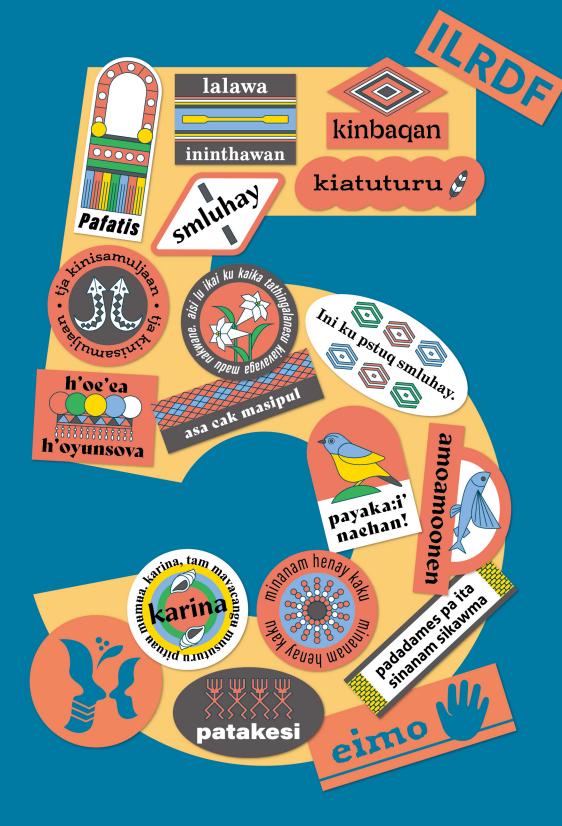


Issue 4



Results of the Indigenous Languages Research and Development Foundation: 5-Year Anniversary Special Edition

We extend our heartfelt gratitude to our dedicated partners that have been with us on the journey to revitalize indigenous languages. It is because of you that our languages and cultures continue to thrive and be passed on.

Contents

tousvusvutu, meaning "conversation" in the Tsou language, serves as the title of this publication. Both the cover text and table of contents are presented in Tsou, while the stickers feature words from the languages of each of Taiwan's 16 indigenous peoples.

Promotion and Implementation

Indigenous Language Promotion the

Indigenous Language Promotion Ac

Preservation and Application

Indigenous Language Corpus
Indigenous Language Online Diction

Research and Development

Indigenous Language Learning Aid

Indigenous Language Publications

Progress in AI Research and Develo

About the Indigenous Languages Research and Development Foundation (ILRDF)

Words from the Chairman	6
Organizational Structure	10
Board of Directors and Supervisors	12
Major Events	14



Key Execution Achievements

Education and Transmission

Taiwan-Austronesian Indigenous Words and Narrations	20
Indigenous Language Proficiency Certification Test (ILPCT)	26
Cultivation of Indigenous Language Professionals	34



Indigenous Language Learning Resources Platform

rough New Media	42
ctivities and Competitions	43

	54
onaries	58

ls	64
	69

opment for	Indigenous	Languages	74
------------	------------	-----------	----

About the
LanguagesIndigenous
Research and
FoundationDevelopmentFoundation(ILRDF)Indigenous



The Indigenous Languages Development Act was enacted to promote the preservation and growth of Taiwan's indigenous languages (hereafter referred to as "indigenous languages") and to safeguard the rights of their use and transmission. At the end of 2019, the Council of Indigenous Peoples established the Indigenous Languages Research and Development Foundation (ILRDF), focusing on four core areas: indigenous language promotion and marketing, research

and development, preservation and innovation, and certification testing. As Taiwan's first organization dedicated solely to the research and development of indigenous languages, the ILRDF has since worked tirelessly to revitalize indigenous languages.



Rooted in Our Land, Letting Our Languages Bloom Beautifully



Siwkolan Amis

Since our founding, the ILRDF has actively undertaken the mission of preserving and promoting indigenous languages, and we have begun to see encouraging results. In 2022, we launched tousvusvutu, a multilingual "Conversation" magazine presenting indigenous languages from all 16 recognized tribes alongside Mandarin. This publication not only showcases the ILRDF's language initiatives but also connects readers with language revitalization efforts across Taiwan and around the world. Through visual design, indigenous language recordings for listening and reading, and other multimedia, the magazine broadens the view of indigenous languages and fosters

Words from the Chairman

opportunities for indigenous peoples to engage with their mother tongues. Our "Taiwan-Austronesian Indigenous Words and Narrations" website, now in its 10th year, has become the primary platform for learners studying indigenous languages and preparing for certification exams, reaching over 34 million users and drawing more than a million monthly views. This clearly demonstrates the growing public appreciation and recognition of indigenous languages.

7

The Online Indigenous Languages Dictionary, first launched in 2007, underwent a major revision under the ILRDF's leadership in 2021. This update included optimized search tools, the addition of Mandarin sorting, and the development of an editing platform for 42 language varieties, significantly enhancing the dictionary's comprehensiveness. To date, the dictionary has been accessed over 30 million times. In 2022, we established the Indigenous Language Corpus, which collects, preserves, and applies both spoken and written language resources, as well as archival materials from each indigenous group. By the end of this year, we aim to catalog 4 million words from Taiwan's 16 indigenous peoples.

To address the demand for certification tests, we introduced the "Indigenous Language Proficiency Certification Test Online Practice Program" (ILPCT Online Practice Program) in 2021. This helps candidates simulate the testing environment and question types, allowing them to build valuable preparation experience. The online practice program has garnered over 10 million clicks, and last year, the pass rate for the ILPCT reached a recordbreaking 78%. All in all, the ILRDF has demonstrated significant results in indigenous language talent development, promotional activities, teaching material development, and publications.

As technology advances, generative AI and related applications have reached new levels of maturity. The emergence of ChatGPT, an AI model powered by deep learning natural language processing, has demonstrated the possibilities for natural language generation and translation. This breakthrough brings new potential and inspiration to indigenous language revitalization efforts. Embracing technological advancements, the ILRDF launched Phase 1 of the "Indigenous Language Corpus and AI Translation Project" at the end of last year. Through the development of foundational AI systems, we aim to reduce the manual workload of transcription and translation, streamlining text conversion and accelerating the accumulation of language data. This foundation will also open new possibilities for AI applications in indigenous languages.

As time passes, the ILRDF is now approaching its 5th anniversary a milestone that inspires us to look to the future with optimism. With the support of innovative technology, we are committed to actively integrating indigenous languages into family life and systematically advancing



research, promotion, preservation, and certification efforts. Our goal is to ensure that these languages, which have thrived on this island for thousands of years, remain deeply rooted and continue to be passed down for generations to come.

Organizational Structure

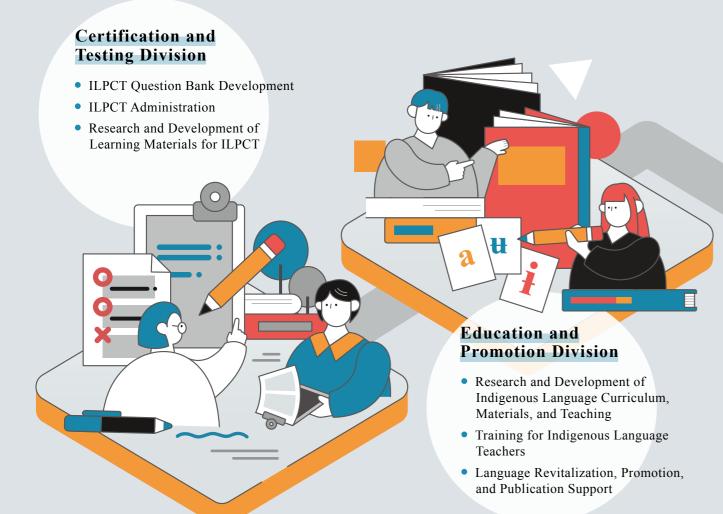


Languages

Takibakha Bunun

About the ILRDF

The Council of Indigenous Peoples has donated funds to establish the Indigenous Languages Research and Development Foundation (ILRDF) in accordance with Article 27 of the Indigenous Language Development Act. The ILRDF is the first organization dedicated to the research and development of indigenous languages in Taiwan. It focuses on indigenous language promotion and marketing, research and development, preservation and innovation, and certification and testing. Embodying the spirit of the Indigenous Language Development Act, the ILRDF comprehensively promotes language revitalization efforts to continue the transmission and sound development of indigenous languages.



Board of Directors

Research and Development Division

- Policy Research and Capacity Survey for Indigenous Languages
- Indigenous Language Dictionary Compilation
- Translation and Editing of Indigenous Languages
- Indigenous Language Corpus Development





List of the first board of directors and supervisors of the ILRDF



List of the second board of directors and supervisors of the ILRDF



12

of Education

Han

Law, National Taipei

University of Education

Truku

Paiwan

Amis

Major Events



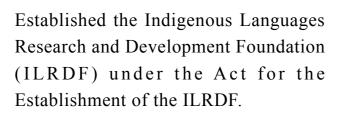
Announced the Act for the Establishment of the Indigenous Languages Research and Development Foundation.



