

# 建立东巴文象形字典语料库的构想

甘 露

(华东师范大学中国文字研究与应用中心, 上海 200062)

**摘要:** 本文以方国瑜《纳西象形文字谱》、李霖灿《么些象形文字字典》和洛克《纳西语—英语百科辞典》为基础, 拟建立一个东巴文象形字典语料库。这个语料库包括九个方面的内容: 1、统一的字号, 为三部字典所收字的新编号; 2、东巴文字形, 对东巴文异体字整理后的标准字形; 3、纳西语读音, 分别列出三部字典各自的读音; 4、释义; 5、字与词的界定, 指出所收字形是字还是词; 6、造字法; 7、词义类别, 根据东巴文的意义分为 12 类; 8、词例, 列出词的不同用法; 9、出处, 注明词汇来源的经典。

**关键词:** 东巴文; 象形字; 语料库

**中图分类号:** H257 **文献类型:** A

在我國少數民族古文字中, 東巴文以其形態原始、內涵豐富, 引起了國內外學者的極大關注。從 1913 年法國人巴克(J·Bacot)第一部研究東巴文的著作(《Les Moso》《麼些研究》)開始, 至今, 中外學者已經從民族學、文字學、人類學、宗教學、文學、藝術、哲學、天文曆法等多學科多角度地對東巴文化進行了大量研究, 所撰寫的學術論文不勝枚舉, 東巴文化研究已取得了令人矚目的成就。

目前, 收錄東巴文的字典主要為方國瑜《納西象形文字譜》、李霖燦《麼些象形文字字典》和洛克《納西語—英語百科辭典》。這三部字典是不同時代、不同國籍的學者編撰的, 因而辭典的收字範圍、釋義體例、讀音標注方面均不盡一致, 各家自有長短。因目前研究的需要, 我們擬建立一個東巴文字典的語料庫, 綜合三家的長處, 彌補各自的不足, 以便更好地為東巴文化的研究服務。

## 一、東巴文辭典的介紹

方國瑜《納西象形文字譜》是三部字典中最容易見到的, 館藏也很豐富, 許多學者研究東巴文字均是從這部書入手。這部字典是方國瑜於 1935 年寫成的,<sup>[1]</sup>該書歷經四十多年, “四易其稿”, 於 1981 年由雲南人民出版社出版。該書分為天象、地理、植物、飛禽、走獸、蟲魚、人稱、形體、服飾、飲食、居住、器用、行止、形狀、宗教和傳說古人名號等 18 類, 收象形字 1340 個及 222 個派生字, 對每個字逐字標音解說。這部書體例完善、釋義簡明、準確可靠, 是研究東巴文和東巴文化的重要參考書之一。

李霖燦《麼些象形文字字典》於 1944 年編印出版, 是第一部正式出版的東巴文字典, 在當時學術界的影響非常巨大, 聞宥評價該書是“取材之富, 實為已往所未有, 每字下之音讀, 精確可信”。<sup>[2]</sup>該書的材料十分豐富, 共收字 2120 個, 分為天文、地理、人文、人體、鳥、獸、植物、用具、飲食、衣飾、武器、建築數目動作、“若喀”字、“古宗”字、宗教、鬼怪、東巴龍王、神等 18 類。該書的發音人是東巴和才, 國際音標注音者是著名語言學家李方桂的高足張琨, 所以讀音可以稱得上是準確可靠。此外, 該書的解釋也很詳盡, 盡可能地列出了各種異體字的寫法, 對假借字的解釋也很詳細, 有些並舉出了例子。雖然已經過去五十多年, 該書仍是不可替代的東巴文化研究的重要工具書。

洛克《納西語—英語百科辭典》分為上下兩冊, 由義大利羅馬東方學研究所於 1963 年和 1972 年出版。該書以讀音為序, 上冊收錄了 3414 條詞, 對每個字都有詳細的解釋, 有些還注明了來源的經書以及詳細的文化背景說明, 下冊是收錄的鬼神、祭祀儀式、星象和地名的詞, 共 4600 多個, 也注明了來源的經書。這部字典花費了洛克三十年心血才完成, 堪稱納西族的百科全書, 對研究納西族的文字、

