Language documentation, description and revitalisation

Peter K. Austin

SOAS, University of London

University of Oxford

Bolzano, Italy, 2019-02-11

Outline

- What is language documentation?
- What is language revitalisation?
- Methods and tools
- Current directions and trends – some critical reflections (see Readings)
- Relationships between documentation, description and revitalisation
- Colonialism in documentation and revitalisation – is it time for a change?

Language Documentation – past

Language documentation 1

- Term widely used in late 19th and early 20th century to refer to the study of indigenous languages in the Boasian tradition, characterised by:

    - brief summer fieldwork
    - collection of texts, vocabulary and grammatical forms
    - part of broad anthropological enterprise to 'save' disappearing cultures
    - training and engagement of informants as data producers and co-authors
    - use of latest technology

- goal: production of 'Boasian trilogy': text collection, grammar, dictionary
- (much material ends up in archives but not as a goal)

Language documentation 2

- "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)
- Features:
  - *Focus on primary data*
  - *Accountability*
  - *Long-term storage and preservation of primary data*
  - *Interdisciplinary teams*
  - *Cooperation with and direct involvement of the speech community*
- Outcome is **annotated and translated corpus** of archived representative materials on a language, cf. DoBeS, ELAR
- Woodbury (2003, 2011) 'transparent records of a language'

Drivers

- developed since 1995 in response to the urgent need perceived by researchers to make an enduring record of the world's many endangered languages and to support speakers of these languages in their desire to maintain them, fuelled also by developments in information, media and communication technologies

- concerned with roles of language speakers and communities and their rights and needs

- is not limited to endangered languages – can be applied to any linguistic variety

Big money – ELDP projects

Big archives – ELAR at SOAS

Big archives – DoBeS at MPI Nijmegen

Components of documentation

- *Planning* – language, funding, fieldwork, equipment
- *Recording* – of media and text (including metadata) in context
- *Transfer* – to data management environment
- *Adding value* – transcription, translation, annotation, notation and linking of metadata
- *Archiving* – creating archival objects, assigning access and usage rights
- *Mobilisation* – creation, publication and distribution of outputs

Language description

Is concerned with questions like:

- What is a language system/grammar?
- To what extent are languages alike and to what extent are they different?
- What does this tell us about the human mind?
- What does this tell us about human communication?
- How does a language system work and how is it acquired?

Language description requires

1. Asking the right questions/collecting relevant data. Rice (2006: 236) argues that formal syntactic theory forces a grammar writer to ask questions that are not very likely to be asked otherwise.

2. Making generalisations and drawing distinctions about the grammar of languages. In other words, descriptions must be generalizable, rather than simply observational, that is, they must represent generalizations about the described linguistic system.

3. Labelling and categorizing the phenomena in one way or another (i.e., you need a 'metalanguage', comparable concepts, terminology …)

4. A theory that underlies this, e.g., 'generative' or 'functional' mechanisms and a model for explanations.

5. Presenting data and analyses in a way that is acceptable and interesting to a wider audience (and 'replicable'?) – a "grammar" or "dictionary" as an academic object

Workflow differences

Description

something happened

something *inscribed*

NOT OF INTEREST

applied knowledge, made decisions

cleaned up, selected, analysed

representations, lists, summaries, analyses

presented, published

FOCUS OF INTEREST

Documentation

recapitulates

something happened

applied knowledge, techniques

recording

made decisions, applied linguistic knowledge

representations, eg transcription, annotation

archived, mobilised

FOCUS OF INTEREST   **FOCUS OF INTEREST**

Documentary methods: how do we collect linguistic data?

Methods

- There are several well-tried data collection methods and each has its own advantages and disadvantages:

    - elicitation

    - staged communication (experiments)

    - participant observation

- Some linguists insist that you have to learn the language being studied and use it as much as possible (monolingual fieldwork) while others rely on lingua francas (eg. Spanish, Thai, Bislama) and translation, and others use a mixture

Collecting data

What you might want to collect in the field as a linguist:

(1) **Non-structural data**

- sociolinguistic situation, multilingualism (information on other languages spoken in the area; the number of speakers; the degree of their competence and the correlation with age, gender etc.)
- basic ethnographic and anthropological information you will need in order to understand language use, including texts and sentence examples (material and non-material culture; samples of folklore)
- historical information

Collecting data

(2) Data for a specific linguistic project (but if the language is highly endangered, it may be a good idea to record everything you can, even if you don't have time to analyse everything immediately - you never know what will be useful later on!)