December



The 1st session of the 1st Board and Supervisors Joint Meeting elected Ying-Jie Wen (Mo'o E'ucna) as the first Chairman and Ching-Tarn Wu as the first Standing Supervisor.



December

Listening to

Indigenous

ILRDF

2019

At the 3rd session of the 1st Board and Supervisors Joint Meeting, real-time indigenous language interpretation was first introduced, a practice that continues today.



2020

Official operations of

the ILRDF began: The

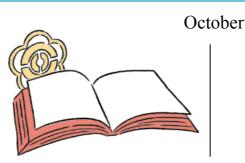
first group of ILRDF

Takibakha Bunun

employees joined in.

April

June



August

Reported to the Legislative Yuan with bilingual written documentation for the first time.



Started management, content development, and optimization of the "Taiwan-Austronesian Indigenous Words and Narrations" website.



Held the first international exchange event, the 2020 Forum on the Revitalization and Exchange of the Austronesian Languages.



December



February

November

September



Completed development of the ILPCT

Online Practice Program, followed by its

launch on the Lokahsu ILPCT website.

021221

Organized various talent training programs, including indigenous language interpretation and translation, language test question design, scriptwriting, Podcast production, and indigenous language MC courses-initiatives that continue to this day.



Promoted digital learning for Taiwan's indigenous languages via the PaGamO platform for gamebased indigenous language learning.

Co-hosted the first "International Mother

Language Day—Indigenous Language

Revitalization Awards Ceremony and

Development Conference" with the CIP,

along with the International Mother

Language Day Series Exhibition: "Let's

Speak Indigenous Languages."



Co-hosted the first national finals for the Indigenous Language Drama Competition with the CIP.



Completed the first-ever translation of indigenous laws and regulations in 16 indigenous languages.



Published the first online indigenous language audio magazine tousvusvutu.



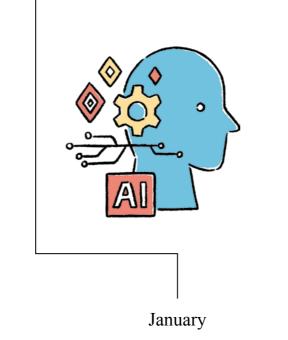


Released the first audio teaching aids

"Indigenous Language Learning."

Passed the ISO 27001 certification for Information Security Management Systems (ISMS) for the first time.

Signed the first domestic MOU with the National Center for High-Performance Computing: Indigenous Languages AI Platform Development and Educational Application.



2023

The 1st session of the 2nd Board and Supervisors Joint Meeting elected Mayaw Kumud as the 2nd Chairman and Hsin-Hui Chien as the 2nd Standing Supervisor.

June

September



May



Co-hosted the first national finals for the Indigenous Language Vocabulary Competition with the CIP.



Launched the first "i loma': Let's Speak Together" Podcast.



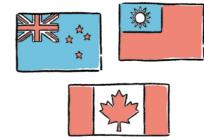
September

Hosted the first book launch event for the Taiwan Austronesian Languages Series III - Parts of Speech and Teaching.



Signed the first MOU overseas with Boston University's Department of Psychology: Research on AI translation for Indigenous languages.





March

000





Hosted the international exchange event: the Indigenous Language AI Exchange Forum.

Participated in the 2024 Taipei International Book Exhibition for the first time, showcasing various indigenous language publications under the theme "Pafatis: Reading Together," and won the Silver Award for booth design.



December

February

Held the first internal staff expanded meeting with reports presented in mother languages, now held bimonthly.

First visit and exchange with foreign organizations: Welcomed representatives from New Zealand's Māori Language Commission and the Director of Canadian Language Museum.

October

Released the ILPCT Guide: Intermediate-High Level in 16 Indigenous Languages.





 $\mathbf{202}$

Signed a domestic MOU with the Indigenous Peoples Cultural Foundation: Collaboration on the Indigenous Language Communication Platform.

Key Execution Achievements Education and Transmission

Taiwan-Austronesian Indigenous Words and Narrations

The Taiwan-Austronesian Indigenous Words and Narrations website (abbreviated as the E-learning website) is Taiwan's digital learning resource hub for indigenous languages. It offers rich and diverse learning materials in 16 indigenous languages, encompassing 42 language varieties.

Originally established and maintained by the Council of Indigenous Peoples in 2013, the website was handed over to the Indigenous Languages Research and Development Foundation (ILRDF) in 2020 for ongoing management and operations, which continue to this day.

Between 2020 and 2023, 40 picture books and teaching plans were created, along with resources for parent-child indigenous language learning, situational indigenous languages, short indigenous texts, reading materials, syntax exercises, teaching modules, interactive modules, and digitization of LIMA audiobook teaching aids. Additional resources include indigenous news reading, grammar structure instructional videos, and video-based indigenous language learning. All materials are shared under a Creative Commons (CC) license.



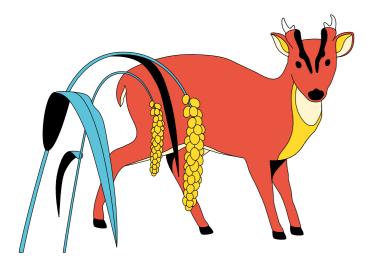
Basic materials

Introductory content for beginners, including nine-level materials, sentence patterns, and co-learning materials.



Supplementary materials

Additional materials like situational language lessons, reading texts, and teaching modules.





Truku Seediq



Four-part curriculum materials

Progressive indigenous language materials, covering alphabets, everyday conversations, and cultural articles.



C Teaching tools and learning aids

Digitized physical teaching aids, including WAWA Touch Games, thematic posters, and LIMA audiobooks.



5 Open platforms

Interactive platforms for diverse functions, including picture books, animations, and customized teaching resources.

6 Indigenous language videos

A wide range of instructional videos, such as everyday indigenous language practice and indigenous studio classroom.

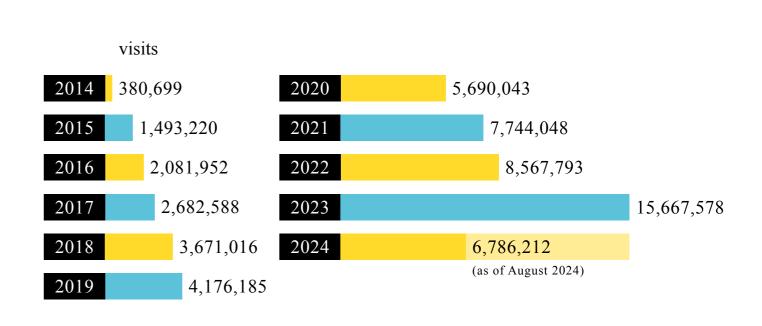




Indigenous language helpers

A wealth of indigenous language resources to assist learners, including vocabulary lookups and resource downloads.

Annual E-learning Website Visits



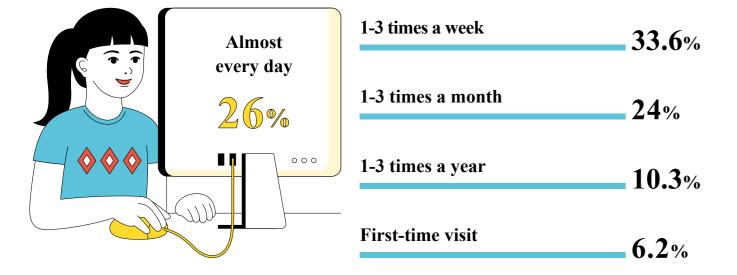
Most Popular Learning Resources (Top 4)

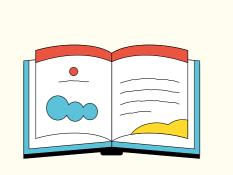


(Data from July 2022 - July 2024)

User Survey Results





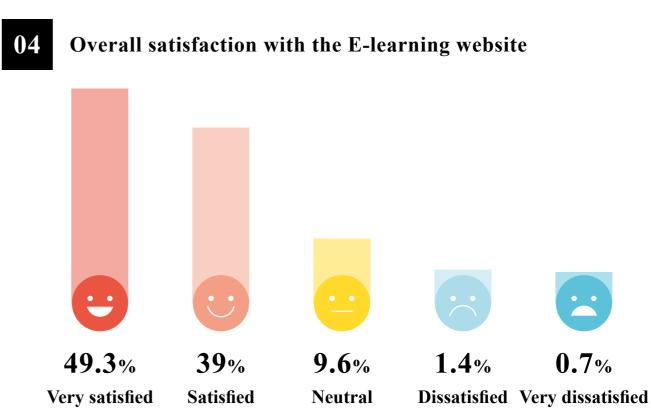


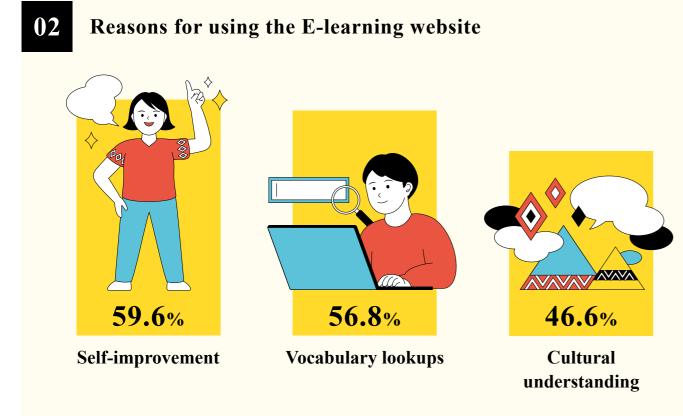
03



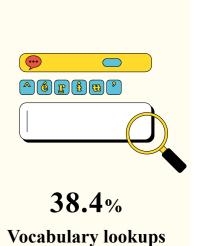
39%

54.8%	
Indigenous language	Eve
materials for the 12-Year	
Basic Education Curriculum	





