

《荀子》舊版造字轉碼說明

中研院資訊所文獻處理實驗室

中研院史語所漢籍電子文獻工作小組

2008/1/4 陳建安 製作

一、《荀子》一書(荀子.xml)使用舊版造字 72 個，字頻 206 次，詳如附件一。這 72 個造字中，61 個可轉成 Windows XP 能顯示的字，字頻 177 次；另外 11 個字必須轉成構字式，字頻 29 次。

二、附件一的造字分析表說明如下：

甲、編號：Big5 造字空間為 6217 個，編號由 1 到 6217。

乙、造字：舊版造字

丙、字頻(txt)：造字在「.txt」文件的出現次數

丁、字頻(xml)：造字在「.xml」文件的出現次數

戊、Big5：造字的 Big5 碼

己、Unicode：造字所對應的 Unicode 碼

庚、WinXP：造字在 Windows XP 的對應字形

辛、備註凡例：

- 1、校對問題，舊版漢籍錯字，可用程式全部取代：在舊版漢籍電子文獻中即存在的錯字，因校對時的疏漏而未更正，持續留存在新版漢籍電子文獻中；若該造字的所有頻次，皆屬於錯誤使用的錯字情形，可以用程式全部取代為正確字形。如編號 1836 的「口𠂔丹」字，原字為「口𠂔丹」。
- 2、Unicode 字型呈現差異：Unicode 字型與舊漢籍造字有些微差異，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。
- 3、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 5749 的「片𠂔庸」。

三、Unicode 目前收錄的漢字總數為 70194，分屬於三個不同區段，詳如表一。目前 Windows XP 只支援 CJK 認同表意文字區的 20902 個字，內碼為 4E00-9FFF。所以造字編號 1339 的「儻」字，Unicode 編碼為 34A9，由於 Windows XP 並不支援，仍須使用構字式「亻𠂔蠡」。

表一、Unicode 的字數及編碼區段

Unicode	字集子集合	新增字數	新增編碼區段	總字數	WinXP
1.1 版	CJK 認同表意文字區	20902	4E00-9FFF	20902	支援
3.0 版	CJK 認同表意文字擴充 A 區	6582	3400-4DFF	27484	不支援
3.1 版	CJK 認同表意文字擴充 B 區	42710	20000-2A6D6	70194	不支援

四、《荀子》中未確定字形「●」共 87 處，詳細處理內容可參考附件二的未確定字形手動取代表。附件二的手動取代表說明如下：

- 甲、檔案原文摘錄：摘錄檔案(荀子.xml)中包含未確定字形「●」的文句段落，其中紅字底線的部分，為需要手動取代的字形。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。
- 乙、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。
- 丙、備註：註明相同文句在檔案中出現的次數，未註明者表示只出現一次。

附件一、《荀子》造字分析表

編號	造字	頻次 (txt)	頻次 (xml)	Big5	Unicode	WinXP	構字式	備註
813	𪛗	1	1	8E5B	65E4	𪛗		
972	𪛘	5	5	8F5D	7287	𪛘		
1039	𪛙	1	1	8FC2	69C0	𪛙		
1059	𪛚	1	1	8FD6	51A3	𪛚		
1077	𪛛	1	1	8FE8	7A3E	𪛛		

編號	造字	頻次 (txt)	頻次 (xml)	Big5	Unicode	WinXP	構字式	備註
1339	儼	2	2	91B4	34A9		彳 𠂔 𧈧	
1565	繇	2	2	92F9	7DDC	繇		
1836		2	2	94CE			口 𠂔 丹	口 𠂔 丹，校對問題，舊版漢籍錯字，可用程式全部取代。
1885	逢	4	4	9540	882D	逢		
1937	詢	1	1	9574	8A3D	詢		
1981	讐	3	3	95C2	8B90	讐		
2146	蛄	1	1	96CA	8739	蛄		
3039	覩	1	1	9C77	89A9	覩		
3061	僂	1	1	9CAF	5044	僂		
3799	刼	1	1	815E	5226	刼		
3831	詰	1	1	817E	5586	詰		
3835	噍	2	2	81A4	564D	噍		
3855	擘	2	2	81B8	5B7C	擘		
3883	徧	10	10	81D4	5FA7	徧		
3888	惠	1	1	81D9	60B3	惠		
3924	朞	1	1	81FD	671E	朞		