地理、動物植物、文化等領域都具有重要意義。

這三部辭書各有長短，方書較為普及，使用者最多；李書解釋更為詳細，且有例證，使用者較方書次之；而洛克開始收集、研究的時間早於方李二人，且又是以一個外國人的眼光觀察納西社會，因此對東巴教的儀式、文字的描寫比較細緻深入，有許多前二者所沒有的優點。較之其他學者，洛克在麗江從事考察工作的時間較早，當時許多知識淵博的大東巴仍然健在，他挑選為他工作的東巴時又非常慎重，因此，在他對東巴教儀式及經書的翻譯和解釋中保留了很多珍貴資料，有許多後世年輕的東巴已解釋不清楚。<sup>[3]</sup>但是，這部辭典是英文版，至今還沒有中譯本，因而國內學者對此作研究或利用它進行研究的學者還不是很多，能見到的也都是一些介紹性的文章。

## 二、建立語料庫的必要性和可行性

### 1、必要性

對大多數研究者來說，目前的研究對象主要還是方國瑜的字典，輔之以李霖燦的字典。而洛克辭典因語言障礙利用率很低，不參照利用洛克的辭典，對東巴文乃至於東巴文化也就不能有一個全面正確的認識。即使要同時參照三部辭書，查閱也十分不便，影響工作效率。再者，由於方國瑜、李霖燦、洛克三人在調查翻譯東巴經時，選用了來自不同地區的東巴，因而同一字的讀音就不盡相同，三部辭書的標音系統也就不一致，形成了三套不同的音韻系統。方國瑜使用的是東部方言區麗江壩的讀音，<sup>[4]</sup>李霖燦採用的是西部方言區魯甸的發音，<sup>[5]</sup>洛克採用的是麗江北部和西部地區刺實的發音。<sup>[6]</sup>不同的音韻系統也給研究工作帶來了不便。對於文字的研究，我們常常要做一些定量分析，這就需要大量的統計工作，而傳統方法的做卡片、人工統計經常會出現誤差。如果利用語料庫的話，不但可以做到量化的精確可靠，而且還能多角度高效率地多次排序。

鑒於以上種種原因，我們認為，當務之急就是建立一個東巴文象形字典的語料庫，為研究者在資料收集、篩選、統計和排序方面提供方便，以促進東巴文和東巴文化的研究。

### 2、可行性

其它古文字在信息化處理方面已經取得了突破性進展，為東巴文的信息化處理提供了寶貴的經驗。如北京師範大學已經成功研制開發了小篆字庫和小篆輸入法；華東師範大學中國文字與應用中心已經掌握了古文字字庫的建設技術，並建成了金文、石刻文字、古璽文、古陶文的字庫和資料庫，甲骨文、貨幣文、簡帛文及各種少數民族古文字的字庫及資料庫也在進一步開發之中。此外，東巴文信息化處理已取得了突破，楊曉輝成功開發了“東巴象形文字計算機處理系統”，又稱作“電子東巴”。<sup>[7]</sup>該系統由一個字庫和兩種輸入法組成，在 Windows95 以上的操作系統下，能對東巴文和哥巴文的字符集進行處理。這個系統取材於方國瑜《納西象形文字譜》，為東巴象形文字和哥巴標音文字建立了標準字庫，收錄了 1562 個東巴文，661 個哥巴文，共 2223 字。這套軟件為我們建立東巴象形文字字典語料庫打下了良好的基礎。

## 三、語料庫建立的前提

### 1、選取標準字形

東巴文中的異體字特別多，每個字的異體字少則一個，多則七八個。這主要是因為東巴文一般由經師東巴所掌握，他們在傳抄經書的過程中，往往為顯示自己的法儀高明而隨意改變字形，因此當地有“自書自讀，行雲流水；人寫我讀，如犁石山”的說法。這三部字典所列代表字的字形略有差異，它們在解釋字義時，也列出了許多異體字。因此，我們首先要歸納並選取通用的標準字形以進入字庫。

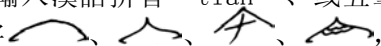
### 2、擴充東巴文字字庫

根據文字的計算機信息化處理要求，我們必須使每一個字形在字庫中擁有一個唯一的內碼碼位，這樣才能在計算機查詢、統計、篩選過程中成為唯一的處理對象，從而保證語料的準確可靠。但是，東巴文的字形是圖畫性的，規範程度很低，異體字也特別多，因此，我們必須將這些字形整理後才放入字庫，否則，計算機信息化就無法實現。建立東巴文字典語料庫的前提是建立一個東巴文字庫，這個字庫分為“基本字形”、“異體字”兩個層次。

