

中央研究院第三屆國際漢學會議

**Third International Conference on Sinology  
Academia Sinica**

漢籍數位典藏組：數位化的語文工具（一）  
**Digital Collections for Sinology Studies:Linguistic  
Digital Tools I**

文字學標籤集之標準的制定與  
國際標準化數位檔案的轉換

John A Lehman (University of Alaska Fairbanks)  
周亞民 (國立台灣大學資訊管理研究所)  
邱台鳳、謝清俊 (中央研究院資訊科學研究所)

中央研究院 主辦  
行政院國家科學委員會 協辦  
Hosted by **Academia Sinica**  
Co-sponsored by **National Science Council**  
June 29—July 1, 2000

# 文字學標籤集之標準的制定與 國際標準化數位檔案的轉換

John A Lehman\* 周亞民\*\* 邱台鳳\*\*\* 謝清俊\*\*\*

\*University of Alaska Fairbanks

\*\*國立臺灣大學資訊管理研究所

\*\*\*中央研究院資訊科學研究所文獻處理實驗室

## 摘 要

在計算機中表達文字學知識，最好的方法之一，就是利用標籤(tags)，但是目前我們面臨兩個問題，第一、缺乏能夠達文字學知識的標籤集(tag set)，第二、文字學知識的標籤集，必須要有共同的標準，才能在不同的計算機環境中，互相交換和處理。本研究的目的，即在於建立文字學的標籤集，並作為制定文字學標籤集之標準的基礎。

本研究的研究方法，是利用漢語大字典，整理和歸納可以描述文字學知識的標籤集合，因為漢語大字典的內容，含蓋了十分豐富的文字學知識，包括字形的流變（如甲骨文、金文、小篆、隸書和楷書等）、上古音、中古音、現代音、本義、引申義與字義的考釋，因此我們以漢語大字典，作為建立文字學標籤集合的基礎。

依據漢語大字典，我們已整理出三十七個標籤，其中有二十一個屬於文字學的標籤，十一個標籤屬於書目學，二個校刊學有關的標籤，其它的標籤有三個。本研究所整理的三十七個標籤，除了可以提供文字學標籤集合的制定之用，還可以提供書目與版本學作為制定標籤集的參考。

我們利用這三十七個標籤，已經初步的完成漢語大字典中三千八百九十三個領頭字下的內容，共一百零六萬字的標誌(markup)工作，其中有三十個標籤可以利用計算機自動進行標誌的工作。我們所選擇的三千八百多個單

字，皆是林樹字集中的最常用字與次常用字，因為最常用字與次常用字的內容較為豐富，故做為我們測試的對象。

國際上，多數的文獻(archives)使用 TEI 指引(Text Encoding Initiative Guidelines)作為標誌文獻的標準，以利文獻的交換和處理，本研究依 TEI 指引，結果有百分之九十的標籤皆可互相對映，並且可以利用 XSLT(eXtensible StyleSheet Language Transformations)，讓計算機自動進行轉換，使本研究標誌後的文獻，很容易就轉換為國際標準化數位檔案。

## 一、研究目的

目前，計算機能夠利用的文字學知識，仍然十分的有限，而造成文字學者無法充分利用計算機的能力，從事相關的研究。爲什麼計算機能夠利用的文字學知識有限？主要的原因就是沒有利用適當的方式，在計算機中表達文字學知識。我們認爲如果要在計算機中表達文字學知識，最好的方法之一，就是利用標籤(tags)，但是目前我們面臨兩個問題，第一、缺乏能夠達文字學知識的標籤集(tag set)，第二、文字學知識的標籤集，必須要有共同的標準，才能在不同的計算機環境中，互相交換和處理。

事實上，利用標籤表達計算機所需的知識，已經使用很久了，但是早期大部份都只用於排版和印刷輸出的需要，後來才用來表達其它的知識。目前最重要的標籤集之一，是西元一九八七年，一項名爲 TEI(Text Encoding Initiative)的國際合作計劃，該計劃的目的是制定數位檔案(digital archives)標籤集與交換的標準。經過多年的努力，已經制定出了很多標籤集的標準，這些標準是以 SGML(Standard Generalized Markup Language)作爲制定標準的工具【1】。TEI 主要是以西方語言(Western Language)的檔案(archives)爲制定標準的對象，但是文字學與語言關係密切，在 TEI 中缺乏文字學所需的標籤集。

我們如果有了文字學標籤集的標準，文字學的學者就可以較現在的環境，更充份利用計算機，搜集與整理所需的資料，不僅能夠節省人力和時間，以便將精神專注於更有創造性的工作上，更可以改善研究成的品質。但是標籤集必須是大家共同所接受和採用，才能有利於彼此的分享，不斷的持續累積研究的成果，本研究的目的，即在於建立文字學的標籤集，並且作爲制定文字學標籤集之標準的基礎，該從事文字學研究的學者，有一個更好的研究工具。

## 二、研究方法

本研究的研究方法，是利用漢語大字典，整理和歸納可以描述文字學知識的標籤集合。漢語大字典為大陸學者所編寫，於西元一九八五年出版，本研究採用繁體字版，由遠東圖書公司所印刷和發行。單字條目的內容包括領頭字、解形、注音、釋義和引證。有古字的領頭字，則列出能字形的演變，並有字形解說。注音包括領頭字的現代音、中古音和上古音，現代音用漢語拼音，中古音以反切為主，且有聲韻調。上古音則只有韻部，出現字則不沒有中古音和上古音，出現於中古的字則沒有上古音。一字如有多義則依本義、引申義和通假義的順序排列【3】。

我們由漢語大字典整理出標籤集後，再撰寫程式以便能夠自動進行標誌(markup)的工作，最後利用 XSLT(eXtensible StyleSheet Language Transformations)，讓計算機自動進行轉換，使本研究標誌後的文獻，很容易就轉換為國際標準化數位檔案。

## 三、標籤集整理結果

本研究利用漢語大字典，整理出來的標籤集，除了文字學有關的標籤外，尚包括書目學與校刊學有關的標籤。因為漢語大字典的內容，不僅包含中文字的形、音、義，還包含了許多引證資料，引證的資料與書目有關，故整理出來的標籤中，有一些與書目學有關。漢語大字典為求正確，也作了部份的校刊的工作，因此，也有與校刊有關的標籤。本節將我們由漢語大字典中所整理出來的標籤集，分為文字學、書目學、校刊學和其它等四類說明。

### 3.1 文字學有關的標籤

#### 1. 單字條目

單字條目的組成包括字頭、解形、注音、釋義、引證，範例如下：

<單字條目>

\*甬

(一) q i 4 《龍龕手鑑》音去厶卅。同“棄”。《龍龕手鑑·不部》：“甬，音去厶卅。”清錢大昕《十駕齋養新錄·宋時俗字》：“《龍龕手鑑》多收鄙俗之字，如……甬爲棄。”