(3) Data for a "comprehensive" linguistic description

(4) Documentation of a language (representation of the observable linguistic behaviour manifested in everyday interaction between members of the speech community, and their metalinguistic understanding of them)

Collecting data

Two main methods of data collection:

1. Observation ("unobtrusive" methods): **non-participant observation** (analysis of written and spoken spontaneous speech events, media etc.) and **participant observation** (e.g. personal participation in a conversation or ritual; recording can be made by members of the community).

2. Experimentation ("obtrusive" methods): linguistic **elicitation**; stimulus-response testing etc. and **introspection**.

Be aware of the **observer's paradox**

Metadata

Description of the recording event:

- item ID
- keywords (content)
- additional information about the topic recorded
- cross-references (links to video, photos, etc.)
- length

etc.

Metadata

Context of the recording event:

- recording person
- recorded person
- place
- time
- participants
- equipment utilised

etc.

Metadata

Information about the consultant(s) (can be filled in gradually)

- name (and possibly a nickname)
- date and place of birth, age
- clan/tribe (if relevant)
- languages
- education, occupation
- nationality of the parents
- marital status

etc.

- Different data types and methods depend on project goals
- Phonetics/phonology research
- Grammatical research
- Semantic and pragmatic research
- Discourse and ethnographic research

- **Phonetics/phonology research:**
  - Variety of speakers needed
  - Wordlists (see also Chelliah & De Reuse)
  - Transcribe in IPA initially
  - Minimal pairs
  - Suprasegmentals: tone, stress, intonation
  - Use of frames (cf. listing intonation)
  - Acoustic analysis
  - Orthography development

Elicitation for grammar

- **Contextualising elicitation**:
  - Speakers are asked to comment on or provide contexts for a given word/construction.

- **Translation equivalent**:
  - Speakers are asked to translate a given word/utterance, often from a questionnaire.

- **Judgement**:
  - Speakers are asked to evaluate the acceptability/grammaticality of a given form.

Data resulting from contextualising elicitation

"How do you say hello to people in the morning?"

- PLUS:
  - Yield phonologically natural utterances.
  - Can be quantified to some extent.
  - Are highly controlled, or at least seem to be.
  - Offer negative evidence

- MINUS:
  - Results depend heavily on the creativity of the researcher and the receptiveness of the consultant
  - Easily lead to misunder-standings that go by unnoticed
  - Can thus yield syntactically, semantically, pragmatically odd utterances

Data resulting from translational equivalent elicitation

"How do you say 'my mother' in Ugong?"

- **PLUS:**
  - Are easy when starting work on an unknown language
  - Give good data to work on phoneme inventory, basic lexicon, and for lexical comparison
  - Are quantifiable and highly controlled
  - Offer negative evidence

- **MINUS:**
  - Yield phonologically odd utterances
  - Give no complete picture of the extension of the word in the target language
  - Can easily lead to misunderstandings due to the lack of context
  - Translatable items are limited in number
  - Hyper-cooperative consultants may create neologisms to be helpful

Data resulting from acceptability judgements

Can I say 'maka na' when the stick is lying over there?

- PLUS:
  - Are controlled and quantifiable
  - Can give results for domains that are difficult to cover otherwise
  - Give comparable results for many fields
  - Offer negative evidence

- MINUS:
  - Very often do not test acceptability of the utterance, but rather of the context provided for it
  - Can therefore very often be contradicted by the same and by different speakers

Participant observation

- other terms: 'naturalistic data', 'spontaneous speech data'

- external interference is limited to the fact that the communicative event is being observed and recorded – attempt to create a 'natural' context of interaction (story telling, ritual, conversation etc)

- generally constitutes the backbone of a language documentation and an important component of a data corpus

Data resulting from monologues

"The elephant went into the forest and waited for the tiger ..."

