



台灣南島語數位典藏

第一期計畫研究成果



計畫主持人：齊莉莎 研究員

版權所有：中央研究院 語言學研究所



中央研究院語言典藏

- 「台灣南島語數位典藏」是國科會NDAP機構計畫下分項主題為「語言典藏」的子計畫之一，整體建置作業之規劃暨主持人為齊莉莎研究員。
- 此計畫之建置目的在於建立一個數位圖書館，期能收集、保存、編輯及透過網路散播語言資源，供使用者存取已錄製及記音的南島語。
- 第一期計畫為：2002-2006，現已進入第二期計畫（2007-2011）。另外在第一期計畫開始的前一年（也就是2001年）先執行實驗研究計畫



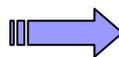
台灣南島語語料庫之主要特色

■ 呈現方式：

1. 中文
2. 英文

■ 三個資料庫：

1. 語料庫
2. 地理資訊系統
3. 書目資料庫





語料庫查詢介面



語料顯示



檢視語料庫 - Windows Internet Explorer

檔案(F) 編輯(E) 檢視(V) 我的最愛(A) 工具(T) 說明(H)

http://formosan.sinica.edu.tw/formosan/ch/text_frame.asp

Google

檢視語料庫

■ 搜尋範圍: 魯凱語/萬山 文體: 全部 語料呈現: 全部 語料尋找: 瀏覽模式

■ 重新檢索

文本別: 1. 我們的話 [後設資料](#) 選擇其他文本

檢視別: 句子 - Sentence 顯示

請選擇您要看的內容:

翻譯(Translation) 中文

原文(Transcription)

單字註解(Word-glosses) 中文

全選 清除 繼續

完成 國際網路 100%

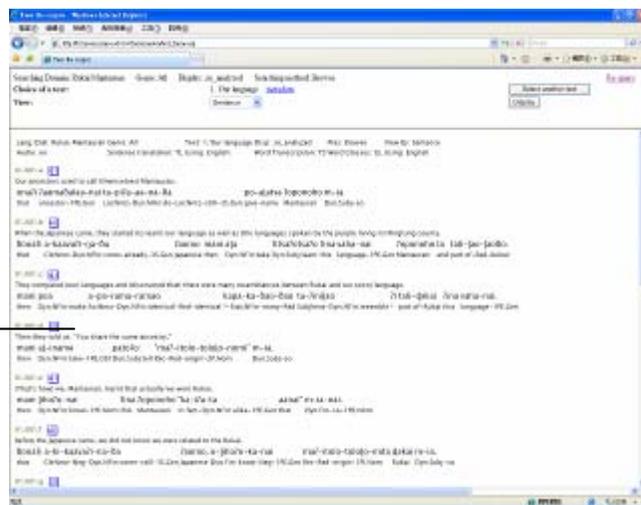
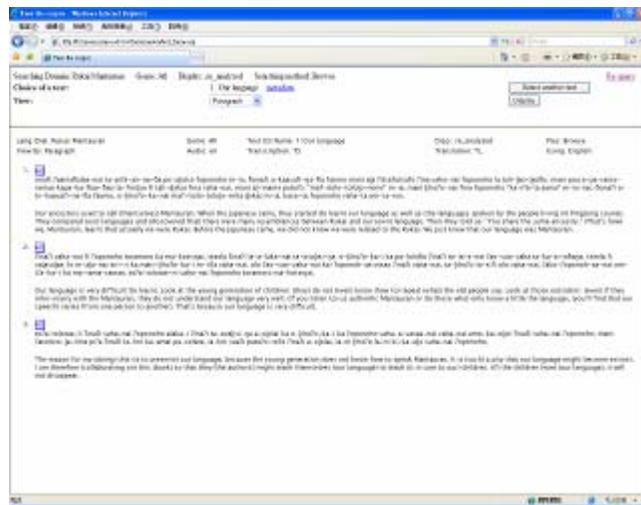


段落/句子顯示

中文顯示



聲音檔



英文顯示



原始 vs. 重新編輯之語料

The screenshot shows a web browser window titled "檢視語料庫 - Windows Internet Explorer". The address bar shows the URL "http://formosan.sinica.edu.tw/formosan/ch/text_frame.asp". The search criteria are set to "魯凱語/萬山" (Rukai/Wanshan) and "句子 - Sentence".

