

# 《左傳》舊版造字轉碼說明

中研院資訊所文獻處理實驗室  
中央研究院語言所語言典藏計畫  
2008/2/26 丁玟伶

《左傳》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 70,194 個字，而 XP 只能支援 20,902 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》使用舊版造字 105 個，字頻 1995 次，這 105 個造字中，83 個可轉成 Windows XP 能顯示的字，字頻 1879 次；另外 22 個字必須轉成構字式，字頻 116 次。

轉碼完成製作轉碼分析表，請參考附件一《左傳》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：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：舊版造字編號 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》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：檔案中使用舊版造字編號 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》轉碼分析表

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
297	𠄎	2	FBED			𠄎	
299	𠄐	1	FBEF			𠄐	
300	𠄑	1	FBF0			𠄑	

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
436	𠄎	1	FCDB			𠄎	士𠄎月，錯字，以程式全部取代。
626	𠄎	1	FDFC			𠄎	𠄎，錯字，以程式全部取代。
763	𠄎	1	FEE8			𠄎	
793	𠄎	1	8E47	97E4	𠄎		
978	𠄎	1	8F63	4CD2		𠄎△鳥	
980	𠄎	12	8F65	72AB	𠄎		
982	𠄎	1	8F67	80F7	𠄎		胸，異體字問題。
988	𠄎	8	8F6D	7F8B	𠄎		
1062		1	8FD9			馬△𠄎	駟，異體字問題。
1077	𠄎	1	8FE8	7A3E	𠄎		𠄎，錯字，以程式全部取代。
1565	𠄎	9	92F9	7DDC	𠄎		
1756	𠄎	1	945C	84D8	𠄎		𠄎，異體字問題。
1857	𠄎	3	94E3	86A0	𠄎		

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
1885	逢	4	9540	882D	逢		
1893	袞	1	9548	886E	袞		袞，異體字問題。
1937	詢	2	9574	8A3D	詢		
2043	蹕	2	9641	8E01	蹕		
2449	雒	3	98BF			名 ㄩ 隹	雒，錯字，以程式全部取代。
2478	鞏	6	98DC	978C	鞏		
2479	鞞	1	98DD	979B	鞞		
2651	鸚	9	99EC	9E1C	鸚		
2836	三	5	9B49			三	
2837	三	14	9B4A			三	
2878	叟	3	9B73	5081	叟		
2906	竽	1	9BB1	659A	竽		竽，異體字問題。
3068	三	8	9CB6			三	
3069	三	10	9CB7			三	
3071	三	10	9CB9			三	
3072	三	9	9CBA			三	
3073	三	7	9CBB			三	

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
3074	彡	17	9CBC			彡	
3770	个	4	8141	4E2A	个		
3780	儻	5	814B	5101	儻		
3787	冢	6	8152	51A2	冢		
3791	况	5	8156	51B5	况		况，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		
3921	皙	23	81FA	6673	皙		
3929	犁	2	8243	68C3	犁		
3930	憵	2	8244	6901	憵		
3935	槩	7	8249	69E9	槩		
3949	泣	29	8257	6D96	泣		
3974	牀	5	8270	7240	牀		
3976	犁	12	8272	7282	犁		部份使用不同字形「犁」，手動修改。
3987	穀	3	827D	7474	穀		

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
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	皞	8	82B7	76A1	皞		皞，異體字問題。
4030	秆	1	82CA	79C6	秆		
4038	竈	17	82D2	7AC8	竈		
4080	羣	96	82FC	7FA3	羣		
4085	耻	1	8342	803B	耻		
4091	皐	18	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			血𠂇𠂇	眾，異體字問題。
4134	衛	777	8373	885E	衛		衛，異體字問題。
4142	褻	1	837B	8943	褻		褻，異體字問題。
4145	羈	5	837E	898A	羈		羈，異體字