- PLUS:
    - Have a high degree of ecological validity
    - Yield phonologically, semantically and syntactically natural utterances
    - Give insight into the culture, if thematically balanced
    - Show high-frequency phenomena

- MINUS:
    - Can seem 'natural' but actually aren't because the cultural settings are not respected
    - Can contain pragmatic oddities
    - Are not very controlled
    - Many features are not quantifiable because a unique performance of one speaker
    - Don't offer negative evidence and are not good for low-frequency phenomena

Data resulting from conversation

A: "you won't believe what I heard on the bus this morning"

B: "are you still catching the 19 to Euston?

- **PLUS:**
  - Often seen as the non-plus-ultra in naturalness
  - Yields data that are naturalistic in every respect
  - Also gives important information about the culture

- **MINUS:**
  - Is not controlled at all
  - Is very difficult to get
  - Is tedious and time-consuming to transcribe
  - Is even more time-consuming to analyse
  - Don't offer negative evidence and insight into low-frequency phenomena

Staged communication

- Other term: 'quasi naturalistic data'
- Communicative events that are enacted for the purpose of recording them for analysis:
  - Telling of a story/joke/the way to do something
  - Description of a picture/acted video/animated video, eg. Frog Story, Pear Story
  - Matching/sorting game that involves non-linguistic categorisation or linguistic interaction, eg. Map Task, Tinker Toy Game
  - MPI Nijmegen videos

Static stimuli
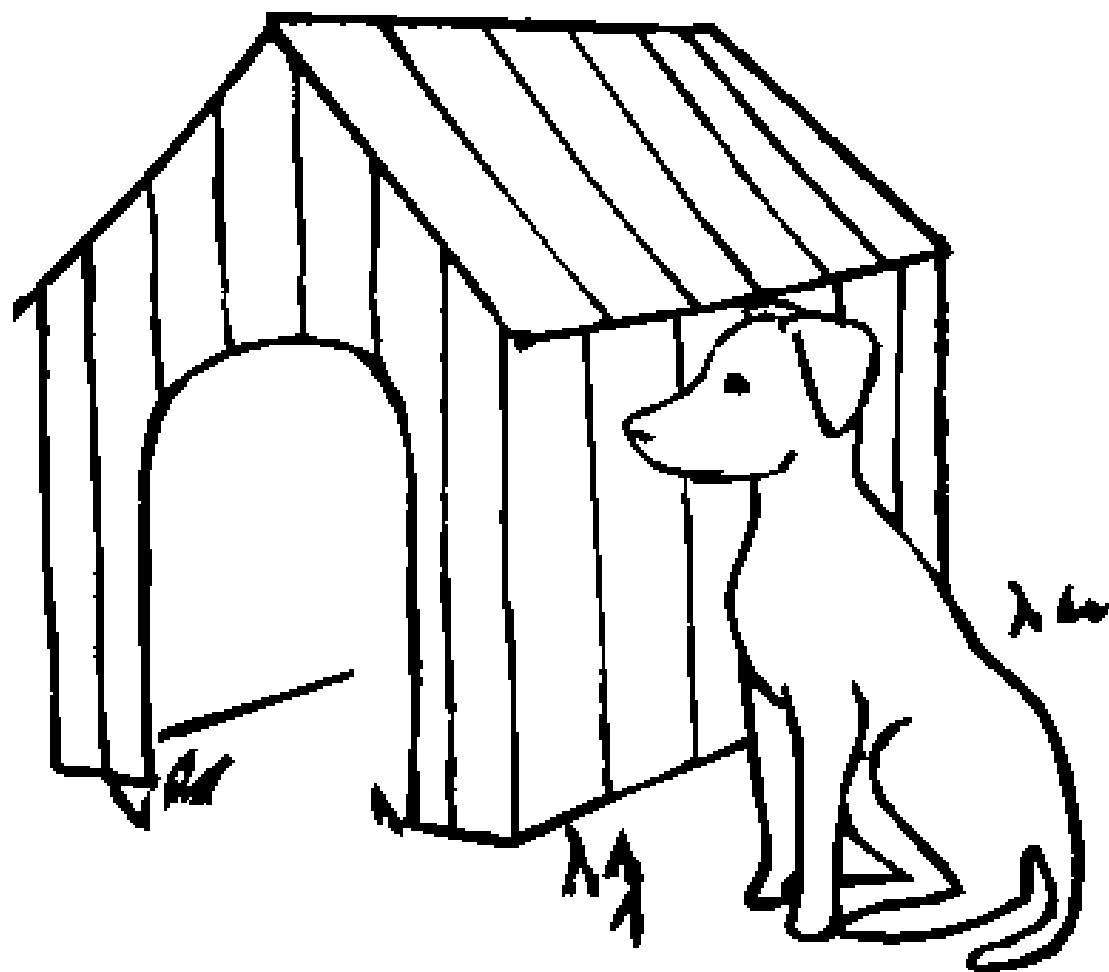
Data resulting from static stimuli