(二) b e n g 2 方言。“不用”的合音。用不；不必。如：甬說了；你甬管。

</單字條目>

## 2. 字頭

標誌領頭字。範例如下：

<字頭>暈</字頭>〔𠄎日一车〇〕

{

說文新附·日部

}

## 3. 簡化字

標誌簡化字，範例如下(註：本例中的簡化字，因爲是缺字，因此以部件序表示之【5】)：

\*暈<簡化字>〔𠄎日一车〇〕</簡化字>

{

說文新附·日部

}

#### 4. 字形流變

標誌字形源流演變的古文字形，範例如下：

\* 暈〔𠄎日一车〇〕

<字形流變>

說文新附·日部

</字形流變>

#### 5. 解形

標誌釋釋文字的構形，範例如下：

\* 暈〔𠄎日一车〇〕

{

說文新附·日部

}

<解形>《說文新附》：“暈，日月气也。 日，軍聲。”鈕樹玉新附考：

“暈即暉之異體。”</解形>

\* 件

{

說文·人部

漢印

}

<解形>《說文》：“件，分也。〇〇人人，〇〇人牛。牛，大物，故可分。”

桂馥義證：“件，大徐所加。”</解形>

## 6. 音義

一個音義標誌包含一個分列音項及其下統率的所有義項，範例如下：

\*暈〔日一車 $\square$ 〕

{  
說文新附·日部

}

《說文新附》：“暈，日月气也。∞人日，軍聲。”鈕樹玉新附考：“暈即暉之異體。”

<音義> (一) y u n 4 《廣韻》王問切，去問云。諄部。

(1) 日月週圍的光圈。《說文新附·日部》：“暈，日月气也。”《廣韻·問韻》：“暈，日月旁氣。”《史記·天官書》：“日月暈適，雲風，此天之客氣，其發見亦有大運。”裴駰集解引孟康曰：“暈，日旁氣也。”唐李白《橫江詞》：“月暈天風霧不開，海鯨東蹙百川迴。”《金史·天文志》：“二十九年正月乙卯巳初，日有暈，左右有珥，上有背氣兩重，其色青赤而厚。”

(2) 光影、色澤四週模糊的部分。唐韓愈《宿龍宮灘》：“夢覺燈生暈，宵殘雨送涼。”宋蘇軾《墨花》：“花心超墨暈，春色散毫端。”《徐霞客遊記·粵西遊日記一》：“後壁紅白紋若縷，圓宕三暈，恰當其首，故曰玩珠。”

(3) 浸潤；擴散。明湯顯祖《牡丹亭·禦淮》：“血暈幾重圍，孤城怎生料？”《警世通言·錢舍人題詩燕子樓》：“鈿暈羅衫色似煙，一回看著一潸然。”朱自清《槳聲燈影裏的秦淮河》：“從兩重玻璃裏映出那幅射著的黃黃的散光，反暈出一片朦朧的火 $\Delta$ 因靄。”

(4) 昏眩。如：暈船；暈車。《正字通·日部》：“暈，方書有暈眩證（症），頭暈、昏暈是也。”唐陸龜蒙《奉酬襲美先輩吳中苦雨一百韻》：“看花雖眼暈，見酒忘肺渴。”《水滸全傳》第三十六回：“（宋



江) 向前扶他，不覺自家也頭暈眼花，撲地倒了。” </音義>

<音義> (二) y u n l

昏迷；失去知覺。如：暈倒；暈厥。 </音義>

## 7. 音項

音項將領頭字的拼音變化組合成一個單位，範例如下：

\*暈〔𠄎日一车〇〕

{

說文新附·日部

}

《說文新附》：“暈，日月气也。〇人日，軍聲。”鈕樹玉新附考：

“暈即暉之異體。”

<音項> y u n 4 《廣韻》王問切，去問云。諄部。 </音項>

日月週圍的光圈。 . . .

光影、色澤四週模糊。 . . .

浸潤；擴散。 . . .

昏眩扶他，不覺。 . . .

<音項> y u n l

昏迷；失去知覺。如：暈倒；暈厥。 </音項>

## 8. 義項

一字可能多義，每個意義以義項表示，範例如下：

\*暈〔𠄎日一车〇〕

{

說文新附·日部

}

《說文新附》：“暈，日月氣也。∞人日，軍聲。”鈕樹玉新附考：“暈即暉之異體。”

(一) y u n 4 《廣韻》王問切，去問云。諄部。

<義項> (1) 日月週圍的光圈。《說文新附·日部》：“暈，日月氣也。”

《廣韻·問韻》：“暈，日月旁氣。”《史記·天官書》：“日月暈適，雲風，此天之客氣，其發見亦有大運。”裴駟集解引孟康曰：“暈，日旁氣也。”唐李白《橫江詞》：“月暈天風霧不開，海鯨東蹙百川迴。”

《金史·天文志》：“二十九年正月乙卯巳初，日有暈，左右有珥，上有背氣兩重，其色青赤而厚。”</義項>

<義項> (2) 光影、色澤四週模糊的部分。唐韓愈《宿龍宮灘》：“夢覺燈生暈，宵殘雨送涼。”宋蘇軾《墨花》：“花心超墨暈，春色散毫端。”《徐霞客遊記·粵西遊日記一》：“後壁紅白紋若縷，圓宕三暈，恰當其首，故曰玩珠。”</義項>

<義項> (3) 浸潤；擴散。明湯顯祖《牡丹亭·禦淮》：“血暈幾重圍，孤城怎生料？”《警世通言·錢舍人題詩燕子樓》：“鈿暈羅衫色似煙，一回看著一潸然。”朱自清《槳聲燈影裏的秦淮河》：“從兩重玻璃裏映出那幅射著的黃黃的散光，反暈出一片朦朧的火△因靄。”</義項>

<義項> (4) 昏眩。如：暈船；暈車。《正字通·日部》：“暈，方書有暈眩證（症），頭暈、昏暈是也。”唐陸龜蒙《奉酬襲美先輩吳中苦雨一百韻》：“看花雖眼暈，見酒忘肺渴。”《水滸全傳》第三十六回：“（宋江）向前扶他，不覺自家也頭暈眼花，撲地倒了。”</義項>

(二) y u n 1

昏迷；失去知覺。如：暈倒；暈厥。

## 9. 現代音

標誌該拼音爲現代音，範例如下：

\*暈〔𠄎日一车〇〕

{

說文新附·日部

}

《說文新附》：“暈，日月气也。∞人日，軍聲。”鈕樹玉新附考：“暈即暉之異體。”

(一) <現代音> y u n 4 </現代音> 《廣韻》王問切，去問云。諄部。

(二) <現代音> y u n 1 </現代音> 昏迷；失去知覺。如：暈倒；

暈厥。

## 10. 調號

於<現代音>標誌中再將調號加以標誌，範例如下：

\*暈〔𠄎日一车〇〕

{

說文新附·日部

}

《說文新附》：“暈，日月气也。∞人日，軍聲。”鈕樹玉新附考：“暈即暉之異體。”