Most frequently used learning materials on the E-learning website



eryday conversations

Source: 2023 E-learning website user survey

Indigenous Language Proficiency Certification Test (ILPCT)



The Indigenous Language Proficiency Certification Test (ILPCT) was established to support the revitalization of Taiwan's indigenous languages, encouraging both indigenous and non-indigenous individuals to test their indigenous language skills and foster a culture of learning indigenous languages. This test helps safeguard indigenous youth's educational rights and promotes multilingual cultural education. The ILPCT is offered at five levels: Elementary, Intermediate, Intermediate-High, Advanced, and Superior. Since 2020, efforts have been made to refine the ILPCT levels, making them more comprehensive and distinctive. The prospect is to continue enhancing the test's precision and accessibility, promoting it toward broader recognition and use.







Education and Transmission

Milestones

2020

Hosted by the CIP and ILRDF, the certification process involved 21 meetings with 430 participants, including experts, scholars, and indigenous community representatives. The test was held in 16 test regions (including outlying islands) at 54 test sites.



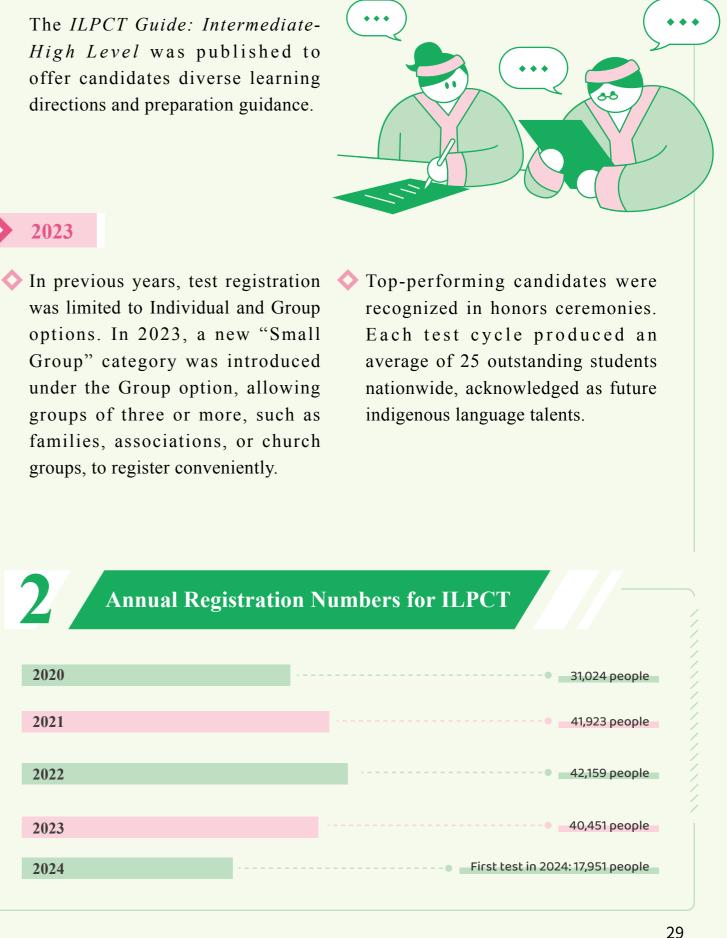
2021

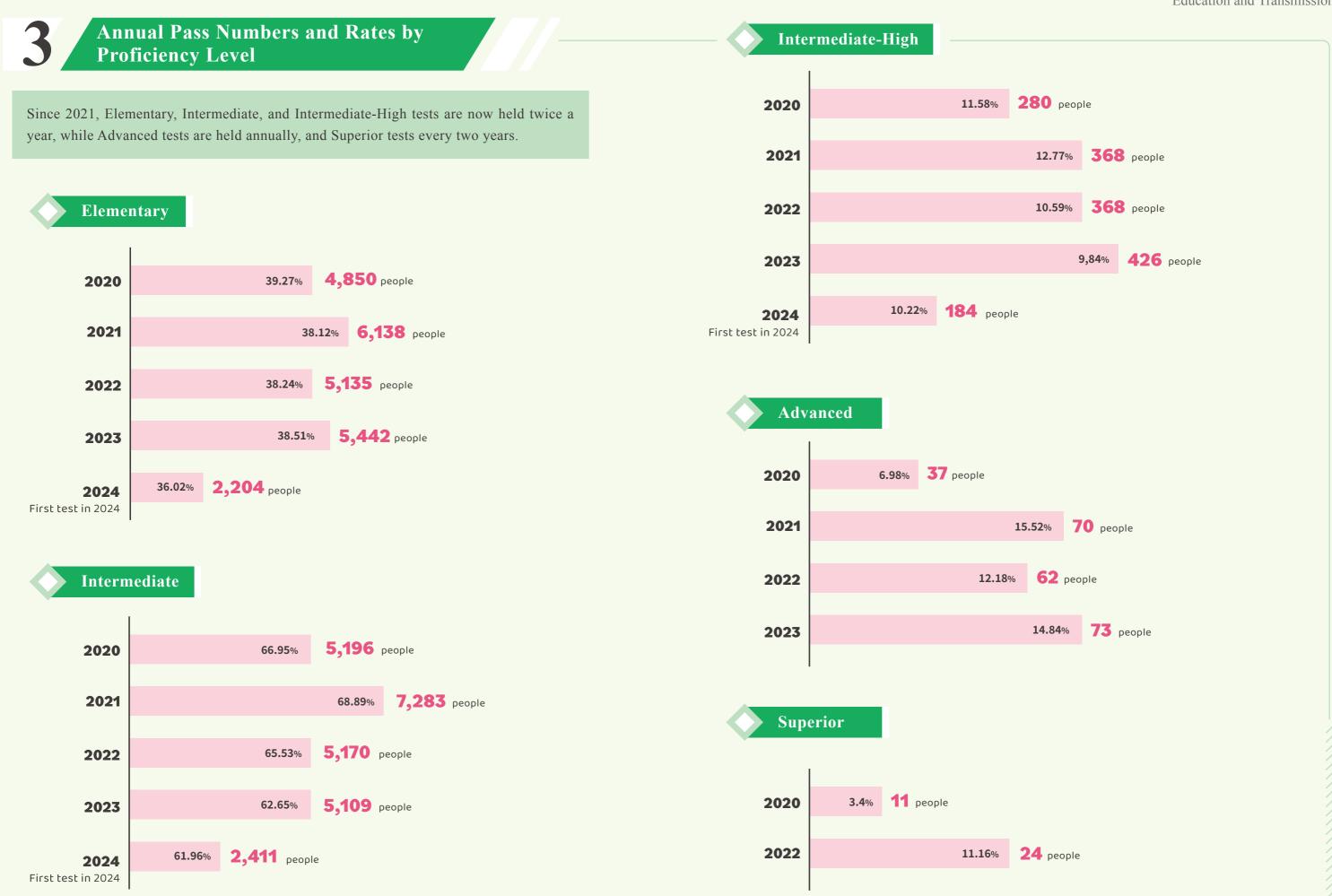
◆ To improve language literacy ◇ and revitalization in listening, speaking, reading, and writing, the ILPCT adopted a revised schedule: Elementary, Intermediate, and Intermediate-High tests are now held twice a year, while Advanced tests are held annually, and Superior tests every two years. This revised schedule has led to over 40,000 applicants each year. An additional 31 test sites were added, with 23 located in indigenous communities, benefiting 1.963 candidates in these areas.