- PLUS:
  - Are highly controlled, quantifiable and comparable
  - Yield phonologically, semantically and syntactically accurate data
  - Are free from linguistic interference of the metalanguage and from misunderstandings of context
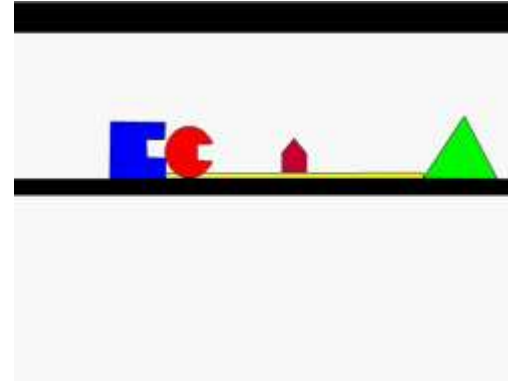 - Can be used for non-linguistic categorisation tasks

- MINUS:
  - Validity of the data depends on coverage of the domain under inspection by the stimulus
  - If gaps in parameters, data can be severely flawed
  - Cross-cultural applicability can be limited
  - Use is limited to visually depictable scenes

Dynamic stimuli – MPI Nijmegen

Data resulting from dynamic stimuli

- **PLUS:**
  - Yield phonologically, syntactically and semantically quantifiable and comparable data etc. (see previous slide)
  - Can be used for non-linguistic categorisation tasks

- **MINUS:**
  - See previous slide and:
  - Require the use of high-tech, which is complicated if not impossible in many field settings
  - if stimulus is abstract and the purpose is unclear, misunderstandings can occur

