

การทำคลังความรู้ภาษาและวัฒนธรรมของ กลุ่มชาติพันธุ์ในประเทศไทย

3 ธันวาคม 2562

ผศ.ดร.ศิริเพ็ญ อังสิทธิ์พูนพร

สถาบันวิจัยภาษาและวัฒนธรรมเอเชีย

มหาวิทยาลัยมหิดล



เป็นส่วนหนึ่งของโครงการวิจัยเรื่อง

โครงการคลังข้อมูลดิจิทัลของกลุ่มชาติพันธุ์: ส่งเสริม
การปกป้อง และดำรงรักษา ความหลากหลายทางภาษา
และวัฒนธรรมในประเทศไทย

สนับสนุนโดย Newton Institutional Links ซึ่งเป็นความร่วมมือระหว่าง

RILCA, MU

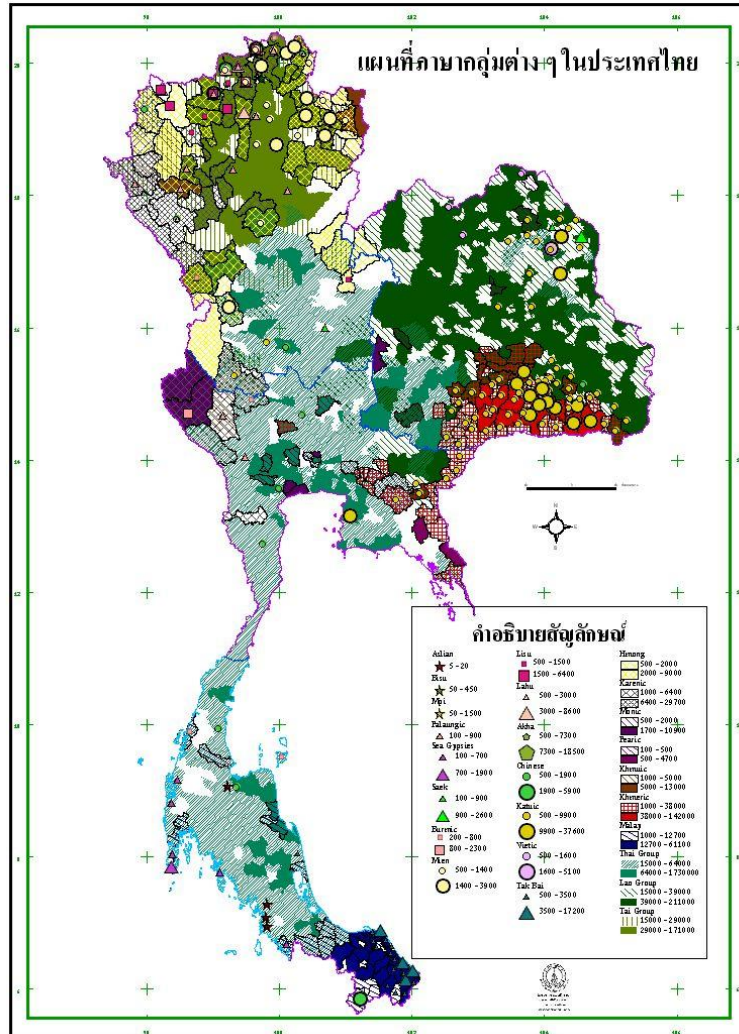


ELAR, SOAS, UK

ระยะเวลา 1 ปี

ที่มาและความสำคัญของโครงการ

ประเทศไทยมีกลุ่มภาษาทั้งสิ้นประมาณ 70 กลุ่ม ในจำนวนนี้มีภาษาที่อยู่ในภาวะวิกฤตถึง 15 กลุ่ม



โครงการวิจัยนี้ให้ความสำคัญในเรื่อง

- การสูญเสียความหลากหลายทางภาษาและภูมิปัญญาทางวัฒนธรรม
- มุ่งเน้นที่จะพัฒนาวิธีการอนุรักษ์ความหลากหลายทางภาษาและภูมิปัญญาทางวัฒนธรรม โดยปราชญ์และคนในชุมชนเองรวมทั้งพัฒนาศักยภาพคนในระดับท้องถิ่น
- พัฒนาค้นข้อมูลดิจิทัลที่ชุมชนในท้องถิ่นสามารถเข้าถึงได้
- แลกเปลี่ยนเรียนรู้งานวิจัยกับนักวิชาการ ทำให้เกิดผลกระทบระดับสังคม

