

《建炎以來繫年要錄》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
2007/12/17 黃榮順 製作

一、《建炎以來繫年要錄》一書(建炎以來繫年要錄.xml)使用舊版造字 24 個，字頻 1892 次，詳如附件一。這 24 個造字中，15 個可轉成 Windows XP 可顯示的字，字頻 1741 次；另外 9 個字必須轉成構字式，字頻 151 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，手動刪除：在舊版漢籍電子文獻中即存在的誤增字，為原書所無，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，則手動將其刪除。如編號 3647 的「」(無對應造字)，字碼對應的部件外字為「阝」，檔案內容使用「阝」，卻未見於原書，應為誤增字，故手動刪除。
- 2、重碼問題，採用部件外字，不轉碼：舊漢籍造字與構字式部件外字，皆是建立在 Big5 的使用者造字區中，兩者所使用的碼區是重疊的，在同時使用構字式和舊漢籍造字的新版漢籍中，這些重疊的字碼可能被用以表達兩個不同字形，我們稱之為「重碼問題」。然而漢籍在舊轉新的過程中，若已將原舊漢籍造字的部分全數改為 Unicode 或構字式，則在 xml 出現的造字頻次皆採用部件外字，此時便不需轉碼。本書 24 個造字中，有 22 個都是這個情況。如編號 5344，字碼為 8B45，對應的舊漢籍造字為「𠂔」，部件外字為「久」，比對原書發現，書中皆取部件外字「久」，故不需轉碼。

- 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 4933，字碼為 88A3，對應的舊漢籍造字為「𡗗」，部件外字為「爪」。原書檔案皆使用部件外字「爪」，而「爪」實即「瓜」之簡體，經查原書，並無專取其形的用例，故本書皆以「瓜」字取代。

- 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 5012 的「媿」字，Unicode 編碼為 36D0，由於 Windows XP 並不支援，仍須改用構字式「女𡗗更」。

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、部件：舊漢籍造字字碼所對應之部件外字。
- 丁、取代：使用字形與造字欄或部件欄不同時，於此欄列出。
- 戊、檔案原文摘錄：摘錄檔案(建炎以來繫年要錄.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。
- 己、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。

- 五、「建炎以來繫年要錄.xml」缺字外發現之「❓」，修正情形詳列附件三，附件三「建炎以來繫年要錄.xml」「❓」手動修改表體例說明如下：

- 庚、編號：流水號，依原書出現順序排列。
- 辛、檔案原文摘錄：摘錄檔案(建炎以來繫年要錄.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。

壬、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。

附件一、「建炎以來繫年要錄.xml」造字分析表

編號	造字	字頻 (txt)	字頻 (xml)	Big5	Unicode	WinXP	構字式	備註
3647		1	1	A063				阨，校對問題，舊版漢籍錯字，手動刪除。
4262	麪	1	1	8456			麥△𠂔	昏，重碼問題，部件外字，不轉碼。
4278	亘	36	355	8466	4E98	亘		扶，重碼問題，部件外字，不轉碼。
4552	趣	8	8	85FD	27F08		走△𠂔	𠂔，重碼問題，部件外字，不轉碼。
4933	尙	1	34	88A3	5CDD	尙		瓜，重碼問題，部件外字，不轉碼。 瓜，異體字問題，以「瓜」字取代。
5005	幫	1	53	88EB	5E2E	幫		并，重碼問題，部件外字，不轉碼。
5012	媵	1	1	88F2	36D0		女△更	茲，重碼問題，部件外字，不轉碼。
5048	婚	1	11	8957	36F0		女△昏	𠂔，重碼問題，部件外字，不轉碼。
5078		31	2	8975			亩△心	𠂔，重碼問題，部件外字，不轉碼。 𠂔，異體字問題，以「𠂔」字取代。
5222	亘	36	708	8A68	4EB6	亘		来，重碼問題，部件外字，不轉碼。

