

《韓非子》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
2008/4/22 丁政伶

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 790 的「𠄎」字，Unicode 編碼是 4039，由於 XP 並不支援，仍需使用構字式「目𠄎夾」。

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：

《韓非子》使用舊版造字 71 個，字頻 244 次，這 71 個造字中，63 個可轉成 Windows XP 能顯示的字，字頻 232 次；另外 8 個字必須轉成構字式，字頻 12 次。

轉碼完成製作轉碼分析表，請參考附件一《韓非子》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 4134 的「衛」字，是「衛」的異體字，以標準字「衛」取代。
2. **部份使用不同字形，手動修改**：檔案中為同一舊版造字，校對時卻發現原書使用兩個以上不同的字形，可能部份舊版造字為錯字或原書使用兩種異體字形，導致同一舊版造字轉碼時對應到兩個以上的字形；將一部份手動修改，剩餘的頻次則用程式取代成對應的 Unicode 或構字式。例：造字編號 4190 的「郟」字頻次為 6，依書將 1 字手動修改為「衛」，其他 5 字皆轉「郟」。手動修改部份請參考附件二《韓非子》手動取代表。
3. **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 2013 的「斐」字，而原書使用之字為「讐」，「斐」為錯字，以程式取代為正確字形「讐」。
4. **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。
5. **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 2811 的「𠂔𠂕田」。

附件一、《韓非子》轉碼分析表

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
790	映	3	8E44	4039		目厶夾	
826	癱	1	8E68	7655	癱		
982	曾	1	8F67	80F7	曾		胸，異體字問題。
1040	聃	1	8FC3	803C	聃		
1044	儻	1	8FC7	511B	儻		
1048	頷	1	8FCB	981F	頷		
1053	傲	2	8FD0	6160	傲		
1067	碁	1	8FDE	68CA	碁		
1097	嶠	1	8FFC	37ED		山厶昏	
1394	啗	3	91EB	35D6		口厶啗	
1583	談	1	934C			言厶夾	
1912	襍	1	955B	894D	襍		
1981	讐	6	95C2	8B90	讐		
1982	調	1	95C3	8B87	調		
2013	悽	1	95E2	291F5		非厶古	讐，錯字，以程式全部取代。
2811		1	9AEF			亞厶田	待造字。
2817	窈	1	9AF5	3AC4		窈	

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
3791	况	2	8156	51B5	况		况，異體字問題。
3806	勳	1	8165	52E0	勳		
3830	咤	1	817D	54A4	咤		
3846	壻	1	81AF	58FB	壻		
3855	孽	2	81B8	5B7C	孽		
3868	卮	11	81C5	5DF5	卮		
3873	廢	2	81CA			广 ㄩ 𠂇	廢，異體字問題。
3883	徧	7	81D4	5FA7	徧		
3929	黎	4	8243	68C3	黎		
3930	憵	2	8244	6901	憵		
3935	槩	1	8249	69E9	槩		Unicode 字型呈現差異。
3937	檝	2	824B	6A9D	檝		
3966	烟	1	8268	70DF	烟		
3976	犁	1	8272	7282	犁		目 𠂇 夾，錯字，以程式全部取代。
3977	狗	1	8273	72E5	狗		
3996	疎	4	82A8	758E	疎		

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
4035	窳	2	82CF	7A7D	窳		
4038	竈	10	82D2	7AC8	竈		
4051	簠	1	82DF	7C35	簠		
4070	縹	1	82F2	7DE4	縹		
4104	莅	7	8355	8385	莅		
4131	劔	2	8370	8842	劔		
4132		1	8371			血𠂇丞	羈，錯字，以程式全部取代。
4134	衛	47	8373	885E	衛		衛，異體字問題。
4136	衰	3	8375	887A	衰		
4139	袴	11	8378	88B4	袴		
4140	衽	1	8379	88B5	衽		衽，異體字問題。
4145	羈	1	837E	898A	羈		羈，異體字問題。
4172	屨	2	83BB	8E77	屨		
4184	迹	9	83C7	8FF9	迹		
4190	邳	6	83CD	90C4	邳		部份使用不同字形「衛」，手動修改。