จุดประสงค์ของโครงการ

- จัดทำระบบคลังข้อมูลดิจิทัลให้เป็นมาตรฐาน เพื่อรวบรวมองค์ความรู้ของกลุ่มชาติพันธุ์ในประเทศไทย โดยเฉพาะกลุ่มภาษาที่อยู่ในภาวะวิกฤต โดยผ่านเว็บไซต์
- จัดทำมาตรฐานการเก็บรักษาข้อมูลด้านภาษาและวัฒนธรรมของกลุ่มชาติพันธุ์ในประเทศไทยให้พร้อมเผยแพร่
- สร้างบุคลากรที่สามารถจัดทำคลังข้อมูลดิจิทัล มีความรู้ด้านการดูแลรักษาข้อมูลให้สามารถใช้งานได้ในระยะยาว
- สร้างเครือข่ายการทำงานด้านกลุ่มชาติพันธุ์ ร่วมกับสถาบันการศึกษาอื่น ๆ
- ฝึกอบรมและจัดการประชุมระดับชุมชนและระดับนานาชาติเผยแพร่การจัดทำคลังข้อมูล

<http://langarchive-th.com/>



Add Content ▾

My account

Dashboard

LOG OUT



เกี่ยวกับ

กลุ่มชาติพันธุ์

ภาษาและวัฒนธรรม

ค้นหา



ยินดีต้อนรับสู่ LangArchive-TH

คลังข้อมูลดิจิทัลของกลุ่มชาติพันธุ์ในประเทศไทย

Archive of Languages and Cultures of Ethnic Groups of Thailand

สนับสนุนการอนุรักษ์ความหลากหลายทางภาษาและภูมิปัญญาทางวัฒนธรรมของกลุ่มชาติพันธุ์ และมุ่งเน้นพัฒนาคลังข้อมูลที่ชุมชนในท้องถิ่นสามารถเข้าถึงได้

ภายใต้โครงการ “คลังข้อมูลดิจิทัลของกลุ่มชาติพันธุ์: ส่งเสริมการปกป้อง และดำรงรักษา ความหลากหลายทางภาษาและวัฒนธรรมในประเทศไทย” ด้วยทุนสนับสนุนจาก Newton Fund

การจัดทำคลังข้อมูลภาษาและวัฒนธรรมของกลุ่มชาติพันธุ์

- แบบออฟไลน์

- จัดการโดยใช้ระบบ **Network-Attached Storage (NAS)** คือ การจัดเก็บข้อมูลของเราบนเครือข่าย
- ใช้สำหรับ เก็บข้อมูลต้นฉบับ (back up ข้อมูล)
- ใช้สำหรับจัดการไฟล์ข้อมูลอย่างเป็นระบบ



- แบบออนไลน์

- ใช้สำหรับนำเสนอข้อมูล ผ่านเว็บไซต์ (Mukurtu)
- ข้อมูลประเภทไฟล์เสียง ให้ upload ผ่าน soundcloud
- ข้อมูลประเภทไฟล์วิดีโอ ให้ upload ผ่าน vimeo
- ข้อมูลประเภทไฟล์ภาพ หรือไฟล์เอกสาร สามารถ upload ผ่าน mukurtu ได้เลย

การจัดเก็บคลังข้อมูลภาษาและวัฒนธรรมของกลุ่มชาติพันธุ์

ตรวจสอบ-คัดแยกข้อมูล

ข้อมูล analogue

- คาสเซ็ทเทป
- ม้วนฟิล์ม
- หนังสือ
- cd-rom

ข้อมูล digital

- ไฟล์ทุกประเภทที่จัดเก็บในคอมพิวเตอร์

การจัดการข้อมูล

แปลงข้อมูลด้วยอุปกรณ์ดิจิทัล

- แปลงเทปด้วยเครื่องบันทึกเสียงดิจิทัล
- สแกนฟิล์ม
- สแกน/ถ่ายภาพหนังสือ
- cd แปลงไฟล์