編號	造字	字頻 (txt)	字頻 (xml)	Big5	Unicode	WinXP	構字式	備註
								来，異體字問題，以「來」字取代。
5299	戟	7	11	8AD7	39B8		卓  戈	𣎵，重碼問題，部件外字，不轉碼。
5338	暎	1	355	8AFE	668E	暎		𣎵，重碼問題，部件外字，不轉碼。
5344	磬	2	24	8B45	55F8	磬		久，重碼問題，部件外字，不轉碼。
5395	覘	7	115	8B78	4093		矢  見	卜，重碼問題，部件外字，不轉碼。
5491	槿	2	41	8BFA	6967	槿		𣎵，重碼問題，部件外字，不轉碼。
5537	离	21	4	8C69	79BC	离		𣎵，重碼問題，部件外字，不轉碼。
5558	龕	1	5	8C7E	9E84	龕		冗，重碼問題，部件外字，不轉碼。
5559	欸	80	10	8CA1	6B35	欸		亾，重碼問題，部件外字，不轉碼。 亾，異體字問題，以「亡」字取代。
5572	埠	2	24	8CAE	57BE	埠		𣎵，重碼問題，部件外字，不轉碼。
5583	殮	6	30	8CB9	98F1	殮		葬，重碼問題，部件外字，不轉碼。
5585	寓	163	1	8CBB	3762		宀  禹	鳥，重碼問題，部件外字，不轉碼。
5588	恠	3	96	8CBE	6061	恠		𣎵，重碼問題，部件外字，不轉碼。
5610	泠	1	1	8CD4	6C75	泠		耳，重碼問題，部件外字，不轉碼。
5859	支	1	1	C6D2	6534	支		

附件二、「建炎以來繫年要錄.xml」手動取代表

編號	造字	部件	取代	檔案原文摘錄	手動取代結果
3647		陔		身在江西 <u>陔</u>	身在江西 (刪除「陔」字)
4933	尙	瓜	瓜	<u>瓜</u> (全部取代)	<u>瓜</u> (全部取代)
5078	亩会心	廿	廿	<u>廿</u> (全部取代)	<u>廿</u> (全部取代)
5222	亶	来	來	<u>来</u> (全部取代)	<u>來</u> (全部取代)
5559	欸	亾	亡	<u>亾</u> (全部取代)	<u>亡</u> (全部取代)

附件三、「建炎以來繫年要錄.xml」「㊦」手動修改表

編號	檔案原文摘錄	手動取代結果
1	無 <u>㊦</u> 志之語	無 <u>鬪</u> 志之語
2	兵革戰 <u>㊦</u> 之事	兵革戰 <u>鬪</u> 之事
3	人無 <u>㊦</u> 志	人無 <u>鬪</u> 志
4	少年善 <u>㊦</u> 嘗犯法	少年善 <u>鬪</u> 嘗犯法
5	<u>㊦</u> 汪伯彥共議	<u>善</u> 汪伯彥共議
6	甚眾人無 <u>㊦</u> 志	甚眾人無 <u>鬪</u> 志
7	士爭死 <u>㊦</u> 敵	士爭死 <u>鬪</u> 敵
8	不習戰 <u>㊦</u> 動	不習戰 <u>鬪</u> 動
9	命皆有 <u>㊦</u> 志	命皆有 <u>鬪</u> 志
10	上下神 <u>㊦</u> 之意	上下神 <u>示</u> 之意
11	踰半日榮 <u>㊦</u> 死	踰半日榮 <u>鬪</u> 死

