

# 《左傳》(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  
2008/2/26 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「戡」字，Unicode 編碼是 4CD2，由於 XP 並不支援，仍需使用構字式「戡△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：

《左傳》(標記)使用舊版造字 106 個，字頻 1624 次，這 106 個造字中，97 個可轉成 Windows XP 能顯示的字，字頻 1607 次；另外 16 個字必須轉成構字式，字頻 108 次。

轉碼完成製作轉碼分析表，請參考附件一《左傳》(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 4237 的「髡」字頻次為 9，原書卻分別寫成「髡」和「髡」兩種異體字字形，依書將代表人名的 2 字手動修改為「髡」，其他 7 字皆轉為標準字「髡」。手動修改部份請參考附件二：《左傳》(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 4353 的「菟」字，而原書使用之字為「菟」，查詢教育部異體字字典確認「菟」並非「菟」之異體字，因此歸為錯字，以程式取代為正確字形「菟」。

附件一、《左傳》(標記)轉碼分析表

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
793	韃	1	8E47	97E4	韃		
978	戡	1	8F63	4CD2		戡△鳥	

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
980	犖	8	8F65	72AB	犖		
982	胷	3	8F67	80F7	胷		胸，異體字問題。
1062		1	8FD9			馬𠂇亞	駟，異體字問題。
1077	稟	1	8FE8	7A3E	稟		稟，異體字問題。
1565	繇	9	92F9	7DDC	繇		
1756	蓑	1	945C	84D8	蓑		蓑，異體字問題。
1857	蚩	3	94E3	86A0	蚩		
1885	螽	4	9540	882D	螽		
1893	袞	1	9548	886E	袞		袞，異體字問題。
1937	詢	1	9574	8A3D	詢		
1981	讐	4	95C2	8B90	讐		讐，異體字問題。
1999	贊	1	95D4	8CDB	贊		贊，異體字問題。
2043	蹕	2	9641	8E01	蹕		
2449	雒	3	98BF			名𠂇隹	雒，錯字，以程式全部取代。

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
2478	鞏	6	98DC	978C	鞏		
2479	鞞	1	98DD	979B	鞞		
2651	鸚	9	99EC	9E1C	鸚		
2836	𠄎	6	9B49			𠄎	部份使用不同字形「𠄎」，手動修改。
2837	𠄎	14	9B4A			𠄎	
2878	𠄎	3	9B73	5081	𠄎		
2885	𠄎	1	9B7A	544C	𠄎		𠄎，異體字問題。
2906	𠄎	1	9BB1	659A	𠄎		𠄎，異體字問題。
2947	𠄎	1	9BDA	8066	𠄎		𠄎，異體字問題。
3068	𠄎	8	9CB6			𠄎	
3069	𠄎	10	9CB7			𠄎	
3071	𠄎	10	9CB9			𠄎	
3072	𠄎	9	9CBA			𠄎	
3073	𠄎	8	9CBB			𠄎	部份使用不同字形「𠄎」，手動

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
							修改。
3074	𠄎	15	9CBC			𠄎	
3770	个	3	8141	4E2A	个		
3780	儻	3	814B	5101	儻		
3787	冢	6	8152	51A2	冢		
3804	効	1	8163	52B9	効		効，異體字問題。
3834	嗶	1	81A3	5637	嗶		
3846	壻	5	81AF	58FB	壻		
3883	徧	8	81D4	5FA7	徧		
3891	慙	3	81DC	6159	慙		
3907	搯	1	81EC	644F	搯		
3921	皙	23	81FA	6673	皙		
3930	捰	2	8244	6901	捰		
3935	槩	7	8249	69E9	槩		
3949	涖	29	8257	6D96	涖		
3974	牀	5	8270	7240	牀		
3976	犁	12	8272	7282	犁		部份使用不同字形「犁」，手動

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
							修改。
3987	穀	3	827D	7474	穀		
4010	皐	3	82B6	7690	皐		皐，異體字問題。 部份使用不同字形「皐」，手動修改。
4011	皦	8	82B7	76A1	皦		皦，異體字問題。
4030	秆	1	82CA	79C6	秆		
4038	竈	17	82D2	7AC8	竈		
4041	豎	1	82D5	7AEA	豎		豎，異體字問題。
4080	羣	96	82FC	7FA3	羣		
4091	皐	25	8348	81EF	皐		部份使用不同字形「皐」，手動修改。
4107	葱	2	8358	8471	葱		
4132		159	8371			血𠂇	眾，異體字問題。

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
4134	衛	778	8373	885E	衛		衛，異體字問題。
4140	衽	1	8379	88B5	衽		衽，異體字問題。
4142	褻	1	837B	8943	褻		褻，異體字問題。
4145	羈	3	837E	898A	羈		羈，異體字問題。
4156	諂	2	83AB	8B1F	諂		
4172	屨	2	83BB	8E77	屨		
4184	迹	3	83C7	8FF9	迹		
4185	邊	1	83C8	9037	邊		
4203	隣	1	83DA	96A3	隣		隣，異體字問題。
4224	顛	2	83EF	983F	顛		
4237	髡	9	83FC	9AE0	髡		髡，異體字問題。 部份使用不同字形「髡」，手動修改。
4238	髡	1	83FD	9AE2	髡		

