## 《顧亭林詩文集》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 2008/4/10 黃榮順 製作

- 一、《顧亭林詩文集》一書(顧亭林詩文集.xml)使用舊版造字 109 個,字頻 331 次,詳如附件一。這 109 個造字中,83 個可轉成 Windows XP 可顯示的字,字頻 235 次;另外 26 個字必須轉成構字式,字頻 96 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、構字式: Windows XP 若無對應字形,則改採用構字式

壬、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 681 的「厂」字,原書字形為「厂」。
- 2、標記問題,新版漢籍新增一標題句:標記問題主要是漢籍電子文獻在舊轉新的過程中,採用了新的標記語言,而在修改標記時,對於某些標題句做了增刪的動作,因為這些標題句的內容包含了缺字,所以這些更改標記的動作造成 txt 檔與 xml 檔缺字頻次不符的情形。如編號4546 的「鵶」,新版漢籍新增一標題句,致使「鵶」字在 xml 檔增加 1 個頻次,由於頻次和 txt 檔不同,必須手動將 xml 檔的頻次都取代為 Unicode 的「鵶」。
- 3、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982的「胷」係「胸」字之異體,故以「胸」字取代。

- 4、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 3846 的「壻」字, Unicode 字型呈現為「壻」,實際上仍為同一字。
- 5、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 2937 的「監会立」。
- 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。所以如造字編號 865 的「蛐」字,Unicode編碼為 27358,由於 Windows XP 並不支援,仍須改用構字式「虫 Ш鱼」。

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、 txt 檔與 xml 檔頻次不同且造字分析表「備註」欄有備註問題情 況者,詳細取代內容可參考附件二的手動取代表。附件二的手動 取代表的欄位說明如下:
  - 甲、編號:Big5 造字空間為 6217 個,編號由 1 到 6217。
  - 乙、造字:舊漢籍造字。
  - 丙、檔案原文摘錄:摘錄檔案(顧亭林詩文集.xml)中包含該造字 的文句段落,其中紅字底線的部分,為需要手動取代的原文。
  - 丁、 手動取代結果:變更過後的檔案內容,其中<u>藍字底線</u>的部分,為修改後的結果。
- 五、「顧亭林詩文集.xml」發現之未確定字形「●」共93處,修正情 形詳列附件三,附件三「顧亭林詩文集.xml」「●」修改表體例說 明如下:
  - 甲、檔案原文摘錄:摘錄檔案(顧亭林詩文集.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

取代的原文。

乙、手動取代結果:變更過後的檔案內容,其中<u>藍字底線</u>的部分,為修改後的結果。

丙、 備註:修改後的字形若屬於「待造字」,則於此欄標示。

## 附件一、「顧亭林詩文集. xml」造字分析表

編號	造字		字頻 (xml)	Big5	Unicode	WinXP	構字式	備註
74	髩	1	1	FAAB	29B06		髟会丐	髩 ,異體字問題。
296		9	9	FBEC				
297	M	21	21	FBED			Δ	
298		9	9	FBEE				
299	形	7	7	FBEF			形	
300	$\odot$	7	7	FBF0			·	
314	7	1	1	FBFE			7	
676	乒	1	1	FE6F			产	
681	Γ	8	8	FE74			٢	厂 , 校對問題 , 舊版漢籍 錯字 , 可用程式全部取代。
758	回	1	1	FEE3			Ш	
759	广	1	1	FEE4			ŗ	

編號	造字		字頻 (xml)	Big5	Unicode	WinXP	構字式	備註
775	۲	4	4	FEF4			٢	
865	幽	1	1	8EB1	27358		虫∭幽	
899	駞	2	2	8ED3	99DE	馬		
921	碝	3	3	8EE9	789D	碝		
969	牐	5	5	8F5A	7250	牐		
971	崩	3	3	8F5C	48D9		崩▲阝	
982	<b></b> 国月	1	1	8F67	80F7	匈月		胸,異體字問題。
1077	高米	1	1	8FE8	7A3E	禀		
1153	抝	2	2	9075	362D		土瓜幻	
1237	矟	1	1	90EB	77DF	稍		
1555	緍	1	1	92EF	7DCD	緍		
1565	緜	4	4	92F9	7DDC	緜		
1928	职	1	1	956B	25133		<sup>四</sup> <b>会</b> 氺	
1937	韵	1	1	9574	8A3D	韵		
1944	砮	1	1	957B	8A67	詧		
2350	傪	1	1	97F9	50AA	傪		

