Constructing a Yami online audiovisual dictionary

Ann Chang
ann.ronald@msa.hinet.net
National Chung Cheng University

Victoria Rau
lngrau@ccu.edu.tw
National Chung Cheng University

Maa-neu Dong
National Museum of Science
1. The significance of audio documentation

2. Language documentation of Endangered indigenous Languages in Taiwan

3. Audio supplementary material for *The Teacher’s Grammar of Yami*
   - Audio collecting and editing
   - Applying Lexique Pro to audio documentation as audiovisual supplementary materials

4. The Yami audiovisual supplementary material vs. Nathan’s (2008) four issues of audio documentation

5. Conclusion
1. The significance of audio documentation

- ‘language documentation’ or ‘documentary linguistics’
  - ‘concerned with the methods, tools, and theoretical underpinnings for compiling a representative and lasting multipurpose record of a natural language or one of its varieties’ (Himmelmann 1998, 2006)

- makes use of modern digital and media technology to make contributions to language maintenance and revitalization (Nathan & Austin 2005)

- brought us from a paper-based era to the digital era (Bird & Simons 2003)
Audio documentation for language maintenance and revitalization

Software tools: audio documentation for language maintenance and revitalization

- Lexicon editors: Toolbox, WeSay, Lexique Pro
- Creating complex textual annotations to multimedia resources: ELAN
- Segmenting audio recordings: Audacity
- Analyzing phonetics and speech: Praat

Four key audio-related issues (Nathan 2008, 2009)

- audio quality
- symbolic data/metadata of the audio data
- Mobilisation: the usage of data for practical purposes
- Protocol – documenters have to enhance the ways they deal with sensitivities and protocol in audio access and distribution.
2. Language documentation of Endangered indigenous Languages in Taiwan

Austronesian languages in Taiwan (Li, 2013)

Never being a dominant language

2% of the total population

- With the most population: Amis (200,000 people)
- With the second population: Atayal and Paiwan (less than 100,000 people)
- With the less population: Thao (only 728 people)

The lower the population of the tribe, the worst conservation it has.

Center for Taiwanese Indigenous Peoples' Education and Enterprise Development, National Chiayi University
http://www.nctu.edu.tw/aptc/content.aspx?site_content_sn=6760
pongso no tao ‘people of island’
- Orchid Island

- Take Yami (Tao) for example
  - About 4000 Tao
  - 6 villages
    - Fluently used Yami > 60 Y; Iraralay: children still used Yami in daily interaction (Rau, 1995)
    - In other five villages the use of Yami by teenagers with their parents continued to decline (Lin, 2007)
    - a language shift (Yami > Mandarin) (Chen, 1998)

Council of indigenous Peoples
(http://www.apc.gov.tw/portal/docList.html?CID=6726E5B80C8822F9)
2. Language documentation of Endangered indigenous Languages in Taiwan

The Formosan Language Digital Archive
- Academia Sinica
- Dr. Elizabeth Zeitoun (2001-2014)

The NTU Corpus of Formosan Languages
(http://corpus.linguistics.ntu.edu.tw/)
- National Taiwan University
- Dr. Shuanfan Huang, Dr. I-wen Su and Dr. Li-May Sung (2001-2003)
- Dr. Li-May Sung (2006-2012)

Digital Archiving Yami Language Documentation
- Dr. Victoria Rau (National Chung-Cheng University) and Dr. Meng-Chien Yang (Providence University)
- Since 2005 ~ present
• Since 2005, Tao research team has been constructing Yami digital archives

• including Yami online dictionary

(http://yamibow.cs.pu.edu.tw)
(http://yamionto.cs.pu.edu.tw/tao_dict/lexicon/main.htm)
Yami online audiovisual dictionary

The “Yami (Tao) Online Dictionary” is an essential website for Chinese speakers learning the Yami (Tao) language. This project was funded by the Council of Indigenous Peoples, Executive Yuan (July 1, 2007-June 30, 2009) to Providence University. Taiwan. The PI and co-PI are Professor D. Victoria Rau at Wheaton College, USA, and Ms. Maa-Neu Dong at the National Museum of Natural Science in Taiwan. The research team consists of the community language consultants, Yong-Quan Xie and Xi-Yue Tseng on Orchid Island, and IT consultant, Professor Meng-Chien Yang, and four research assistants, Ann Hui-Huan Chang, Huei-Shuan Guo, Yung-Sheng Tai, and Chia-Ying Tseng at Providence University. The Yami (Tao) Learning Dictionary was completed at the end of the first year, while the Yami (Tao) Encyclopedia will be completed the second year. All vocabulary and example sentences used for the dictionary entries were recorded by our Yami (Tao) Language consultants, whereas special cultural objects and events of Orchid Island are accompanied by photographs.
Search