Interactive stimuli

parked van

east lake
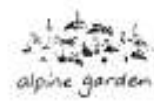
picket fence

level crossing

alpine garden

telephone box

disused monastery

west lake

START

X

flight museum

Camera Shop

allotments

yacht club

--1f

Map 0f

parked van

picket fence

east lake

FINISH

youth hostel

thatched mud hut

alpine garden

telephone box

disused monastery

START

Camera Shop

allotments

parked van

yacht club

--1g
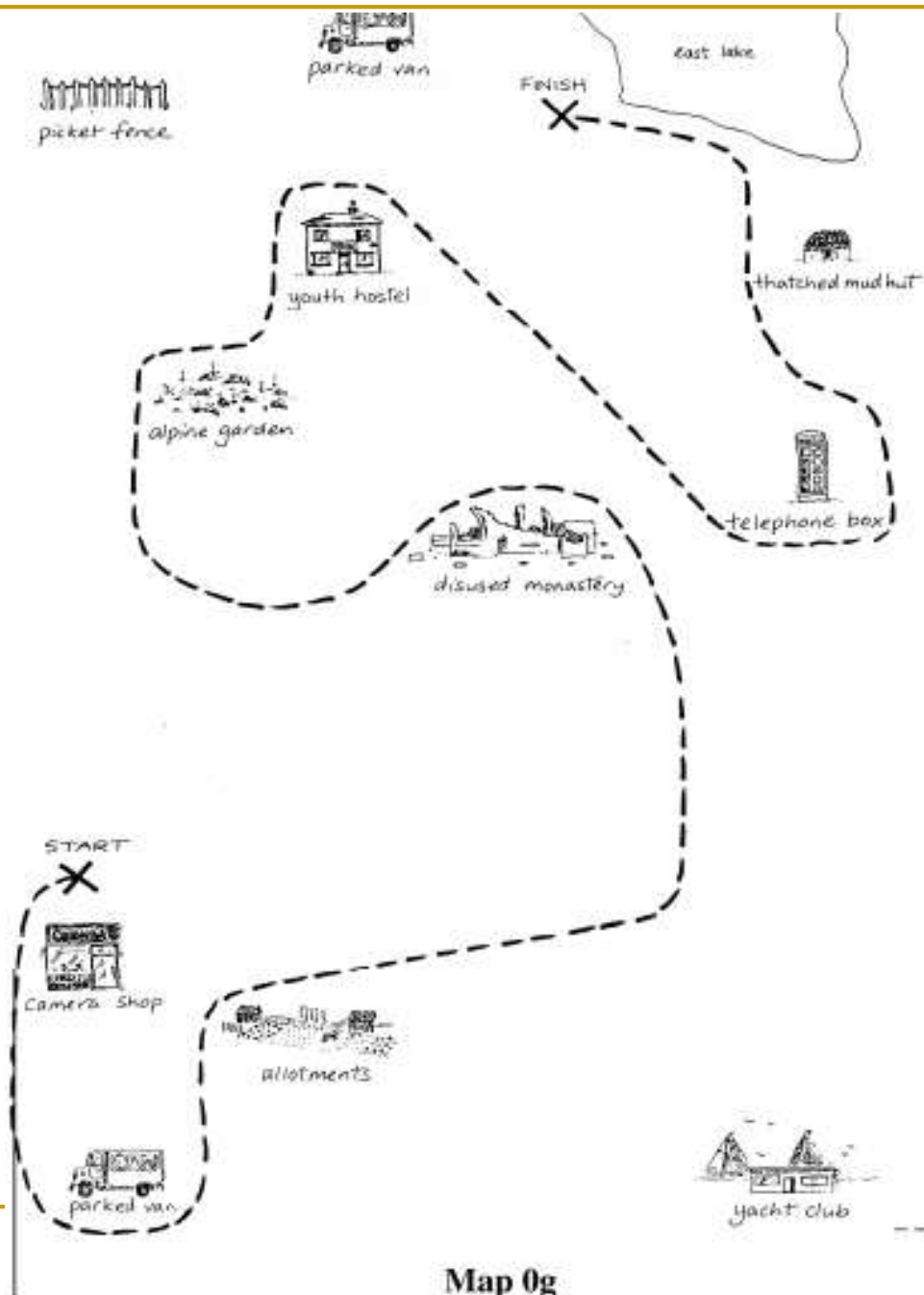
Map 0g

Data resulting from interactive stimuli
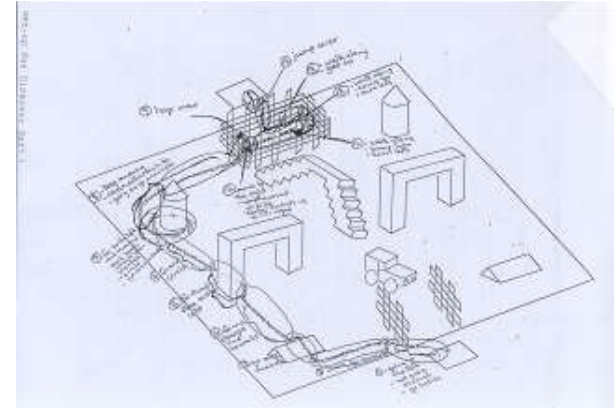
- **PLUS:**
  - Allow controlled interaction of two or more speakers
  - Yield quantifiable and comparable data
  - Can be used for non-linguistic categorisation tasks

- **MINUS:**
  - May create culturally inappropriate or strange situations.
  - Since the true purpose of the interaction is normally not known to the consultants, misunderstandings occur easily

Documentary tools – software

After you make a recording

- You probably need to transcribe it.
- You may need to translate it.
- You may want to add other information.