編號	造字		字頻 (xml)	Big5	Unicode	WinXP	構字式	備註
2567	稡	2	2	9976	7A21	稡		
2733	犀	1	1	9AA1	21C5D		尸▲羊	
2937		2	2	9BD0			₽₩☆立	豎,異體字問題。
3801	剗	1	1	8160	5257	剗		
3802	劚)	1	1	8161	529A	劚		
3805	勅	5	5	8164	52C5	勅		
3809	匳	1	1	8168	5333	匳		
3816	厓	10	10	816F	5393	厓		
3826	叶	11	11	8179	53F6	叶		
3846	壻	1	1	81AF	58FB	壻		Unicode 字型差異。
3852	婣	2	2	81B5	5A63	婣		
3860	尠	1	1	81BD	5C20	尠		
3869	币	1	1	81C6	5E00	币		
3883	徧	20	20	81D4	5FA7	徧		
3901	捓	2	2	81E6	6353	捓		
3931	椀	1	1	8245	6900	椀		

編號	造字		字頻 (xml)	Big5	Unicode	WinXP	構字式	備註
3935	既	5	5	8249	69E9	旣		概,異體字問題。
3937	楫	3	3	824B	6A9D	檝		
3977	狗	6	6	8273	72E5	狗		
3990	甎	1	1	82A2	750E	甎		
3996	疎	11	11	82A8	758E	疎		
4013	夗	2	2	82B9	76CC	盌		
4025	礴	1	1	82C5	7934	礴		
4034	稺	2	2	82CE	7A49	穉		
4038	空黽	2	2	82D2	7AC8	空電		
4040	竝	3	3	82D4	7ADD	竝		
4077	罇	1	1	82F9	7F47	罇		
4089	脇	1	1	8346	8107	脇		
4115	慧	8	8	8360	85C1	蒙泉		
4129	虬	2	2	836E	866C	虬		
4139	袴	1	1	8378	88B4	袴		
4140	袵	1	1	8379	88B5	袵		

編號	造字		字頻 (xml)	Big5	Unicode	WinXP	構字式	備註
4146	覩	7	7	83A1	89A9	覩		
4158	譌	4	4	83AD	8B4C	譌		Unicode 字型差異。
4184	迹	13	13	83C7	8FF9	迹		跡,異體字問題。
4186	遡	3	3	83C9	9061	遡		
4188	邨	2	2	83CB	90A8	邨		
4209	鞉	1	1	83E0	9789	鞉		
4213	華	3	3	83E4	97BE	鞾		
4222	顇	1	1	83ED	9847	顇		
4225	顦	1	1	83F0	9866	顦		
4232	隲	1	1	83F7	96B2	隲		
4240	影弗	3	3	8440	9AF4	影		
4242	斷	7	7	8442	9B2D	斸		
4251	鰕	1	1	844B	9C15	鰕		
4259	麛	2	2	8453	9E95	鹰		Unicode 字型差異。
4265	圭黽	1	1	8459	9F03	畫黽		
4266	唱	1	1	845A	9F02	旦黽		

編號	造字		字頻 (xml)	Big5	Unicode	WinXP	構字式	備註
4267	鼈	2	2	845B	9F08	<b>幣</b>		
4342	倮	1	1	84C8	502E	倮		
4371	僴	1	1	84E5	50F4	僴		
4378	陟馬	1	1	84EC	28F65		陟▲馬	鷹,異體字問題。
4546	鵐	2	3	85F7	9D76	鵶		<b>鵶</b> ,標記問題。
4575	縕	1	1	8655	7E7F	繿		
4627	礟	3	3	86AB	791F	礟		
4696	㸌	1	1	86F0	3E0C		火▲霍	
4735	壠	2	2	8758	58E0	壠		
4816	1	2	2	87CB			1	
4925	輔	1	1	8879	8F65	轥		
4996	斵	2	2	88E2	65B5	斵		
5013	·~	4	4	88F3			ì	
5038	儿口	1	1	894D	3563		凸	
5169	戹	3	3	89F2	6239	戹		
5292	甫方	2	2	8AD0	65C9	甫方		