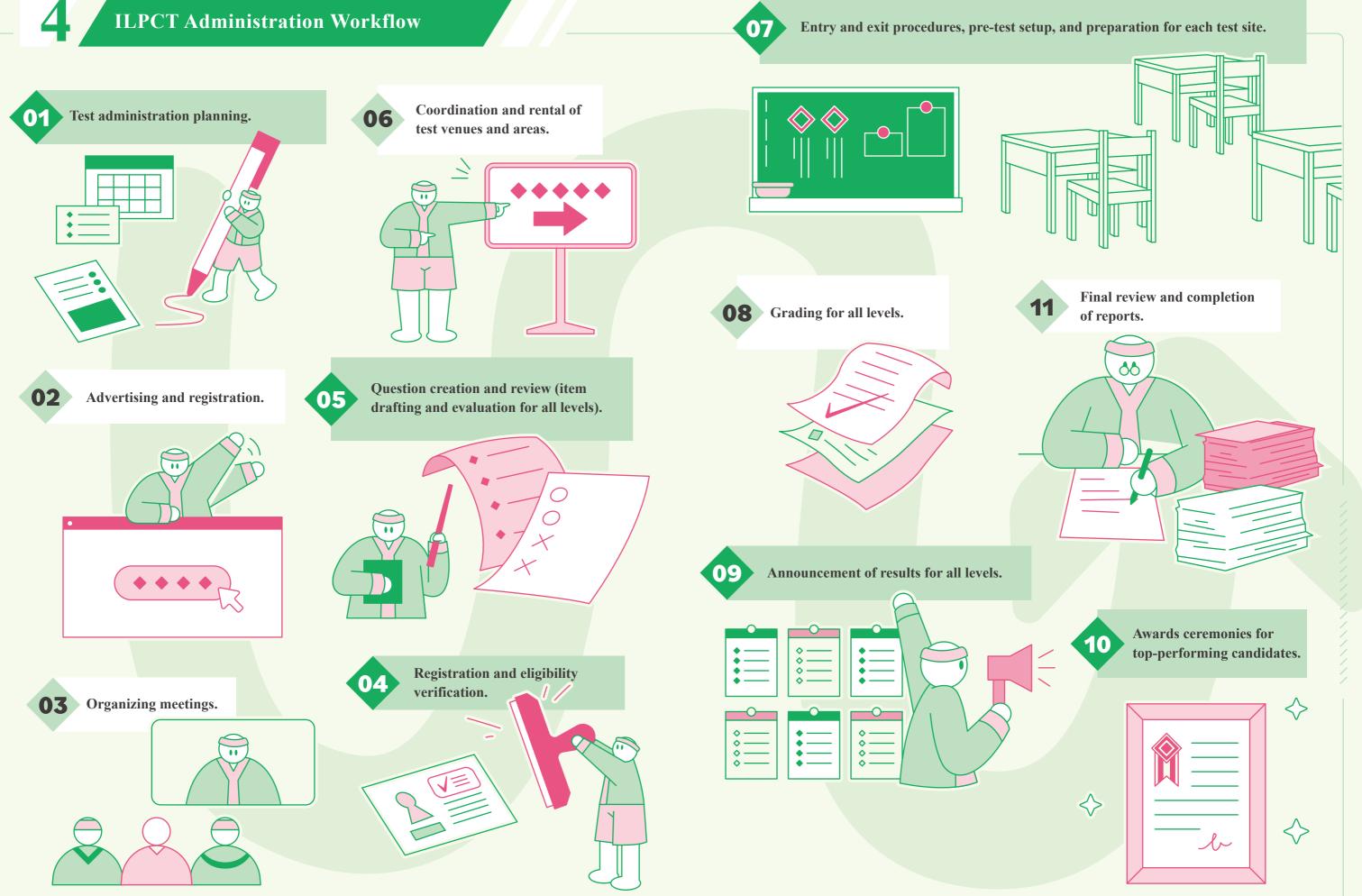
Advanced test questions were adapted and revised based on other local languages, doubling the pass rate and fostering more talent for indigenous language question development.

2022

was limited to Individual and Group options. In 2023, a new "Small Group" category was introduced under the Group option, allowing groups of three or more, such as families, associations, or church groups, to register conveniently.







2020 Indigenous Language Professional Talent Training Workshop



The Indigenous Language Professional Talent Training Workshop ran from November 2020 to May 2021, spanning a period of six months. The workshop was held across different regions (Eastern, Northern, and Southern Taiwan), with locations at National Dong Hwa University, National Tsing Hua University, and National Pingtung University. A total of six classes were organized as follows: Eastern Region: Indigenous Language Question Design and Practice; Northern Region: Indigenous Language Dubbing Talent, Language Archiving and Application, Indigenous

Language Question Design and Practice; Southern Region: Indigenous Language Teaching and Multimedia Application, Indigenous Language Question Design and Practice



2023 Indigenous Language Podcaster Talent Empowerment Program

In line with the current trend toward online education through social media channels to share language and cultural knowledge, the 2023 Indigenous Language Podcaster Talent Empowerment Program was established to encourage the creation of podcasts focused on indigenous languages or bilingual content. This initiative not only preserves invaluable language resources but also fosters public awareness, understanding, and motivation for active learning. The program includes a 12-hour



Education and Transmission



beginner course and a 16-hour advanced practical course. With 141 applicants aged 23 to 72, a total of 77 participants completed the program.



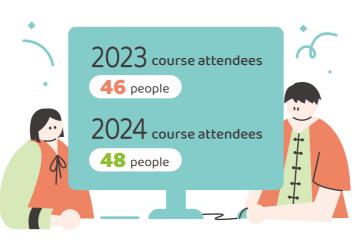
2023-2024 Intermediate-High Level Test Question Developer Training and Excellent Test Question Selection

In 2023, the ILRDF launched its first series-based training program for indigenous language question developers, offering courses on "Indigenous Language Structure and Test Question Development" in 2023 and "Intermediate-High Level Test Question Development Techniques" in 2024. In addition, participants were encouraged to submit entries for the Intermediate-High Level Excellent Test Question Selection. This program received 94 applications, ultimately training 58 question developers, including 37 new talents. Of the graduates, 13 joined the 2024 test question developing team, with 5 hailing from 3 endangered languages: Thaw, Katatipul Puyuma, and Sakizaya.

Each Intermediate-High Level test paper includes 9 main question types for a total of 42 questions. In 2024, the question selection activity collected 479 questions across 10 languages and 9 question types.

sakizaya

Participant Numbers by Year





2021 and 2023 Indigenous Language Translation and Interpretation Professionals Training

sakizaya

C Listening to Indigenous

To promote a language-friendly environment for indigenous languages across government agencies, private businesses, and organizations, the ILRDF hosted Indigenous Language Translation and Interpretation Professionals Training courses in 2021 and 2023. Designed by translation experts and experienced indigenous practitioners, the courses provided foundational and advanced training for interpretation and translation, as well as indigenous language seed teacher training. Over both years, the program successfully trained 191 participants across 16 indigenous peoples. Graduates were added to the Indigenous Language Talent Database, creating a network of translation professionals available through a matching mechanism to meet

7	_
Participation Statistics	
Graduates in 20	21
8	1 people
Graduates in 2023	
110 people	7N

various translation needs. The ILRDF will continue to enhance curriculum design and establish a professional indigenous language translation community, furthering the visibility and quality of translation services.

2022-2023 Indigenous Language Scriptwriting Workshop

sakizaya	
	C Listening to Indigenous Languages

To cultivate talent for future drama competitions, indigenous language instructors, and language advocacy organizations, the ILRDF organized a scriptwriting workshop designed to enhance the creation of indigenous language scripts and foster diverse cultural preservation and growth. Twenty-five participants were enrolled for each session, with a cumulative 37 graduates in 2022 and 2023. Graduates were added to the ILRDF talent database, providing a valuable resource for organizations seeking skilled individuals in this field. Education and Transmission



39

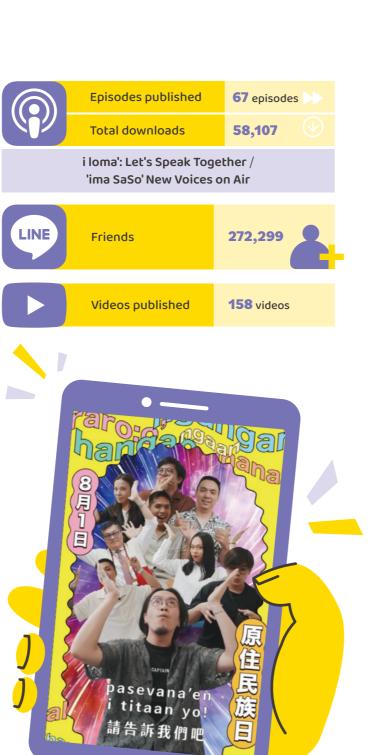
Key Execution Achievements Promotion and Implementation



Indigenous Language Promotion through New Media



Since its establishment in 2020, the ILRDF has launched an official Facebook page to actively engage with the younger generation. From 2021, it has expanded its presence across YouTube, Podcast, Instagram, and LINE official accounts to share indigenous language content and event updates, using various creative media formats to enhance indigenous language promotion.



Thaw

LRDF Rela

2021-2023 Indigenous Language Drama Competition

To inspire greater enthusiasm for indigenous language development and ignite a movement for "speaking and learning indigenous languages," a competition featuring short plays and stage performances was organized. Families and indigenous community groups were encouraged to unleash their creativity, crafting engaging scripts themed around family dialogs, community life, and cultural history. Through immersive, scenario-based performances, participants deepened their connection with language and culture, expanding the everyday use and presence of indigenous languages within family and village (community) settings. From 2021 to 2023, the competition attracted approximately 1,510 participants, divided into Family, Community, and Student categories, with a total of 87 teams participating.



