

《爾雅》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
2011/02/11 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 4867 的「𪛗」字，Unicode 編碼是 4D79，由於 XP 並不支援，仍需使用構字式「𪛗知于𪛗」。

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：

《爾雅》使用舊版造字 86 個，字頻 156 次，這 86 個造字中，78 個可轉成 Windows XP 能顯示的字，字頻 143 次；另外 8 個字必須轉成構字式，字頻 13 次。

轉碼完成製作轉碼分析表，請參考附件一《爾雅》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，除非是專有名詞或特殊情形，如：人名、地名等。例：造字編號 4132 的「血𠂇𠂇」字，是「眾」的異體字，以標準字「眾」取代。
 2. 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 5739 的「冗𠂇古」字，而原書使用之字為「罟」，查詢教育部異體字字典確認「冗𠂇古」並非「罟」之異體字，因此歸為錯字，以程式取代為正確字形「罟」。
 3. Unicode 字型呈現差異：Unicode 字型與舊版造字有些微差異，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 4091 的「臯」字，Unicode 字型呈現為「臯」，實際上仍為同一字。

附件一、《爾雅》轉碼分析表

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
800	逕	1	8E4E	905D	逕		
886	攝	3	8EC6	6B07	攝		
915	遲	3	8EE3	905F	遲		
1002	鳩	1	8F7B	9D02	鳩		
1062		3	8FD9			馬𠂇亞	駢，異體字問

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
							題。
1416	埃	1	9242	7AE2	埃		
1565	絲	3	92F9	7DDC	絲		
1815	藻	2	94B9	85BB	藻		
1837	蘼	1	94CF		蘼		
1871	蟲	1	94F1	87C1	蟲		
1885	逢	3	9540	882D	逢		
1893	袞	1	9548	886E	袞		
1981	讐	1	95C2	8B90	讐		
2424	藥	4	98A6	6909	藥		
2616	鱸	1	99C9	9C6F	鱸		
2762	碩	1	9ABE	78D2	碩		
2865	蕓	1	9B66	8608	蕓		
2866	藹	1	9B67	8649	藹		
2870	績	1	9B6B	8C44	績		
2903	启	1	9BAE	542F	启		
2965	畫	1	9BEC	87D7	畫		
2991	麋	1	9C47			鹿△各	

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
3787	冢	3	8152	51A2	冢		
3816	厓	6	816F	5393	厓		
3846	壻	5	81AF	58FB	壻		
3855	孽	2	81B8	5B7C	孽		
3866	崐	2	81C3	5D10	崐		
3883	徧	1	81D4	5FA7	徧		
3891	慙	1	81DC	6159	慙		
3936	棹	1	824A	6A70	棹		Unicode 字型 呈現差異。
3949	涖	1	8257	6D96	涖		
3965	裁	1	8267	70D6	裁		
3974	牀	1	8270	7240	牀		
4018	矚	1	82BE	77AF	矚		
4040	竝	1	82D4	7ADD	竝		
4071	纓	1	82F3	7DF5	纓		
4091	臯	3	8348	81EF	臯		Unicode 字型 呈現差異。
4097	芑	1	834E			艹 厶 巳	

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
4102	苜	1	8353	82E2	苜		
4103	苣	1	8354	831D	苣		
4107	葱	2	8358	8471	葱		
4132		9	8371			血𠄎𠄎	眾，異體字問題。
4134	衛	2	8373	885E	衛		衛，異體字問題。
4150	訕	2	83A5	8A29	訕		
4156	諂	1	83AB	8B1F	諂		
4165	獾	1	83B4	8C9B	獾		
4172	屨	1	83BB	8E77	屨		
4180	臯	1	83C3	8FA0	臯		
4184	迹	1	83C7	8FF9	迹		
4185	邊	1	83C8	9037	邊		
4222	頰	1	83ED	9847	頰		
4237	髡	1	83FC	9AE0	髡		
4241	萑	1	8441	84B5	萑		

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
4243	𩺰	1	8443	9B34	𩺰		
4250	𩺱	1	844A	9C0C	𩺱		
4251	鰕	3	844B	9C15	鰕		
4252	鰓	1	844C			魚𩺱眾	鰓，異體字問題。
4253	鳧	10	844D	9CF8	鳧		
4256	鷄	1	8450	9DC4	鷄		
4259	麇	4	8453	9E95	麇		
4261	麇	1	8455	9EA2	麇		
4267	鼈	1	845B	9F08	鼈		
4268	鼈	1	845C	9F17	鼈		
4322	備	1	84B4	4FCC	備		
4353	菟	2	84D3	83AC	菟		
4371	倜	1	84E5	50F4	倜		
4395	霽	1	84FD	9721	霽		
4409	回	1	854C	518B	回		
4461	勗	1	85A2	52D4	勗		
4752	厯	2	8769	61D5	厯		

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
4866	鼃	5	87FD	9F04	鼃		
4867	鼃	4	87FE	4D79		形知于 黽	
4902	充	1	8862	5156	充		
4922	鷓	1	8876	9DC9	鷓		
4950	嶮	1	88B4	5D18	嶮		
5111	旆	1	89B8	65BE	旆		
5302	昏	1	8ADA	662C	昏		
5382	弄	2	8B6B			馬会升	
5404	鴨	1	8BA3	4CED		即厶鳥	
5442	栱	1	8BC9	67BF	栱		
5501	龔	2	8C45	4AA5		龔	
5529	芑	1	8C61	4497		艹会乞	
5564	槩	1	8CA6	6ABF	槩		
5739		6	8DB8			冗会古	畧，錯字，以程式全部取代。
5793	麇	1	8DEE	2A2A8		鹿厶旨	
5854	幺	1	C6CD	5E7A	幺		