編號	造字		字頻 (xml)	Big5	Unicode	WinXP	構字式	備註
5302	昬	5	5	8ADA	662C	昬		
5344	嗸	2	2	8B45	55F8	嗸		
5361	隝	1	1	8B56	969D	隝		
5439	栢	2	2	8BC6	6822	栢		
5455	黎	2	2	8BD6	48DB		泰▲阝	
5464	煗	1	1	8BDF	7157	煗		
5509	敤	2	2	8C4D	9A00	騀		
5510	駊	2	2	8C4E	99CA	胰		
5549	芸品系	4	4	8C75	85F3	营禾		
5555	逈	2	2	8C7B	9008	逈		
5584	愞	2	2	8CBA	611E	愞		
5591	杷	2	2	8CC1	233CC		木▲□	杞,異體字問題。
5678	画町	1	1	8D59	757A	画田		
5695	幐	3	3	8D6A	3B3A		朕▲巾	
5705	廐	1	1	8D74	5ED0	廐		<b>殿</b> ,異體字問題。
5730	院	4	4	8DAF	9638	阸		

編號	造字	字頻 (txt)	字頻 (xml)	Big5	Unicode	WinXP	構字式	備註
5764	築	3	3	8DD1	7C5E	築		
5772	濙	1	1	8DD9	6FD9	濙		

## 附件二、「顧亭林詩文集. xml」手動取代表

編號	造字	檔案原文摘錄	手動取代結果
		東二十里有西島	東二十里有西 <u>鵶</u>
4546	鵶	東二十里有西島	東二十里有西 <u>鵶</u>
		前月西 <u>島</u> 禁不開	前月西 <u>鵶</u> 禁不開

## 附件三、「顧亭林詩文集. xml」「●」修改表

檔案原文摘錄	手動取代結果	備註
●江故家得手帖	<u>吳</u> 江故家得手帖	
其元帙舛	其元帙舛 <u>譌</u>	
病弟●孫懸望殊切	病弟 <u>穉</u> 孫懸望殊切	
以充僮	以充僮 <u>竪</u>	
我則終於韞●而已	我則終於韞 <u></u> 而已	
為大吏添一●穴也	為大吏添一 <u>鑛</u> 穴也	

粗具草●	粗具草藁	
史●之成	史藁之成	
不敢不●縷以報也	不敢不覷縷以報也	
●天下都是我去依人	<u>编</u> 天下都是我去依人	
幾●詞壇	幾編詞壇	
敵難●攻	敵難 <u>徧</u> 攻	
前赴●右	前赴 <u>壠</u> 右	
指日 之	指日 <u>竢</u> 之	
上●國課	上 <u>誤</u> 國課	
已具一	已具一 <u>藁</u>	
又可識●事之精敏矣	又可識 <u>蒞</u> 事之精敏矣	
輒敢 <u>●</u> 縷	輒敢 <u>覼</u> 縷	
寄●他氏者乎	寄餐他氏者乎	
寸●無遺	寸 <u>畝</u> 無遺	
凡我●友	凡我 <u>姻</u> 友	
所與居者●舊鄉曲也	所與居者 <u>姻</u> 舊鄉曲也	
未得會●成帙	未得會秤成帙	
靡所●止	靡所 <u>厂▲氏</u> 止	
伊于胡	伊于胡 <u>厂▲氏</u>	
皆是●字	皆是 <u>厂▲氏</u> 字	
本●字有下一畫	本 <u>厂▲氏</u> 字有下一畫	
●音同而古亦可通	<u>厂▲氏</u> 音同而古亦可通	
自惟●劣	自惟 <u>謭</u> 劣	
臨楮不盡瞻●	臨楮不盡瞻 <u>遡</u>	
●風即路	<u>遡</u> 風即路	
伏惟矜●	伏惟矜 <u>詧</u>	
得拜●函	得拜 <u>珤</u> 函	
●其素心	<u>餐</u> 其素心	
弟以 <u>●</u> 陋	弟以 <u>謭</u> 陋	
引領●之矣	引領 <u>埃</u> 之矣	
沈天甫詐騙●中翰	沈天甫詐騙 <u>吳</u> 中翰	
天●其衷	天 <u>片</u>	待造字
但垂●馬之慈	但垂掛馬之慈	
期以秋杪往●其事	期以秋杪往 <u>蒞</u> 其事	
足徵政事文章大●	足徵政事文章大概	
奸人構●	奸人構 <u>旤</u>	