Chinese - Online Dictionary - Mozilla Firefox

Search

vala [vala] 紅色, red.
mivala [mivala] mivahay o ayob na. 他的衣服是紅色的。 His clothes are red.
valag [valag] o vayo pa nimosino am, mivahay do valag. 對未婚的男的叫 valag. People who are just married live in a valag.
valag am, vahay no abo so vahay, valag is having a house in the village.
valak [valak] 陰間, 喝咒, curse (obscure), rebuke.
manvalavalak [manvalavalak] manvalavalak. 對丈夫, 在某人, yell at someone, scold someone. To manirsiring am, am ni manvalavalak. 他在跟人講話時, 好像在侮辱某人一樣, 很聲。
valalakan [valalakan] na valalakan o anak na. 他大聲的罵他的妻子。 He scolded his child loudly.
valakawan [valakawan] vahay o anaka; 紅殼蟹, type of sea crab. Grapsus torresianus.
valang [valang] 炎熱, sunshine.
kavalangan [kavalangan] kavalangan o ayob na. 他在陽光下。 Be exposed to the sun. Kavalangan kana. Come indoors, or the sun will shine on you. Pakamemen dang, ta ko kavalangan. 千萬不要去, 因為會融掉的。 Move over a bit because the sun is shining on me.
manvalang [manvalang] manvalang o ayob na. 太陽曬到他們。
The sun is baking me.

尋找: hongse
Colors

颜面 Color

bazangbang [bazangbang] 他的手红了; red. He is hurt and the red.

mabazangbang [mabazangbang] 他的嘴红的. His lips are red.

lavang [lavang] 白色的; white. Your hair is grey now.

alavang [alavang] 他的衣服很干净. His clothes are very clean.

malavang [malavang] 你的衣服很干净. Your clothes are really clean.

lin [lin] 上色; stain. It is used as dye.

malino [malino] 红色的; red. It is used as dyes. Red mud is used as dyes for painting boats.

manilo [manilo] 在船上, 在船上. The patterns on the boat.

mabek [mabek] 要求; ask. He asks me for something.
Loan words

bozo [bozo] 本・球, ball
ko mangay a manazang so bozo. 我去買球。I am going to buy a ball.

ipibozo [ipiboza] 本・打打球的原因, 打球的時間, the reason for playing ball, time to play ball.
ìkong nìo ipibozo. 你们為什麼打球? Why are you playing ball?
no sia miwalam em. or da ipiboza. 他們只有在下雨時才打球。They only play ball during the breaks.

mibozo [miboza] 本・打球, play ball, mi ta mibozo. 我們去打球。Let's go play ball.
ko mangay mibozo do gako. 我要去學校打球。I am going to school to play ball.

pibozoa [pibozaon] 本・打球的地方, place to play ball
ajin o mi ta pibozoa? 我們去哪裡打球? Where are we going to play ball?

pingk [pingk] 手機, flashlight
ina, ya mian so dingk? 妹妹, 有沒有手機?
Mom, do we have a flashlight?
mi ko manazang so dingk. 我去買手機電筒。I am going to buy a flashlight.

pingwa [pingwa] 手機, phone
ya abo o pingwa mo? 你沒有電話? You don't have a phone?
mi na makawa o pingwa namen. 我們的手機壞了。Our phone is broken.

pingdingwa [pingdingwa] 手機, 打電話給人; call someone
sino mi pingdingwa ang? 你打電話給誰? Who are you calling?
ko pingdingwa si kaka. 我打電話給我哥哥。I am calling my older brother.

midingwa [midingwa] 手機, phone, make a phone call
si makaranes ko am, midingwa ka jiaken. 你打了電話給誰? 誰打電話給你? You called who? Who called you?

naka [naka] 學校, school
ko mangay do naka. 我要去學校。I am going to school.