編號	造字	頻次 (txt)	頻次 (xml)	Big5	Unicode	WinXP	構字式	備註
3930	攄	4	4	8244	6901	攄		
3935	槩	2	2	8249	69E9	槩		Unicode 呈現差異。
3937	檝	1	1	824B	6A9D	檝		
3951	湏	1	1	8259	6E4F	湏		
3974	牀	6	6	8270	7240	牀		
3978	猪	2	2	8274	732A	猪		Unicode 呈現差異。
3981	珉	3	3	8277	73C9	珉		
4014	蓋	1	1	82BA	8462	蓋		
4040	竝	12	12	82D4	7ADD	竝		
4103	茝	1	1	8354	831D	茝		
4104	莅	1	1	8355	8385	莅		
4124	螽	3	3	8369	8771	螽		
4146	覩	2	2	83A1	89A9	覩		
4156	諂	1	1	83AB	8B1F	諂		
4172	屢	4	4	83BB	8E77	屢		Unicode 呈現差異。
4184	迹	4	4	83C7	8FF9	迹		
4190	邾	1	1	83CD	90C4	邾		

編號	造字	頻次 (txt)	頻次 (xml)	Big5	Unicode	WinXP	構字式	備註
4222	顙	1	1	83ED	9847	顙		
4226	𩚑	1	1	83F1	98E1	𩚑		
4250	鰭	7	7	844A	9C0C	鰭		
4256	鷄	3	3	8450	9DC4	鷄		
4265	鼃	1	1	8459	9F03	鼃		
4267	鼈	8	8	845B	9F08	鼈		
4287	纖	1	1	846F	7E8E	纖		
4312	皞	1	1	84AA	66AD	皞		
4342	倮	1	1	84C8	502E	倮		
4458	勸	4	4	857D	52CC	勸		Unicode 呈現差異。
4768	妬	8	8	8779	59AC	妬		
4855	汩	1	1	87F2	3CC5		汩 𣶒 不	
4996	斲	2	2	88E2	65B5	斲		
5034	弃	8	8	8949	5F03	弃		
5133	憊	1	1	89CE	6199	憊		
5169	𪔐	1	1	89F2	6239	𪔐		
5289	慤	1	1	8ACD	6164	慤		

編號	造字	頻次 (txt)	頻次 (xml)	Big5	Unicode	WinXP	構字式	備註
5299	戟	2	2	8AD7	39B8		卓𠂔戈	
5300	鬪	27	27	8AD8	9B2A	鬪		
5302	昏	1	1	8ADA	662C	昏		
5351	肱	1	1	8B4C	80AC	肱		
5355	𡗗	1	1	8B50	20AE4		𡗗	
5434	湟	1	1	8BC1	5E7C		形 汨 白 工 口	
5655	𪔐	1	1	8D42	56BB	𪔐		
5687	𧢲	9	9	8D62	4AC9		𧢲	
5689	𡗗	1	1	8D64	20087		𡗗	
5701		8	8	8D70				徧，校對問題，舊版漢籍錯字，可用程式全部取代。
5705	廐	1	1	8D74	5ED0	廐		Unicode 呈現差異。
5706	敍	2	2	8D75	654D	敍		
5717	陘	1	1	8DA2	967F	陘		
5722		1	1	8DA7				暮，校對問題，舊版漢籍錯字，可用程式全部取代。
5730	阨	4	4	8DAF	9638	阨		

編號	造字	頻次 (txt)	頻次 (xml)	Big5	Unicode	WinXP	構字式	備註
5749		1	1	8DC2			片厶庸	待造字。
5752	摠	1	1	8DC5	63D4	摠		