จัดระบบฐานข้อมูล
Metadata (excel)

ไฟล์ดิจิทัล
เก็บใน hdd ของระบบ NAS

- การจัดระบบไฟล์ข้อมูล..เปลี่ยนชื่อไฟล์ทั้งหมดให้เป็นระบบเดียวกัน
- ใส่ข้อมูลทั้งหมดเกี่ยวกับไฟล์นั้นลง metadata sheet

Mukurtu

- ระบบที่สร้างเว็บไซต์มาเพื่อรองรับการนำเสนอข้อมูลด้านภาษาและวัฒนธรรมของกลุ่มชาติพันธุ์
- ออกแบบแพลตฟอร์มที่ง่ายต่อการนำเข้าข้อมูลด้วยตนเอง
- ผู้ใช้สามารถออกแบบหน้าตาของเว็บไซต์ได้เองตามรูปแบบที่มีหลากหลาย เลือกชมได้ที่ <http://mukurtu.org/showcase/>
- เว็บไซต์จะรองรับไฟล์ที่มีขนาดไม่ใหญ่
- เว็บไซต์รองรับภาษาไทย
- รองรับแพลตฟอร์มของมือถือ แท็บเล็ต



[Learn](#) [About](#) [Get](#) [FAQs](#) [Showcase](#) [Mobile](#) [News](#)

Our Mission

Mukurtu (MOOK-oo-too) is a grassroots project aiming to empower communities to manage, share, narrate, and exchange their digital heritage in culturally relevant and ethically-minded ways. We are committed to maintaining an open, community-driven approach to Mukurtu's continued development. Our first priority is to help build a platform that fosters relationships of respect and trust.

Mukurtu Beginnings

In 2007, Warumungu community members collaborated with Kim Christen and Craig Dietrich to produce the Mukurtu Wumpurrarni-kari Archive. Mukurtu is a Warumungu word meaning 'dilly bag' or a safe keeping place for sacred materials. Warumungu elder, Michael Jampin Jones chose Mukurtu as the name for the community archive to remind users that the archive, too, is a safe keeping place where Warumungu people can share stories, knowledge, and cultural materials properly using their own protocols. Growing from this community need, Mukurtu CMS is now an open source platform flexible enough to meet the needs of diverse communities who want to manage and share their digital cultural heritage in their own way, on their own terms.

Implementation and Sustainability

Mukurtu CMS has been developed as an access platform and is intended to serve as one component in your larger digital stewardship ecosystem (access, preservation, sustainability). If you are planning to include use of Mukurtu CMS in a grant proposal, keep this in mind. A successful digital project plan will address the tools, resources, staff, and workflows for digitization (if relevant), preservation, description, access, and long-term sustainability.

See our [FAQs page](#) for more information on digital stewardship, see our [GitHub repository](#) for the Mukurtu CMS development roadmap and please contact us at support@mukurtu.org if you have questions about Mukurtu CMS features, functions, capacity or sustainability.

ขั้นตอนในการจัดการข้อมูล

- ข้อมูลเดิม (ข้อมูลเก่า)

- แปลงไฟล์จาก analog ให้เป็น digital

- Cassette tape ---- .mp3

- VDO Tape ---- .mp4

- Scan ภาพถ่าย / หนังสือเล่มเล็ก ----- .jpg

- reversal films ---- .jpg

- negative films ----- .jpg

- หนังสือเล่มยักษ์ ใช้วิธีการถ่ายภาพ -----
.jpg

- Back up ข้อมูล ลงใน NAS

- จัดระบบไฟล์ข้อมูล

- ข้อมูลปัจจุบัน (ที่เป็นไฟล์ digital มาแล้ว)

