



# **Basic Oral Language Documentation in the Philippine Context**

Levi C. Cruz

[levi\\_cruz@sil.org](mailto:levi_cruz@sil.org)

Translators Association of the Philippines

---

- **Language endangerment** has become a mainstream issue among linguists
- **Basic Oral Language Documentation (BOLD)** method was designed to increase the speed and efficiency of language documentation

background

---



Ayta Abellen [abp]



Klata Bagobo [bgi]

- **Himmelman** (1998) *"The aim of language documentation is to provide a comprehensive record of the linguistics practices characteristic of a given speech community"*
- **Woodburry** (2011) *"the creation, annotation, preservation and dissemination of transparent records of a language."*

## LANGUAGE DOCUMENTATION

---

- “the coming century will either see the death or the doom of 90% of mankind’s languages”  
(Krauss 1992)

- We must document languages before they are gone forever

-Language documentation aims to record endangered languages before they become inaccessible

# LANGUAGE DOCUMENTATION

---

- The product is the primary data —a corpus of recorded speech events that document the language in actual use
- Uses both audio and video recordings.
- Not an alternative to description, but a complement to it.

## **LANGUAGE DOCUMENTATION**

---

## THREE BASIC TASKS

- “*Language Documentation is concerned with compiling, commenting on, and archiving language documents.*” — Himmelmann
- **1. Compile** - a sample of recordings of a full range of speech event types
- **2. Comment** - on those recordings  
E.g., transcription, translation, discussion, situational context, informed consent to share
- **3. Archive** - the complete corpus of recordings and commentary with an institution that will provide long-term access

# LANGUAGE DOCUMENTATION

---

- **Basic Oral Language Documentation (BOLD)** method was designed to increase the speed and efficiency of language documentation
  - By using an oral approach to annotation (transcription, translation, analytical comments)

## **BOLD METHOD**

---



# Comment orally:

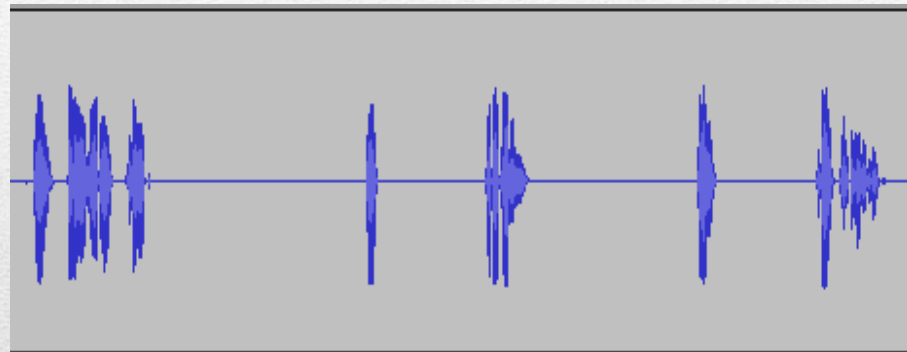
- Many languages that need documentation have neither an orthography nor a literate community. BOLD method focuses on Oral Commenting, or *Oral Data Annotation*
- **Oral Data Annotation:** recording language information about a media file through audio (and/or video) recordings of MT speakers, tagged to the original media files.