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
							問題。
4156	諂	1	83AB	8B1F	諂		
4172	屜	2	83BB	8E77	屜		
4184	迹	3	83C7	8FF9	迹		
4185	邊	1	83C8	9037	邊		
4224	頤	2	83EF	983F	頤		
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		
4242	鬪	58	8442	9B2D	鬪		
4250	鮓	6	844A	9C0C	鮓		
4252	鰓	3	844C			魚𩺰眾	鰓，異體字問題。
4259	麇	1	8453	9E95	麇		
4265	鼃	1	8459	9F03	鼃		
4307	罵	1	84A5	99E1	罵		罵，異體字問題。
4308	劫	5	84A6	5227	劫		劫，異體字

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
							問題。
4312	皞	1	84AA	66AD	皞		皞，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		
4507	吳	419	85D0	5434	吳		吳，異體字問題。
4598	鞞	2	866C	292FE		革△韋	鞞，異體字問題。
4753		3	876A			尗△比△大	尗△大，錯字，以程式全部取代。
4762	旃	1	8773	3ACB		舛△冉	旃，異體字問題。
4768	妬	1	8779	59AC	妬		
4822	尗	1	87D1	5C29	尗		尗，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。



編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
5035		3	894A			具 	郟，錯字，以程式全部取代。
5085	芊	4	897C	26B12		艹  艸	芊，錯字。部份使用不同字形「芊」，手動修改。
5156	𠂔	14	89E5	353B		𠂔	
5248		1	8AA4			扌  間	擱，異體字問題。
5290	携	8	8ACE	643A	携		攜，異體字問題。
5299	𠂔	2	8AD7	39B8		卓  戈	𠂔，異體字問題。
5302	昏	1	8ADA	662C	昏		昏，異體字問題。
5310	烏	1	8AE2	8204	烏		
5321		1	8AED			圭  隹	隼，錯字，以程式全部取代。
5404	鴟	1	8BA3	4CED		即  鳥	
5456		1	8BD7			韋  未	韋，異體字問題。

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
5505	耆	1	8C49	264BE		老厶占	考厶口，錯字，以程式全部取代。
5622	邛	5	8CE0	48B5		云厶卩	
5652	蝨	2	8CFE	27312		或厶虫	
5655	躑	3	8D42	56BB	躑		躑，異體字問題。
5705	廐	5	8D74	5ED0	廐		廐，異體字問題。
5713	躡	2	8D7C	481E		足厶戚	
5752	揔	1	8DC5	63D4	揔		

## 二、手動取代表說明：

《左傳》檔案中之舊版造字，因字形部份錯誤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：

- (一) 編號：Big5 造字空間為 6217 個，編號由 1 到 6217。
- (二) 檔案原文摘錄：摘錄檔案《左傳》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。
- (三) 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。
- (四) 備註：記錄修改相同辭句的次數或其他事項。

附件二、《左傳》手動取代表

編號	檔案原文摘錄	手動取代結果	備註
3976	秋邠 <u>島</u> 來來朝(全部取代)	秋邠 <u>犁</u> 來來朝(全部取代)	有 2 處
	伯州 <u>島</u> 奔楚(全部取代)	伯州 <u>犁</u> 奔楚(全部取代)	有 5 處
	使沈尹壽與師祁 <u>島</u> 讓之	使沈尹壽與師祁 <u>犁</u> 讓之	
	晉人執宋行人樂祁 <u>島</u>	晉人執宋行人樂祁 <u>犁</u>	
4010	公曰行也·及 <u>当</u> 鮑	公曰行也·及 <u>皐</u> 鮑	
4091	以肥子 <u>肩</u> 歸	以肥子 <u>皋</u> 歸	
4237	之師于雞父·胡子 <u>差</u> (全部取代)	之師于雞父·胡子 <u>髡</u> (全部取代)	有 2 處
	宋樂 <u>差</u> 帥師伐曹	宋樂 <u>髡</u> 帥師伐曹	
5085	國民信之· <u>丕</u> 姓有亂	國民信之· <u>半</u> 姓有亂	