕		
3831	詰	1	817E	5586	詰		
3891	慙	2	81DC	6159	慙		

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
3976	犁	5	8272	7282	犁		
4040	竝	2	82D4	7ADD	竝		
4080	羣	15	82FC	7FA3	羣		
4104	莅	1	8355	8385	莅		
4132		20	8371			血𠂇𠂇	眾，異體字問題。
4134	衛	8	8373	885E	衛		衛，異體字問題。
4146	覩	7	83A1	89A9	覩		
4150	訥	2	83A5	8A29	訥		
4184	迹	4	83C7	8FF9	迹		
4186	遡	1	83C9	9061	遡		
4342	倮	2	84C8	502E	倮		
4455	劒	4	857A			僉𠂇刀	
4458	勑	1	857D	52CC	勑		
4507	吳	4	85D0	5434	吳		吳，異體字問題。
5034	弃	3	8949	5F03	弃		Unicode 字型呈現差異。

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
5065	忼	1	8968	5FFC	忼		
5134	愷	2	89CF	619E	愷		
5200		1	8A52			𠃉ノ門 米大𠃉	奧，異體字問題。
5305	晒	1	8ADD	661E	晒		
5354	昊	3	8B4F	23173		日𠃉昊	
5575	峇	2	8CB1	8A14	峇		

二、手動取代表說明：

《法言》檔案中出現「●」字形，共計12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《法言》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、《法言》「●」字手動取代表

頁碼	檔案原文摘錄	原書字形	備註
p. 9	螟●之子	虫𠃉需	
p. 34	甚於●斯	與𠃉鳥	

頁碼	檔案原文摘錄	原書字形	備註
p. 68	山●之蹊	山 𡵓 逕	
p. 160	著古昔之●●	口 𡵓 昏 口 𡵓 昏	
p. 278	●之而已	木 𡵓 敬	
p. 330	又終●	敝 𡵓 大	
p. 369	始皇方虎●	捩	
p. 420	魯仲連●	亻 𡵓 湯	
p. 420	制而不●	亻 𡵓 湯	
p. 550	鷹隼●●	羽 𡵓 堯 羽 𡵓 堯	