- Back up ข้อมูล ลงใน NAS

- แปลงไฟล์ข้อมูล ให้มีขนาดเล็กลง

- ไฟล์ภาพ ---- .jpg

- ไฟล์เสียง ----- .mp3

- ไฟล์วิดีโอ ----- .mp4

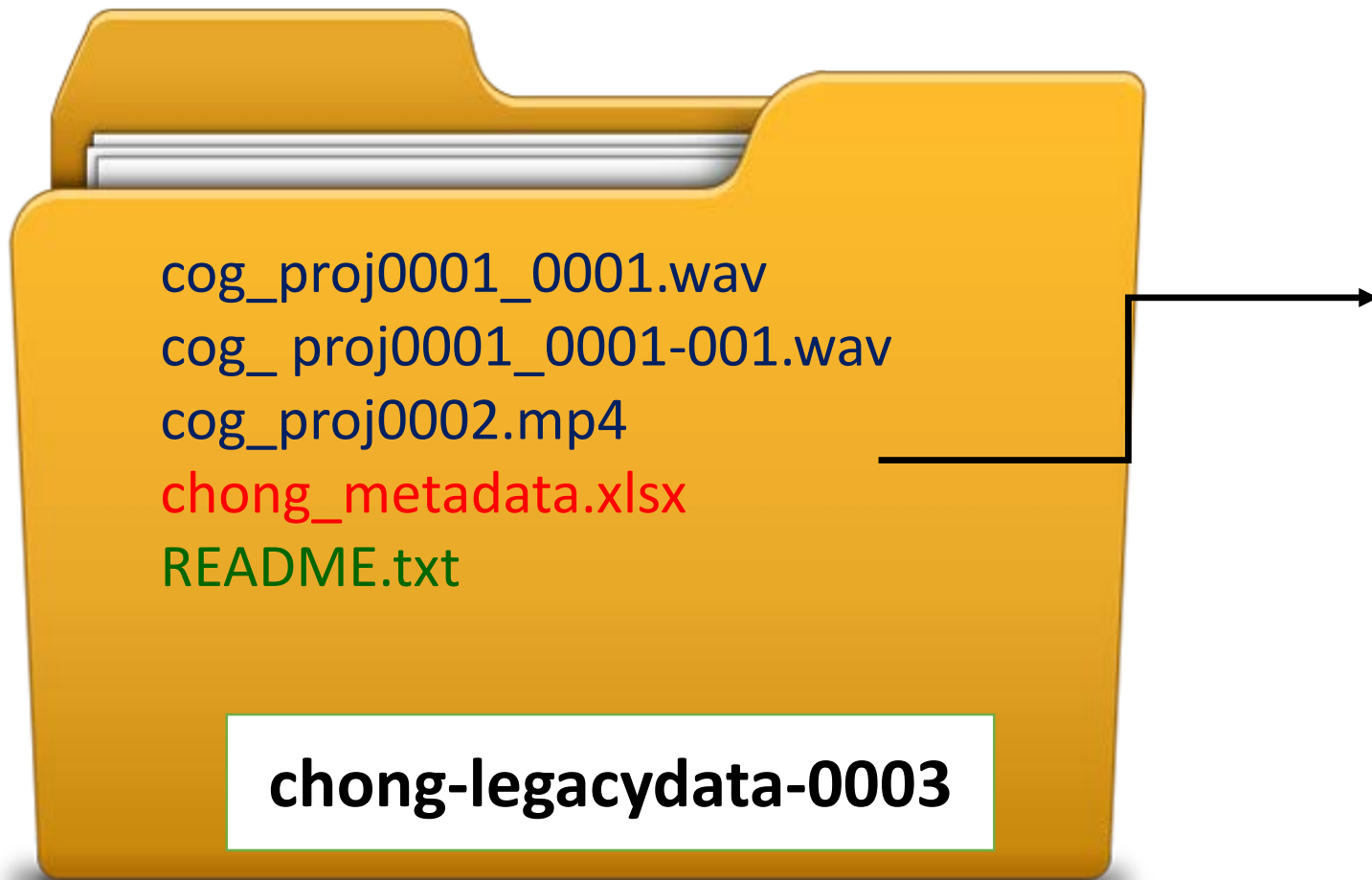
- ไฟล์เอกสาร ---- .pdf

- eaf

- Flex

- จัดระบบไฟล์ข้อมูล

การจัดระบบไฟล์ข้อมูล



ภายใน folder ประกอบไปด้วยข้อมูล

- 1) เสียง, ภาพ, วิดีโอ, ข้อมูลเอกสาร ฯลฯ โดยข้อมูลแต่ละประเภทมีการตั้งชื่อไฟล์อย่างเป็นระบบ
- 2) ข้อมูล metadata ในรูปแบบ spreadsheet
- 3) ความหมายของ code