(一) <現代音> y u n <調號> 4 </調號> </現代音> 《廣韻》王問切，去問云。諄部。

(二) <現代音> y u n <調號> 1 </調號> </現代音> 昏迷；失去知覺。如：暈倒；暈厥。

## 11. 中古音

標誌中古音，範例如下：

\*暈〔𠄎日一車〇〕

{

說文新附·日部

}

《說文新附》：“暈，日月气也。〇人日，軍聲。”鈕樹玉新附考：  
“暈即暉之異體。”

(一) y u n 4 <中古音>《廣韻》王問切，去問云。</中古音>諄部。

(二) y u n 1 昏迷；失去知覺。如：暈倒；暈厥。

## 12. 反切上部、13. 反切下部、14. 聲韻調

於<中古音>中再標誌<反切上部>、<反切下部>和<聲韻調>，<反切上部>標誌反切之聲母，<反切下部>標誌反切之韻母，<聲韻調>標誌中古音中之聲韻調，範例如下：

\*暈〔𠄎日一車〇〕

{

說文新附·日部

}

《說文新附》：“暈，日月气也。〇人日，軍聲。”鈕樹玉新附考：  
“暈即暉之異體。”

(一) y u n 4 《廣韻》<中古音><反切上部>王</反切上部><反切下部>問</反切下部>切，<聲韻調>去問云</聲韻調>。</中古音>諄部。

(二) y u n 1 昏迷；失去知覺。如：暈倒；暈厥。

## 15. 上古音

標誌上古音，範例如下：

\*暈〔𠄎日一車〇〕

{

說文新附·日部

}

《說文新附》：“暈，日月气也。∞人日，軍聲。”鈕樹玉新附考：  
“暈即暉之異體。”

(一) y u n 4 《廣韻》王問切，去問云。<上古音>諄部。</上古音>

(二) y u n 1 昏迷；失去知覺。如：暈倒；暈厥。

## 16. 後起同音代替字

標誌異體關係中，表示後起同音代替字之“用同某”，範例如下：

\*咨

z i 1 《廣韻》即夷切

(7) <後起同音代替字>用同“齟”。</後起同音代替字>露出（牙齒）。《西遊記》第五回：“大聖喝了一口，即咨牙 嘴道：‘不好喫！不好喫！’”

## 17. 異體字

異體關係中，異體字分別以“同某”、“後作某”、“也作某”表示者，標誌為異體字，範例如下：

### \*單

(五) chān 3 《集韻》齒善切，上獮昌。元部。

(1) <異體字>同“𦉳”。</異體字>聲音舒緩。《集韻·牙ㄤ璽韻》：“𦉳，聲緩也。或省。”

### \*單

(八) tān 2 《集韻》唐干切，平寒定。

姓。<異體字>也作“檀”。</異體字>《集韻·寒韻》：“單，闕。姓也，鄭有櫟邑大夫單伯。通作檀。”

### \*哇

(四) wā 2

小孩。<異體字>後作“娃”。</異體字>元佚名《抱粧盒》第二折：“你道是別人家的哇哇，他是西宮李美人生的太子。”明湯顯祖《牡丹亭·閨塾》：“有指證，姜嫄產哇；不嫉妒，后妃賢達。”

## 18. 通假字

異體關係中，通假義項用“通某”表示，標誌為通假字，範例如下：

### \*單

(五) chān 3 《集韻》齒善切，上獮昌。元部。

(2) <通假字>通“繹”。</通假字>舒緩貌。《馬王堆漢墓帛書·老子乙本·德經》：“單而善謀。”按：今本《老子》第七十三章作“繹經”：“單而善謀”。朱謙之集釋：“繹，音闡，舒緩貌。”

## 19. 本義

標誌本義，範例如下：

\*巨

(一) j u 4 《廣韻》其呂切， (1) <本義>規矩，木工的方尺  
<本義>。後作“矩”。《說文·工部》：“巨，規巨也。”邵瑛 經正字：  
“巨，今經典多作矩，正字當作巨。”《管子·宙合》：“成功之術，  
必有巨矦。”趙守正注：“巨矦，規矩。”

(2) 大。

(3) 最；極。

(4) 通“距”。距離。

(5) 通“詎”。

## 20. 引申義

標誌引申義，範例如下：

\*巨

(一) j u 4 《廣韻》其呂切， (1) 規矩，木工的方尺。

(2) <引申義>大。</引申義>《方言》卷一：“巨，大也。齊宋之間曰巨。”《墨子·非樂》：“三者，民之巨患也。”唐李白《古風五

十九首》之三十三：“北溟有巨魚，身長數千里。”魯迅《華蓋集·忽然想到》：“好像炎熱的巨刃從天而降。”

(3) <引申義>最；極。<引申義>三國魏曹植《辯道論》：“言不盡於此，頗難悉載，故粗舉其巨怪者。”《宋書·江夏王義恭傳》：“乾靈降禍，二凶極逆，深酷巨痛，終古未有。”

(4) 通“距”。距離。

(5) 通“詎”。相當於“豈”。

## 21. 通假義

標誌通假義，範例如下：

### \*巨

(一) j u 4 《廣韻》其呂切，(1) 規矩，木工的方尺。後作“矩”

(2) 大。《方言》卷一：“巨，大

(3) 最；極。

(4) 通“距”。<通假義>距離。<通假義>《墨子·備高臨》：“守爲臺城，以臨羊黔，左右出巨，各二十尺。”孫詒讓閒詁：“巨，當爲距之段字。”

(5) 通“詎”。<通假義>相當於“豈”<通假義>。清朱駿聲《說文通訓定聲·豫部》：“巨，發聲之詞，字亦作詎。”《漢書·高帝紀上》：“沛公不先破關中兵，公巨能入乎？”顏師古注：“巨，讀曰詎。詎，猶豈也。”

(4) 昏眩。如：暈船；暈車。《正字通·日部》：“暈，方書有暈眩證（症），頭暈、昏暈是也。”唐<人名>陸龜蒙<人名>《奉酬襲美先輩吳中苦雨一百韻》：“看花雖眼暈，見酒忘肺渴。”《水滸全傳》第三



十六回：“（<人名>宋江<人名>）向前扶他，不覺自家也頭暈眼花，撲地倒了。”

### 3.2 書目學有關的標籤

#### 1. 書名

標誌書名，範例如下：

##### \*九

（一）j i u 3 《<書名>廣韻</書名>》舉有切，上有見。幽部。

（1）數詞。八加一的和。

（2）泛指多數。

（3）《<書名>周易</書名>》以陽爻爲九……

（4）時令名。∞人冬至起……

（5）通“久”。老。《<書名>莊子</書名>·至樂》：“黃軫生乎九猷。”陸德明釋文：“李云，九宜爲久。久，老也。猷，蟲名也。”

#### 2. 書目細目

於<書名>標誌之後，如有更細的篇章結構，則以<書名細目>標誌之，範例如下：

##### \*九

（一）j i u 3 《<書名>廣韻</書名>》舉有切，上有見。幽部。

（1）數詞。八加一的和。

（2）泛指多數。

（3）《<書名>周易</書名>》以陽爻爲九……

(4) 時令名。∞人冬至起。 . . .

