

《晏子春秋》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

2008/1/11 陳建安 製作

一、《晏子春秋》一書(晏子春秋.xml)使用舊版造字 46 個，字頻 156 次，詳如附件一。這 46 個造字中，39 個可轉成 Windows XP 能顯示的字，字頻 119 次；另外 7 個字必須轉成構字式，字頻 37 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、異體字問題：新版漢籍考量到使用者檢索及使用時的便利性，將用字原則改為除專詞等特殊情形之外，一律改用標準字呈現。如編號 4132 的「血𩇑𩇑」係「眾」字之異體，故以「眾」字取代。

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 4043 的「筭」字，Unicode 字型呈現為「筭」，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 4532 的「口𩇑𩇑」。

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 4855 號的「汩」字，Unicode 編碼為 3CC5 由於 Windows XP 並不支援，仍須使用構字式「汩𩇑不」。

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、《晏子春秋》中未確定字形「●」共 21 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(晏子春秋.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，不一一列舉所有修改後結果。
- 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、《晏子春秋》造字分析表

編號	造字	頻次 (txt)	頻次 (xml)	Big5	Unicode	WinXP	構字式	備註
842	𣎵	1	1	8E78	3AB7		𣎵	
852	猻	1	1	8EA4	72E2	猻		
957	遨	1	1	8F4E	906C	遨		
972	犇	5	5	8F5D	7287	犇		
994	𪗇	7	7	8F73	8E75	𪗇		
1039	𪗈	3	3	8FC2	69C0	𪗈		

編號	造字	頻次 (txt)	頻次 (xml)	Big5	Unicode	WinXP	構字式	備註
1668	臍	1	1	93C3	81C8	臍		
1871	蟲	2	2	94F1	87C1	蟲		
1981	讐	2	2	95C2	8B90	讐		
3858	寃	3	3	81BB	5BC3	寃		
3883	徧	3	3	81D4	5FA7	徧		
3891	慙	9	9	81DC	6159	慙		
3921	皙	1	1	81FA	6673	皙		
3924	朥	1	1	81FD	671E	朥		
3930	攄	1	1	8244	6901	攄		
3994	畊	5	5	82A6	754A	畊		
4014	蓋	1	1	82BA	8462	蓋		
4029	褖	1	1	82C9	799D	褖		
4040	竝	6	6	82D4	7ADD	竝		
4043	筭	1	1	82D7	7B53	筭		Unicode 呈現差異。
4077	罇	4	4	82F9	7F47	罇		
4080	羣	20	20	82FC	7FA3	羣		Unicode 呈現差異。
4104	莅	13	13	8355	8385	莅		

編號	造字	頻次 (txt)	頻次 (xml)	Big5	Unicode	WinXP	構字式	備註
4132		28	28	8371			血𠂇𠂇	眾，異體字問題。
4139	袴	1	1	8378	88B4	袴		
4146	覩	1	1	83A1	89A9	覩		
4156	諂	2	2	83AB	8B1F	諂		
4172	屨	1	1	83BB	8E77	屨		Unicode 呈現差異。
4227	飪	1	1	83F2	98E6	飪		
4239	髯	2	2	83FE	9AE5	髯		
4242	鬪	6	6	8442	9B2D	鬪		
4252	鰈	2	2	844C			魚𩺰眾	鰈，異體字問題。
4267	鼈	1	1	845B	9F08	鼈		
4342	倮	1	1	84C8	502E	倮		
4455	劒	1	1	857A			僉𠂇刀	
4460	勅	4	4	85A1	52D1	勅		
4532		1	1	85E9			口𠂇𠂇	待造字。
4855	汩	2	2	87F2	3CC5		汩𠂇不	
5169	𪔐	1	1	89F2	6239	𪔐		
5310	烏	1	1	8AE2	8204	烏		

編號	造字	頻次 (txt)	頻次 (xml)	Big5	Unicode	WinXP	構字式	備註
5437	并	2	2	8BC4	5E77	并		
5447	躡	1	1	8BCE	8E83	躡		
5572	埠	1	1	8CAE	57BE	埠		
5588	恹	1	1	8CBE	6061	恹		
5666	汚	1	1	8D4D	6C5A	汚		
5739		2	2	8DB8			冗𡗗古	𡗗，異體字問題。

附件二、《晏子春秋》「●」字形手動取代表

檔案原文摘錄	手動取代結果	備註
若何●●去此而死乎	若何𡗗𡗗去此而死乎	
孽星在●	孽星在𡗗	
衣不務于隅●之削	衣不務于隅𡗗之削	
涕●而去	涕𡗗而去	
接一搏●而再	接一搏𡗗而再	
其●久乎	其𡗗久乎	
睹●冠反裘負芻	睹𡗗冠反裘負芻	
臣受命●邑之君	臣受命𡗗邑之君	
若夫●車駕馬以朝	若夫𡗗車駕馬以朝	
力●矣	力𡗗矣	
民之死近矣。●其	民之死近矣。𡗗其	
衣食●薄（全部取代）	衣食𡗗薄（全部取代）	2
●車駕馬（全部取代）	𡗗車駕馬（全部取代）	3
乘●車，駕駕馬	乘𡗗車，駕駕馬	
景公●于路寢之宮	景公𡗗于路寢之宮	
身老，子●	身老，子𡗗	
國門外宇●之下	國門外宇𡗗之下	