## **BOLD METHOD**

---

## Oral Annotation: An Example from Ayta Abellen [abp]

I recorded a  
communicative  
event in Ayta Abellen

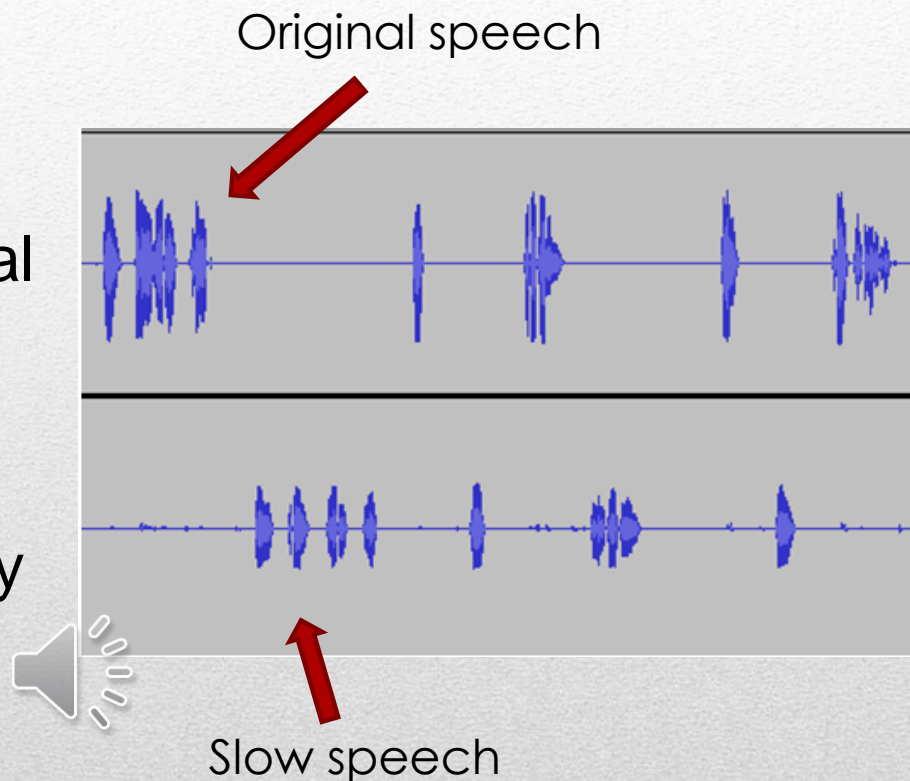


# **BOLD METHOD**

---

## Comment orally: An Example from Ayta Abellen [abp]

In a field “studio”, a native speaker listens to the original recording from a player, hits pause button at natural breaks, then repeats the segment slowly and carefully onto a recorder



# BOLD METHOD

---

- **The form of the corpus**

- A well-formed BOLD corpus contains:
- A document introducing the language, people, project, coverage, methods
- A table of contents listing each item
- A set of fully commented items
- A fully commented item consists of:
  - Recording
  - Informed consent
  - Situational metadata
  - Oral transcription
  - Oral translation

# **BOLD METHOD**

---



**People Group:** Ayta Abellen [abp]

**Population:** 3,500 (Stone 2005)

**EGIDS:** 5 “developing”

**Bilingual in:** Tagalog, Ilokano, Kapampangan

**When:** June 2010

**Purpose:** To test BOLD Method, collect more text for further linguistic analysis

**Communicative Events:**

narratives

songs

conversations

procedural discourse

cultural description

**Lists:** 207 Word list

---

**Ayta Abellen**

# Audio Capture ( 3 weeks)



- Initial visit to the community to meet the contacts, community leaders
- looked for speakers with the help of one LRP
- Explain the purpose to each speaker
- Recording of original speech
- Recorded with Zoom H2 at 24-bit / 48kHz
- Record metadata
- Logbook entry
- Obtain informed consent



- Wordlist with audio and video
- Using mirrored recliner

**Pros:** provides visual data, of simultaneous  
• frontal and profile face views of articulations

**Cons:** inconvenient to carry if traveling alone

Ayta Abellen

---

# Oral Transcription / slow-speed (1 week)



- Took two native speakers to Manila for slow-speed recording
- Used a make-shift studio inside the office
- Used two Zoom H2 recorders (one as player, another as recorder)



- Written transcription done by one native speaker while the other was recording (using their orthography)