(5) 通“久”。老。《<書名>莊子</書名><書名細目>至樂</書名細目>》：“黃軛生乎九猷。”陸德明釋文：“李云，九宜爲久。久，老也。猷，蟲名也。”

### 3. 卷次

標誌卷次，範例如下：

\*不

(一) b u 4 《廣韻》分物切。 . . .

(1) 非；不是。清王引之《經傳釋詞》<卷次>卷十</卷次>：“不，非也。”

### 4. 回數

標誌回數，範例如下：

(一) y u n 4 《廣韻》

(4) 昏眩。如：暈船；暈車。《正字通·日部》：“暈，方書有暈眩證（症），頭暈、昏暈是也。”唐陸龜蒙《奉酬襲美先輩吳中苦雨一百韻》：“看花雖眼暈，見酒忘肺渴。”《水滸全傳》<回數>第三十六回</回數>：“（宋江）向前扶他，不覺自家也頭暈眼花，撲地倒了。”

### 5. 本數

標誌本數，範例如下：

\*低

d i 1 《廣韻》都奚切，平齊端。

(3) 指聲音細微。宋蘇軾《趙成伯家有麗人僕忝鄉人不肯開樽徒吟春雪美句次韻一笑》：“試問高吟三十韻，何如低唱兩三杯。”元王實甫《西廂記》<本數>第一本</本數>第三折：“去△口纔箇悄悄相問，他那裏低低應。”

## 6. 折數

標誌本數，範例如下：

\*低

d i 1 《廣韻》都奚切，平齊端。

(3) 指聲音細微。宋蘇軾《趙成伯家有麗人僕忝鄉人不肯開樽徒吟春雪美句次韻一笑》：“試問高吟三十韻，何如低唱兩三杯。”元王實甫《西廂記》第一本<折數>第三折</折數>：“去△口纔箇悄悄相問，他那裏低低應。”

## 7. 齣數

標誌齣數，範例如下：

\*喬

(一) q i a o 2 《廣韻》

(4) 做作；裝假。如：喬裝打扮。元佚名《連環計》第三折：“父子每都要帽光光，做出這喬模樣。”明凌濛初《紅拂擇配》〈齣數〉第一齣〈齣數〉：“分明是長噓氣，喬說是歌遶畫梁飛。”《紅樓夢》第八十回：“寶蟾又喬裝躲閃，連忙縮手。”

## 8. 部數

標誌部數，範例如下：

\* 咐

(二) f u 4

〔吩咐〕囑托；安排。《紅樓夢》第七十一回：“東府裏的奶奶立等一位奶奶，有話吩咐。”《暴風驟雨》〈部數〉第一部〈部數〉十九：“蕭隊長吩咐把所有子彈全收集攏來。”

## 9. 章數

標誌章數，範例如下：

\*哄 (一) h o n g 4 《廣韻》胡  
(2) 吵鬧；喧囂。《集韻·送韻》：“口ㄩ官，血ㄩ瓜聲。或作哄。”  
《水滸全傳》第三回：“血ㄩ瓜人見是魯提轄，一哄都走了。”柳青《創業史》第一部<章數>第十九章</章數>：“突然間，她聽見校院裏一群女同學噉噉噉噉起哄了。”

## 3.3 校刊學有關的標籤

### 1. 校訂

校訂原著的脫誤，範例如下：

\*今  
j i n 1 《廣韻》居吟切，平侵見  
(5) 連詞。表示假設關係，相當於“若”、“假如”。清王引之《經傳釋詞》卷五：“今，家大人曰：‘今，猶若也。’”《孟子·梁惠王下》：“此無他，與民同樂也。今王與百姓同樂，則王矣。”《史記·項羽本紀》：“今不急下，吾烹太公。”《漢書·項籍傳》：“章將軍<校訂>〔等〕</校訂>詐吾屬降諸侯，今能入關破秦，大善；即不能，諸侯虜吾屬而東，秦又盡誅吾父母妻子。”

### 2. 增補

爲使前後文義貫通，在引文中增補字句，範例如下：。

\*暈〔𠄎日一車〇〕  
(一) y u n 4 《廣韻》

(4) 昏眩。如：暈船；暈車。《正字通·日部》：“暈，方書有暈眩證（症），頭暈、昏暈是也。”唐陸龜蒙《奉酬襲美先輩吳中苦雨一百韻》：“看花雖眼暈，見酒忘肺渴。”《水滸全傳》第三十六回：“（宋江）向前扶他，不覺自家也頭暈眼花，撲地倒了。”

### 3.4 其它標籤

#### 1. 朝代

標誌朝代，範例如下：

\*暈〔𠄎日一車〇〕

(一) y u n 4 《廣韻》王問切，(1) 日月週圍的光圈。《說文新附·日部》：“暈，日月氣也。”《廣韻·問韻》：“暈，日月旁氣。”《史記·天官書》：“日月暈適，雲風，此天之客氣，其發見亦有大運。”裴駟集解引孟康曰：“暈，日旁氣也。”<朝代>唐</朝代>李白《橫江詞》：“月暈天風霧不開，海鯨東蹙百川迴。”《金史·天文志》：“二十九年正月乙卯巳初，日有暈，左右有珥，上有背氣兩重，其色青赤而厚。”

(2) 光影、色澤四週模糊的部分。<朝代>唐</朝代>韓愈《宿龍宮灘》：“夢覺燈生暈，宵殘雨送涼。”<朝代>宋</朝代>蘇軾《墨花》：“花心超墨暈，春色散毫端。”《徐霞客遊記·粵西遊日記一》：“後壁紅白紋若縷，圓宕三暈，恰當其首，故曰玩珠。”

(3) 浸潤；擴散。<朝代>明</朝代>湯顯祖《牡丹亭·禦淮》：“血暈幾重圍，孤城怎生料？”《警世通言·錢舍人題詩燕子樓》：“鈿暈羅衫色似煙，一回看著一潸然。”朱自清《樂聲燈影裏的秦淮河》：“從兩重玻璃裏映出那幅射著的黃黃的散光，反暈出一片朦朧的火△因靄。”

(4) 昏眩。如：暈船；暈車。《正字通·日部》：“暈，方書有暈眩證（症），頭暈、昏暈是也。”<朝代>唐</朝代>陸龜蒙《奉酬襲美先輩

吳中苦雨一百韻》：“看花雖眼暈，見酒忘肺渴。”《水滸全傳》第三十六回：“（宋江）向前扶他，不覺自家也頭暈眼花，撲地倒了。”