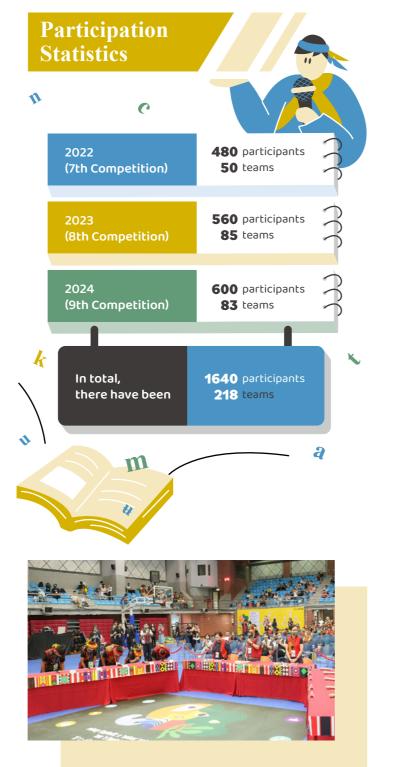
Participation Statistics



2022-2024 **Indigenous** Language **Vocabulary Competition**

Kanakanavu

The Indigenous Language Vocabulary Competition fostered an engaging and competitive environment for indigenous language learners to master vocabulary effectively. This event ignited enthusiasm for language learning and helped improve listening, speaking, reading, and writing skills across all levels. The grand finals were broadcast live on television and online, allowing indigenous communities, families, and schools across Taiwan to watch, amplifying the reach and impact of indigenous language revitalization efforts. From the 7th to the 9th editions of the competition, a total of 1,640 participants took part, divided into 4 categories: General Elementary, General Junior High, Endangered Elementary, and Endangered Junior High, with 218 teams participating overall.



International Mother Language Day-Indigenous Languages Development Conference

In alignment with the UN's designation of February 21 as International Mother Language Day, the Council of Indigenous Peoples (CIP) initiated the Indigenous Language Development Conference in 2020. Since 2021, the ILRDF has co-hosted the event with the CIP annually for 4 consecutive years. Themes from past conferences were: Indigenous Language Development (2021), Traditional Leadership (2022), Indigenous Language in Churches (2023), and Community-Driven Language Promotion (2024). This conference not only emphasizes the importance of language preservation and transmission but also honors individuals, organizations, and agencies that have made outstanding contributions to the revitalization of indigenous languages.

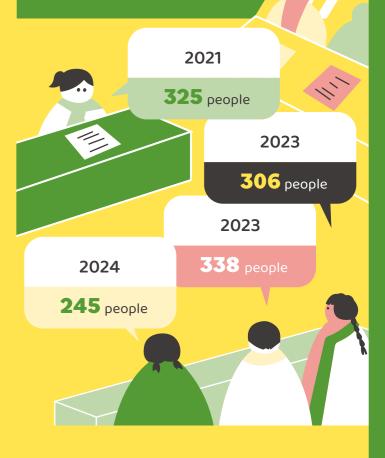






Kanakanavu

Participation Statistics



2020-2023 Austronesian Language Revitalization Forum



Kanakanavu

As the homeland of Austronesian languages, Taiwan has been actively expanding linguistic and cultural collaborations with other Austronesian countries. Through the Council of Indigenous Peoples' "Austronesian Forum: A Six-Year Plan," Taiwan is promoting international exchanges to revitalize indigenous languages. In 2020, the ILRDF held the first Forum on the Revitalization and Exchange of the Austronesian Languages, focusing on themes of revitalization, practice, connection, and innovation. This inaugural event attracted envoys from 9 countries, alongside experts, scholars, and indigenous language advocates. In 2022, the forum focused on the origins of Austronesian



languages, indigenous language education, and online learning applications, featuring simultaneous interpretation in 8 languages and a live online broadcast. In 2023, the forum's theme shifted to "Indigenous Language Development Policies of 3 Countries," with scholars from Canada and New Zealand sharing insights on the current state of indigenous language revitalization.



ILRDF x PaGamO Platform

In 2021, the ILRDF started to establish an online pathway for learning indigenous languages (via PaGamO platform), leveraging gamified learning, challenge-based progression, and real-time feedback to spark learner motivation. A database featuring vocabulary from 16 indigenous peoples and 42 language varieties was created as the platform's game-based question bank, attracting nearly 160,000 participants.

> 2021-2022 Participant Numbers in PaGamO Public World Missions





In 2022, 3 key initiatives were accomplished: executing language learning tasks for each indigenous group with a total of 344,327 participants; launching the online indigenous language game "Battle Against the Grand Demon King" with 410 players; and actively promoting and sharing relevant content on various media platforms to enhance engagement and interaction in gamified learning.



2023 World Reading Day-Reading and Illustration: Indigenous Language Picture Book Exhibition

In its first celebration of World Reading Day, the ILRDF hosted the 2023 World Reading Day - Reading and Illustration: Indigenous Language Picture Book Exhibition at Huashan 1914 Creative Park in Taipei. This exhibition incorporated digital learning resources from the Taiwan-Austronesian Indigenous Words and Narrations (E-leaning website), organizing 6 thematic areas focused on "Entering Stories," "Discovering Language," "Immersive Reading," and "Interactive Experiences" to promote indigenous language literacy and learning. The exhibition showcased a collection of 74 picture books and 16 animated adaptations in 16 indigenous languages from the E-learning website. After the

Hla'alua



event, a complete set of indigenous language picture books was donated to 9 local elementary schools and libraries in Hualien to further promote indigenous language literacy through picture book reading.









Promotion and Implementation



2024 Taipei International Book Exhibition -Pafatis: Reading Together



To advance Taiwan's indigenous language publishing and cultural content, facilitate exchange with domestic and international publishers, and deepen public understanding of indigenous language promotion, the ILRDF participated for the first time in the 2024 Taipei International Book Exhibition. Centered around the Amis concept of "Pafatis" (sharing), the exhibition's visual design used indigenous text and geometric patterns to embody the core spirit of "Reading Together."

The showcase featured 92 indigenous language books and teaching aids published over 4 years by the ILRDF, complemented by indigenous language promotion activities that attracted 3,816 participants. In its debut, the exhibition won the Silver Award for Best Booth Design – Small to Medium Booth Category.