การตั้งชื่อไฟล์

- ชื่อไฟล์ประกอบด้วย

ชื่อภาษา
• Melayu
• Chong

ลำดับเลข ระบุชุดไฟล์ข้อมูล
• _0001
• _0002

ลำดับเลข ระบุจำนวนไฟล์
• _00001
• _00002

นามสกุลไฟล์
• .jpg .pdf
• .wav .doc
• .mp4

melayu_0001_00001.jpg
melayu_0001_00002.jpg
melayu_0001_00001.pdf
melayu_0001_00001.mp4

melayu_0020_00001.jpg
melayu_0020_00002.jpg
melayu_0020_00001.pdf
melayu_0020_00001.mp4

melayu_0310_00001.jpg
melayu_0310_00002.jpg
melayu_0310_00001.pdf
melayu_0310_00001.mp4

- ข้อกำหนด

- ✓ ตัวอักษรควรเป็น ภาษาอังกฤษ ตัวเล็ก
- ✓ ตัวเลขควรนำด้วย 0 หรือ 00
- ✓ สัญลักษณ์คั่น – (hyphen) หรือ _ (underscore)

- ข้อห้าม

- ✗ ตัวอักษรเกิน 10
- ✗ ตัวอักษรภาษาไทย
- ✗ ช่องว่าง
- ✗ อักขระพิเศษอื่น ๆ

ข้อมูล Metadata

- Title หัวเรื่อง / หัวข้อ
- Summary ข้อมูลที่สรุปอย่างสั้น
- Category ประเภทข้อมูล
- Creator ผู้ผลิตข้อมูล ทุกคนที่เกี่ยวข้องกับการผลิตข้อมูล
- Contributor ผู้ให้ข้อมูล เจ้าของข้อมูล
- Original Date วันที่บันทึกข้อมูล / วันที่ผลิตข้อมูล
- Description คำอธิบายรายละเอียดของข้อมูล
- Traditional Knowledge อธิบายรายละเอียดของข้อมูลเฉพาะที่ต่างจากกลุ่มอื่นๆ
- Keywords คำสำคัญ
- Publisher ผู้เผยแพร่ข้อมูล
- Rights ผู้มีสิทธิเข้าถึงข้อมูล
- Format รูปแบบของไฟล์ เช่น MP4 ; JPG ; PDF
- Type ประเภทของข้อมูล เช่น Video ; Image ; Text
- Identifier ได้จาก website www.gguid.com
- Language เนื้อหาของภาษา
- People ผู้เกี่ยวข้องทั้งหมด
- อื่นๆ เช่น ที่มาของไฟล์ (หนังสือ หรือวัตถุข้อมูล) วัตถุ/ไฟล์อื่นที่เกี่ยวข้องกัน

ประโยชน์ของคลังข้อมูล

- เป็นการจัดการฐานข้อมูลขนาดใหญ่
- เป็นที่รวบรวมฐานข้อมูล และเป็นแหล่งเรียนรู้ทางภาษาและวัฒนธรรมของกลุ่มชาติพันธุ์ในประเทศไทย
- นักวิชาการสามารถใช้ข้อมูลจากคลังนี้ไปศึกษาเปรียบเทียบ หรือศึกษาต่อ
- เจ้าของภาษาสามารถศึกษาค้นคว้า และเป็นแหล่งเรียนรู้ของลูกหลานต่อไป
- ชำรงไว้ซึ่งความหลากหลายของสังคมพหุวัฒนธรรม
- เจ้าขององค์ความรู้เกิดความภูมิใจ ที่มีการบันทึกสิ่งที่มีค่าไม่ให้สูญหาย
- เป็นพันธกิจหนึ่งของสถาบันฯ และมหาวิทยาลัยมหิดล