## 2. 人名

標誌人名，範例如下：

\*暈〔日一車〇〕

（一）y u n 4 《廣韻》王問切，（1）日月週圍的光圈。《說文新附·日部》：“暈，日月氣也。”《廣韻·問韻》：“暈，日月旁氣。”《史記·天官書》：“日月暈適，雲風，此天之客氣，其發見亦有大運。”裴駟集解引<人名>孟康</人名>曰：“暈，日旁氣也。”唐<人名>李白</人名>《橫江詞》：“月暈天風霧不開，海鯨東蹙百川迴。”·《金史·天文志》：“二十九年正月乙卯巳初，日有暈，左右有珥，上有背氣兩重，其色青赤而厚。”

（2）光影、色澤四週模糊的部分。唐<人名>韓愈</人名>《宿龍宮灘》：“夢覺燈生暈，宵殘雨送涼。”宋<人名>蘇軾</人名>《墨花》：“花心超墨暈，春色散毫端。”《徐霞客遊記·粵西遊日記一》：“後壁紅白紋若縷，圓宕三暈，恰當其首，故曰玩珠。”

（3）浸潤；擴散。明<人名>湯顯祖</人名>《牡丹亭·禦淮》：“血暈幾重圍，孤城怎生料？”《警世通言·錢舍人題詩燕子樓》：“鈿暈羅衫色似煙，一回看著一潸然。”<人名>朱自清</人名>《槳聲燈影裏的秦淮河》：“從兩重玻璃裏映出那幅射著的黃黃的散光，反暈出一片朦朧的火△因靄。”

### 3. 化學元素符號

元素符號前後標誌化學元素符號，範例如下：

\* 鋅

(二) x i n l

化學元素。符號<化學元素符號>Zn</化學元素符號>，原子序 30。藍白色金屬，在潮濕的空氣中容易氧化，表面形成白色保護層，質地脆，溶於酸和強 中，大多用來製合金或者鍍鐵板。舊稱亞鉛。(新拉 Zincum)

#### 四、標誌的過程

1. 選字：以林樹字集中的最常用字與次常用字為選字的根據，最常用字共一八五七字，次常用字共二〇六八字【2】。林樹字集最常用字與次常用字中，有三十二個字是漢語大字典所沒有的，因此，扣除三十二個字後共選出三八九三字，作為本研究標誌的對象。
2. 打字規則：以倚天中文倉頡碼和 BIG-5 內碼進行打字，格式原則上依照原書輸入，如遇缺字則以內碼 A1B4 表示。
3. 打字與校對：本研究將選出的三八九三字，分送兩家打字行打字，再將兩家打字後的結果，利用校對程式進行互相的比對，如果有不同之處，再以人工比對，以減輕校對的人力與時間。缺字的部份則用字根序表示之。
4. 標誌的過程：
  - (1)工具：Microsoft Word 97 VBA 與 Windows 98。
  - (2)找出標籤之間的相互關係，決定標誌的先後順序。
  - (3)找出每個標籤在要插入它所在的漢語大字典的位置時，與漢語大字典相對應的關鍵字、詞、符號或規則，以做為自動標誌的依據。再利用 Word 97 VBA 及正規表示式(regular expression)來標誌所有的標



籤。

## 五、標誌的結果

1. 遠東漢語大字典總共七卷 89 個檔案，標完 30 個 TAG，約需 11 小時。
2. 遠東漢語大字典七卷中的缺字已補上。
3. 目前共設計 37 個 TAG，已完成 30 個。尚未加標誌者還有本義、引申義、通假義、書證、人名、朝代、校訂等 7 個，因為在凡例中找不到明顯的規則可循，而在打字輸入時，亦未加以標明，故無法自動標誌，須以人工為之。
4. 標誌前與標誌後的結果，請參考圖一與圖二。
5. 會產生混淆之標誌：

(1) 校訂：校訂在漢語大字典中用方括號〔〕表明。簡化字也是用〔〕表明；而另外有些字，在某些特殊讀音下，必與某字合用，成為一詞。例如：「薺」，當讀為 j i 4 時，為〔薺菜〕。而此時所使用之標點符號亦是用〔〕，會產生混淆，故無法標誌，範例如下：

標籤名稱	範例
字形流變	* 暈〔𠄎日一车〇〕 { 說文新附·日部 }
校訂	* 今 (5) 連詞。……《漢書·項籍傳》：“章將軍〔等〕詐吾屬降諸侯，今能入關破秦，大善；即不能，諸侯虜吾屬而東，秦又盡誅吾父母妻子。”

無	<p>*薺</p> <p>(二) j i 4 《廣韻》徂禮切，上薺從。脂部。</p> <p>〔薺菜〕十字花科。一年生或二年生草本。花小，白色。嫩葉可食。全草入藥。《玉篇·艸部》：“薺，甘菜。”《詩·邶風·谷風》：“誰謂荼苦，其甘如薺。”《宋書·顧覲之傳》：“松柳異質，薺荼殊性。”</p>
---	--

(2) 增補：漢語大字典中用圓括號（）表示增補，但音項與義項也會用到圓括號（），故產生混淆。

標籤名稱	範例
增補	<p>*暈〔𠄎日一车〇〕</p> <p>(一) y u n 4 《廣韻》</p> <p>(4) 昏眩。如：暈船；暈車。《正字通·日部》：“暈，方書有暈眩證（症），頭暈、昏暈是也。”唐陸龜蒙《奉酬襲美先輩吳中苦雨一百韻》：“看花雖眼暈，見酒忘肺渴。”《水滸全傳》第三十六回：“（宋江）向前扶他，不覺自家也頭暈眼花，撲地倒了。”</p>
音項	<p>*暈〔𠄎日一车〇〕</p> <p>(一) y u n 4 《廣韻》王問切，去問云。諄部。 . . .</p> <p>(二) y u n 1 昏迷；失去知覺。如：暈倒；暈厥。</p>
義項	<p>*暈〔𠄎日一车〇〕</p> <p>(一) y u n 4 《廣韻》王問切，去問云。諄部。</p> <p>(1) 日月週圍的光圈 . . .</p> <p>(2) 光影、色澤四週模糊的 . . .</p> <p>(3) 浸潤；擴散 . . .</p>

(4) 昏眩扶他，不覺……

圖一 標誌範例—暈(標誌前)

\*暈〔日一車<sup>〇</sup>〕

{

說文新附·日部

}

《說文新附》：“暈，日月气也。〇人日，軍聲。”鈕樹玉新附考：“暈即暉之異體。”

(一) y u n 4 《廣韻》王問切，去問云。諄部。

(1) 日月週圍的光圈。《說文新附·日部》：“暈，日月气也。”《廣韻·問韻》：“暈，日月旁氣。”《史記·天官書》：“日月暈適，雲風，此天之客氣，其發見亦有大運。”裴駟集解引孟康曰：“暈，日旁氣也。”唐李白《橫江詞》：“月暈天風霧不開，海鯨東蹙百川迴。”《金史·天文志》：“二十九年正月乙卯巳初，日有暈，左右有珥，上有背氣兩重，其色青赤而厚。”

