

《韓非子》(標記)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 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》(標記)使用舊版造字 63 個，字頻 148 次，這 63 個造字中，55 個可轉成 Windows XP 能顯示的字，字頻 131 次；另外 8 個字必須轉成構字式，字頻 17 次。

轉碼完成製作轉碼分析表，請參考附件一《韓非子》(標記)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，如：人名、地名等特殊情形。例：造字編號 982 的「胃」字，是「胸」的異體字，以標準字「胸」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 298 的「△」構字符號頻次為 3，依書將 2 個符號手動修改為「▲」，其他 1 字轉為「△」。手動修改部份請參考附件二《韓非子》(標記)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》(標記)轉碼分析表

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
170		1	FB4C				棄，錯字，以程式全部取代。
297		4	FBED			厶	
298		3	FBEE			厶	部份使用不同構字符號「厶」，手動修改。
717		2	FEBA		疋		
770		1	FEEF		扌		
790	映	3	8E44	4039		目厶夾	
826	癱	1	8E68	7655	癱		
982	胃	1	8F67	80F7	胃		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	

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
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		
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	擘	2	81B8	5B7C	擘		
3868	卮	11	81C5	5DF5	卮		
3873	廩	2	81CA			广厶段	廩，異體字問題。
3883	徧	7	81D4	5FA7	徧		
3924	菴	5	81FD	671E	菴		

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
3929	黎	4	8243	68C3	黎		
3930	攄	2	8244	6901	攄		
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 𠂔 夾，錯字，以程式全部取代。
3977	狗	1	8273	72E5	狗		
3996	疎	4	82A8	758E	疎		
4035	穿	2	82CF	7A7D	穿		
4038	竈	10	82D2	7AC8	竈		
4051	籛	1	82DF	7C35	籛		
4070	縹	1	82F2	7DE4	縹		
4134	衛	1	8373	885E	衛		衛，異體字問題。
4184	迹	5	83C7	8FF9	迹		
4190	邳	5	83CD	90C4	邳		
4501	獮	1	85CA	7356	獮		

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
4533	唵	1	85EA	552B	唵		
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	2	8949	5F03	弃		棄，異體字問題。
5091	鄒	1	89A4	92E3	鄒		
5108	惛	5	89B5	60FD	惛		
5300	鬪	1	8AD8	9B2A	鬪		鬪，異體字問題。
5302	昏	2	8ADA	662C	昏		昏，異體字問題。
5396	鎔	1	8B79	938B	鎔		
5578	弛	5	8CB4	38AE		弓  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		
5726	呂	1	8DAB		呂		呂，異體字問題。

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
5752	摠	1	8DC5	63D4	摠		
5860	无	1	C6D3	65E0	无		

二、手動取代表說明：

《韓非子》(標記)檔案中需手動修改的情況分為以下幾種：

(一) 《韓非子》(標記)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. 標記行碼：語言所標記檔裡設定的行碼。
5. 檔案原文摘錄：摘錄檔案《韓非子》(標記)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。
6. 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。
7. 備註：記錄修改相同辭句的次數或其他事項。

附件二、《韓非子》(標記)手動取代表

編號	造字	取代	標記行碼	檔案原文摘錄	手動取代結果	備註
298	△	▲	13772	龐米 <u>閒</u> 氏	龐米 <u>間</u> 氏	有 2 處

編號	造字	取代	標記行碼	檔案原文摘錄	手動取代結果	備註
				(NB3) (全部取代)	(NB3) (全部取代)	

(二)《韓非子》(標記)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，請參考附件三。欄位說明如下：

1. 頁碼：「●」字形所在位置之原書頁碼。
2. 標記行碼：語言所標記檔裡設定的行碼。
3. 檔案原文摘錄：摘錄《韓非子》(標記)檔案有「●」字形的文句。
4. 原書字形：「●」在原書中之字形，以此字取代檔案原文的「●」。
5. 備註：記錄修改相同詞句的次數或其他事項。

附件三、《韓非子》(標記)「●」字手動取代表

頁碼	標記行碼	檔案原文摘錄	原書字形	備註
p. 249	3361	姦●(VH1)[+co]	衰	

(三)《韓非子》(標記)檔案中出現「☆」字形，共計 14 字，比對原書後以 Unicode 字形或構字式手動取代，並製作「☆」字形手動取代表，請參考附件四。欄位說明如下：

