

《戰國策》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
2008/3/20 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 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》使用舊版造字 114 個，字頻 743 次，這 114 個造字中，83 個可轉成 Windows XP 能顯示的字，字頻 721 次；另外 9 個字必須轉成構字式，字頻 22 次。

轉碼完成製作轉碼分析表，請參考附件一《戰國策》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 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》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「𠄎」，查詢教育部異體字字典確認「𠄎」並非「𠄎」之異體字，因此歸為錯字，以程式取代為正確字形「𠄎」。

附件一、《戰國策》轉碼分析表

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
813	旣	1	8E5B	65E4	旣		
826	癱	1	8E68	7655	癱		
828	𠄎	1	8E6A	8DD4	𠄎		

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
832	沱	1	8E6E	6CB2	沱		
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	螽	5	94E3	86A0	螽		
1925	羈	4	9568	8989	羈		羈，異體字問題。
2044	踈	1	9642	8E08	踈		
2053	躡	1	964B	8EA7	躡		
2154	魀	4	96D2	9B40	魀		
2162	醜	1	96DA	918E	醜		
2415	鬪	6	987B	95D8	鬪		
2501	颯	1	98F3	98C3	颯		飄，異體字問題。
2733	屮	14	9AA1	21C5D		尸△羊	
2783	蚤	1	9AD3	8689	蚤		蚊，異體字問題。
3791	况	9	8156	51B5	况		况，異體字問題。

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
							題。
3794	湊	1	8159	51D1	湊		湊，異體字問題。
3796	決	19	815B	51B3	決		決，異體字問題。
3801	剗	2	8160	5257	剗		
3804	効	6	8163	52B9	効		効，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		
3883	徧	4	81D4	5FA7	徧		
3891	慙	2	81DC	6159	慙		
3924	朞	3	81FD	671E	朞		朞，異體字問題。
3937	檝	2	824B	6A9D	檝		

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
3967	炳	1	8269	712B	炳		
3976	犁	1	8272	7282	犁		
3981	珉	24	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「𪗇」，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		
4043	筭	1	82D7	7B53	筭		
4044	筓	1	82D8	7B6F	筓		
4045	箒	1	82D9	7B92	箒		
4070	縹	11	82F2	7DE4	縹		
4080	羣	66	82FC	7FA3	羣		

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
4085	耻	5	8342	803B	耻		
4091	臯	17	8348	81EF	臯		
4104	莅	3	8355	8385	莅		
4107	葱	5	8358	8471	葱		
4109	蒟	3	835A	84AD	蒟		
4132		79	8371			血会𠂇	眾，異體字問題。
4134	衛	124	8373	885E	衛		衛，異體字問題。
4144	霸	13	837D	8987	霸		霸，異體字問題。
4146	覩	4	83A1	89A9	覩		睹，異體字問題。 部份使用不同字形「覩」，手動修改。
4167	賚	1	83B6	8CEB	賚		
4184	迹	8	83C7	8FF9	迹		
4186	遡	1	83C9	9061	遡		
4190	邳	11	83CD	90C4	邳		
4193	鈞	2	83D0	920E	鈞		鈞，異體字問題。

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
4203	隣	8	83DA	96A3	隣		鄰，異體字問題。
4226	凜	3	83F1	98E1	凜		
4237	髡	22	83FC	9AE0	髡		
4242	鬪	16	8442	9B2D	鬪		
4250	鱻	2	844A	9C0C	鱻		
4256	鷄	3	8450	9DC4	鷄		雞，異體字問題。
4265	鼃	1	8459	9F03	鼃		
4266	鼃	1	845A	9F02	鼃		
4267	鼃	2	845B	9F08	鼃		
4307	罵	1	84A5	99E1	罵		罵，異體字問題。
4309	刼	3	84A7	523C	刼		劫，異體字問題。
4417	●	1	8554		●		半，錯字，以程式全部取代。
4450	勑	1	8575	524F	勑		
4458	勑	1	857D	52CC	勑		
4558	廐	1	8644	53A9	廐		廐，異體字問題。

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
4561		1	8647			形 厂 曲比	麤，錯字，以程式全部取代。
4660	崱	39	86CC	57CA	崱		
4735	壠	1	8758	58E0	壠		壠，異體字問題。
4768	妬	15	8779	59AC	妬		妒，異體字問題。
4792		1	87B3			竹 虚	待造字
4887	醯	1	8853	4908		酉 益	
4904	尅	2	8864	5C05	尅		尅，異體字問題。
5034	弃	1	8949	5F03	弃		棄，異體字問題。
5051	輦	1	895A	4870		車 輦	
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	21	89CE	6199	憇		
5282	敷	2	8AC6	656B	敷		
5293		1	8AD1			冫 幕	冫，異體字問

