

Populist platforms and strategic documentation in a changing Southeast Asia

Daniel Kaufman, ELA

May 29, 2013



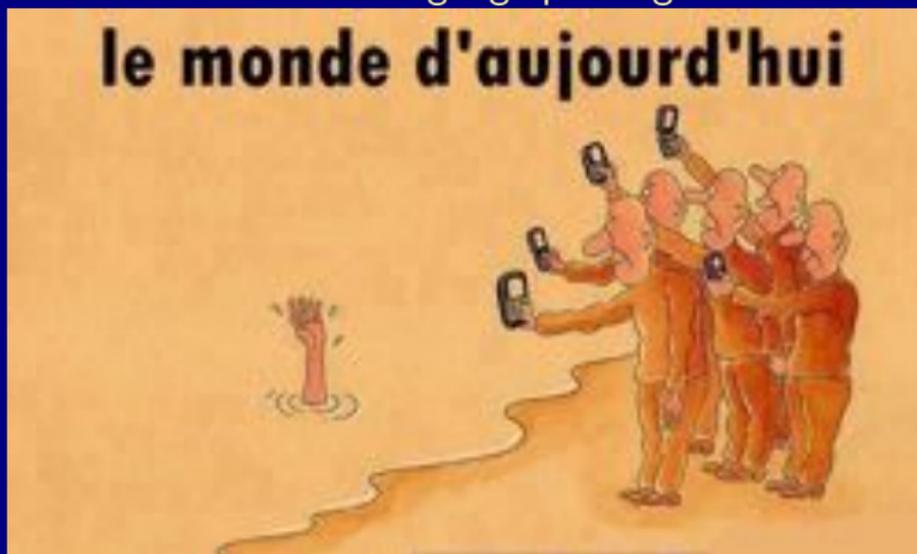
- 1 Background
 - A changing (Island) Southeast Asia
- 2 Traditional methods
 - Descriptive fieldwork
 - Documentation workshops
- 3 Alternatives for the future
 - Two ugly truths
 - The ELP funded Timor project
- 4 Conclusion

Endangerment

- Ethnologue has recently posted new statistics on the vitality of the languages of the world following the EGIDS scale.
- In many areas, there appears to be a disconnect between the ratings and the “reality on the ground”.
- Indonesia as a whole represents an endangerment zone.
 - There is little to no support for local languages – no press, no radio, no status. Music cassettes and karaoke are the only media produced and circulated.
 - Until recently, corporal punishment was the norm for children who were caught speaking their home language in school.
 - In most of the country, there has been a rapid break in language transmission between generations.
 - Unlike in the Philippines, the introduction of the internet does not appear to have given a substantial boost to local languages. (Although local language music videos are now more accessible.)

Endangerment

- If trends continue, Indonesia could easily be among the world's epicenters of language loss.
- In this talk I will assume, that some agreement that documentation should be more than language pickling for scholars.



“The world today”

The Lone Ranger

- The “lone ranger” descriptivist remains the dominant fieldwork paradigm in linguistics, and for good reason.
- The lone ranger can write excellent grammars and dictionaries efficiently. But, this can require anywhere from 4 to 10 years with steady institutional support.
- So can this approach make a significant impact on the endangerment crisis?
 - Positive scientific impact, but keep in mind that documentation \neq conservation.
 - Many fieldworkers now build community projects like “community dictionaries” and pedagogical materials into their descriptive fieldwork.
 - But there are still not nearly enough fieldworkers in places like Eastern Indonesia to address the issue where it matters most.
- The growing consensus is that our fieldwork needs *multiplier effects*.

Looking for multiplier effects

- Sustainable community-based projects.
- Can documentary projects sustain themselves without a linguist?
- There are certainly examples of this, but for every good example there are many failures.

Documentation workshops

- A recent approach has been to teach speakers digital documentary skills in the hope that they will go on to create and archive their own documentation.
- Such workshops typically focus on hardware and software training and have been described as “capacity building”.
- Unfortunately, there are very few examples of such workshops leading to sustained community based documentation, as the driving force too often comes from the outside.