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
4242	鬪	58	8442	9B2D	鬪		
4250	鱒	6	844A	9C0C	鱒		
4252	鰓	3	844C			魚厶眾	鰓，異體字問題。
4259	麇	1	8453	9E95	麇		
4265	龜	1	8459	9F03	龜		
4307	罵	1	84A5	99E1	罵		罵，異體字問題。
4308	劫	5	84A6	5227	劫		劫，異體字問題。
4312	睥	1	84AA	66AD	睥		睥，異體字問題。
4353	菟	1	84D3	83AC	菟		菟，錯字，以程式全部取代。
4469	勻	41	85AA	5304	勻		
4598	韡	2	866C	292FE		革厶韡	韡，異體字問題。
4753		3	876A			角厶比 厶大	角厶大，錯字，以程式全部取代。
4762	旃	2	8773	3ACB		舛厶冉	旃，異體字問題。

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
4768	妬	1	8779	59AC	妬		妒，異體字問題。
4822	尫	1	87D1	5C29	尫		尫，異體字問題。
4878	宀	1	884A	5B82	宀		宀，異體字問題。
4894		1	885A			宀 ㇀ 心 ㇀ 冉	甯，異體字問題。
5034	弃	9	8949	5F03	弃		棄，異體字問題。
5035		3	894A			具 ㇀ 卩	郟，錯字，以程式全部取代。
5085	苒	1	897C	26B12		艹 ㇀ 丰	半，錯字，以程式全部取代。
5111	旆	1	89B8	65BE	旆		旆，異體字問題。
5150	伏	1	89DF	3432		亻 ㇀ 大	伏，異體字問題。
5156	丕	13	89E5	353B		丕	部份使用不同字形「丕」，手動修改。

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
5248		1	8AA4			扌厶間	擱，異體字問題。
5286	窻	1	8ACA	41AB		穴厶忽	窗，異體字問題。
5299	戟	1	8AD7	39B8		卓厶戈	戟，異體字問題。
5404	鴟	1	8BA3	4CED		即厶鳥	
5456		1	8BD7			韋厶未	韎，異體字問題。
5505	耆	1	8C49	264BE		老厶占	考厶口，錯字，以程式全部取代。
5591	杞	61	8CC1	233CC		木厶巳	杞，異體字問題。
5622	邛	5	8CE0	48B5		云厶卩	
5627	涑	1	8CE5	6D13	涑		涑，錯字，以程式全部取代。
5652	蝥	2	8CFE	27312		或厶虫	
5655	囀	5	8D42	56BB	囀		囀，異體字問題。
5705	廐	5	8D74	5ED0	廐		廐，異體字問題。

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
5713	躡	2	8D7C	481E		足𠂔戚	
5752	摠	1	8DC5	63D4	摠		
5860	无	5	C6D3	65E0	无		無，異體字問題。

## 二、手動取代表說明：

《左傳》(標記)檔案中需手動修改的情況分為以下幾種：

(一) 《左傳》(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《左傳》(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、《左傳》(標記)手動取代表

編號	標記行碼	檔案原文摘錄	手動取代結果	備註
2836	30649	坤(NB2)[+prop] <u>  </u> (U) 亭(U)	坤(NB2)[+prop] <u>  </u> (U) 亭(U)	

編號	標記行碼	檔案原文摘錄	手動取代結果	備註
3073	29583	比(NB2)[+prop] 𠄎(U) 𠄎(U)	比(NB2)[+prop] 𠄎(U) 𠄎(U)	
3976	2590	邠(NB3)[+prop] - 邠來(NB1)[+prop] (全部取代)	邠(NB3)[+prop] 邠來(NB1)[+prop] (全部取代)	有 2 處
	16264	伯州邠 (NB1)[+prop] 奔 (VAN) 楚(全部取 代)	伯州邠 (NB1)[+prop] 奔 (VAN) 楚(全部取 代)	有 5 處
	22461	與(C) 師邠邠 (NB1)[+prop]	與(C) 師邠邠 (NB1)[+prop]	
	37739	樂邠邠 (NB1)[+prop]	樂邠邠 (NB1)[+prop]	
4010	及(VAN) 𠄎𠄎 (NB4)[+prop]	及(VAN) 𠄎𠄎 (NB4)[+prop]		
4091	肥子 𠄎 (NB1)[+prop]	肥子 𠄎 (NB1)[+prop]		
4237	34073	胡子 𠄎 (NB1)[+prop] (全 部取代)	胡子 𠄎 (NB1)[+prop] (全 部取代)	有 2 處
	39528	宋(NB3)[+prop] 樂 𠄎(NB1)[+prop]	宋(NB3)[+prop] 樂 𠄎(NB1)[+prop]	
5156	27899	𠄎(NB1)[+prop] 之(T) 難(NI)	𠄎(NB1)[+prop] 之(T) 難(NI)	

(二)《左傳》(標記)檔案中出現「●」字形，共計 7 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《左傳》(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、《左傳》(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. 447	7699	魏●(NB1)[+prop]	𠄎	共 4 處
p. 456	7882	子 玉 (NB1)[+prop] 使 (VF) 伯 ● (NB1)[+prop]	𠄎	
p. 1148	24583	慶 ● (NB1)[+prop] 為(VG) 上獻	𠄎	
p. 1724	42579	君(NH) 將(DD) ● (VC1) 之(NH)	𠄎士月爿	

(三)《左傳》(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《左傳》(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、《左傳》(標記)「？」字形手動取代表

標記行碼	檔案原文	原書字形	備註
32341	?鳩氏(NA1)[+others]	鴟	

(四)《左傳》(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《左傳》(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、《左傳》(標記)缺字外錯字修改表

標記行碼	檔案原文摘錄	手動取代結果	備註
5512	<u>平</u> 鄭(NB1)[+prop]	<u>丕</u> 鄭(NB1)[+prop]	
5561	<u>平</u> 豹(NB1)[+prop]	<u>丕</u> 豹(NB1)[+prop]	