



圖一編輯部
文一蔡佩珍
(政治大學原住民族研究中心專任助理)

四方に響く声 民族語朗讀文章データベースの計画と創設
Planning and Construction of the Database of Aboriginal
Languages Oral Reading Articles, Reading Text Corpus

朗聲四起 族語朗讀文章的規劃與建置

政大 原住民族研究中心受教育部(終身教育司)委託,首次有系統規劃編纂族語朗讀文章,歷時2年,於去年(2016)5月完成了國小組、國中組、高中組42語別舊稿匯整與新稿編纂工作,總計匯整出舊稿252篇,新撰稿1,008篇,並將匯整成果(含歷屆稿件和新撰稿件)計1,260篇,經過「朗聲四起」平台的建置,成為數位化資料,使全國語文競賽參賽者以及對族語學習有興趣的讀者,可以透過這個網站資料庫(上網搜尋「朗聲四起」),進行各族族語朗讀文章檢索、瀏覽、下載,相當便利。

網站設計風格簡約

「朗聲四起」數位化平台不同其他平台那樣繽紛繁雜,而是以簡潔方便的設計為原點,輔以大地色與拉阿魯哇族的貝殼圖紋,呈現出簡約的風格。「眼睛」、「嘴巴」、「書本」等icon,隱喻族語「視」、「說」、「讀」多種能力的展現,國小組以樹木萌芽圖示、國中組以樹木成長圖示、高中組以樹木茁壯圖示,清晰而生動地標誌著各組學生族語能力的學習歷程。

從選文到公告的艱辛歷程

2013年以前,全國語文競賽原住民各族語言的文章,僅收集族語文稿,缺漏甚多,且品質不均。鑑於當前族語環境弱勢,提供適切可讀的族語文章,以及漢語對譯,讓參賽者參考,比較容易有效地達成學習與練習的目的。從2014年開始,我們進行舊稿匯整、族語譯稿、新稿編纂等3項工作,配合全國語文競賽的期程,編



「朗聲四起」的首頁 (<https://alcd-con.tw/alr/>)。



纂期間歷經繁複的排版、校對、選文、公告的作業，為能簡化執行流程並完成文稿數位化作業，建置「朗聲四起」數位化平台，透過程式設計，讓撰稿委員及校閱委員都不受時間與空間限制，線上撰寫與校閱，實為時勢所趨。

朗讀文章公告分為兩個階段。第一階段為初稿公告，提供WORD 與PDF 兩種檔案格式下載，方便讀者糾正文章內容的錯誤，公告期間為1 個月，公告期間透過平台，匯整各方意見並修訂族語文稿。第二階段則為定稿公告，定稿公告後即不再修訂朗讀文稿，成為全國語文競賽的正式用稿。

從選文到公告，過程相當艱辛，除了委員們的密切配合上網交作業之外，校訂期間還要不斷校閱，反覆閱讀確認。所幸撰稿委員們都絞盡腦汁，在有限的時間內提供文稿，同時在校閱過程中商榷各篇的文字及標點。

體例和題材都嚴謹規範

各族的文章基本上參酌族語的基本常用詞彙表（千詞表），並有完整的結構且文體適合朗讀，內容則以生活化、學習之趣、敘事抒情

「朗聲四起」網頁設計簡潔，大地色與拉阿魯哇族的貝殼圖紋，呈現出簡約風格。「眼睛」、「嘴巴」、「書本」等icon，隱喻族語「視」、「說」、「讀」多種能力的展現，國小組、國中組、高中組分別以樹木從幼苗茁壯成大樹的圖示，標誌著學生族語能力的學習歷程。



「朗聲四起」可依據國小、國中、高中等組別，並根據年度、族語別、語別搜尋朗讀文章。

或現今部落生活及原住民族各族文化核心價值為主。字數方面，國小組200 至300 詞；國中組250 至350 詞；高中組300 至400 詞，並附有漢語參考譯文。

撰寫文章的初衷是以族語原創稿為主，並輔以漢語對譯。然而目前的撰寫方式多採翻譯的方式進行（先有漢語版，才有族語版），可能產生文章太難的情況，但撰稿委員能力優秀，大多能轉譯，使語意對等，不致脫離漢語原文太遠。例如，「伊索寓言」故事，轉譯為族語後，必須再修正語氣及若干詞彙，使其適



現今各民族語言的保存，有相當大一部分需要憑藉數位化，族語朗讀文章資料庫的建立，提供族語使用者多元學習的管道。假以時日，這個資料庫不斷累積文章的量與質，應可成為政府推展族語文字化的一項重要成果。



105 年度全國語文競賽原住民族語朗讀題目雙語選文拉阿魯哇語朗讀文章〈ihlahlamu ia Hla'alua 我們是拉阿魯哇〉(族語版)。

稽。大族採聯合命題相互轉譯，例如卑南族、布農族、排灣族、阿美族、賽德克族。小族則各自獨力撰稿，但大多不約而同地以傳統文化或祭典敘述為題材。然而，撰稿者提供文稿頗為多元，有時仍須克服文章過難或詞數過多的問題，這方面就需要校閱委員的把關。