Ugly truth #1

- To have a maximum impact on obsolescence and even documentation, we need to stop being linguists!
- If our goal is maximum documentation with an eye towards conservation, we should be creating a *language corpus for younger generations that is **useful**, **accessible** and **meaningful***.
- At the outset, such a corpus “only” requires transcriptions and translations, which can and should be produced by the speakers themselves.
- This leaves us as facilitators:
 - helping speakers create the best recordings possible (both technically and content-wise)
 - helping speakers interview each other
 - helping speakers transcribe and translate their recordings in a regular fashion
 - putting all this material together

Ugly truth #2

- “Capacity building” only works in cases where participants are really eager to use technology and where easy access to technology exists.
- Few people get excited about the prospects of learning complex and often confusing software.
- Software changes so rapidly that even if linguistic capacity building was proven to work, it would not be very sustainable in rural areas.
- At the risk of perpetuating the unfortunate technological asymmetries that exist today, let everyone do what they’re comfortable with and already good at!

The Endangered Language Project (ELP)

- Initiated by Google.org as a collaborative project with linguist list (Eastern Michigan University), University of Hawai'i at Mānoa, the First Peoples' Cultural Council (British Columbia, Canada). Governing council includes members of several other organizations including the Endangered Language Alliance (ELA).
- The main ideas:
 - A catalogue of endangered languages (ELCat) based on the most recent data and complete with citations (cf. Ethnologue and UNESCO's Atlas of Languages in Danger).
 - An aggregation of documentary materials on endangered languages.
 - A forum for speakers of endangered languages to create and upload their own media.

Timor 2012

- In 2012, ELA secured funding and support from ELP (Google.org), together with Mark Donohue (Australian National University) and Charles Grimes (Unit Bahasa dan Budaya, Indonesia), to run a two week “documentation workshop” with 15 speakers of local languages from the Timor-Rote area.
- The two previous ugly truths formed the basis of our approach, meaning:
 - (i) Minimum attempts at technology transfer and (ii) minimum linguistics.
- This resulted in 271 audio recordings, many with accompanying video and almost as many transcriptions + translations.
- Unfair advantage: many of our participants were experienced translators who had been working with UBB and were comfortable transcribing their language (but had never done so before).

The participants



The workshop



Making recordings



Transcribing their own stories



Making time-aligned transcripts using Praat



Timor 2012

- The results were “processed” and converted to web videos after the workshop.
- Because of the ELP funding, it was decided from the beginning that the main presentation of our output would be on the ELP site via YouTube.

The “front page” for Helong

Create your account · Sign in · Change Language: EN

Endangered Languages
A project by the Alliance for Linguistic Diversity

Explore Knowledge sharing Blog Search on this site

Helong

(aka Helon, Semau, Kuyang)

Classification: Austronesian · vulnerable

Description Samples Resources Activity Bibliography Subscribe

SAMPLES

The Helong language in school

Helong - Local languages in the school

COMMUNITY RATING

☆☆☆☆☆

Flag as inappropriate

02:18 / 12:36

Metadata

Language metadata

ALSO KNOWN AS	Hetón, Semau, Kupang, Kapong
CLASSIFICATION	Austronesian, Malayo-Polynesian, Central-Eastern Malayo-Polynesian, Central Malayo-Polynesian, Timor, Hetong
CODE AUTHORITY	ISO 639-3
LANGUAGE CODE	heg
VARIANTS & DIALECTS	<ul style="list-style-type: none"> Hetong Darat Hetong Pulau Funal
ISO KNOWLEDGE	As CV+
MORE RESOURCES	OLAC search



LOCATION INFORMATION

COORDINATES +10.1905, 123.5247

COMPARE SOURCES (1)

Information from: "Atlas of the World's Languages in Danger", Christopher Mosley - UNESCO Publishing

Language information by source

COMPARE SOURCES (5)

Information from: "Ethnologue: Languages of the World", M. Paul Lewis - SIL International

Vulnerable

Uncertain

14,000

Native speakers worldwide in 1997 (Grimes, Theria, Grimes, jac08)

ETHNIC POPULATION
15,000-20,000

PLACES
Indonesia:

Want to spread the word and help raise awareness?