●洄之思	遡洄之思	
已焚●矣	已焚 <u>藁</u> 矣 兰阳楼玉料工	
益耽情于●石	益耽情于 <u>漱</u> 石	
●信于同人	<u>编</u> 信于同人	
不能●呈	不能 <u>编</u> 呈	
物緣漸●	物緣漸 <u>疏</u>	
及前朝旌表之●	及前朝旌表之 <u>概</u>	
江湖沸●	江湖沸 <u>洶</u> 	
因●字即筌	因 <u>罤</u> 字即筌	
終是●人之學	終是 <u>狗</u> 人之學	
	一瞻丘 <u>壠</u>	
每●不過四五金	每編不過四五金	
終	終 <u>无</u> 咎也	
趙兄名●來一書	趙兄名 <u>艺</u> 來一書	
志●為其兄乞文	志 <u>**会洍</u> 為其兄乞文	待造字
而●●數十百舟艤	而 <u>僅僅</u> 數十百舟艤	
●贈承德郎	<u>敕</u> 贈承德郎	
洽不知可通●否	洽不知可通 <u>叶</u> 否	
文 二首呈正	文 <u>黨</u> 二首呈正	
●好其聲	<u></u> 最好其聲	
●● 乎其泓	<u>戛戛</u> 乎其泓	
●兮害其有平	<u>鬱</u> 兮害其有平	
并審出●起章丘地	并審出鲜起章丘地	
粗得古人之陳●	粗得古人之陳跡	
復●酉山之祕	復睹酉山之祕	
●中無可發揮處	胸中無可發揮處	
京兆尹●監	京兆尹 <u>勑</u> 監	
殆非也。●者	殆非也。 <u>勑</u> 者	
皆日●,亦作敕	皆曰 <u>勑</u> ,亦作敕	
何●之遣武吏	何並之遣武吏	
 魯相韓 <mark>●</mark> 修孔子廟		
人臣不當名●	人臣不當名敕	
以為●音賚	以為敕音賚	
後人借為●字	後人借為勑字	
<ul><li>●我五典五惇哉</li></ul>	<b>勃我五典五</b> 惇哉	
●天之命	<b>勃天之命</b>	
<u>'</u>	<u> </u>	l .

來,也	來, <u>勑</u> 也	
●爾瞽,率爾眾	<u>勅</u> 爾瞽,率爾眾	
何嘗不知 為敕哉	何嘗不知 <u>勑</u> 為敕哉	
而言京兆尹●遣之	而言京兆尹 <u>勑</u> 遣之	
夫誰●之	夫誰 <u>尼</u> 之	
將覽前古●	將覽前古 <u>跡</u>	
劉一●、韓爌	劉一燝、韓爌	
未盡厥●	未盡厥 <u>辠</u>	
所共側目之 <u>●</u>	所共側目之 <u>郄</u>	
史繼偕、沈●	史繼偕、沈 <u>邢》一</u> 佳。	待造字
且卿等侍先帝●几	且卿等侍先帝 <u>憑</u> 几	
親●自有分別	親 <u>疏</u> 自有分別	
後之傳者日遠日●	後之傳者日遠日 <u>譌</u>	