Yami









Promotion and Implementation





Key Execution Achievements Preservation and Application





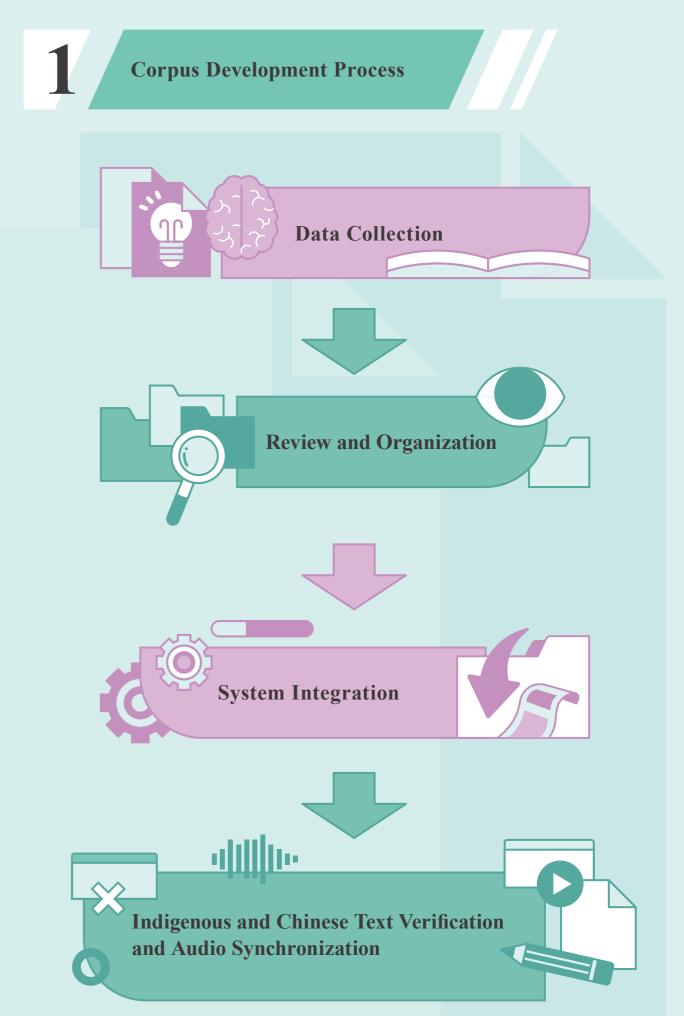
Indigenous Language Corpus

Ke na cquliq Tayal

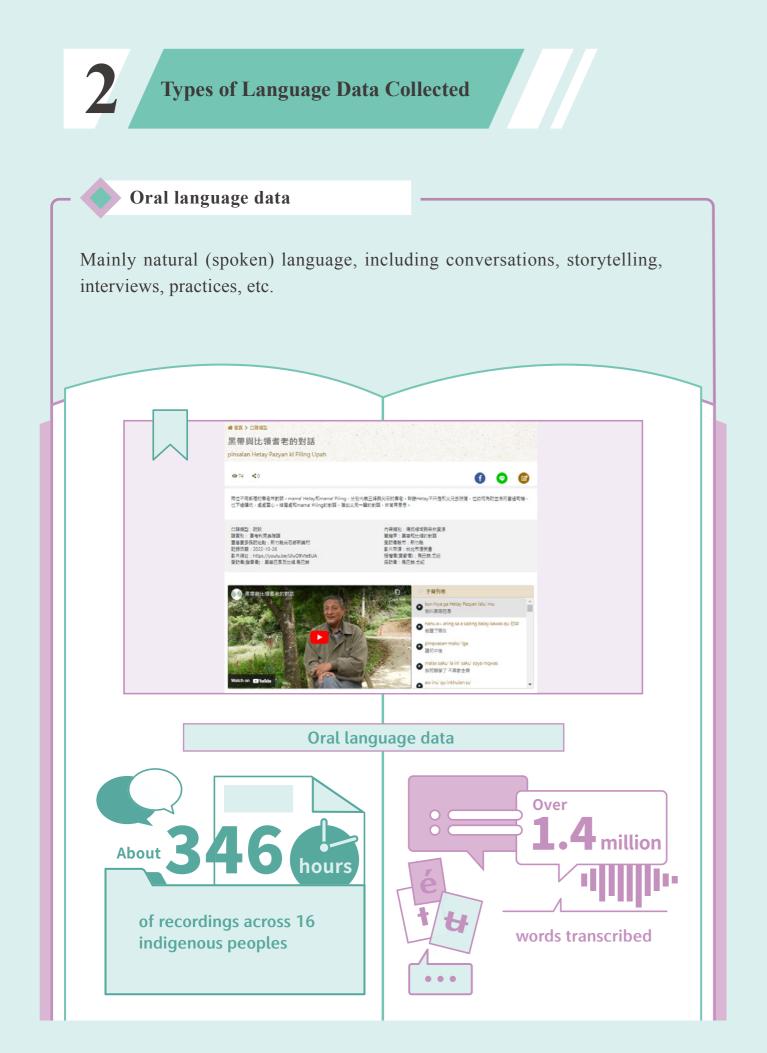


As part of the "Indigenous Language Revitalization Plan," the Council of Indigenous Peoples commissioned the ILRDF in 2017 to begin developing a research project aimed at creating a digital indigenous language corpus.

Since its establishment in 2019, the ILRDF has focused on preserving indigenous language data with an eye on future AI applications. To expand the Indigenous Language Corpus, the ILRDF adjusted its data collection approach, collaborating with language promotion partners and organizations. This networked effort has allowed the ILRDF to compile an indigenous language corpus that includes oral, written, and archival language data gathered from various projects. In 2022, the Indigenous Language Corpus was officially launched, providing public access to selected indigenous language audio-visual interview materials.



Preservation and Application



····	en lang	uugu	
Mainly te publication		ed ma	terials, such
			至 110年5月28日止,礼 anini a remiad tangasa 110pa
為防治嚴重特殊傳染性肺炎,即 日起至110年5月28日止,花蓮難 斯坦疫情三級響或管制招牌,請 民眾持續客實過感並從合,顯守	族語 = vivv	草語	族語・華語並列 ■ / > - @ms + CD &
社區防線。 	The second		Mislany : Nposukal a remini ? Conflex a miske 110pt Nposukal a shi ? Invisien sha 10010050A Tariwana ! subject : U saplawa ta sapangalev tu remini dangaa 110pay 250 vidad 28 (rel; ? palayniyay ru sakanket tu tu' sikil ru saka la yu na saka for un
Archi	val dat	a	
Mainly un	ndigenc	sed ma	aterials, includ aguage historic
Mainly un digitized i	ndigenc	sed ma	
Mainly un digitized i	nprocess ndigenc	sed ma	
Mainly un digitized i	ndigenc	sed ma	
Mainly un digitized i * #耳 > ## 青蛙的故 @33 <0 Watana	ndigence ndigence 和 家科 文事	sed ma bus lar	nguage historic
Mainly un digitized i * =東 > 未編 青蛙的故 @33 <0 Wallsha () Wallsha	ndigence ndigence 和 家科 文事	sed ma bus lar	
Mainly un digitized i 《 = 用 > 未編 青蛙的故 @ 33 <0 收羅區內容	process ndigence の	sed ma bus lar	nguage historic





Indigenous Language Online Dictionaries



Since 2007, the Council of Indigenous Peoples has been compiling dictionaries for Taiwan's 16 indigenous languages in stages. Starting in 2020, the ILRDF took over their management, gradually enhancing the online dictionaries' functionality with features such as "word suggestion" to expand and refine entries for each language. The dictionaries now serve as a valuable platform for learning, teaching, research, and promotion that is available to both indigenous communities and the wider public interested in indigenous languages.

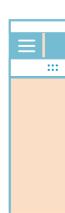


1

Key Features of the Indigenous Language Online Dictionaries

Online Search

Full-text search is available in both indigenous languages and Chinese, along with crosslanguage search capabilities, allowing quick access to content for learners and supporting crosslinguistic analysis for researchers.





Users can download the full vocabulary for each language, as well as image cards and complete documents, enabling offline reading and learning.





Interactive online tests with theme-based learning modules provide a richer, more diverse learning experience, and enhance both enjoyment and practicality, making vocabulary mastery more comprehensive and effective for users.



Preservation and Application



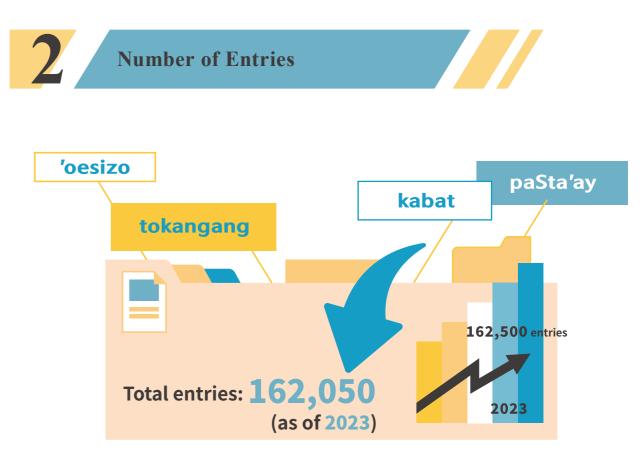




New Word Announcement

When administrators announce new terms, users can access them through the "New Words Announcement" feature to view each term's primary category, subcategory, indigenous language term, and Chinese translation. This feature also allows users to search by year.

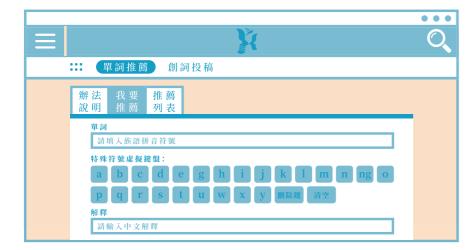
=	:			X		•	0
	🕕	2年度	1114	耳度 108年度	107年度 10	6年度 105年度	
	公告新言	司					
	NO.	主類別	次類別	族語	中文	發音	
	1	學校、 校園 相關		sa:pae' baboeloe'	一般垃圾		
	2	學校、 校園 相關		mayaka:i' talekaw	口試		
	3	學校、 校園 相關		kamatilhaehael ka kamatortoroe'	小老師		



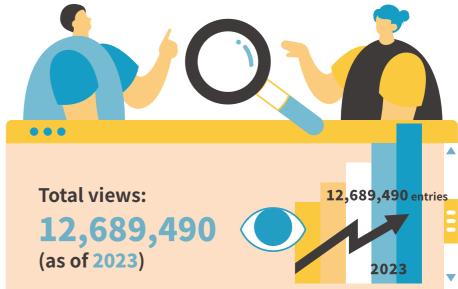


Recommended Indigenous Terms

Word recommendations primarily focus on the inclusion of "traditional indigenous vocabulary." User-submitted word recommendations undergo an initial review by dictionary administrators. Those meeting the review criteria are then forwarded to a committee of indigenous language experts for a detailed expansion review. Upon completion, the results are communicated to the contributors and published in the Indigenous Language Online Dictionaries.