1. 標記行碼：語言所標記檔裡設定的行碼。
2. 檔案原文：摘錄《韓非子》(標記)檔案有「☆」字形的文句。
3. 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。

附件四、《韓非子》(標記)「☆」字形手動取代表

標記行碼	檔案原文摘錄	原書字形	備註
372	☆然(VI)[+poly]	𠂇 𠂇亥	
1483	☆☆(VI)[+poly]	口 𠂇顏	
4095	而(C) ☆(U) 小變(NI)	車 𠂇辰	
5417	☆器(NA2)[+attr]	𠂇 𠂇窆	
8938	☆(NA2) 者(NH)	穴 𠂇舛	
9815	☆擻(VA)[+co]	扌 𠂇騫	
10314	☆危(NA1)[+others, +on]	𠂇	
10354	與(C) ☆危(NA1)	𠂇	
10391	所(NH) ☆(VC1Y)	𠂇	
10401	問(VE) ☆危(NA1)	𠂇	
10402	親(DH) ☆(VC1) 子	𠂇	
10673	將(DD) 何(NH) ☆	尸 𠂇橋	
17578	不(DC) ☆(VC1) 瘞	扌 𠂇副	

(四)《韓非子》(標記)檔案中出現「?」字形，共計 20 字，比對原書後以 Unicode 字形或構字式手動取代，並製作「?」字形手動取代表，請參考附件五。欄位說明如下：

5. 標記行碼：語言所標記檔裡設定的行碼。
6. 檔案原文：摘錄《韓非子》(標記)檔案有「?」字形的文句。
7. 字形：「?」在原書中之字形，以此字取代檔案原文的「?」。
8. 備註：記錄修改相同詞句的次數或其他事項。

附件五、《韓非子》(標記)「?」字形手動取代表

標記行碼	檔案原文摘錄	原書字形	備註
615	不(DC) ?(VA) 而(C) 死	鬪	
1483	其(NH) ?(NI)	鬪	

標記行碼	檔案原文摘錄	原書字形	備註
1664	邑(DG) ?(VA) 之(T) 勇	鬪	
4927	戰?(VA)[+co]	鬪	
7925	怒(VH2) 則(C) ?(VA)	鬪	
7926	?(VA) 必(VM) 三族	鬪	
9846	與(P) 之(NH) ?(VA)	鬪	
11271	衢(NA4) 而(C) ?(VP)	飡	
11277	乃(DL) ?(VP)[+nv] 之	飡	
11284	而(C) ?(VP)[+nv] 民	飡	
11291	女(NH) 之(T) ?(VP)	飡	
11306	而(C) ?(VP)[+nv] 之	飡	
15139	?洧(NA4)[+co]	鬪	有 2 處
15573	而(C) 戰?(VA)[+co]	鬪	
16449	作(VP) ?(NI) 以(P) 散	鬪	
17339	?爭(VA)[+co]	鬪	
17406	家(DN) ?(VA) 之(T) 勇	鬪	
17407	無(VG) 私?(NI)[+attr]	鬪	
17591	力(NI) 疾(DV) ?(VA)	鬪	

(五)《韓非子》(標記)檔案中，對於舊版造字以外發現的錯字，以手動修改，並製作錯字修改表，請參考附件六。欄位說明如下：

1. 標記行碼：語言所標記檔裡設定的行碼。
2. 檔案原文摘錄：摘錄檔案《韓非子》(標記)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。
3. 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。