(2) 光影、色澤四週模糊的部分。唐韓愈《宿龍宮灘》：“夢覺燈生暈，宵殘雨送涼。”宋蘇軾《墨花》：“花心超墨暈，春色散毫端。”《徐霞客遊記·粵西遊日記一》：“後壁紅白紋若縷，圓宕三暈，恰當其首，故曰玩珠。”

(3) 浸潤；擴散。明湯顯祖《牡丹亭·禦淮》：“血暈幾重圍，孤城怎生料？”《警世通言·錢舍人題詩燕子樓》：“鈿暈羅衫色似煙，一回看著一潸然。”朱自清《樂聲燈影裏的秦淮河》：“從兩重玻璃裏映出那幅射著的黃黃的散光，反暈出一片朦朧的火△因靄。”

(4) 昏眩。如：暈船；暈車。《正字通·日部》：“暈，方書有暈眩證（症），頭暈、昏暈是也。”唐陸龜蒙《奉酬襲美先輩吳中苦雨一百韻》：“看花雖眼暈，見酒忘肺渴。”《水滸全傳》第三十六回：“（宋江）向前扶他，不覺自家也頭暈眼花，撲地倒了。”

(二) y u n l

昏迷；失去知覺。如：暈倒；暈厥。

圖二 標誌範例—暈(標誌後)

<單字條目>

<字頭>暈</字頭>

<簡化字>〔𠄎日一車〇〕</簡化字>

<字形流變>說文新附·日部</字形流變>

<解形>《<書名>說文新附</書名>》：“暈，日月气也。∞人日，軍聲。”鈕樹玉新附考：“暈即暉之異體。”</解形>

<音義>

<音項><現代音>y u n<調號>4</調號></現代音>《<書名>廣韻</書名>》<中古音><反切上部>王</反切上部><反切下部>問</反切下部>切，<聲韻調>去問云</聲韻調>。</中古音><上古音>諄部。</上古音></音項>

<義項>日月週圍的光圈。《<書名>說文新附</書名><書名細目>日部</書名細目>》：“暈，日月气也。”《<書名>廣韻</書名><書名細目>問韻</書名細目>》：“暈，日月旁氣。”《<書名>史記</書名><書名細目>天官書</書名細目>》：“日月暈適，雲風，此天之客氣，其發見亦有大運。”裴駰集解引孟康曰：“暈，日旁氣也。”唐李白《<書名>橫江詞</書名>》：“月暈天風霧不開，海鯨東蹙百川迴。”《<書名>金史</書名><書名細目>天文志</書名細目>

名細目>》：“二十九年正月乙卯巳初，日有暈，左右有珥，上有背氣兩重，其色青赤而厚。” </義項>

<義項>光影、色澤四週模糊的部分。唐韓愈《<書名>宿龍宮灘</書名>》：“夢覺燈生暈，宵殘雨送涼。”宋蘇軾《<書名>墨花</書名>》：“花心超墨暈，春色散毫端。”《<書名>徐霞客遊記</書名><書名細目>粵西遊日記一</書名細目>》：“後壁紅白紋若縷，圓宕三暈，恰當其首，故曰玩珠。” </義項>

<義項>浸潤；擴散。明湯顯祖《<書名>牡丹亭</書名><書名細目>禦淮</書名細目>》：“血暈幾重圍，孤城怎生料？”《<書名>警世通言</書名><書名細目>錢舍人題詩燕子樓</書名細目>》：“鈿暈羅衫色似煙，一回看著一潸然。”朱自清《<書名>槳聲燈影裏的秦淮河</書名>》：“從兩重玻璃裏映出那幅射著的黃黃的散光，反暈出一片朦朧的火△因靄。” </義項>

<義項>昏眩。<例證>如：暈船；暈車。</例證>《<書名>正字通</書名><書名細目>日部</書名細目>》：“暈，方書有暈眩證<增補>（症）</增補>，頭暈、昏暈是也。”唐陸龜蒙《<書名>奉酬襲美先輩吳中苦雨一百韻</書名>》：“看花雖眼暈，見酒忘肺渴。”《<書名>水滸全傳</書名><回數>第三十六回</回數>》：“<增補>（宋江）</增補>向前扶他，不覺自家也頭暈眼花，撲地倒了。” </義項>

</音義>

<音義>

<音項><現代音>y u n<調號>l</調號></現代音>昏迷；失去知覺。<例證>如：暈倒；暈厥。</例證></音項>

</音義>

</單字條目>

## 六、轉換為 TEI 相容的標籤集

如果要將本研究所整理出來的標籤集，轉換為 TEI 的標籤，會產生三種可能的情形：(1)直接可以對映到 TEI 的標籤，(2)可以對映到 TEI 的標籤，但是必須要有額外的屬性描述，(3)可能無法對映到 TEI 的標籤。可以直接對映的有單字條目、字頭、解形、音義、義項、現代音、書名細目、例證、書證、書名和人名。

轉換上的問題，很多來自於表意文字，與拼音文字【7】，及用來描述古時後的拼音系統。因為西方語言的字典(這也是 TEI 所處理的對象)【4】，並沒有能標籤集能夠描述中文字與如何書寫之間(字形)，所存在的多對多關係。如果要將本研究的有關字形的標籤(如異體字、通假字、後起同音代替字等)，利用 TEI 的標籤表示，則必須要用 TEI 的<orth>標籤，並加以延申其表達的含義，增加新的屬性使它能夠表達除了拼音字以外的語言。由於在 TEI 中增加新的屬性，需要改變 TEI 的 DTD(Document Type Definition)，而我們所整理出來的標籤集，還需要擴充和測試，因此，我們希望所整理出標籤集穩定下來後，再著手修改 TEI 的 DTD，或者是作為 TEI 新的標籤集，故暫不改變 TEI 的 DTD。

在拼音部份，可能產生問題的標籤包括音項、現代音、調號、中古音、反切上部、反切下部、聲韻調和上古音，但是，經過我們分析後，音項是唯一的問題，音項將中文字上古音、中古音和現代音群集在一起，但是因為 TEI 假設每一個字，都只有一個拼音，但是，中文字可能有一個以上的拼音。古法語(Old French)在時間上相當於中古的漢語，拉丁語(Latin)相當於上古漢語，但法文字典不會列出現代音是”roi”的古法語拼音是”reis”，拉丁語拼音是”rex”，也不會列出普羅旺斯語的(Provençal)拼音，因為它們的拼字不同。

但是，中文字並不相同，因為中文字並非拼音文字，而是表意文字，同一個中文字，可能因為地區或時代的不同，而有不同的拼音，但是 TEI 對每一個字，並沒有一群不同的拼音，以說明其變化，如果要用 TEI，就必須改用幾個單獨的標籤(即 pron 標籤)，表示很多個拼音，但是如此一來，就無法清楚表示這些拼音之間的關係。