二、手動取代表說明：

《爾雅》檔案中出現「●」字形，共計86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，請參考附件二。欄位說明如下：

1. 頁碼：「●」字形所在位置之原書頁碼。
2. 檔案原文摘錄：摘錄《爾雅》檔案有「●」字形的文句。
3. 原書字形：「●」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。

附件二、《爾雅》「●」字手動取代表

頁碼	檔案原文摘錄	原書字形	備註
p. 1	愷。康。●。般	𡗗	
p. 1	於也。●。郤	合𡗗文	
p. 2	癩。癩。●。痒	疒𡗗里	
p. 2	痒。●。疵	疵	
p. 2	瘥。痒。●。瘵	疒𡗗亶	
p. 2	也。●。埃	𠄎𠄎止頁𠄎	
p. 2	危也。●。汽也	豈𡗗幾	
p. 2	仍。●。埤	𡗗	
p. 2	劉。●。樂也	𡗗	
p. 2	嗟。咨。●也	𡗗𡗗差	
p. 2	休。苦。●。黷	𡗗	
p. 2	具也。●。憐	𡗗𡗗某	

頁碼	檔案原文摘錄	原書字形	備註
p. 2	剡。●。利也	剡 厶 各	
p. 3	逸。●。過也	侃 厶 言	
p. 3	幽也。●。屬也	𠂔 牙 攴 厂 毛 𠂔	
p. 3	猷。●。可也	肯	
p. 3	潰也。●。闇也	陪	
p. 3	闇也。●。膠也	黍 厶 刀	
p. 4	足也。●	足 厶 非	
p. 6	謂之●。扶	宀 厶 交	
p. 6	謂之●。垝	戶 厶 巳	
p. 7	斫。謂之●	鑄	
p. 7	●。謂之黓	廌 厶 斗	
p. 7	衣梳。謂之●	衤 厶 兒	
p. 7	謂之鬲。●	鬲 厶 曾	
p. 7	謂之●。鏤	鞏	
p. 7	●也。卣。	金 厶 突	
p. 8	大墳謂之●	呂 厶 丿 厶 呂	
p. 10	峯者。厓●	厂 厶 義	
p. 11	紼。●也	索 厶 率	
p. 11	黍蓬。●。鼠莞	艹 厶 庫	
p. 11	菑。●。熒	艹 厶 富	
p. 11	夫須。萼薊●	蒨	

頁碼	檔案原文摘錄	原書字形	備註
p. 11	●●。離南	形 ^艸 女麥 ^口 ^艸 彡縷	
p. 11	活●。龍	^艸 彡允	
p. 11	芄蘭。●。茫	蓐 ^艸 彡	
p. 11	鹿●。其實菴	^艸 彡霍	
p. 12	牡●箭。山莓	^艸 彡贊	
p. 12	盜庚。●。麻母	苧	
p. 12	仲無筮。●	竹 ^彡 彡息	
p. 12	蕨。●。蕎	^艸 彡驚	
p. 12	棘●。策。刺	^艸 彡 ^宀 彡兔	改為菟
p. 12	菡。●。海藻	蓐 ^艸 彡	
p. 12	黏。●。椒	扌 ^彡 廢	
p. 12	柜●。栩	柳	
p. 12	杞。枸●。杭	櫪	
p. 12	棧。●其。劉	木 ^彡 速	
p. 13	蟪●。入耳	行 ^彡 虫	
p. 13	茅蝟。●。馬蝟	虫 ^彡 面	待造字
p. 13	蝎。蝓●。蟻	虫 ^彡 屈	
p. 13	王蚊。蝓●	施 ^彡 虫	
p. 13	強●。不過	形 ^虫 彡干 ^口	
p. 13	蝮蝟。蟋蟀。●	虫 ^彡 爻	

頁碼	檔案原文摘錄	原書字形	備註
p. 13	蠅。馬●。皇蟲	虫△爻	
p. 13	莫●。螻蛄	𧈧	
p. 13	●。羅。翰	我△虫	
p. 13	打螳。●。飛	尉△虫	
p. 13	●蛸。長跣	𧈧	
p. 13	黑●。鱈	魚△茲	
p. 13	鱈●。鰕鱈	𧈧	
p. 13	龜●。蟾諸	酋△黽	
p. 13	蛭。●。蚌	𧈧	
p. 13	●。大而險	虫△困	
p. 13	●。小而橢蠓蠓	虫△責	
p. 14	鴉鷂。●鳩	鴟	
p. 14	鳩。王●	鴟	
p. 14	●。鴟	名△鳥	
p. 14	鴟●。鷂	其△鳥	
p. 14	糜鴟。●	名△鳥	
p. 14	烏●。舒鴈	暴△鳥	
p. 14	鷂。●老。鴈	其△鳥	
p. 14	鷂鷂。●。天鷂	卓△鳥	
p. 14	鷂。●。鷂	鷂	
p. 14	狂。●鳥	癡	

頁碼	檔案原文摘錄	原書字形	備註
p. 14	南方曰●	曷	
p. 14	其子。●。其跡	鹿△天	
p. 14	絕有力●	鹿△开	
p. 14	其子●。其跡	鹿△旦	
p. 14	豬。●。豮	豕△隋	
p. 14	四●皆白	獮	
p. 14	絕有力。●	豕△厄	
p. 14	謂之●貓	號	
p. 15	麇。狸子。●	豸△隶	
p. 15	●子。貍	貍	
p. 15	●獮。類貍	豸△契	
p. 15	●貓。食虎豹	號	
p. 15	狸狐獮●醜	貍	
p. 15	其跡●。蒙頌	𠂔	
p. 15	●。黃白雜毛	𠂔	
p. 15	●牛。擺牛	牛△暴	
p. 15	五尺為●	豕△厄	