3. Audio supplementary material for *The Teacher’s Grammar of Yami*

*The Teacher’s Grammar of Yami* (Rau and Dong, forthcoming)

- All entries, example sentences, and story narratives
  - Uttered by the native speaker of Yami by Ms. Maa-neu Dong
  - Recorded and edited by Ann Hui-Huan Chang

Lexique Pro

- < Export as Web Page > ➔ online edition
- < Distribute Lexicon > ➔ personal edition
The website of *The Teacher’s Grammar of Yami*

http://www.ccunix.ccu.edu.tw/~Ingrau/TAO_Teaching_Web/taoteaching.html
explore

(1) how to select suitable audio equipment for the highest possible quality recording of speech

(2) how to use a digital audio editor to edit sound recording

(3) how to edit the audio data and integrate them with the text to construct high quality audio learning materials.
3.1 Audio collecting and editing

- **The recording environment**
  - quiet space
  - the speaker’s working studio: ‘National Museum of Natural Science’
  - familiar with the environment => High Quality
  - to meet the speaker’s convenience
    - on weekends
  - finished recording by July of 2014
The way of recording

- Every entry and example of the book
  - uttered twice or three times
  - Once finished recording the whole data, editing of the sound files began

- Re-recording
  - Questionable sound files
  - A revised part of the book

- Re-recorded in August the same year
When recording...

Figure 1. The speaker (Ms. Maa-neu Dong)

Figure 2. The speaker (Ms. Maa-neu Dong) and the recordist (Ann Hui-Huan Chang)
The audio file format

◆ Sound Quality
  ◆ WAV > MP3

◆ WAV
  ◆ it takes a lot of storage space

◆ The project used MP3 format

◆ The filename
  ◆ the number of its example sentence (e.g., 1_2; chapter one, second sentence)
  ◆ named by its lexicon (e.g., paha.wav).
Recording equipment

- Roland R-09 24bit/96kHz WAVE/MP3 Recorder
  - high-resolution and low-noise
  - a screen
    - the situation of volume
    - a digital distortion crackling and cracking signals
- microphone: CS-15 high quality stereo microphone
- ear-enclosing headphones
  - monitor the recording
- Batteries
  - in case there was a power failure
Audio Editor

- Use Audacity to edit sound files
- http://audacity.sourceforge.net/
- free open source digital audio editor
# Table 1 MDF: The frequently-used field markers

<table>
<thead>
<tr>
<th>Field marker</th>
<th>English</th>
<th>sf</th>
<th>sfx</th>
<th>pc</th>
<th>mr</th>
<th>sy</th>
<th>an</th>
<th>sd</th>
<th>is</th>
<th>cf</th>
<th>ce</th>
<th>cn</th>
<th>mn</th>
<th>va</th>
</tr>
</thead>
<tbody>
<tr>
<td>lx</td>
<td>lexeme</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ps</td>
<td>part of speech</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ge</td>
<td>gloss (English)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>gn</td>
<td>gloss (national language)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>de</td>
<td>description (English)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>dn</td>
<td>description (national language)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>re</td>
<td>reversal (English)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rn</td>
<td>reversal (national language)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>xv</td>
<td>example</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>xe</td>
<td>example gloss (English)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>xn</td>
<td>example gloss (national language)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rf</td>
<td>example reference</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>bw</td>
<td>borrowed word</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Examples of audio data

(1) apeleh ‘very few’ from the 500 Most Commonly Used Yami Words

\lx apeleh lexeme
\wx apéreh variant
\gn 一點點 gloss (national language)
\dn 一點點 definition (national language)
\rn yidiandian reversal (national language)
\sf wordlist/apeleh.mp3 sound file for pronunciation of lexeme (relative to database directory such as wordlist)
\sd 1 semantic domain (Chinese stroke number: 1 )
\sd wordlist semantic domain (the list of frequently-used Yami)
\sd y semantic domain (Mandarin Phonetic Symbols/Bopomofo Symbols: Yi)
\sd bu 1 semantic domain (Chinese radical: yi)
\dt 18/Aug/2014 Date entry was last edited

(2) nionay a libangbang

\vx (2) tosy a, ji mo pavilanga ori. example
\sf nionay a libangbang\nionayalibangbang2.mp3 sound file of example
\dt 23/Aug/2014 date entry was last edited
2). Export to Web Page

Workflow

Lexique Pro
<Export to Web Page>

Upload to the Internet

Online edition of Dictionary

<Frame> or <Frameless> version

Create a ‘htm’ Folder

Open to the public

which fields & which language to display

filename of the web & a location to save the file
Audiovisual supplementary material