4. 備註：記錄修改相同詞句的次數或其他事項。

附件六、《韓非子》(標記)缺字外錯字修改

標記行碼	檔案原文摘錄	手動取代結果	備註
4876	以(P) 道(NI) <u> </u> (VAN) 天下	以(P) 道(NI) <u>莅</u> (VAN) 天下	
5377	<u> </u> (VAN) 天下(NA3) (全部取代)	<u>莅</u> (VAN) 天下(NA3) (全部取代)	有 2 處
5775	<u> </u> (VAN) 政(NI) 三 (S) 年	<u>莅</u> (VAN) 政(NI) 三(S) 年	
7889	以(P) 嚴(NI) <u> </u> (VAN) 人	以(P) 嚴(NI) <u>莅</u> (VAN) 人	
10151	戮(NI) 以(P) <u> </u> (VAN)	戮(NI) 以(P) <u>莅</u> (VAN)	
12646	<u> </u> (VAN) 政(NI)	<u>莅</u> (VAN) 政(NI)	
6328	血(VH1)[+nv] 而(C) <u> </u> (VH1)	血(VH1)[+nv] 而(C) <u>衄</u> (VH1)	
6329	已(DD) <u> </u> (VH1) 而	已(DD) <u>衄</u> (VH1) 而	
13491	僖負 <u> </u> (NB1)[+prop]	僖負 <u>羈</u> (NB1)[+prop]	
1847	<u> </u> 靈公(NB1)[+prop] (全部取代)	<u>衛</u> 靈公(NB1)[+prop] (全部取代)	有 2 處
2381	<u> </u> 公子開方(NB1) (全部取代)	<u>衛</u> 公子開方(NB1) (全部取代)	有 2 處
2384	<u> </u> (NB3) (全部取代)	<u>衛</u> (NB3) (全部取代)	有 23 處
2424	<u> </u> (NB1)[+prop]	<u>衛公子開方</u>	

標記行碼	檔案原文摘錄	手動取代結果	備註
		(NB1)[+prop]	
2951	%君(NB1)[+prop] (全部取代)	衛君(NB1)[+prop] (全部取代)	有 10 處
2952	%國(NB3) (全部取代)	衛國(NB3) (全部取代)	有 2 處
5275	以(P) 慈(NI) % (VC1)	以(P) 慈(NI) 衛 (VC1)	
7818	%嗣君(NB1) (全部取代)	衛嗣君(NB1) (全部取代)	有 5 處
12132	%侯(NB1)[+prop]	衛侯(NB1)[+prop]	
3195	姦'(NI)	姦衰(NI)	
3386	禁(VE) 其(NH) '(NI)	禁(VE) 其(NH) 衰 (NI)	
7607	弊*(NA2) (全部取代)	弊袴(NA2) (全部取代)	有 4 處
8236	今(NA5) 夫(NH) * (NA2)	今(NA5) 夫(NH) 袴 (NA2)	
8237	*(NA2) 之(T) 與 (P)	袴(NA2) 之(T) 與 (P)	
9824	為(VC1) *(NA2)	為(VC1) 袴(NA2)	
9826	今(NA5) *(NA2)	今(NA5) 袴(NA2)	
9828	故*(NA2)[+attr] (全部取代)	故袴(NA2)[+attr] (全部取代)	有 2 處
10787	不(DC) 失(VJ) * (NA2)	不(DC) 失(VJ) 袴 (NA2)	
15665	懷±(NI)[+co]	懷紆(NI)[+co]	

標記行碼	檔案原文摘錄	手動取代結果	備註
11634	<u>Q</u> 旅(NA1)	<u>羈</u> 旅(NA1)	
6703	<u>K</u> 融(NB1)[+prop]	<u>歷</u> 融(NB1)[+prop]	
15289	<u>K</u> 然(VI)[+poly]	<u>歷</u> 然(VI)[+poly]	
2703	今(NA5) 襲(VC1) <u>W</u> (NI)	今(NA5) 襲(VC1) <u>迹</u> (NI)	
4075	亂(NI) 之(T) <u>W</u> (NI)	亂(NI) 之(T) <u>迹</u> (NI)	
10588	<u>w</u> 繫(NA2)	<u>鞞</u> 繫(NA2)	
8195	怒 <u>..</u> (NA1)[+attr]	怒 <u>鼃</u> (NA1)[+attr]	
8204	見(VK) 怒(VH2) <u>..</u>	見(VK) 怒(VH2) <u>鼃</u>	
8208	<u>..</u> (NA1) 有(VG)	<u>鼃</u> (NA1) 有(VG)	
8211	<u>..</u> (NA1) 有(VG) 氣	<u>鼃</u> (NA1) 有(VG) 氣	
9852	買(VC1) <u>a</u> (NA1)	買(VC1) <u>鼃</u> (NA1)	
9856	亡(VP) 其(NH) <u>a</u> (NA1)	亡(VP) 其(NH) <u>鼃</u> (NA1)	
11192	龜 <u>a</u> (NA1)[+co]	龜 <u>鼃</u> (NA1)[+co]	