Ayta Abellen

---

## Stage 3: Oral Translation (1 week)



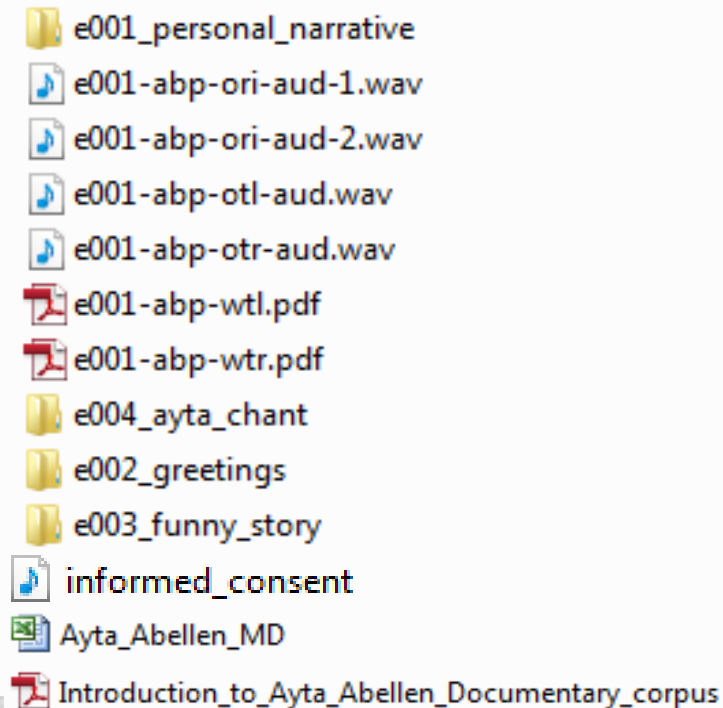
- Another whole week of recording for Oral translation; done by only one native speaker.
- Followed the same process of oral transcription

Ayta Abellen

---



# Preparing archive-ready corpus



- Made sure each file has the appropriate, unique file name
- Organizing files for the archive ready corpus:
  - The complete index of items (with metadata) in the corpus
  - The complete index of files (with metadata) that comprise the items
  - The complete index of contributors (with metadata)
  - The informed consent documentation (whether community level or by individuals or both)
  - The complete set of files that comprise the items



**People Group:** Klata Bagobo [bgi]

**Population:** 55,000 (1990 Census)

**EGIDS:** 7 “shifting”

**Shifting to:** Cebuano

**When:** July 2011

**Communicative Events:**

narratives

greetings & leave takings

conversations

procedural discourse

cultural description

**Lists:** 207 Word list

## Klata Bagobo

---

# Community Leaders Consultation



- Met with the tribal leaders for consultation, rather than simple “courtesy call”
  - Introduced ourselves, explained our purpose
  - Tried to apply a more community-centric approach by asking questions such as “do you want to document your language?” and “what do you want to record”?

Klata Bagobo

---

# Audio Capture



- Most speakers we recorded were sent by the tribal leaders
- Recorded with Zoom H2 at 24-bit / 48kHz
- Record metadata
- Logbook entry
- Obtain informed consent

Klata Bagobo

---

# Oral Transcription / Translation



- Oral transcription and translation done in the field
- Used **Saymore**, a Language Documentation Management Software
  - No need to use two recorders
  - More native-speaker friendly
  - Manages files automatically
  - Follows the BOLD method
  - Easier to tag metadata

The screenshot shows the SayMore software interface. At the top, there is a menu bar with "Project", "Session", "Person", "Progress", and "Help". Below the menu bar, there is a "Sessions" tab. The main area is divided into two panes. The left pane shows a list of sessions with columns for "Id", "Title", "Stages", and "Status". The right pane shows a detailed view of a session, including fields for "Name", "Type", "Date Modified", "Size", and "Duration". Below this, there are tabs for "Session", "Status & Stages", and "Notes". The "Session" tab is active, showing fields for "Id", "Date", "Title", "Setting", "People", "Location", "Genre", "Access", "Situation", and "Abstract".