附件二、《荀子》「●」字形手動取代表

檔案原文摘錄	手動取代結果	備註
●、水為之	<u>冰</u> 、水為之	
雖有稿●	雖有稿 <u>暴</u>	
●●咎咎	<u>喻喻</u> 咎咎	
而不●，柔從	而不 <u>暴</u> ，柔從	
非驕●也	非驕 <u>暴</u> 也	
大心則慢而●	大心則慢而 <u>暴</u>	
乳●不遠遊	乳 <u>豸</u> 厶 <u>苟</u> 不遠遊	
曾●處之不若也	曾 <u>豸</u> 厶 <u>苟</u> 處之不若也	
有●處之勇者	有 <u>豸</u> 厶 <u>苟</u> 處之勇者	
●●然惟利飲食之見	<u>恡恡</u> 然惟利飲食之見	
是●處之勇也	是 <u>豸</u> 厶 <u>苟</u> 處之勇也	
●●然惟利之見	<u>恡恡</u> 然惟利之見	
條●者	條 <u>魚</u> 厶 <u>本</u> 者	
惕悍憍●	惕悍憍 <u>暴</u>	
方知畜鷄●猪處	方知畜鷄 <u>豸</u> 厶 <u>苟</u> 猪處	
●唯則節	<u>瞻</u> 唯則節	
而忍●詢	而忍 <u>言</u> 厶 <u>隹</u> 詢	
而●國安自化矣	而 <u>暴</u> 國安自化矣	
可炊而●也	可炊而 <u>僇</u> 也	
以弟誅兄而非●也	以弟誅兄而非 <u>暴</u> 也	
●回天下於	<u>口</u> 厶 <u>月</u> 回天下於	
●●兮其能長久	<u>猷猷</u> 兮其能長久	
制彊●，則非大儒也	制彊 <u>暴</u> ，則非大儒也	
答桎●國	答桎 <u>暴</u> 國	
舉其上客，●然若	舉其上客， <u>彳</u> 厶 <u>患</u> 然若	
無所擬●	無所擬 <u>作</u> 厶 <u>心</u>	

檔案原文摘錄	手動取代結果	備註
衛弱禁●	衛弱禁暴	
天下脅於●國	天下脅於暴國	
以觀夫●國	以觀夫暴國	
●國之君	暴國之君	
可以誅●禁悍矣	可以誅暴禁悍矣	
彼將日日●露	彼將日日暴露	
立身則憍●	立身則憍暴	
不足以禁●勝悍	不足以禁暴勝悍	
●●如丘山	暴暴如丘山	
事強●之國難（全部取代）	事強暴之國難（全部取代）	2
使強●之國事我易（全部取代）	使強暴之國事我易（全部取代）	2
名聲足以●炙之	名聲足以暴炙之	
而強●之國	而強暴之國	
百姓賤之如●	百姓賤之如仁 厶 匡	
●然而齒墮矣	輝然而齒墮矣	
事●君者	事暴君者	
窮居於●國	窮居於暴國	
是事●君之義也	是事暴君之義也	
不敢●虎	不敢暴虎	
且夫●國之君	且夫暴國之君	
冠●帶劍	冠革 厶 由帶劍	
拱挹指麾，而彊●	拱挹指麾，而彊暴	
彼兵者所以禁●除害也	彼兵者所以禁暴除害也	
宛鉅鐵●	宛鉅鐵鉞	
●悍勇力之屬	暴悍勇力之屬	
有●察之威者	有暴察之威者	
然而其禁●也察	然而其禁暴也察	
夫是之謂●察之威	夫是之謂暴察之威	
●察之威成乎危弱	暴察之威成乎危弱	
然而●國獨侈	然而暴國獨侈	
誅●國之君	誅暴國之君	
●屨，殺、赭衣	紕 厶 封屨，殺、赭衣	
禁●惡惡	禁暴惡惡	
是謂惠●而寬賊也	是謂惠暴而寬賊也	
夫征●誅悍	夫征暴誅悍	
故盜不●	故盜不竊	

檔案原文摘錄	手動取代結果	備註
●貪利，是辱	暴貪利，是辱	
舌●，是辱之由外	舌絲△舉，是辱之由外	
●慢恣孽輕俗	暴慢恣孽輕俗	
●笑、哭泣	漱笑、哭泣	
怒而●亂畏之	怒而暴亂畏之	
帶甲嬰●	帶甲嬰革△由	
●杙、拊臠	靴杙、拊臠	
眾積意●●乎	眾積意言△犀言△犀乎	
<qa>尚</qa>●机	<qa>尚</qa>崩机	
而歸於●	而歸於暴	
眾者●寡而譁之	眾者暴寡而譁之	
不得排●	不得排木△敬	
莊君之●	莊君之勿△口	
充盈乎大●	充盈乎大△△禹	
●泮殺內	氷泮殺內	
皋如也，●如也	皋如也，嶼如也	
藜羹不●	藜羹不糲	