收入字庫中的字數應該低於國際標準字庫系統要求的 6763 字。而三部東巴文字典中收錄的字和詞共計 7096 個，東巴文實際只有 1400 多個字，<sup>[8]</sup>減去三部字典中重複的字形約 1500 個，還有 5596 個，正好在國標允許的範圍之內，因此，我們能夠建立一個新的字庫並在 Windows 等平臺上使用。

已建成的“電子東巴”的字庫收象形字 1562 個，但是它的材料範圍僅限於方國瑜的辭典，遠遠不能滿足研究者的需要。因此，我們可以參照這個字庫，補充未收錄的另外兩部字典的字形。

### 3、建立輸入法

為了調用所建字庫中的字，我們還必須建立一種特殊的輸入法以調用這些新造的字。“電子東巴”採用的是兩種輸入法，一種是納西語讀音，另一種是中文拼音。但是，他所用的納西語的讀音既不是 1957 年制定的“納西文字方案”，也不是納西語的國際音標，因而這種輸入法沒有多大的實際意義，實質上只有中文拼音可以用來輸入。此外，“電子東巴”在輸入上還存在一個問題，我們在選定所需要的字後，出現的東巴文是與基本字形相關的一系列的字，如輸入“我”字，出現的是“我”的象形字和形聲字兩種字形，這對於非專業人士來說就不太容易選擇。在此基礎上，我們要改進這種一對多的關係，做到漢字與東巴文一一對應。我們將建立三種輸入法，一是漢語拼音輸入法，二是五筆字形輸入法，三是國際音標輸入法。以東巴文“天”字為例，若輸入漢語拼音“tian”、或五筆字形編碼“gd”、或國際音標“m<sup>33</sup>”，就會出現“天”的四個異體字，研究者根據需要再進行取舍。此外，還準備設計一種字形輸入法，這種輸入法與五筆字形輸入法有點類似，是根據東巴文的字形結構，按照字素分析的原則，將東巴文分成若干字素，然後按字素輸入。

## 四、語料庫的內容

建立這個語料庫的首要目的就是合併三部東巴文辭典。目前還有一些書收錄的字是上述三部字典中未收錄的，如趙淨修的《東巴象形文常用字詞譯注》中有 45 個字詞是字典中未收到的、<sup>[9]</sup>王超鷹的《東巴文字》中也有一些字典中未收錄的詞。<sup>[10]</sup>我們不但要增加上述兩本書中的字，還將利用最新的研究成果，對這三部字典中字詞進行補充。同時，這個語料庫還是一個開放性的系統，隨著對《納西東巴古籍譯注全集》及其它文獻的深入研究，我們還將進一步增加新發現的字。

這個語料庫包括以下內容：

### 1、統一的字號

方國瑜辭典中，有編號的字共 1340 個，有的字下有派生字，全書共有 222 個派生字，我們使用時為作區別常標作“譜 XX”；李霖燦的字典中每個字均有編號，共 2120 個，標作“象 XX”；而洛克辭典未標號，我們標作“XX.X”，“.”前面的數字為頁碼，後面則為從上至下第幾個字。該語料庫為三部字典所收的字形統一重新編號，同時列出原有的編號，以方便查閱原文。

### 2、字形

東巴文的異體字特別多，我們將整理出一個標準字形作為字頭。在這個界面下，只要點擊一下鼠標，就可以看到其他的異體字，並注明出自哪本字典或經典。

### 3、讀音

方國瑜、李霖燦及洛克的音韻系統不同，李霖燦採用的是西部方言區魯甸的發音，有 43 個聲母，12 個韻母，4 個聲調，調值分別是：低調 31、中平調 33、高平調 55、低升調 13。方國瑜使用的是東部方言區大研鎮麗江壩的讀音，有 31 個聲母，21 個韻母，4 個聲調，調值分別是：低調 21、中調 33、高調 55、低升調 24。洛克採用的是麗江北部和西部地區刺實的發音，他使用了自創的一套音標，稱作“洛克音標”，洛克音標與國際音標對應關係複雜，不便於閱讀。洛克音標有 37 個聲母，29 個韻母，4 個聲調，調值與方國瑜相同。這個界面下將同時列出各自的讀音，並把洛克音標改作國際音標。

