

《新刊大宋宣和遺事》(標記語料)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
2012/08/07 丁玟伶

一、舊版造字轉碼分析：

《新刊大宋宣和遺事》(標記語料)使用舊版造字 11 個，字頻 43 次，這 11 個造字中，8 個可轉成 Windows XP 能顯示的字，字頻 28 次；另外 3 個字必須轉成構字式，字頻 15 次。

轉碼完成製作轉碼分析表，請參考附件一《新刊大宋宣和遺事》(標記語料)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。

附件一、《新刊大宋宣和遺事》(標記語料)轉碼分析表

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
296	厶	1	FBEC			厶	
297	厶	13	FBED			厶	
899	馳	2	8ED3	99DE	馳		
1100	身	3	9040	8EAB	身		

編號	造字	頻次	Big5	Unicode	WinXP	構字式	備註
3797	凭	1	815C	51ED	凭		
3835	噍	1	81A4	564D	噍		
4074	縹	1	82F6	7E27	縹		
4226	飡	1	83F1	98E1	飡		
4461	勛	17	85A2	52D4	勛		
5075	犛	1	8972	72AD		犛	
5551	費	2	8C77	8CF7	費		

二、手動取代表說明：

《新刊大宋宣和遺事》素語料中之「●」字，為日後線上檢索方便，在標記語料多修正為古今通用字，與文獻處理實驗室較以原書為準修改其轉碼後素語料檔案的「●」字，兩者用字略有不同，其中多為異體字差異，特製作用字差異記錄表，以供未來參考。請參考附件二《新刊大宋宣和遺事》素語料與標記語料「●」字差異記錄表。欄位說明如下：

1. 頁碼：修改字形所在位置之原書頁碼。
2. 素語料原文摘錄：摘錄《新刊大宋宣和遺事》素語料檔案的文句段落，其中紅字部分，是經由文獻處理實驗室轉碼修改後的「●」字。
3. 標記語料原文摘錄：摘錄《新刊大宋宣和遺事》標記語料檔案的文句段落，其中藍字的部分，為標記語料修正素語料中之「●」字形。

4. 備註：記錄修改相同辭句的次數或其他事項。

附件二、《新刊大宋宣和遺事》素語料與標記語料

「●」字差異記錄表

頁碼	素語料原文摘錄	標記語料原文摘錄	備註
p. 5	左衽吟	左衽吟	
p. 10	引用壻蔡卞	引用(VC) 婿(Na) 蔡卞	
p. 10	易子而殮者	易(VC) 子(Na) 而 (Cbb) 殮	
p. 11	王安石之壻	王安石(Nb) 之(DE) 婿	
p. 12	玉璣玉卮	玉(Na) 璣(Na) 玉(Na) 卮	
p. 15	布炆十餘里	布(VC) 氣(Na) 十 (Neu) 餘	
p. 15	人畜在炆	人(Na) 畜(Na) 在 (VCL) 氣	
p. 16	皆奉上帝勅命	奉(VF) 上帝(Na) 敕命	
p. 18	不堪礮石	不(Dc) 堪(VL) 砲石	
p. 22	僕就試	僕(Na) 就(VJ) 試	
p. 23	勅問道錄	敕(VE) 問(VE) 道錄	
p. 28	有甄出血	有(V_2) 磚(Na) 出 (VC) 血	
p. 29	二炆	二(Neu) 氣	

頁碼	素語料原文摘錄	標記語料原文摘錄	備註
p. 29	陰炁之盛	陰氣(Na) 之(DE) 盛	
p. 43	才 ㄩ棄命三郎	拼命三郎	
p. 57	鳥 ㄩ甲鳩催	鴨(Na) 鳩(Na) 催(VF)	
p. 64	張穀	張珏(Nb)	有 17 處
p. 93	何稟	何栗(Nb)	有 3 處
p. 117	形 ㄨ ㄨ ㄨ ㄨ 二喪	囊(VC) 二(Neu) 喪	
p. 137	張穀 ㄩ心厲聲	張愨(Nb) 厲聲	