Post it to your profile

[Ret](#)
[Tweet](#)
[Like](#)

Activity ranking

The collection

Description **Samples** Resources Activity Bibliography Subscribe

Videos

Sort by: Recently Added | Best Rated | Most Viewed

The Helong language I...
0:0

The Helong language I...
0:0

The Helong language I...
0:0

Swadesh list (full)
0:0

Interview
0:0

Helping someone give...
0:0

How to make wrapped...
0:0

The History of PNEEL c...
0:0

Helong - Discussion of...
0:0

[View more video samples](#)

Add language samples

If you have any Helong samples you would like to share please upload them to the platform.

[Upload new sample](#)

Activity ranking

The YouTube advantage

- Offers the best discoverability and accessibility by far.
- Subtitles are easily produced from textgrids and ELAN files (via ELAN's export to subtitles function). These can be uploaded directly with a video.
- Each subtitle file supports two tiers but multiple files with translations into different languages can be uploaded as well.
- YouTube also makes available the entire transcript, which can be copied and searched as well as used to navigate the video.

Transcript and translation



Koda - Fishing

dzami haku ja?akuçe
Does Jami catch fish?

0:21 / 0:51

The image shows a video player interface. At the top left, the title "Koda - Fishing" is displayed. The main video area shows two men in a conversation. The man on the left is wearing a patterned shirt, and the man on the right is wearing a red and white striped shirt. Yellow subtitles are overlaid on the video, reading "dzami haku ja?akuçe" and "Does Jami catch fish?". At the bottom of the video player, there is a progress bar showing the video is at 0:21 out of 0:51. Various control icons like play, volume, and settings are visible.

Switching to gloss

Koda - Fishing

Settings...

Translate Captions

English

English - Gloss

Turn Captions Off

dzami haku ja?aku

Does Jami catch fis

0:21 / 0:51

Gloss

Koda - Fishing

dzami haku ja?a-aku-ʔe
Jami fish hold-3pl-PRES

0:21 / 0:51

Things you can't do on YouTube/ELP

- YouTube shouldn't substitute for an actual archive.
- You can't search across multiple transcripts as you would want in a corpus (although most archives do not allow this either).
- Solution: Separate media popular presentation from corpus functions.

Gloss-a-tron

- Together with Raphael Finkel, head of the CS department at U. Kentucky, we are developing a web front-end to SIL's Fieldworks software.
- This allows for searching over multiple texts and lexicons using regular expressions, similar to Fieldworks.
- Each line will be linked to a sound clip, but not to the entire media file.

Gloss-a-tron

17 ●
 he ● dzailategen ● haku ● jaʔakutaku ●
 ● N ● N ● ●
 he ● dzailate ● haku ● jaʔa ●
 yes ● fishermen ● fish ● hold ●
 -gcti ● -aku ●
 -FOC ● 3.pl ●
 -ta ●
 PRES ●
 -aku ●
 3.pl ●

Koda text dict

General search

Sequence number

Word: Text

Word: Gloss

WORD: PART OF SPEECH

Morpheme: Text

Morpheme: CF

Morpheme: Gloss

Morpheme: MSA

Morpheme: Type

Ignore accent marks in the document

Conclusions

- Two major changes in Southeast Asia with major implications for linguists: Increasing access to technology and decreasing diversity.
- How can we use the former to stem the latter?
- Documentation that is inviting and engaging leads to wider interest among the speaker community itself as well as potential funders.
- While archives have focused on how to keep data protected, we should be focusing on how our work can reach as wide an audience as possible.

Conclusions

- Workshops of the type described here provide an ideal place to create documentation with speakers of multiple languages.
- Workshops in the rural SEA context can produce far more results if we scale back on:
 - (i) the elusive goal of technology transfer
 - (ii) the goal of analyzing as we go

Thank you

ขอบคุณมาก!