Yami (Tao) Grammar Reference
- 16 chapters
- Audio files of entries and example sentences

Two story narratives
Sound files of
- Full text
- Paragraphs

The list of 500 frequently-used Yami
- Tao
- Bopomofo Symbols
- Chinese radical
- Chinese total strokes

Online edition
Example of **Online edition** of Audiovisual supplementary material -1
Example of Online edition of Audiovisual supplementary material -2
3). Distribute lexicon Wizard

- **Database**
- **Lexique Pro**
  - <Distribute lexicon Wizard>
  - If including images or sounds files in the results?
- **If letting others edit, export or print the data?**
- **If encrypting the data?**
- **If including introduction to SIL?**
- **choosing the filename and where to put the files**
- **install program of the data (an executable file with .exe format) = personal edition (e-book)**
- **Distribute the database to others**
Example of personal edition of Audiovisual supplementary material -1
Example of personal edition of Audiovisual supplementary material -2
4). Problems using Lexique Pro

◆ .wav OR .mp3

<table>
<thead>
<tr>
<th>WAV format</th>
<th>MP3 format</th>
</tr>
</thead>
<tbody>
<tr>
<td>Uncompressed format</td>
<td>compressed format</td>
</tr>
<tr>
<td>high quality lossless audio format; faithful to the original</td>
<td>loss audio format; drop off in sound quality</td>
</tr>
<tr>
<td>take a lot of storage space</td>
<td>don’t need much space</td>
</tr>
</tbody>
</table>

◆ Online edition (e-book)

◆ Take a lot of audio files
◆ convert .wav > .mp3
◆ Save space
◆ Make Lexique Pro run faster

◆ The place to put sound files

◆ put at the same directory folder as the database
◆ Avoid “unable to find the sound files”
### 4. The Yami audiovisual supplementary material vs. Nathan’s (2008) four issues of audio documentation

<table>
<thead>
<tr>
<th></th>
<th>The four issues of audio documentation (Nathan 2008)</th>
<th>evaluate</th>
<th>The Yami audiovisual supplementary material</th>
</tr>
</thead>
<tbody>
<tr>
<td>Audio quality</td>
<td>+</td>
<td></td>
<td>• Experienced recordist</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>• Using professional audio equipment</td>
</tr>
<tr>
<td>Symbolic data (i.e. metadata)</td>
<td>+/−</td>
<td></td>
<td>• Speaker contains only one person, and state it on the introductory webpage;</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>• Documented the date the entry was last edited, but not the collecting date.</td>
</tr>
<tr>
<td>Mobilization: the usage of data for practical purposes</td>
<td>+</td>
<td></td>
<td>• It is for Yami people to gain literacy in their mother tongue.</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>• It also documents and preserves the fast-fading language and culture.</td>
</tr>
<tr>
<td>Protocol: the need for respect language speakers’ rights and sensitivities in audio access and distribution</td>
<td>+</td>
<td></td>
<td>• It contains no sensitive information.</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>• Everyone has equal access to the materials.</td>
</tr>
</tbody>
</table>
The audiovisual supplementary material contains three editions

- **personal edition (e-book)**
  - allows a learner to follow the text and listen to audio entries and example sentences pertaining to the book

- **online edition**
  - does not include the text, but only provides sound files of every entries and example sentences pertaining to the book

- **simple sound files**
  - don’t have any paper work; an original way to keep the data

As important **digital resource** because they allow anyone to use, reuse, and access to the language, and increases the value of a language documentation.
Project Workflow

1. how these three authors cooperated during the project, what work we have completed
2. what audio tools/software we could use for recording, editing and documenting the language to construct an online audio dictionary (learning material)
Project Workflow

From a paper-based era to the digital era

The Teacher's Grammar of Yami

The Yami speaker
Ms. Maa-neu Dong

Audio data recording
Roland R-09 Recorder
CS-15 Microphone

Recordist
Hui-Huan Chang

Audacity
Editing sound files

Recording
Simple sound files

Using Lexique Pro for Editing


Discussion

Finish and Export Database
1. <Export as Web Page> = online edition
2. <Distribute Lexicon> = personal edition (e-book)

Online edition

personal edition (e-book)

Simple sound files

Recordist
Hui-Huan Chang

Dr. Victoria Rau

http://www.ccunix.ccu.edu.tw/~Ingrau/TAO_Teaching_Web/taoteaching.html


Reference-2

Ayoy!
Thank You!