Preservation and Application

Key Execution Achievements Research and Development



Indigenous Language Learning Aids



Between 2015 and 2018, the Council of Indigenous Peoples developed a total of 8 sets of indigenous language learning aids:





2016





Preschool indigenous language learning aids: Stackable Blocks, Three-Part **Cards, Flip Cards**



2018

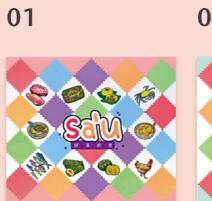
06 Mukasa Language Platter



08 **Mitaling Connection Cards**



indigenous language aids:







Development of teaching aids:



Designed to focus on indigenous language learning centered around village foods and ingredients, Village Kitchen includes interactive matching cards that connect images with indigenous vocabulary in an engaging learning format. This tool can be flexibly paired with the Mukasa Language Platter created in 2018 and Mimio's Treasure Hunt in the Village teaching aids, allowing teachers to creatively expand on students' learning experience.

Healthy Baby





Designed to focus on indigenous language learning centered around scenarios of early childhood, self-care, and problem-solving, Healthy Baby includes interactive matching cards that connect images with indigenous vocabulary in an engaging learning format. This tool can be flexibly paired with the Mukasa Language Platter created in 2018 and Mimio's Treasure Hunt in the Village teaching aids, allowing teachers to creatively expand on students' learning experience.

////	Learning chart:	WAWA 7
Year of Public	ation	
2020		
Quantity Prod	luced	
1,000 sets		
Content Overv	view	
This interact	ive audio-based te	aching too

This interactive audio-based teaching tool emphasizes game-based interaction, offering children a chance for autonomous learning. Covering 16 indigenous peoples across 42 language varieties plus English (43 languages total), it features 6 themed scenarios from children's daily experiences. The content includes relevant dialogs, games, and songs, each focused on 10 key vocabulary words, enabling repetitive learning through chart-based game interactions.

//// LIMA Sound Book, MATAS Indigenous Alphabet Learning Box



The teaching aids introduce 7 themes to facilitate indigenous language learning through vocabulary recognition, language exchanges, and interactive games. Incorporating images and audio, they create an engaging learning experience. They can be combined with previously developed teaching aids, such as the Stackable Blocks (2016), and Village Kitchen, Healthy Baby, and Wawa Touch Games (2020). Teachers have the flexibility to use the MATAS Indigenous Alphabet Learning Box or to develop their own extended learning activities.

Touch Games (Reissue)



1111





Distribution Targets

These learning aids are distributed to key indigenous schools and institutions involved in indigenous language education, including immersion preschools, indigenous language caretakers, village-based preschool education support centers, indigenous language preschool educators, language promotion personnel and organizations, indigenous language teachers, learning centers, and the public attending regional information sessions.







Indigenous Language Publications



Eastern Paiwan

$\mathbf{02}$

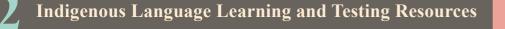
Proceedings of the 2022 Forum on the Revitalization of Austronesian Languages



Indigenous Language News: A Cultural Glossary and **Stories Behind Words**

05

Proceedings of the 2021 International Mother Language Day -**Indigenous Languages Development Conference**



名装通用的



Indigenous Language **Publications**



From 2019 to 2023, the ILRDF has published a total of 704 online and print materials, including books, picture books, and teaching aids. These publications are categorized as follows:

- ◆ Indigenous Language Picture Books: Including 40 themed picture books translated into the languages of 16 indigenous peoples.
- ◆ Indigenous Language Research and Development: Including 8 titles, such as Taiwan's Takituduh Bunun Language: Thirty Years of Heritage, Customs, History, and Life Collection and Indigenous Language News: A Cultural Glossary and Stories Behind Words.



 Indigenous Language Learning and Testing Resources: Covering a total of 8 items, including instructional books for parts of speech across the 16 indigenous languages, ILPCT Guide: Intermediate-High Level, and additional teaching materials and teaching aids.

◆ Indigenous Language Magazines: tousvusvutu magazine has published 4 issues to date, each containing translations in the 16 indigenous languages.

2019-2023 Publication List

//////// Indigenous Language Research and Development /////////

- 01 Taiwan's Takituduh Bunun Language: Thirty Years of Heritage, Customs, History, and Life Collection (in four languages)
- 02 Indigenous Language News: A Cultural Glossary and Stories Behind Words
- **03** Proceedings of the 2022 Forum on the Revitalization of Austronesian Languages
- 04 Indigenous Language Writing Guide for Official Documents (Trial Edition, Third Version)
- **05** Reference Guide for Agenda Writing in Indigenous Languages for Council of Indigenous Peoples Meetings (Trial Edition)
- 06 Proceedings of the 2021 International Mother Language Day Indigenous Languages Development Conference
- 07 Proceedings of the 2022 International Mother Language Day Indigenous Languages Development Conference
- 08 Proceedings of the 2023 International Mother Language Day Indigenous Languages Development Conference

Indigenous Language Magazines

01 tousvusvutu: Issues 1 through 4

- 01 Face ID
- **02** Froggie's Big Move
- **03** The Little Girl Who Loved Trees
- **04** Tales of the Austronesians
- **05** The Nurse's Scooter
- **06** Off to Space!
- 07 Who Took My Egg?
- **08** The Girl and Her Rose
- **09** The Giant Tree on the Hill
- **10** NASA's Female Scientists
- **11** Big Brother's Wedding Day
- **12** Can I Try Weaving Too?
- **13** Where Did My Apple Go?
- 14 The Piggy Who Loved to Paint
- **15** The Super Archer
- **16** The Passing of A Brave Warrior
- **17** The Tale of Lonckjouw
- **18** The Lost Generation
- **19** The Flower of Courage
- **20** The Indian Girl Who Dared to Fly
- **21** Have You Seen Panay?
- **22** Journey Through the Woods
- **23** The Colorful Clothes
- **24** Two Little Pictures
- **25** The Pandemic
- **26** The Stars of the Bunun

- **27** The Guardian of Eyes
- **28** The Power of Knowledge
- 29 Tukucu's Gift
- **30** Big Sister's Visit Home
- 31 Let's Sing Our Song!
- **32** The Roaring Story
- **33** My Dream Job
- **34** The Magical Thread
- **35 Our Names**
- **36** The Salty Story
- **37** Selep's Magic Seeds
- **38** The Hero Who Stood for Justice
- **39** Scam Rings of Cambodia
- 40 Crypto

Indigenous Language Learning and Testing Resources

- 01 Parts of Speech and Teaching in Kanakanavu
- **02** ILPCT Guide: Intermediate-High Level (General Edition for All Indigenous Peoples)
- 03 ILPCT Guide: Intermediate-High Level (in 16 Indigenous Languages)
- 04 Animal and Number Flip Book
- **05** LIMA Touch Book
- 06 MATAS Indigenous Alphabet Learning Box
- 07 Sa'u Village Kitchen
- **08 Iintasu** Healthy Baby

Can ChatGPT Speak Indigenous Languages? The Hurdles and Hopes of **Building AI for Taiwan's Indigenous Tongues**



Kavalan

According to UNESCO's Atlas of the World's Languages in Danger, Taiwan's indigenous languages are classified at various levels of endangerment, reflecting the current state of language vitality and activity. Language is at the heart of indigenous culture, and to combat the crisis of language loss among Taiwan's indigenous communities, the Indigenous Languages Research and Development Foundation (ILRDF) established the Indigenous Language Corpus (referred to as the Corpus) in 2022. This ambitious project aims to use digital technology to collect and preserve invaluable linguistic data across 16 Indigenous languages.

The groundbreaking advancements in generative AI, particularly with the rise of large language models (LLMs) like ChatGPT, have introduced exciting new possibilities for indigenous language revitalization. However, building an AI system to support these languages presents a fundamental challenge: the extensive collection and meticulous organization of linguistic data.

Enriching Language Data to Create a Basis for AI

Akiw Chung-Wen Hsu, Deputy Director of the Research and Development Division, ILRDF, highlighted that the potential to develop an AI language model using today's technology hinges on the availability of large, high-quality language datasets for in-depth machine learning. "Collecting indigenous language data isn't easy," she explains. "With limited data available, accuracy becomes even more crucial for training AI effectively. That's really the foundational reason behind ILRDF's initiative to build an Indigenous Language Corpus in the first place."

Since the establishment of the Corpus, many stakeholders have eagerly anticipated access to language data for future development and application. Yet Akiw cautioned that building a robust corpus is complex and time-intensive, with essential tasks such as organizing, proofreading, and converting data into a machine-readable format. For example, the initial stage of Taiwan's Hakka Language Corpus took over at least 5 years of focused effort.

The Corpus currently sources its data from three primary streams: the local language advocates working at township offices, audiovisual recordings of elders by language promotion groups, and data from the E-learning website. Plans are

underway to expand the database with indigenous publications and official documents. However, challenges persist—without consistent standards for the written form of many indigenous languages and given diverse translation approaches, each entry for AI training must be carefully reviewed. This process involves verifying that indigenous and Mandarin text align sentence by sentence with the audio recordings, as well as ensuring no discrepancies exist between the original meaning and its translation.

