

《戰國策》(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
2008/3/21 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 978 的「戣」字，Unicode 編碼是 4CD2，由於 XP 並不支援，仍需使用構字式「戣△鳥」。

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：

《戰國策》(標記)使用舊版造字 96 個，字頻 390 次，這 96 個造字中，84 個可轉成 Windows XP 能顯示的字，字頻 372 次；另外 12 個字必須轉成構字式，字頻 18 次。

轉碼完成製作轉碼分析表，請參考附件一《戰國策》(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 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 4012 的「鼓」字頻次為 3，依書將地名的 1 字手動修改為「鼓」，其他 2 字皆轉為標準字「鼓」。手動修改部份請參考附件二：《戰國策》(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 4561 的「𠄎」字，而原書使用之字為「鹿」，查詢教育部異體字字典確認「𠄎」並非「鹿」之異體字，因此歸為錯字，以程式取代為正確字形「鹿」。
 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 4792 的「𠄎」。

附件一、《戰國策》(標記)轉碼分析表

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
59	𠄎	1	FA7A	356D		口𠄎天	

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
296	𡗗	2	FBEC			𡗗	
628	繚	2	FDFE			繚	
813	旤	1	8E5B	65E4	旤		
826	癱	1	8E68	7655	癱		
828	跣	1	8E6A	8DD4	跣		
832	淹	1	8E6E	6CB2	淹		
860	裊	10	8EAC	890F	裊		
1024	虵	2	8FB3	8675	虵		
1067	碁	2	8FDE	68CA	碁		
1513	糧	1	92C5	7CAE	糧		糧，異體字問題。
1565	繇	2	92F9	7DDC	繇		
1857	𧈧	2	94E3	86A0	𧈧		
1925	羈	4	9568	8989	羈		羈，異體字問題。
2044	踈	1	9642	8E08	踈		
2053	躑	1	964B	8EA7	躑		
2154	𧈧	1	96D2	9B40	𧈧		
2162	醜	1	96DA	918E	醜		
2415	鬪	6	987B	95D8	鬪		

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
2501	颺	1	98F3	98C3	颺		飄，異體字問題。
2733	羴	5	9AA1	21C5D		尸△羊	
2783	蚤	1	9AD3	8689	蚤		蚊，異體字問題。
3791	況	9	8156	51B5	況		況，異體字問題。
3794	湊	1	8159	51D1	湊		湊，異體字問題。
3796	決	15	815B	51B3	決		決，異體字問題。
3801	剗	2	8160	5257	剗		
3804	効	6	8163	52B9	効		效，異體字問題。
3806	勳	1	8165	52E0	勳		
3814	却	10	816D	5374	却		
3817	廁	2	8170	53A0	廁		廁，異體字問題。
3834	嗥	1	81A3	5637	嗥		
3866	崐	1	81C3	5D10	崐		
3868	卮	4	81C5	5DF5	卮		
3877	迺	1	81CE	5EFC	迺		

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
3883	徧	4	81D4	5FA7	徧		
3891	慙	2	81DC	6159	慙		
3924	碁	3	81FD	671E	碁		期，異體字問題。
3937	檝	2	824B	6A9D	檝		
3967	炳	1	8269	712B	炳		
3976	犁	1	8272	7282	犁		
3981	珉	15	8277	73C9	珉		
3992	嘗	1	82A4	751E	嘗		
3996	疎	1	82A8	758E	疎		
4012	鼓	3	82B8	76B7	鼓		鼓，異體字問題。 部份使用不同字形「鼓」，手動修改。
4015	眦	1	82BB	7726	眦		眦，異體字問題。
4029	稷	4	82C9	799D	稷		稷，異體字問題。
4038	竈	6	82D2	7AC8	竈		
4041	豎	2	82D5	7AEA	豎		

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
4043	筭	1	82D7	7B53	筭		
4044	筓	1	82D8	7B6F	筓		
4045	箒	1	82D9	7B92	箒		
4070	縹	8	82F2	7DE4	縹		
4080	羣	63	82FC	7FA3	羣		
4085	耻	5	8342	803B	耻		
4091	皐	16	8348	81EF	皐		
4186	遡	1	83C9	9061	遡		
4190	郟	5	83CD	90C4	郟		
4417	●	1	8554		●		半，錯字，以程式全部取代。
4450	𠂇	1	8575	524F	𠂇		
4458	勸	1	857D	52CC	勸		
4558	廐	1	8644	53A9	廐		廐，異體字問題。
4561		1	8647			𠂇 厶 𠂇 比 𠂇	廐，錯字，以程式全部取代。
4660	塏	38	86CC	57CA	塏		
4735	壠	1	8758	58E0	壠		壠，異體字問