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
4216	鞮	1	83E7	97C8	鞮		
4226	凜	5	83F1	98E1	凜		
4242	鬪	13	8442	9B2D	鬪		
4243	鬪	2	8443	9B34	鬪		
4265	鼃	4	8459	9F03	鼃		
4267	鼃	3	845B	9F08	鼃		
4501	獫	1	85CA	7356	獫		
4533	唸	1	85EA	552B	唸		
4768	妬	6	8779	59AC	妬		妒，異體字問題。
4904	尅	2	8864	5C05	尅		
4996	斲	3	88E2	65B5	斲		斲，異體字問題。
5034	弃	3	8949	5F03	弃		棄，異體字問題。
5091	鄒	1	89A4	92E3	鄒		
5108	昏	5	89B5	60FD	昏		
5300	鬪	1	8AD8	9B2A	鬪		鬪，異體字問題。
5302	昏	2	8ADA	662C	昏		昏，異體字問題。

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
5396	鎔	1	8B79	938B	鎔		
5578	弛	5	8CB4	38AE		弓𠂔屯	弛，異體字問題。
5705	廐	3	8D74	5ED0	廐		廐，異體字問題。
5709	蟀	1	8D78	872F	蟀		
5722		5	8DA7				期，錯字，以程式全部取代。
5726	呂	1	8DAB			呂	呂，異體字問題。
5752	摠	1	8DC5	63D4	摠		

二、手動取代表說明：

《韓非子》檔案中需手動修改的情況分為以下幾種：

(一) 《韓非子》檔案中之舊版造字，因字形部份不同而無法全部以程式取代，皆手動修改成Windows XP能顯示的字或構字式，並製作手動取代表，請參考附件二。欄位說明如下：

1. 編號：Big5 造字空間為 6217 個，編號由 1 到 6217。
2. 造字：舊版造字字形。
3. 取代：實際使用字形與舊版造字不同時，於此欄列出。
4. 檔案原文摘錄：摘錄檔案《韓非子》中的文句段落，其中紅字底線的部分，為需要手動取代的原文。文句後方標示（全部取代）者，表示執行全部取代的動作，不一一列舉所有原文。

5. 手動取代結果：變更過後的檔案內容，其中藍字底線的部分，為修改後的結果。文句後方標示（全部取代）者，表示執行全部取代的動作，不一一列舉所有原文。
6. 備註：記錄修改相同辭句的次數或其他事項。

附件二、《韓非子》手動取代表

編號	造字	取代	檔案原文摘錄	手動取代結果	備註
4190	郟	衛	<u>郟</u> 州吁	<u>衛</u> 州吁	

(二)《韓非子》檔案中出現「☆」字形，共計 22 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，請參考附件三。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《韓非子》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。

附件三、《韓非子》「☆」字手動取代表

頁碼	檔案原文摘錄	原書字形	備註
P. 37	虛處則☆然	𠂇 厶 亥	
P. 124	其鬪☆☆	口 厶 顏	
P. 298	而☆小變而失長便	車 厶 辰	
P. 387	漆其首以為☆器	彳 厶 寗	
P. 594	有燒倉廩☆☆者	穴 厶 舛	
P. 645	築社者，☆☆	扌 厶 騫	
P. 670	☆☆危坐子皋	𠂇	
P. 670	故昭卯五乘而履☆☆	尸 厶 橋	

頁碼	檔案原文摘錄	原書字形	備註
P. 673	☆危子榮衣	𠂔	
P. 677	所☆者守門	𠂔	
P. 677	☆危引之而	𠂔	
P. 677	子皋問☆危曰	𠂔	
P. 677	親☆子之足	𠂔	
P. 677	☆危曰	𠂔	
P. 691	吾將何☆以履之	尸 △ 橋	
P. 762	成竅☆下	擻	
P. 775	☆者，巖穴之士徒也	娟	
P. 844	吾聞龐☆氏之子不孝	米 △ 問	
P. 844	問龐☆氏子	米 △ 問	
P. 932	☆尊嬖臣而匹上卿	无	
P. 1103	不☆痊則	扌 △ 副	