การวิเคราะห์ข้อมูลทางภาษาศาสตร์

ELAN 4.9.4 - story16 (ลูกคนโต) (wong).eaf

File Edit Annotation Tier Type Search View Options Window Help

Grid Text Subtitles Lexicon Comments Recognizers Metadata Controls

Volume: 100

Rate: 100

00:00:38.757 Selection: 00:00:38.751 - 00:00:38.757 6

story16 (ลูกคนโต) (wong).wav

WON_Transcription-txt-ha

it̚ iá: m̚ hé: jón̚ jón̚ ján̚ kuó̚ n̚àj̚, ɲá: ké: láj̚ ʔè tshiw̚ ján̚ kuó̚ n̚àj̚ ké: láj̚ ʔè

WON_Words-txt-hak

it̚ iá: m̚ hé: jón̚ jón̚ ján̚ kuó̚ n̚àj̚ ɲá: ké: láj̚ ʔè tshiw̚ ján̚ kuó̚ n̚àj̚ k láj̚ ʔè

WON_morph-txt-zh-CN

也 唔 係 样样 蒸 过 吾 个 孖子 就 蒸 过 个 孖子

WON_word-gls-th

แ่น ก็ ไม้ ไม้ ทุกอย่าง หนือ กวา ฉั้น ของฉั้น นั้น ลูกชา ก็ เ่ง กวา ของฉ ลูกชา

WON_Translation-gls-th

มีไข่ทุกอย่างจะแห้งกว่าฉั้น อย่างน้อยลูกของฉั้นก็แห้งกว่าลูกของพ่อ

Interlinear-title-hak

Hakka archive - FieldWorks Language Explorer

File Send/Receive Edit View Data Insert Format Tools Parser Window Help

English Default Paragraph

Texts & Words

- Interlinear Texts
- Concordance
- Complex Concordance
- Word List Concordance
- Word Analyses
- Bulk Edit Wordforms
- Statistics

Lexicon

Texts & Words

Grammar

Texts

Text

Title Chn story 63

Info Baseline Gloss Analyze Tagging Print View Text Chart

2.1 Word ʔón tsáw tsɿŋ kuét ɲin jit liŋ31 tshien̚ ɲán̚ tshien̚.

Morphemes base ʔón tsáw tsɿŋ kuét ɲin jit liŋ31 tshien̚ ɲán̚ tshien̚.