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
							題。
4768	妬	15	8779	59AC	妬		妒，異體字問題。
4792		1	87B3			𠂇 厶 虛	待造字。
4887	醯	1	8853	4908		酉 厶 益	
4904	尅	2	8864	5C05	尅		剋，異體字問題。
5034	弃	1	8949	5F03	弃		棄，異體字問題。
5051	輦	1	895A	4870		般 厶 車	
5065	忼	1	8968	5FFC	忼		
5108	惛	1	89B5	60FD	惛		昏，異體字問題。
5119	博	1	89C0	613D	博		
5133	憇	15	89CE	6199	憇		
5282	敷	2	8AC6	656B	敷		
5293		1	8AD1			冗 厶 幕	幕，異體字問題。
5300	鬪	1	8AD8	9B2A	鬪		
5318	蔭	1	8AEA	4516		艹 厶 陵	
5328		1	8AF4			目 厶 巧	眇，異體字問題。

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
							題。
5335	𧈧	1	8AFB	459F		亡𧈧虫	
5374	盼	1	8B63	670C	盼		
5395	規	4	8B78	4093		矢𧈧見	規，異體字問題。
5396	鎔	1	8B79	938B	鎔		
5408	紕	1	8BA7	7D4D	紕		紕，異體字問題。
5437	并	26	8BC4	5E77	并		
5466	倏	1	8BE1	5010	倏		倏，異體字問題。
5514	𧈧	1	8C52	25095		執𧈧皿	
5541	蠹	3	8C6D	8827	蠹		蠹，異體字問題。
5621	泪	1	8CDF	6CEA	泪		淚，異體字問題。
5695	𧈧	1	8D6A	3B3A		朕𧈧巾	
5697	𧈧	2	8D6C	7151	𧈧		
5704	𧈧	1	8D73	6C0A	𧈧		𧈧，異體字問題。
5705	廐	4	8D74	5ED0	廐		廐，異體字問題。

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
5723	𦍋	1	8DA8	4367		牛 𦍋 羊	
5724	𦍋	2	8DA9	7A38	𦍋		
5809	躄	1	8DFE	8E79	躄		

二、手動取代表說明：

《戰國策》(標記)檔案中需手動修改的情況分為以下幾種：

(一) 《戰國策》(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. 標記行碼：語言所標記檔裡設定的行碼。
3. 檔案原文摘錄：摘錄檔案《戰國策》(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。
4. 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。
5. 備註：記錄修改相同辭句的次數或其他事項。

附件二、《戰國策》(標記)手動取代表

編號	行碼	檔案原文摘錄	手動取代結果	備註
4012	7418	殺(VC1) 閔王 (NB1)[+prop] 於(P) <u>𦍋</u> 里	殺(VC1) 閔王 (NB1)[+prop] 於 (P) <u>鼓</u> 里	

		(NB4)[+prop]	(NB4)[+prop]	
--	--	--------------	--------------	--

(二)《戰國策》(標記)檔案中出現「●」字形，共計 43 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《戰國策》(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、《戰國策》(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. 81	1005	伐(VC1) ●遂 (NB3)[+prop]	𠄎 𠄎甫	
p. 81	1021	文士(NA1)[+co] 並(DA) ●(U)	食 𠄎芳	
p. 186	2745	至(VA) 於(P) ● 水(NB4)[+prop]	𠄎 𠄎菱	改為 ⁺⁺ 𠄎凌
p. 235	3762	周●(NB1)[+prop] 為(P) 楚王	𠄎	有 6 處
p. 273	4250	樓●(NB1)[+prop] 約(VE) 秦	吾 𠄎午	
p. 276	4301	處(VA) 於(P) ●城(NB4)[+prop]	𠄎戶夕 𠄎	
p. 289	4626	狀(NI) 如(VG) 振(U) ●(NA2)	才 𠄎因	