The main content area is split into two columns:

- Left Column (Original):** Labeled "原始" (Original). It shows the raw linguistic data for "01.001.a" through "01.001.g". The text is in a raw format, such as "ona?i ?aama?alaə-nai ta-pi?a-aə-na-?a po-a|atsə ?oponco".
- Right Column (Revised):** Labeled "重新編輯" (Revised). It shows the same linguistic data but in an edited format. For example, "01.001.a" is translated as "我們的祖先萬山自稱是萬山人。" (Our ancestors Wanshan call themselves Wanshan people.).

Both columns include a header with search parameters: "Lang/Dial: 魯凱語/萬山 Genre: 全部 Text: 1/我們的話 Disp: 全部 Format: raw Audio: on Sentence translation: TL Using: Chin".



詞彙分佈

檢視語料庫 - Windows Internet Explorer

檔案(F) 編輯(E) 檢視(V) 我的最愛(A) 工具(T) 說明(H)

http://formosan.sinica.edu.tw/formosan/ch/text_frame.asp

檢視語料庫

■ 搜尋範圍 - 語言別: 魯凱語/萬山 文體: 全部 語料呈現: ; 語料尋找: 瀏覽模式 ■ 重新檢索

文本別: 1. 我們的話 [後設資料](#) 選擇其他文本

檢視別: 詞語索引 - Concordance 顯示

a-kaava7i-rja- ða	1	001.b ðona7i a-kaava7i-rja-ða 7aomo mani a[a 7itsa7otsa7o 7ina vaha-nai 7oponoho la tali-[ao-[aoðo.
a-ki-kaava7i- na-ða	1	001.c mani poa a-pa-rama-ramao kapa-ka-ðao-ðao ta-7ini[ao 7i tali-qəkai 7ina vaha-nai.
alaka-i	1	001.d mani a[-inamə pato7o: "ma7-itolo-tolo[o-nomi" m-ia.
amo-ka-o[pi	1	001.e mani [iho7o-nai 7ina 7oponoho "ka-i7a-ta aanai" m-ia-nai.
a-o[olai	2	003.c mani 7atoloro-[a-ilina pi7a 7ina7i la-lini ka-amai pa-solata, la-lini vaa7i pato7o-to7o 7ina7i a-o[olai, la-ni [iho7o la-ni ki-ka-o]o va
a-pa-rama- ramao	1	
a-vasaa-nai	1	
ka	6	
ka-7oponoh-aa	1	
ka-amai	1	
ka-i7a-ta	1	
kapa-ka-ðao- ðao	1	
kasa-ta	1	
ki-ka-o]o	1	
la	1	
la-a-laka-nai	1	
la-lini	2	
la-ni	2	
lo	1	
m-a[a-nai	1	
ma7-itolo- tolo]o-mita	1	
ma7-itolo- tolo]o-nomi	1	
ma-kotsinai	2	
mani	5	
ma-rama-ramao	1	
mati-[iho7o-ka- i	1	
m-ia	3	
m-ia-nai	1	
m-ii7a	1	
o-[iho7o-ka-i	2	
o-liho7o-ka-nai	1	

完成



關鍵詞

台灣南島語語料庫 - Windows Internet Explorer

http://fomosan.sinica.edu.tw/fomosan/ch/arch_form.asp

內容檢索

回上頁 使用說明 回首頁

搜尋範圍一語言別: 魯凱語/萬山 文體: 全部 語料搜尋方式: 搜尋模式

關鍵詞(原文):

關鍵詞(註解):

詞類選單:

人稱代名詞選單:

執行 清除

關鍵詞示例:

1. 關鍵詞以原文輸入時,若使用到 IPA 音標,請點選對應的代碼如下:

IPA	ð	ŋ	ɔ	a	l	ʔ	i	θ	ʌ	ɤ
代碼	240	331	598	601	621	660	616	952	654	620

IPA	∅	f	æ	œ	β	ɬ	ʎ	j	ʈ
代碼	216	643	230	339	946	649	404	607	648

2. 關鍵詞以註解輸入時,請輸入可能出現的字(如:酒、芒草、書)

國際音標，可繼續增加新符號

搜尋結果 - Windows Internet Explorer

http://fomosan.sinica.edu.tw/fomosan/ch/arch.asp

搜尋結果

下一頁 最後一頁 選擇頁次: 第1頁

- 魯凱語-萬山-口述 (01_001_b)

原文: ðonaʔi a-kaavaʔi-ŋa-ða ʔaomo **mani** əla ʔitsaʔotsaʔo ʔina vaha-nai ʔoponoho la tall-ʔao-ʔaoðo.

註解: 那 分句名物化-動態-非限定 來-已經-他-屬格 日本人 就 動態-非限定-拿 動態-虛擬式-舉 話-我們-屬格 萬山 和 屬於-重疊-下面

中譯: 日本人來了以後就(開始)學我們萬山和屏東縣(排灣及魯凱族)的話。

英譯: When the Japanese came, they started (to learn) our language as well as (the languages spoken by the people living in) Pingtung county.
- 魯凱語-萬山-口述 (01_001_c)

原文: **mani** poa a-pa-rama-ramao kapa-ka-ðao-ðao ta-ʔiniʔao ʔi tall-ɔəʔai ʔina vaha-nai.

註解: 就 動態-非限定-弄 動態名物化-動態-非限定 相同-重疊-相同 *-狀態-非限定-重疊-多 主語名物化-動態-非限定 像 * 屬於-魯凱 話-我們屬

中譯: 他們比較兩種語言後, (就發現) 我們的話有許多 (和) 魯凱族是一樣的。

英譯: They compared (our) languages and (discovered that) there were many resemblances between Rukai and our (own) language.
- 魯凱語-萬山-口述 (01_001_d)

原文: **mani** əl-iname patoʔo: ʔmaʔ-itolə-toʔo-nomi m-ia.

註解: 就 動態-非限定-拿-我們 斜格 動態-虛擬式-告訴 相互-重疊-來源-你們-主格 動態-虛擬式-這樣

中譯: 然後他們告訴我們說: 「你們是同宗 (的)。」

英譯: Then they told us: "You share the same ancestry."
- 魯凱語-萬山-口述 (01_001_e)

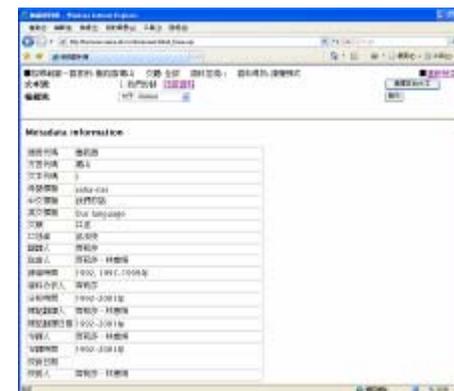
原文: **mani** lihoʔo-nai ʔina ʔoponoho "ka-iʔa-ta aanaŋ" m-ia-nai.

搜尋結果



後設資料

- 所謂「後設資料」乃有關資料的結構性資料，它一方面能夠幫典藏者利用結構化的典藏方式、標記及文本編目加以管理所收集的資料；另一方面，使用者也能夠更容易從不同的典藏品存取所需的資料。
- 後設資料的描述主要為超連結的資訊。換句話說，在南島語典藏裡，它包含如下的資訊：（一）語料的側寫（如：題目、語言及方言、文體），（二）田野調查（如：發音人、田野工作者/謄寫者/編輯者/翻譯者、資料收集/編輯/翻譯的日期）以及（三）管理策略（版權宣告）。
- 為了統一術語起見，以及降低每個文本編目的負擔，分別建立語言資料庫與個人資訊的資料庫。語言資料庫包含的資訊為地理分佈、方言及方言變體、人口以及語言使用狀況，而個人資訊的資料庫則提供了發音人姓名、所屬的語言族群、年齡、語言能力、出生地等等，以及其他關於田野工作者/謄寫者/編輯者/翻譯者的基本資訊。「超連結」進一步使使用者能夠從文本後設資料內的「語言」、「發音人」及「參與者」等元素取得更多資訊。