Morphemes Chn 按照 中国 人

Lex. Gloss according to China people around one or two thousand ago

Word Gloss ตามที่ ประเทศจีน คน พันกว่าสองพันปีก่อน

Free ตามธรรมเนียมที่สืบทอดมาจากคนจีน พันกว่าสองพันปี

2.2 Word tshòn há: lòj kέ: fūŋ siók

Morphemes base tshòn há: lòj kέ: fūŋ siók

Morphemes Chn 存下 来 个 风俗

Lex. Gloss inherit, carry on come that custom, tradition

Word Gloss สืบทอด ถ่ายทอดต่อ มา ที่ ธรรมเนียมประเพณี

Free เป็นธรรมเนียมที่สืบทอดมา

3 Word kuó̚ sín ɲièn kín há: hám tshún tsiet

Morphemes base kuó̚ sín ɲièn kín há: hám tshún tsiet

20:33 5/3/2017

ตัวอย่างพจนานุกรม

Hakka Chinese Lexicon [Compatibility Mode] - Word

FILE HOME INSERT DESIGN PAGE LAYOUT REFERENCES MAILINGS REVIEW VIEW

Clipboard Paste Font Paragraph Styles Editing

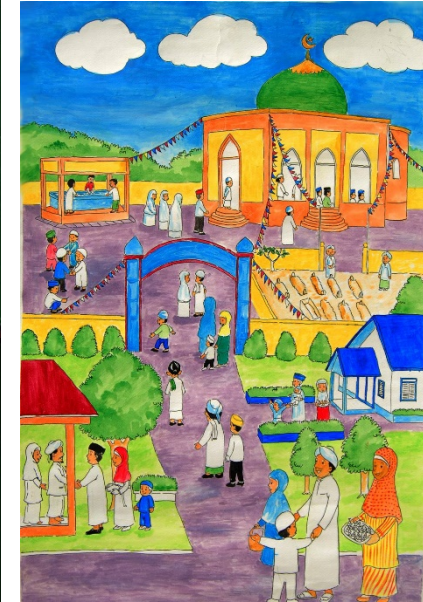
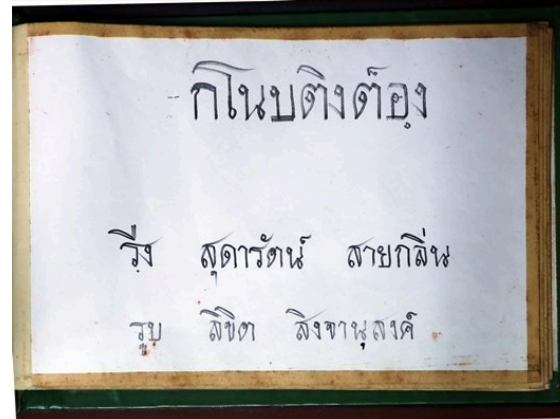
nā:
nā₂ 拿 *Transitive verb.* หยิบ ถือ; *pick, hold.*
nà₂ 泥 *Noun.* ดิน; *soil.*
náj fàn 耐烦 *Verb.* อุตทน; *patient.*
náj 7è, 哪里 *Variant: 7áj 7è. Interrogative pro-form.* ที่ไหน; *where.*
nàm 男 *Noun.* ผู้ชาย; *man.*
nàm sēn 男生 *Noun.* เด็กผู้ชาย; *boy.*
nàn 难 *Adjective.* ยาก; *difficult.*
naw31 si31 脑屎 *Noun.* สมอง; *brain.*
-ne₂ -子
-nè₁ -子 *See main entry: 1è.[nè] Variant: -mè. post.*
nè₃ 哩 *Verbal particle.* แล้ว; *already.*

nian31 sīn
nē 呢 *Verbal particle.* คำลงท้าย.
-nè₂ -哩 *verbprt.*
nèn kéw₂ 能够 *Adverb.* สามารถ; *be able.*
nèn kíw, 能够 *Adverb.* สามารถ; *able.*
nèn lít 能力 *Noun.* ความสามารถ; *ability.*
nī: 蚊 *Noun.* มด; *ant.*
nōŋ 浪 *Verb.* สิ้นเปลือง สูญเสีย; *waste.*
nòŋ fū: 农夫 *Noun.* ชาวนา; *farmer.*
núj jùŋ, 内容 *Noun.* เนื้อหา; *content.*
núj jùŋ, 内容 *Noun.* เนื้อหา; *content.*

N - ŋ

ŋ 鱼 *Noun.* ปลา; *fish.*
ŋ 五 *Cardinal numeral.* ห้า; *five.*
ŋ kuət 早国 *Proper Noun.* 早国; *name of China.*
ŋí: 遇 *Verb.* เจอ พบ; *to meet, to encounter.*
ŋí: 疑 *Verb.* สงสัย; *to doubt, suspect.*
ŋí: 你 *Variant: 你. Pronoun.* you.

ตัวอย่างการแปลงไฟล์ หนังสือเล่มยักษ์หรือจากภาพวัฒนธรรม



ตัวอย่างการเก็บข้อมูลภาคสนาม



ตัวอย่างการเก็บข้อมูลภาคสนาม



ตัวอย่างการเก็บข้อมูลภาคสนาม



ตัวอย่างการจัดอบรมการบันทึกข้อมูลและทำคลังข้อมูล



ตัวอย่างการจัดอบรมการบันทึกข้อมูลและทำคั้งข้อมูล



ข้อดีของการทำงานภาคสนาม: ทำบุญ ทัศนศึกษา



ขอบคุณค่ะ

Thank you

谢谢！