在文章的來源方面，歷年來的文章收集以邀稿為主，將來除了邀稿之外，也希望朝向徵稿方式同步進行，讓稿源及文章內容更為多元。

文章匯集程序的簡化

朗讀文章之所以需要建議數位化資料庫，係由於撰稿與校閱流程可以大大地簡化。過去，以E-mail 郵件進行撰稿與校閱的方式，容易發生漏信，且不易匯整。現在，撰稿者與校閱者直接在「朗聲四起」進行撰稿與校閱的功能，大幅提升了文稿匯整的效率。撰稿者不但可查詢歷年文章，避免重覆。校閱者可以輕鬆複製詞句、段落或全文來進行校閱，輔以顏色標記，更利於撰稿者參考修訂。

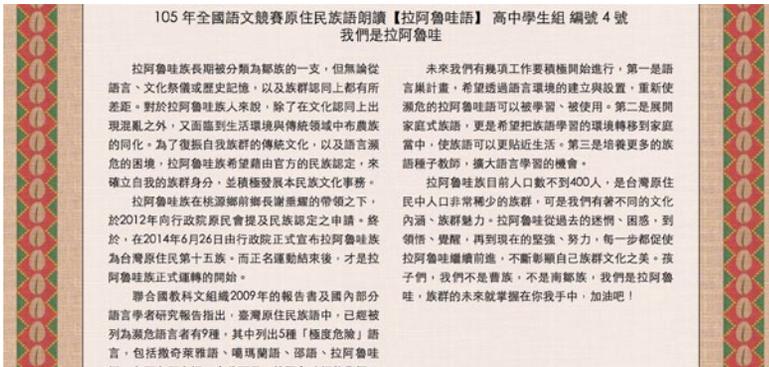
另一方面，資料庫管理系統不僅是朗讀文稿存放的場域，更提供了管理者系統公告、選文、匯整意見等功能，增加了互動性，亦簡化

合「國小組」來朗讀。2015 年，阿美族與撒奇萊雅族合作，以共同討論並投票選文的模式，選出適合的漢語文章並轉譯為各語別族語文章。另外，許多民族的族語書面化過程與族語聖經密切相關，不難發現有些文章題材是源自於《聖經》經文內容，再經改寫而成。

42 語文章逐年改善

面對年年增加的新稿（2014 年126 篇、2015 年378 篇、2016 年504 篇），懂得族語書寫符號的撰稿者，與其他本土語言（台語及客語）相比，人才其實仍相當缺乏，再加上語別多，涉及人力也多，邀稿及催稿過程也千辛萬苦。

雖然是分族分語別撰稿，但也有規則可



105 年度全國語文競賽原住民族語朗讀題目雙語選文拉阿魯哇語朗讀文章〈ihlahlamu ia Hla'alua 我們是拉阿魯哇〉（漢語對照）。

本計畫的執行流程。

與其他本土語言相比，原住民族語言的朗讀文章數量龐大，稿件較為繁瑣，但由於「朗聲四起」資料庫的規劃與建置不斷精益求精，使得使用者一致認為檢索方便，有助於比賽。事實上，其他本土語言的朗讀文章並未設計資料管理及文章檢索的功能，使用起來較為遜色。「朗聲四起」具備公告、編輯、檢索（族語關鍵詞或漢語關鍵詞）、校閱、回應、選文、儲存、下載等多種實用功能。是故，這個資料庫可做為未來全國語言競賽的整合性平台的重要參考。

允許個別化文稿的產出

「朗聲四起」裡的族語文章，以「原住民族語言書寫系統」書寫，由於此系統的文字含有特殊符號，為了避免使用者電腦作業系統不同造成符號顯示出現問題，而導致使用者判讀錯誤，同時考量原住民族語言的內部差異，以不改變文章內容及句子本意，均可接受縣市參賽代表，進行個別化文稿修正。「朗聲四起」文章的下載格式並未鎖定，使用滑鼠右鍵即可

複製，並依格式進行文稿的編排。

文稿數位化不再只是單方面下載，還特別設計了「疑義回報功能」，提供使用者與管理者互動與溝通管道，使用者的意見提供可以為本計畫的執行帶來加乘作用。然而，過去收到的疑義或建議，有許多並非資料庫本身的問題，而是地方語言差異（詞形變體或音音變體）或詞彙書寫方式（正詞法）的不同意見。這些問題尚待學者專家商討方能解決。

族語學習的多元管道

現今各民族語言的保存，有相當大一部分需要憑藉數位化，透過族語朗讀文章資料庫的建立，可以使族人隨時隨地使用電腦、手機或平板進行檢索、瀏覽、下載。實用的資料庫，不僅做為教育部朗讀文章專屬平台，更提供族語使用者多元學習的管道。假以時日，這個資料庫不斷累積文章的量與質，應可成為政府推展族語文字化的一項重要成果。◆



蔡佩珍

嘉義縣Holo人，1979年生。政治大學原住民族研究中心專任助理，2015-2016年執行「原住民族語朗讀文章編輯暨數位化工作」計畫。