表一 文字學中與拼音有關的標籤與 TEI 標籤的對照表

文字學中與拼音有關的標籤集		TEI 的標籤
音項	pronunciation section	無對映的標籤
現代音	modern pronunciation (pinyin)	<pron lang="ch">
調號	tone number	<pron extent="tone">
中古音	archaic Chinese pronunciation	無對映的標籤
反切上部	fanqie initial	<pron lang="archaic-chinese" extent="initial" >
反切下部	fanqie final	<pron lang="archaic-chinese" extent="final" >
聲韻調	rhyme class	<pron lang="archaic-chinese" extent="tone" >
上古音	ancient Chinese pronunciation	<pron lang="ancient-chinese">

表二 本研究的標籤集與 TEI 標籤的對照表

(註：音譯部份為羅馬字拼音，採用遠東漢英大辭典,1992)

本研究的標籤	TEI 的標籤
單字條目 individual character entry	<entry>
字頭 character shape	<form>
簡化字 simplified character construction	<etym>
字形流變 shuowen-derived radical 的屬性	<orth> *需要增加新
解形 shuowen derivation	<def>
音義 underlying word	<hom>
音項 pronunciation section	無對映的 TEI 標籤
義項 underlying meaning	<sense>
現代音 modern pronunciation (pinyin)	<pron lang="ch">
調號 tone number	<pron extent="tone">
中古音 archaic Chinese pronunciation	無對映的 TEI 標籤
反切上部 fanqie initial chinese"	<pron lang ="archaic- chinese" extent= "initial">
反切下部 fanqie final chinese"	<pron lang ="archaic- chinese" extent="final">
聲韻調 rhyme class chinese"	<pron lang ="archaic- chinese" extent="tone">



上古音	ancient Chinese pronunciation	<pron lang = "ancient-chinese" >
書名	book title	<title>
書名細目	first citation body ?	<biblScope>
卷次	jeuan number	<biblScope
	type="jeuan">	
回數	chapter number	<biblScope
	type="hwei">	
本數	been number	<biblScope type="been"
	>	
折數	jer number	<biblScope type="jer">
齣數	chu chapter number	<biblScope type="chu">
部數	buh number	<biblScope type="buh">
章數	jang number	<biblScope
	type="jang">	
後起同音代替字	homonym substitute	<orth> *需要增加新
的屬性		
異體字	variant character	<orth> *需要增加新
的屬性		
通假字	substitute character	<orth> *需要增加新
的屬性		
增補	supplementary character	<orth> *需要增加新
的屬性		
例證	example	<eg>
書證	book cited	<cit><title>
本義	basic meaning	無對映的 TEI 標籤

引申義 extended meaning	無對映的 TEI 標籤
通假義 borrowed meaning	無對映的 TEI 標籤
人名 person name	<name>
朝代 dynasty name	<date value="">
校訂 revision	無對映的 TEI 標籤
化學元素符號 chemical element name	無對映的 TEI 標籤

圖三 轉換為 TEI 標籤的結果

```

<entry>
<form>暈</form>
<etym>〔 暈日一車〔 〕 〕 </etym>
<orth>說文新附·日部</orth>
<def>《<title>說文新附</title>》<gloss>：“暈，日月气也。∞人日，軍聲。”
鈕樹玉新附考：“暈即暉之異體。” </gloss></def>
<hom>
<pron lang="ch"> y u n </pron><pron extent="tone"> 4 </pron> 《<title>廣韻
</title>》<pron lang="archaic-chinese" extent="initial">王</pron><pron lang
="archaic-chinese" extent="final">問</pron>切， <pron lang="archaic-chinese"
extent="tone">去問云</pron>。</pron> <pron lang="ancient-chinese"> 諄部。
</pron><sense n="1"> 日月週圍的光圈。<eg><cit><bibl>《<title>說文新附
</title><biblScope>日部</biblScope>》：<quote> “暈，日月气也。”
</quote></bibl></cit></eg> <eg><cit><bibl>《<title>廣韻</title><biblScope>問韻
</biblScope>》<quote>：“暈，日月旁氣。” </quote></bibl></cit></eg>
<eg><cit><bibl>《<title>史記</title><biblScope>天官書</biblScope>》<quote>：
“日月暈適，雲風，此天之客氣，其發見亦有大運。” 裴駰集解引孟康曰：

```

“暈，日旁氣也。” </quote> <author><date value="Tang"> 唐</date>李白</author></bibl></cit></eg> <eg><cit><bibl>《<title>橫江詞</title></bibl></cit></eg> <quote>：“月暈天風霧不開，海鯨東蹙百川迴。” </quote> </bibl></cit></eg>

<eg><cit><bibl>《<title>金史</title><biblScope>天文志</biblScope></bibl></cit></eg> <quote>：“二十九年正月乙卯巳初，日有暈，左右有珥，上有背氣兩重，其色青赤而厚。” </quote></bibl></cit></eg></sense>

<sense n="2"> 光影、色澤四週模糊的部分。唐韓愈 <eg><cit><bibl>《<title>宿龍宮灘</title></bibl></cit></eg> <quote>：“夢覺燈生暈，宵殘雨送涼。” </quote>

<author><date value="Song"> 宋</date>蘇軾</author> </bibl></cit></eg>

<eg><cit><bibl>《<title>墨花</title></bibl></cit></eg> <quote>：“花心超墨暈，春色散毫端。” </quote> </bibl></cit></eg> <eg><cit><bibl>《<title>徐霞客遊記</title><biblScope>粵西遊日記一</biblScope></bibl></cit></eg> <quote>：“後壁紅白紋若縷，圓宕三暈，恰當其首，故曰玩珠。” </quote> </bibl></cit></eg> </sense>

<sense n="3"> 浸潤；擴散。明湯顯祖 </bibl></cit></eg> <eg><cit><bibl>

《<title>牡丹亭</title><biblScope>禦淮</biblScope></bibl></cit></eg> <quote>：“血暈幾重圍，孤城怎生料？” </quote> </bibl></cit></eg> <eg><cit><bibl>《<title>警世通言</title><biblScope>錢舍人題詩燕子樓</biblScope></bibl></cit></eg> <quote>：“鈿暈羅衫色似煙，一回看著一潸然。” </quote> <author>朱自清</author> </bibl></cit></eg>

<eg><cit><bibl>《<title>槳聲燈影裏的秦淮河</title></bibl></cit></eg> <quote>：“從兩重玻璃裏映出那幅射著的黃黃的散光，反暈出一片朦朧的火<sup>△</sup>因靄。” </quote>

</bibl></cit> </eg> </sense><sense n="4"> 昏眩。<eg>如：暈船；暈車。

</eg><cit><bibl>《<title>正字通</title><biblScope>日部</biblScope></bibl></cit></eg> <quote>：“暈，方書有暈眩證<增補>（症）</增補>，頭暈、昏暈是也。” </quote>