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
							題。
5300	鬪	1	8AD8	9B2A	鬪		
5318	陵	1	8AEA	4516		艹 厶 陵	
5328		1	8AF4			目 厶 丐	眇，異體字問題。
5335	蚤	1	8AFB	459F		亡 厶 虫	
5374	盼	1	8B63	670C	盼		
5395	規	4	8B78	4093		矢 厶 見	規，異體字問題。
5396	鐸	1	8B79	938B	鐸		
5408	紕	1	8BA7	7D4D	紕		紕，異體字問題。
5437	并	32	8BC4	5E77	并		
5466	倏	1	8BE1	5010	倏		倏，異體字問題。
5514	塾	1	8C52	25095		執 厶 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	勝	1	8D6A	3B3A		朕 厶 巾	

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
5697	煇	2	8D6C	7151	煇		
5704	羶	1	8D73	6C0A	羶		𪗇，異體字問題。
5705	廐	4	8D74	5ED0	廐		廐，異體字問題。
5723	𦍋	1	8DA8	4367		牛 𦍋 羊	
5724	𦍋	2	8DA9	7A38	𦍋		
5809	踰	1	8DFE	8E79	踰		

二、手動取代表說明：

《戰國策》檔案中需手動修改的情況分為以下幾種：

(一) 《戰國策》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. 檔案原文摘錄：摘錄檔案《戰國策》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：記錄修改相同辭句的次數或其他事項。

附件二、《戰國策》手動取代表

編號	檔案原文摘錄	手動取代結果	備註
4012	於是殺閔王於 <u>兰</u> 里	於是殺閔王於 <u>鼓</u> 里	
4146	文侯謂 <u>桀</u> 師贊曰	文侯謂 <u>覩</u> 師贊曰	

(二)《戰國策》檔案中出現「●」字形，共計 58 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《戰國策》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、《戰國策》「●」字手動取代表

頁碼	檔案原文摘錄	原書字形	備註
p. 81	昔者神農伐●遂	禘 𠂔甫	
p. 81	兵革不藏；文士並●	食 𠂔芳	
p. 105	乃取欺於亡●	国	簡轉繁為國
p. 186	夜行而晝伏，至於●水	； 𠂔菱	改為 𠂔 凌
p. 235	周●為楚王曰	寂	有 12 處
p. 273	樓●約秦魏	吾 𠂔午	有 4 處
p. 276	秦子異人質於趙，處於●城	𠂔戶夕 𠂔	
p. 289	以示韓倉，狀如振●	扌 𠂔因	
p. 292	必為言之曰：「嗟●乎，司空馬！」	嗟	
p. 380	如使而弗及也。」勝●曰	股 𠂔目	
p. 557	淹乎大沼，府喙●鯉	魚 𠂔卷	
p. 557	方將脩其●盧	符 𠂔升	

頁碼	檔案原文摘錄	原書字形	備註
p. 579	臣請為君●其胸殺之	童△刂	
p. 617	令韓陽告上黨之守靳●曰	黃△重	有 3 處
p. 678	無鈞●鐔蒙須之便	𠂔△干	有 2 處
p. 711	夫膠漆，至●也	黍△刃	
p. 718	●之所以事王者，知也	𠂔△骨	有 5 處
p. 751	車甲羽毛●敝	衤△列	
p. 773	趙王疑之，使趙●	𠂔△忝	
p. 773	使趙 ^𠂔 △忝及顏●代將	寂	有 2 處
p. 822	韓氏因圍●	𠂔 ^𠂔 △田	
P. 824	冰●齧其墓	樂	有 2 處
p. 852	臣聞明王不●中而行	骨	
p. 884	所用者，樓●	广△鼻	有 4 處
p. 930	鐵幕、革抉、●芮	口△天	
p. 963	王●臣反，乃行	骨	
p. 994	聶政驚，愈●其厚	恠	
p. 1095	猶剗●者也	爻△刂	改為刈
p. 1095	使齊犬馬●而不言燕	馬△爻	
p. 1157	宋康王之時有雀生●	鳥△旗	有 4 處