頁碼	標記行碼	檔案原文摘錄	原書字形	備註
p. 292	4648	必為言之曰：「嗟● 乎，司空馬！」	嗟	
p. 380	5982	勝●(NB1)[+prop] 曰	股𠄎目	
p. 557	9714	俯(VA) 嚼(VC1) ●鯉(NA1)[+co]	魚𠄎卷	
p. 557	9722	脩(VP) 其(NH) ●盧(NA2)[+co]	符𠄎升	
p. 579	10141	臣(NH) 請(VF) 為(P) 君(NA1) ●(VC1)	童𠄎卩	
p. 617	10965	靳●(NB1)[+prop] 曰(VE)	黃𠄎重	有 3 處
p. 678	12235	無(VG) 鈞(NA2) ● (NA2)	𠄎𠄎干	有 2 處
p. 711	13029	至(DF) ●(VH1) 也	黍𠄎刃	
p. 718	13185	●(NB1)[+prop] 之(T) 所(𠄎𠄎骨	有 5 處
p. 751	13789	羽毛(NA2)[+co] ●敝(VH1)[+co]	𠄎𠄎列	
p. 773	14213	使(VF) 趙● (NB1)[+prop]	𠄎𠄎念	
p. 773	14213	及(C) 顏● (NB1)[+prop]	冢	有 2 處
p. 822	15144	因(DL) 圍(VC1)	𠄎𠄎立田𠄎	

頁碼	標記行碼	檔案原文摘錄	原書字形	備註
		●(NB2)[+prop]		
p. 852	15752	不(DC) ●(U) 中(NA1)	𠂔	
p. 884	16480	樓●(NB1)[+prop]	广△鼻	有 4 處
p. 963	17927	王(NA1) ●(U) 臣(NH)	𠂔	
p. 994	18457	愈(DF) ●(VH2N) 其(NH) 厚(NI)	恠	
p. 1095	20616	猶(VG) 剗● (VC10)[+co]	爻△卩	改為刈
p. 1095	20637	犬馬(NA1)[+co] ●(U)	馬△𠂔	
p. 1157	21969	有(VG) 雀(NA1) 生(VP) ●(U)	鳥△旗	

(三)《戰國策》(標記)檔案中出現「？」字形，共計 49 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《戰國策》(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、《戰國策》(標記)「？」字形手動取代表

標記行碼	檔案原文	字形	備註
1983	諍(VC1) 人(NA1) 而	鬪	有 2 處

標記行碼	檔案原文	字形	備註
	(C) ?(VA)		
3109	毋(DC) 相與 (DH)[+others] ?(VA)	鬪	
3131	大(DV) 相與 (DH)[+others] ?(VA)	鬪	
3889	兩(S) 虎(NA1) 相 (DH) ?(VA)	鬪	
5327	?(VP) 雞(NA1)	鬪	
5912	淳于?(NB1)[+prop] 為(P) 齊	髡	有 13 處
6032	魚?(NA1)[+co]	鼃	
6153	今(NA5) ?(NB1)[+prop] 賢者	髡	有 6 處
6207	今(NA5) 又(DL) ?(VC1) 趙	劫	
7711	徐子(NB1)[+prop] ?(VA)	鬪	
8693	獻(VC2) ?駭(A)[+others] 之 (T)	雞	
10231	魏(NB4)[+prop] 而(C) ?(VP) 兵	鬪	
10251	白竈(NA2)[+co] 生(VP) ?	鼃	
11187	而(C) 相(DH) ?(VA) 兩(S)	鬪	
11190	決(VA) 不(DC) 相 (DH) ?(VA)	鬪	
11192	而(C) 尚(DD) 相(DH) ?(VA)	鬪	
16729	趙(NB3)[+prop] 必(VM) 復 (DD) ?(VA)	鬪	
17288	韓?(NB1)[+prop] 曰(VE)	鼃	
17367	寧(VK) 為(VG) ?口(NA1)	雞	

標記行碼	檔案原文	字形	備註
17736	私徒(NA1)[+attr] 以 (P) ?(VA)	鬪	
17913	魏(NB4)[+prop] 而(C) ?(VP) 兵	鬪	
17957	楚(NB3)[+prop] ?(VA) 之(T) 日	鬪	
19433	不(DC) 取(VC2) 素?	凵	
19531	都(NA2) 而(C) ?(VC1) 大王	劫	
20069	史?(NB1)[+prop]	鱈	有 2 處
20338	則(C) ?(VC1) 魏	劫	
21834	魚?(NA1)[+co] 龜鼉	鼉	
22529	壺?(NA2)[+attr] 餌	凵	
22541	以(P) 一(S) 壺?(NA2)[+attr]	凵	
22662	莫(NH) 有(VG) ?志(NI)	鬪	

(四)《戰國策》(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，請參考附件五。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《戰國策》(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。