<author><date value="Tang">唐</date>陸龜蒙</author> </bibl></cit></eg>

<eg><cit><bibl>《<title>奉酬襲美先輩吳中苦雨一百韻</title></bibl></cit></eg> <quote>：“看花雖眼暈，見酒忘肺渴。” </quote> </bibl></cit></eg> <eg><cit><bibl>《<title>

```

水滸全傳</title>》 <biblScope type="chapter">第三十六回</biblScope>
<quote>：“（宋江）向前扶他，不覺自家也頭暈眼花，撲地倒了。” </quote>
</bibl></cit></eg> </sense>
</hom>
<hom>
<pron lang="ch"> y u n</pron> <pron extent="tone"> 1</pron> 昏迷；失
去知覺。<eg>如：暈倒；暈厥。</eg>
</hom>
</entry>

```

上述對由漢語大字典所整理出來的標籤集，與 TEI 中字典的標籤集的對照，目的是爲了要發展出一套能夠利用電腦程式，將本研究的標籤集，轉換爲 TEI 的標籤，轉換爲 TEI 相容的標籤集，可以讓能夠讀取 TEI 檔案的應用系統，能夠處理使用本研究所整理的標籤集的檔案。這個轉換用的程式，利用 XSLT 很容易就能夠進行兩者之間的轉換。通常，XSLT 用來將 XML 的檔案轉換爲 HTML，以便讓流覽器處理後，給讀者閱讀 XML 檔案的內容。但是，事實上 XSLT 能夠用來作爲不同標誌語言之間的轉換，因此，我們可以利用 XSLT，將本研究的標籤集，轉換爲 TEI 的標籤。

圖四 利用 XSLT 轉換文字學標籤爲 TEI 相容標籤的例子(部份的內容)

```

<XSLT:stylesheet xmlns:XSLT="http://www.w3.org/1999/XSLT/Transform">
<XSLT:template match="單字條目">
  <entry><XSLT:apply-templates/></entry>

```

</XSLT:template>

<XSLT:template match="字頭">

    <form><XSLT:apply-templates/></form>

</XSLT:template>

<XSLT:template match="簡化字">

    <etym><XSLT:apply-templates/></etym>

</XSLT:template>

<XSLT:template match="解形">

    <def><XSLT:apply-templates/></def>

</XSLT:template>

<XSLT:template match="中古音">

    <pron lang="archaic-chinese" extent="initial"><XSLT:apply-  
        templates/></pron>

</XSLT:template>.....

## 七、結論與未來研究方向

依據漢語大字典，我們已整理出三十七個標籤，其中有二十一個屬於文字學的標籤，十一個標籤屬於書目學，二個校刊學有關的標籤，其它的標籤有三個。本研究所整理的三十七個標籤，除了可以提供文字學標籤集合的制定之用，還可以提供書目與版本學作為制定標籤集的參考。

我們利用這三十七個標籤，已經初步的完成漢語大字典中三千八百九十

三個領頭字下的內容，共一百零六萬字的標誌(markup)工作，其中有三十個標籤可以利用計算機自動進行標誌的工作，尙未加標誌者還有本義、引申義、通假義、書證、人名、朝代、校訂等 7 個。

我們所整理出來的文字學標籤集，還並不完備，仍需要進一步的研究，而 TEI 的標籤集，是經由數十位學者，且經歷了六年以上的時間，才完成初步的標籤集，而且現在仍然不斷的持續進行中，使其更加完備，本研究若要使文字學標籤集更加的完備，也必須投入更多的人力，並嘗試利用其它的字書，來建立文字學標籤集。文字學標籤集的制定，必須要有更多文字學方面的學者參與，而且本研究也只是一個開始，我們所整理出來的標籤集，也不是十分完備的文字學標籤集，因為本研究是利用漢語大字典，作為建立文字學標籤集的工具，但畢竟漢語大字典無法含蓋所有可能的文字學知識，因此，我們所整理出來的標籤集只能含蓋有限的文字學知識。我們目前正在進行說文解字詁林的分析與標籤集整理的工作，檢視我們利用漢語大字典所整理的標籤集，如果缺乏適當的標籤，則加以擴充，使得文字學的標籤集更加完備。

當文字學的標籤集能夠更加的完備後，我們希望能夠成為 TEI 的標準之一，不僅能夠與 TEI 的標籤集相容，也能藉 TEI 的力量，將文字學的標籤集推廣，讓更多的人能夠接觸和採用，使得文字學的標籤集發揮傳遞和交換的優點，因為如果文字學的標籤集只為少數人所使用，它所能貢獻的效果就被限制了。除了將文字學的標籤集成為 TEI 的標準之一，有利於普及，如果能夠開發應用的系統，也能夠幫助文字學標籤集能被接受，而應用系統的出現，也能讓文字學的學者，實際看到文字學標籤集對於研究工作上的助益，節省資料搜集與整理的時間，而將注意力放在更有創意的研究工作上。

缺字的問題，本研究以部件序與構字式表示之，但是未來我們應該要建

立一個大字都能接受且很容易使用的機置，利用這個機置描述表達缺字的方法，讓不同的計算機系統能夠處理缺字的問題。因為缺字的表達方法，已經有很多人提出來，但是尚未達成共識，因此，我們如果要所有的人都接受某一種缺字表達方法並不容易，但是如果我們能夠利用標誌語言，描述所用的缺字表達方法是什麼，使得各種缺字的表達方法都可以使用，解決缺字的問題【5】【6】【7】。

除了文字學的標籤集外，我們也整理出部份的書目學與校刊學的標籤，這些標籤可以作為書目學與校刊學學者的使用，雖然我們在字書中能夠整理出與書目學與校刊學有關的標籤集，但是如果要建立足夠用來標誌書目學知識和校刊學知識的標籤集，則必須要由與書目學或校刊學有關的學者，由相關的書籍中去整理所需的標籤，本研究的研究方法，也可以提供制定書目學，或校刊學標籤集標準的參考。

## 八、參考文獻

1. Sperburg-McQueen, S., and Burnard, L. (ed.), *Guidelines for Electronic Text Encoding and Interchange (TEI P3)*, 1994.
2. 林樹，中文電腦基本用字研究，國立交通大學，科技研究報告電子計算機科學 CC-601 號，1972。
3. 徐中舒，《漢語大字典》，遠東圖書公司出版，1990。
4. Ide, N, and Veronis, J., “Encoding Dictionaries” from Ide, N, and Veronis, J. (ed.) *Text Encoding Initiative, Background and Context*, 1995.
5. 謝清俊，〈電子古籍中的缺字問題〉，《第一屆中國文字學會學術討論會論文集》，天津，1996。
6. 謝清俊，〈從缺字問題，談漢字交換碼的重新設計〉，漢字字碼與資料庫國際研討會，京都，東京，1996。
7. 謝清俊，〈A Descriptive Method for Re-engineering Hanzi Information Interchange Codes〉（從缺字問題，談漢字交換碼的重新設計），漢字字碼與資料庫國際研討會，京都，東京，1996。