#### 4、釋義

該語料庫在列出三部字典各自釋義的同時，指出辭典中有疏誤的地方，增加最新的研究成果。在這個界面下，我們可以通過意義查詢字詞。

#### 5、字與詞的界定

這些字典所收錄的內容不完全是字，還包括字與字的組合（即字組或詞）。字的數量是有限的，而字組的則不同，它的數量很大。該語料庫劃分字組的原則依據喻遂生師歸納的方法，即，獨體為字，合體單音節為字，合體多音節中有表義的形符為形聲字，合體多音節中形體與詞語不能對應的為會意字。<sup>[1]</sup>以前的研究對象往往是字和詞混淆不清，通過這個用戶界面，我們就能準確知道這三部字典中所收的字與詞的具體數字，並可對字與詞分別研究。

#### 6、造字法

東巴文雖然比較原始，但是它已經具有完備的造字法。我們將分析每一個字的造字法，並在語料庫中標明。

#### 7、分類

方國瑜和李霖燦均把東巴文分為 18 類，但是他們所分的類別不同，並且對字的歸屬也有不當之處，這個語料庫把東巴文分為天文、地理、動物、植物、人文、服飾、飲食、用具、建築、人名、地名和若喀字等 12 類。

#### 8、詞例

李霖燦、洛克辭典在解釋詞義時，常常會舉出詞例，特別是對於假借字的用法，往往舉出不同的例子。該語料庫在列出原有詞例的同時，還將參照利用《納西東巴古籍譯注全集》增加一些詞例。

#### 9、出處

洛克辭典的特點之一就是他注明了所釋詞的來源，這對於我們查閱原始資料很有幫助。該語料庫將發揚這一優點，對一些特殊辭彙注明來源經典。

---

#### 注释

[1] 方國瑜《納西象形文字譜·弁言》1—3 頁，云南人民出版社，1981。按：“弁言”原作“三六年成書”，誤。

[2] 聞宥《評〈麼些象形文字字典〉》，《燕京學報》第 30 期，1946 年 6 月；又《東巴文化論集》480—481 頁，雲南人民出版社，1985 年。

[3] 楊福泉《西方納西東巴文化研究述評》，《雲南社會科學》1991 年 4 期。

[4] 方國瑜《納西象形文字譜·緒論·納西語的音標說明》81 頁，雲南人民出版社，1995 年。

[5] 李濟《麼些象形文字標音文字字典·序》，台灣文史哲出版社，1972 年；李國文《人神之媒——東巴祭司面面觀》145-147 頁，雲南人民出版社，1993 年。

---

[6] J.F.Rock. *Introduction. Classification of the Na-Khi Language. A Na-Khi –English Encyclopedic Dictionary*, Roma Istituto Italiano Peril Medio ed Estreme Orientale. Pe.I 1963.

[7] 這套軟件於 2001 年 6 月問世，2002 年 10 月，作者又將哥巴文收錄進來。可以在 <http://www.edongba.com> 下載。

[8] 喻遂生《納西東巴字字和字組的劃分及字數的統計》，《語苑擷英》，北京語言文化大學出版社，1998 年。

[9] 趙淨修《東巴象形文常用字詞譯注·前言》，雲南人民出版社，1995 年。

[10] 王超鷹『トンバ文字』，株式會社ホームマル社，1996 年。

[11] 喻遂生《納西東巴字字和字組的劃分及字數的統計》，《語苑擷英》，北京語言文化大學出版社，1998 年。

## Conception of Building a Data Base of Dong-ba Pictograph Dictionaries

GAN Lu

(Center for the study of Chinese Characters and Their Applications, East China Normal University, Shanghai 200062)

**Abstract:** Basing on Fang Guo-yu's *the Manual of Na-xi Pictograph*, Li Lin-chan's *Mo-so Hieroglyphics*, and Rock's *A Na-khi-English Encyclopedic Dictionary*, This paper will build a data base about these three dictionaries. This data base includes 9 aspects: 1.the number of Dong-ba pictograph; 2.the form of Dong-ba pictograph; 3.the pronunciation; 4.explation; 5.the definition of pictograph; 6.the types of forming the pictograph; 7.the classifier of the meaning of pictograph; 8.the example of pictograph; 9.the reference of pictograph.

**Key words:** Dongba writing; pictograph; database