12	而士無 ^② 志未	而士無 ^鬪 志未
13	所部與格 ^② 因入	所部與格 ^鬪 因入
14	塵不曾戰 ^② 何故	塵不曾戰 ^鬪 何故
15	籌策而不預戰 ^②	籌策而不預戰 ^鬪
16	武帥類皆 ^②	武帥類皆 ^鬪
17	含怒必 ^② 之志	含怒必 ^鬪 之志
18	使自黃 ^② △替 ^②	使自黃 ^② △替 ^善
19	多驍雋敢 ^② 慶	多驍雋敢 ^鬪 慶
20	黑虎者尤善格 ^②	黑虎者尤善格 ^鬪
21	士卒輕財善 ^②	士卒輕財善 ^鬪
22	^② 自辰及酉賊	^鬪 自辰及酉賊
23	皆饑困無 ^② 志	皆饑困無 ^鬪 志
24	原有擄掠無戰 ^②	原有擄掠無戰 ^鬪
25	天上帝皇地 ^②	天上帝皇地 ^示
26	且勇於戰 ^②	且勇於戰 ^鬪
27	以至賞刑 ^②	以至賞刑 ^鬪
28	勝本初無 ^② 心	勝本初無 ^鬪 心
29	兵法不習戰 ^②	兵法不習戰 ^鬪
30	自子城出春 ^②	自子城出春 ^鬪
31	戰 ^② 可得而用	戰 ^鬪 可得而用
32	其人勇悍喜 ^②	其人勇悍喜 ^鬪
33	降川陝 ^② 殺情	降川陝 ^鬪 殺情
34	高第黃潛 ^②	高第黃潛 ^善
35	以下諸將戰 ^②	以下諸將戰 ^鬪
36	有謀略不止於 ^②	有謀略不止於 ^鬪
37	措畫數與敵 ^② 屢	措畫數與敵 ^鬪 屢
38	曰今士無 ^② 志	曰今士無 ^鬪 志
39	兩虎安得私 ^②	兩虎安得私 ^鬪
40	私 ^② 使耳餘渾	私 ^鬪 使耳餘渾
41	等以臨敵無 ^② 志	等以臨敵無 ^鬪 志
42	出澗準備戰 ^②	出澗準備戰 ^鬪
43	來戰則人無 ^② 志	來戰則人無 ^鬪 志
44	戰之處敘其 ^②	戰之處敘其 ^鬪
45	諸生寧能 ^② 乎	諸生寧能 ^鬪 乎
46	之虞皆非 ^② 戰者	之虞皆非 ^善 戰者
47	可乘以戰 ^② 人	可乘以戰 ^鬪 人
48	可憫若因 ^②	可憫若因 ^鬪

49	兩軍之士交 ^②	兩軍之士交鬪
50	下囚一等內 ^②	下囚一等內鬪
51	罪已下囚及 ^②	罪已下囚及鬪
52	^② 因得進用懼	善因得進用懼
53	銷縮莫有 ^② 心	銷縮莫有鬪心
54	激勵豈無 ^② 心	激勵豈無鬪心
55	以來習於戰 ^②	以來習於戰鬪
56	河南用兵投 ^②	河南用兵投鬪
57	我師皆殊死 ^②	我師皆殊死鬪
58	中人勇於私 ^②	中人勇於私鬪
59	在紹興初交 ^②	在紹興初交鬪
60	激軍中使無 ^②	激軍中使無鬪
61	縣夜兩軍交 ^②	縣夜兩軍交鬪
62	移潮與洲 ^② 怒	移潮與洲鬪怒
63	折納既免 ^②	折納既免鬪
64	正副非盜賊 ^②	正副非盜賊鬪
65	惡實犯故殺 ^②	惡實犯故殺鬪
66	有不因財氣 ^②	有不因財氣鬪
67	如此則故殺 ^②	如此則故殺鬪
68	其黨持兵毆 ^②	其黨持兵毆鬪
69	閱之使有 ^② 心	閱之使有鬪心
70	坦平可以戰 ^②	坦平可以戰鬪
71	簽軍實無 ^② 志	簽軍實無鬪志
72	率所部合 ^② 殺	率所部合鬪殺
73	勳 ^② 士倍於晉	勳鬪士倍於晉
74	但與金人戰 ^②	但與金人戰鬪
75	若奮身前 ^② 萬	若奮身前鬪萬
76	登軍皆死 ^②	登軍皆死鬪
77	迎擊士皆死 ^②	迎擊士皆死鬪
78	此夫豈有 ^② 志	此夫豈有鬪志
79	之前不得 ^② 退	之前不得鬪退
80	死亦須窮 ^② 及	死亦須窮鬪及
81	望言諸軍 ^② 志	望言諸軍鬪志
82	其人即殊死 ^② 敵	其人即殊死鬪敵
83	既免 ^② 訟亦稀	既免鬪訟亦稀
84	乃仲牙兵死 ^②	乃仲牙兵死鬪
85	陣士卒死 ^② 敵	陣士卒死鬪敵