Experts estimate that developing a robust AI translation system requires between 3 to 5 million words of data per language, and if the goal is high-quality translation output, at least 10 million words per language would be ideal. Currently, the Corpus has amassed over 2 million words, with projections to reach 4 million words across 16 languages by late 2024. Achieving the ultimate ambition of a fullscale indigenous language AI translation model, or even a generative AI system, will require continued commitment, particularly given the linguistic diversity of Taiwan-16 indigenous peoples encompassing 42 varieties. For now, ILRDF is pursuing a phased approach, focusing on pilot research and development, gradually advancing toward a more comprehensive system.

AI-Powered Advancement: Accelerating Indigenous Language Data Collection and Application

In response to the pressing need for documenting and preserving indigenous language data, alongside a growing anticipation for AI's application in indigenous languages, the ILRDF launched the initial phase of its "Indigenous Language AI Project" in late 2023. This endeavor aims to deliver a machinereadable indigenous language corpus, voice recognition system, text-to-speech synthesis, and prototype AI translation model by the end of 2025. Given the experimental nature of this project, outcomes across languages may vary due to differences in data volume and accuracy.

As Akiw elaborates, this project's goal is to build a foundational AI system that eases the manual labor involved in transcription and proofreading, significantly speeding up data generation. "The process of gathering indigenous language data, from initial preparation to transcription and organization, is incredibly meticulous. Most of our language data collectors aren't particularly tech-savvy, so by using AI speech recognition to convert audio into text, we're looking at a huge boost in data collection efficiency."

Beyond these goals, we also envision this pilot project as a launchpad for diverse, future applications of indigenous language AI. Akiw explains, when young people learn indigenous languages, they may hesitate to speak them out of fear of making mistakes. A voice system like Google Voice could encourage them to engage with the language more actively. Similarly, as shortform video dominates social media, AI textto-speech could enable real-time voiceovers in indigenous languages, offering young people a more dynamic way to creatively express and share their heritage, thus expanding the presence of indigenous languages on new media platforms.

"While the work being done by the ILRDF may seem basic at this stage, it's actually an essential cornerstone for AI development," Akiw emphasizes. With a robust collection of machine-readable indigenous language data, the applications could span various sectors-from education and exams to media, literature, and healthcare. The possibilities include voice assistants, spoken language evaluation systems, real-time translation subtitles, text proofing, creative writing support, and even interactive healthcare robots in indigenous languages. Our hope is that AI will enable indigenous languages to spread rapidly through new generations and emerging technologies, making them accessible in every aspect of daily life, touching every corner, and readily available to all.

Indigenous Language AI Translation Project

The Indigenous Language Corpus and Indigenous Language AI Project align with the Indigenous Language Development Act, the Act for the Establishment of the Indigenous Languages Research and Development Foundation, and the National Languages Development Program (2022-2026).



Research and Development



- Confirm text data formats, such as words, sentences, articles, and audiovisual materials.
- Ensure text aligns with audio in audiovisual data, with proper audio file tagging.
- Reconfirm data accuracy.





Machine-Readable Corpus



Organize readable format data into the machine-readable corpus.

Enable future access to the corpus through an application and authorization process.

Speech Recognition

- Convert indigenous language audio into text.
- Enable future applications such as verbatim transcripts and AI-driven dialogs in indigenous languages.

Generate audio content from indigenous language text, commonly used for video dubbing.

(C g

voice assistants.

Research and Development



This technology can also support the development of indigenous language

AI Translation Beta Development

- 1 Conduct a trial-based approach to assess the current state of language data development.
- 2 Use the "Google Translate" system as a reference model to support applications such as indigenous language subtitle translation.



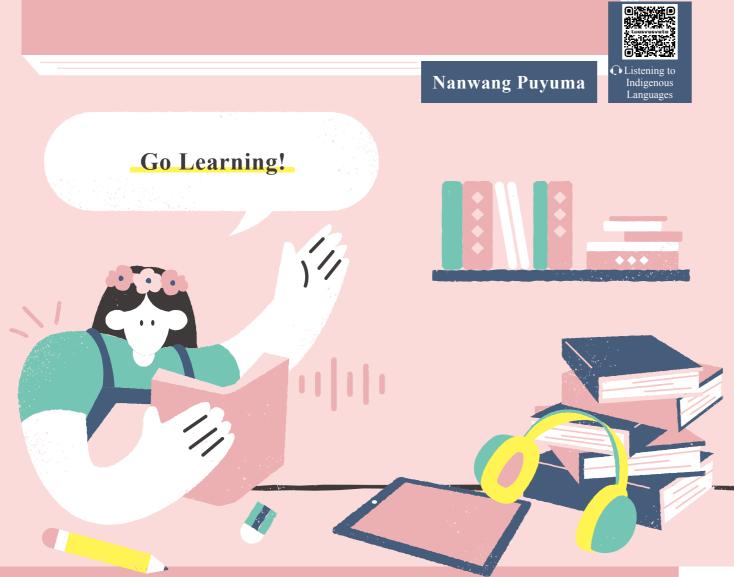
Learning Resources

AI Application Website

Develop an AI application system website, allowing users to participate in trials and provide feedback on their experience.

Ę

Indigenous Language Learning Resources Platform



Go Learning! Curious about learning Taiwan's indigenous languages? Wondering where to start? We've compiled all available resources to help you quickly find the best tools for your language learning journey! 1

Taiwan-Austronesian Indigenous Words and Narrations



A digital indigenous language learning platform offering resources across 16 indigenous peoples and 42 language varieties, with rich and interactive content tailored for preschoolers, K-12 students, adults, and indigenous language teachers.



Learning Resources

Indigenous Language Online Dictionaries

2



Featuring online vocabulary search, downloads, online quizzes, new word updates, and word suggestions. Learners can access definitions and pronunciations for terms across 16 indigenous languages, and test their knowledge through online interactive games, making learning more enjoyable.



Indigenous Language Vocabulary Lists



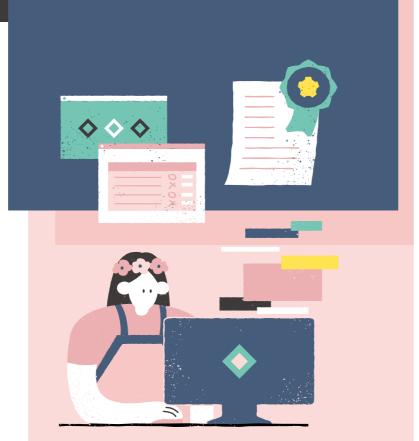
A foundational tool for indigenous language learning, vocabulary competitions, ILPCT question preparation, and learning material compilation and development, featuring vocabulary lookup, audio learning, and interactive exercises.

4

Indigenous Language Proficiency Certification Test (ILPCT)



Provides essential information and resources for the ILPCT. Users can access online study materials, practice tests, review test formats, and engage in online drills that simulate the real test environment.





Indigenous Language Corpus



A platform for preserving indigenous language materials, housing a range of recorded spoken language resources, written texts, and unprocessed archival materials (primary sources include recordings from language advocates and documentation from language promotion organizations.) Most of the oral data displayed consists of natural spoken language recorded by language advocates, transcribed into indigenous language text and translated into Mandarin. This serves as a foundational resource for AI development in indigenous languages

6

Indigenous Language Learning Center Course Platform



An integrated course portal covering Taiwan's seven indigenous language learning centers (2018-2023), offering search tools for topics, locations, and keywords to help indigenous language teachers, learners, and enthusiasts easily find nearby learning center resources. Specialized courses aimed at indigenous language skill enhancement and learning are also provided.



tousvusvutu

Consultants

lalo' biyae' Pilin • Yapu

Editorial Advisors

Siwkolan Amis Takibakha Bunun Truku Seediq Labuwan Rukai Sakizaya | Thaw Kanakanavu | Hla'alua | Yami | Ke na cquliq Tayal SaySiyat | Truku | Eastern Paiwan Kavalan | Nanwang Puyuma Tsou

Sing 'Olam Atul Mangogo' Lituk Teymu Muakai Takivalrithi Buting Nukay Hudun Lhkatanamarutaw Pani Kanapaniana Vanau Salapuana sinan Sakayan Yawi • Nomin toway a kalih titiyon Mimig Karaw dremedreman tjamatjakan ti pipis Akawyan Pakawyan sayungu 'e noacachiana yavai 'e noacachiana

Issuing Agency: Indigenous Languages Research and Development Foundation (ILRDF) Address: No. 63, Sec. 1, Roosevelt Rd., Zhongzheng Dist., Taipei City 10093 Telephone: 02-2341-8508 Website: http://www.ilrdf.org.tw/ Publisher: Mayaw•Kumud Editor-in-Chief: Lowking Nowbucyang Deputy Editor-in-Chief: Mulas Ismahasan Executive Editor: Maital Manghuhu Tanapima \ Aman Fuyan Executive Coordinator: Decent Idea English Translation by: Linguitronics Publication Date: December 2024 Price: NT\$240 Funding Agency: Council of Indigenous Peoples ISSN 3005-4761