地理資訊系統 (GIS)

■ 在地理上：

- 可以呈現台灣南島語言分佈
- 可以比較族名、及中英文對每一個族的稱呼。

■ 在語言學上：

- 可以比較不同南島語語言/方言
- 可以進一步建立台灣南島語之分支



地理分佈

臺灣南島語空間資料查詢系統 使用說明

語言分佈 辭彙查詢

LayerLi...
套疊/隱藏
關閉視窗

- 賽夏
- 巴則海
- 噶瑪蘭
- 道卡斯
- 邵
- 卡那卡那富
- 沙阿魯阿
- 鄒
- 阿美
- 卑南
- 猴猴
- 賽德克
- 布農
- 魯凱
- 排灣
- 雅美
- 凱達格蘭
- 泰雅

圖層
套疊選擇
 縣市界
 鄉鎮界
 村里界
重畫地圖

GIScience

台北縣 [請選擇鄉鎮] 放大至所選區域

「縣市界」為現在啟動的圖層



詞彙分佈

臺灣南島語空間資料查詢系統

query_item - Microsoft Internet Explorer

語言分布 辭彙查詢

身體
血
【請選擇語言】

語言 (Language)	方言 (Dialect)	方言變體 (Dialectal Variant)	詞彙	同詞
魯凱	大南		aray	
魯凱	霧台		aray	

大南分布範圍

縣市編碼	縣市名稱	鄉鎮編碼	鄉鎮名稱	村里編碼	村里名稱
10014	台東縣	1001404	卑南鄉	1001404012	東興村

放大至分布範圍

圖層
數量選擇

- 縣市界
- 鄉鎮界
- 村里界

重畫地圖

1) 如何使用IPA/聲音檔?
2) 如何比較不同方言
3) 如何對照語料與地圖

1) 時間/空間
2) 不同記錄
3) 43方言



書目資料庫

1. 語言學



持續更新及擴增中

2. 語言教學



3. 文學



4. 音樂

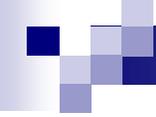




台灣南島語言數位典藏團隊

• 語料分析	魯凱語	齊莉莎、林惠娟、邢天馨*
	賽夏語	朱黛華、齊莉莎
	泰雅語	葉郁婷*、齊莉莎
	布農語	齊莉莎、劉秋雲*、林秀蓮、林聖賢、曾思奇、李文甦
	排灣語	華加婧*、齊莉莎
	卑南語	Josiane Cauquelin、齊莉莎、鄧芳青
• 語料輸入	布農語	林志憲
	鄒語	
	巴宰語	
	卡那卡那富語	
	西拉雅語	
	排灣語	
	魯凱語(大南方言)	
阿美語	林翠緯*、林秀蓮	
• 地理系統	華加婧、白璧玲*	
• 系統工程師	余清華	
• 後設資料小組	翁翠霞*、沈漢聰*	
• 書目資料庫	齊莉莎、劉秋雲、華加婧	

* 已不在此計畫底下服務。



出版文獻

- Zeitoun, Elizabeth, Yu Ching-hua and Weng Cui-xia. 2003. The Formosan Language Archive: development of a multimedia tool to salvage the languages and oral traditions of the indigenous tribes of Taiwan. *Oceanic Linguistics* 42.1: 218-232.
- Zeitoun, Elizabeth and Yu Ching-hua. 2005 Language analysis and language processing. *Computational Linguistics and Chinese Language Processing* 10.2: 167-200.
- Zeitoun, Elizabeth and Yu Ching-hua. 2007. The Formosan Language Archive. In Chang H. (ed.) *The Indigenous Languages [of Taiwan]: Theory and Practice*, 28-42. Taitung: Council of Indigenous Peoples, Executive Yuan. [In Chinese]