Id	Title	Stages	Status
E001-gbw	History and some observations	□□□□	○
E002-gbw	event 2	□□□□	○
E003-gbw	event 3	□□□□	○
E004-gbw	Why Klata language declined	□□□□	○
E005-gbw	Funny Story	□□□□	○
E006-gbw	Akang's testimony	□□□□	○
E007-gbw	Klata dish	□□□□	○
E008-gbw	Translated Hymn	□□□□	○
E009-gbw	Superstitious belief	□□□□	○
E010-gbw	event 10	□□□□	○
E011-gbw	Pekpek	□□□□	○
E012-gbw	Klata Marriage	□□□□	○
E013-gbw	being a preschool teacher	□□□□	○
E014-gbw	Greetings and Leave-takings	□□□□	○
E015-gbw		□□□□	○
E016-gbw	Accident story	□□□□	○
E017-gbw		□□□□	○
E018-gbw		□□□□	○
E019-gbw	Klata House Discussion	□□□□	○
E020-gbw	Klata Justice	□□□□	○
E021-gbw	Klata Justice Discussion	□□□□	○
E022-gbw	Klata Wedding	□□□□	○
E023-gbw	Klata Wedding Part 2	□□□□	○
E024-gbw	Preliminary Wedding	□□□□	○
E025-gbw	Monkey and the Turtle	□□□□	○
E026-gbw	Death of Datu	□□□□	○

Name	Type	Date Modified	Size	Duration
E001-gbw.sessi...	Session	5/2/2013 3:48:16 ...	489 B	
E001-gbw_oriv...	Audio	1/1/2007 4:11:36 ...	64.52 ...	00:03:54
E001-gbw_...	Annotations	12/11/2011 8:21:...	22 KB	
E001-g...	OralAnnotations	12/13/2011 2:40:...	95.36 ...	00:00:00
labels.txt	Text Document	12/11/2011 8:21:...	1 KB	

Id	Date
E001-gbw	7/23/2011

Title	Setting
History and some observations	under a shed

People	Location
Cesar Betil	Brgy. Tamayong

Genre	Access
Report	

Situation	Abstract
Datu is sitting, infront of 5 people	

Klata Bagobo

# Archiving

Work is not done until the corpus is committed to an archive for long term preservation and access.

At this time, neither Aytá Abellen nor Klata Bagobo LD project have been made publicly available through an archive that is publicly searchable and accessible on the internet

Klata Bagobo

---

	<b>AYTA ABELLEN</b>	<b>KLATA</b>
<b>EGIDS Level</b>	5- Developing, but could be 6a based on a recent visit.	7- Shifting
<b>Documentation Status</b>	Grammar sketch available Audio recordings- none	None
<b>Purpose</b>	To test BOLD method and gather more text for discourse analysis	Preliminary documentation (build oral corpus)
<b>Community Involvement</b>	minimal	Community leaders are aware, and willing to help
<b>Prior Relationship with community</b>	high	low
<b>Personnel</b>	1 Language Documenter; 1 paid LRP	A team of 4; without a paid LRP
<b>Oral Annotation</b>	-Using two recorders -done after audio capture	-Using Saymore Software -done simultaneously with audio capture
<b>Word List</b>	Audio and video, with mirrored recliner	Audio only

## SUMMARY OF COMPARISON

---

### **Personnel**

A team of 2 or 3 members rather than a single researcher in a Language Documentation project. A trained linguist in the team is ideal, but not necessary. Intensive training should be given to all members, on BOLD protocols, recording techniques, softwares, metadata, etc.

Presence of a full-time, paid cultural insider in the team.

### **Community involvement**

Community leaders were aware of the purpose, and involved in decision-making processes. We saved time looking for speakers as they were provided/recommended by the community leaders.

### **Saymore Software vs. Two-recorders**

Use of Saymore software rather than two recorders. It is user-friendlier and avoids pauses made by the annotator, which do not correspond to natural pauses in the original speech during oral annotation.

Facilitates written transcription, file-naming and file management.

### **Word List**

Use of mirrored recliner, as it provided simultaneous views of articulation.

## What worked well

---





January 2012

Language Revitalization effort started by Klata leaders. Klata speakers in Calinan gathered during the “**Klata Awareness Day**”, where they as a community planned how to intentionally promote language use especially among young generation.



outcome

---

1. Language communities in the Philippines in many ways can benefit from an oral approach of documentation, which can produce a corpus in a short period of time with a minimal budget required.
2. with an oral approach, preliterate communities can easily participate in the documentation process.
3. With this simple technique, and technology available, anyone with proper training and interest in preserving languages can do so. This will ensure that languages are documented before they are gone.

## CONCLUSION

---

**Salamat po**

---