附件五、《戰國策》(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
14278	文侯 (NB1)[+prop] 謂 (VF) <u>1</u> 師贊 (NB1)[+prop	文侯 (NB1)[+prop] 謂 (VF) <u>覩</u> 師贊 (NB1)[+prop	
7910	終身 (NA1)[+others] 不 (DC) <u>1</u> (VKX)	終身 (NA1)[+others] 不 (DC) <u>睹</u> (VKX)	
7912	不 (DC) <u>1</u> (VKX) 之 (T) 故 (NI)	不 (DC) <u>睹</u> (VKX) 之 (T) 故 (NI)	
22709	未 (DC) <u>1</u> (VK) 其 (NH) 利 (NI)	未 (DC) <u>睹</u> (VK) 其 (NH) 利 (NI)	
4551	而 (C) <u>1</u> (VP) 兵 (NA1)	而 (C) <u>郤</u> (VP) 兵 (NA1)	
6151	則 (C) <u>1</u> (VP) 車 (NA2)	則 (C) <u>郤</u> (VP) 車 (NA2)	
14104	之 (T) 有 (VG) 所 (NH) <u>1</u> (VH1)	之 (T) 有 (VG) 所 (NH) <u>郤</u> (VH1)	
17863	以 (P) <u>1</u> (U) 齊 (NB3)[+prop]	以 (P) <u>郤</u> (U) 齊 (NB3)[+prop]	
20773	寡人 (NH)[+attr] 有 (VG) <u>1</u> (NI)	寡人 (NH)[+attr] 有 (VG) <u>郤</u> (NI)	
21308	趙 (NB3)[+prop] 有 (VG) <u>1</u> (NI) (全部 取代)	趙 (NB3)[+prop] 有 (VG) <u>郤</u> (NI) (全 部取代)	有 2 處
6397	遣 (VF) 太傅 (NA1)[+attr] <u>F</u> (VD)	遣 (VF) 太傅 (NA1)[+attr] <u>賈</u> (VD)	

標記行碼	檔案原文摘錄	手動取代結果	備註
5827	以(P) 滅(VP) <u>W</u> (NI) 也(T)	以(P) 滅(VP) <u>迹</u> (NI) 也(T)	
11409	流言(NI)[+attr] 之(T) <u>W</u> (NI)	流言(NI)[+attr] 之(T) <u>迹</u> (NI)	
11542	遠(VH1N) <u>W</u> (NI) 不(DC) 服	遠(VH1N) <u>迹</u> (NI) 不(DC) 服	
11677	襄(NB1)[+prop] 之 (T) <u>W</u> (NI)	襄(NB1)[+prop] 之 (T) <u>迹</u> (NI)	
14432	能士(NA1)[+attr] 之(T) <u>W</u> (NI)	能士(NA1)[+attr] 之(T) <u>迹</u> (NI)	
20890	吳王(NB1)[+prop] 遠(VP) <u>W</u> (NI)	吳王(NB1)[+prop] 遠(VP) <u>迹</u> (NI)	
20898	先王(NA1)[+attr] 之(T) <u>W</u> (NI) 者	先王(NA1)[+attr] 之(T) <u>迹</u> (NI) 者	
14253	因(DL) 索(VE) 蔡 (NB2)[+prop] 、 梁(NB2)[+prop]	因(DL) 索(VE) 蔡 (NB2)[+prop] 、 <u>皇</u> 梁(NB2)[+prop]	
2611	明主(NA1)[+attr] (VAN) 正(NI)	明主(NA1)[+attr] <u>莅</u> (VAN) 正(NI)	
11455	<u> </u> (VAN) 國(NA3) 之(T) 日(NA5)	<u>莅</u> (VAN) 國(NA3) 之(T) 日(NA5)	
11966	是以(C)[+others] <u> </u> (VAN) 國(NA3) 者	是以(C)[+others] <u>莅</u> (VAN) 國(NA3) 者	
15647	龐 <u> </u> (NB1)[+prop] 與(P) 太子	龐 <u>莅</u> (NB1)[+prop] 與(P) 太子	

標記行碼	檔案原文摘錄	手動取代結果	備註
15661	龐 (NB1)[+prop] 曰(VE)	龐 <u>葱</u> (NB1)[+prop] 曰(VE)	
3628	而(C) 立(VP) 負_ (NB1)[+prop]	而(C) 立(VP) 負 <u>葛</u> (NB1)[+prop]	
3629	負_(NB1)[+prop] 立(VH1)	負 <u>葛</u> (NB1)[+prop] 立(VH1)	
3637	負_(NB1)[+prop] 必(VM) 以(P) 魏	負 <u>葛</u> (NB1)[+prop] 必(VM) 以(P) 魏	
4616	緝(NB1)[+prop] 病 (VH2N) 𠄎(NI)	緝(NB1)[+prop] 病 (VH2N) <u>鈞</u> (NI)	
4830	𠄎(NA2) 不(DC) 能 (VM) 牽	<u>鈞</u> (NA2) 不(DC) 能(VM) 牽	