Some tools will help you transcribe.

ELAN, Transcriber and Praat are three that documentary and descriptive linguists are using these days (also SayMore, to be discussed later)

ELAN

- "ELAN (EUDICO Linguistic Annotator) is an annotation tool that allows you to create, edit, visualize and search annotations for video and audio data."
- links text annotations with audio and/or video data.
- one audio stream, up to four video streams
- ELAN files can be exported in a variety of formats (including to Shoebox/Toolbox for interlinearisation, then reimported)

What can't ELAN do?

- It can't do your transcription
- It can't do your analysis
- It can't keep you organised
- It can't (by itself) make a viewer for community members
- It isn't (unfortunately) very easy to learn

What can ELAN do?

- It can help with transcription and translation
- It can help with your analysis by presenting your data
- It can help keep you organised by linking the media and data files together
- It can help you find things in your data
- It can help if making a product for community members (text, subtitled video)

Tiers

Tiers

- Tiers are where you put your annotations

- Tiers can contain many kinds of annotations, some of the most obvious are:
    - IPA transcription
    - practical orthographic transcription
    - free translations into languages of wider communication
    - morphemes and gloss
    - gesture annotation
    - grammar notes
    - socially significant information
    - any other information which seems relevant

ELAN – plus and minus

- Handles most audio and video formats
- Powerful for annotating and searching
- Good compatibility with Toolbox/ (FLEX)
- Good exports for web video etc via CUPED or other tools
- Prospects for development
- Multi-platform, open-source

- Difficult to get started – steep learning curve
- No inbuilt tools for interlinearising or lexicon building
- *Too* powerful/flexible – temptation to add zillions of tiers, gets cluttered and confusing

Transcriber

- Transcriber is a tool for assisting the manual annotation of speech signals.

- It provides a user interface for segmenting long duration speech recordings, transcribing them, and labeling speech turns, topic changes and acoustic conditions.

- http://trans.sourceforge.net/en/presentation.php

File   Edit   Signal   Segmentation   Options   Help

report

speaker#2
○ ((Yeah)).

speaker#1
○ {inhale} He's hilarious. {laugh}

speaker#2
○ He's great.

speaker#1 + speaker#2
○ 1: {inhale} He's really a trip.
   2: I know. But it really shows you,

speaker#2
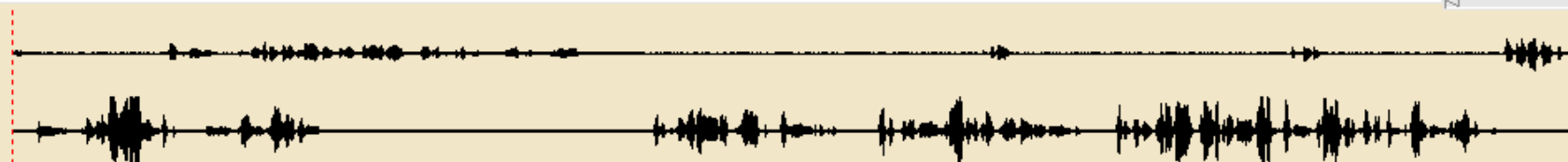○ I mean, you know, you really don't have to put up with the Anthony's of the world.

speaker#1
○ ((I-)) You know what, Ann, it's like, I mean, {exhale}

speaker#1 + speaker#2

⏮ ⏪ ⏸ ▶ ⏩ ⏭ i   know

Resolution

report

speaker#1	s.	speaker.	speaker#2	speaker#1	speaker#1 +...	speaker#1	.s	speaker#1	speak
{inhale} ...	H.	{inhale}.	I mean, you know, you...	((I-)) You know ...	I just didn't know. ...	And the thing is,.	.{	You know ...	{laugh
... {laugh}	..at.	I know...	... the world.	... mean, {exhale}	I know.	... {laugh}	.{	... just-	}

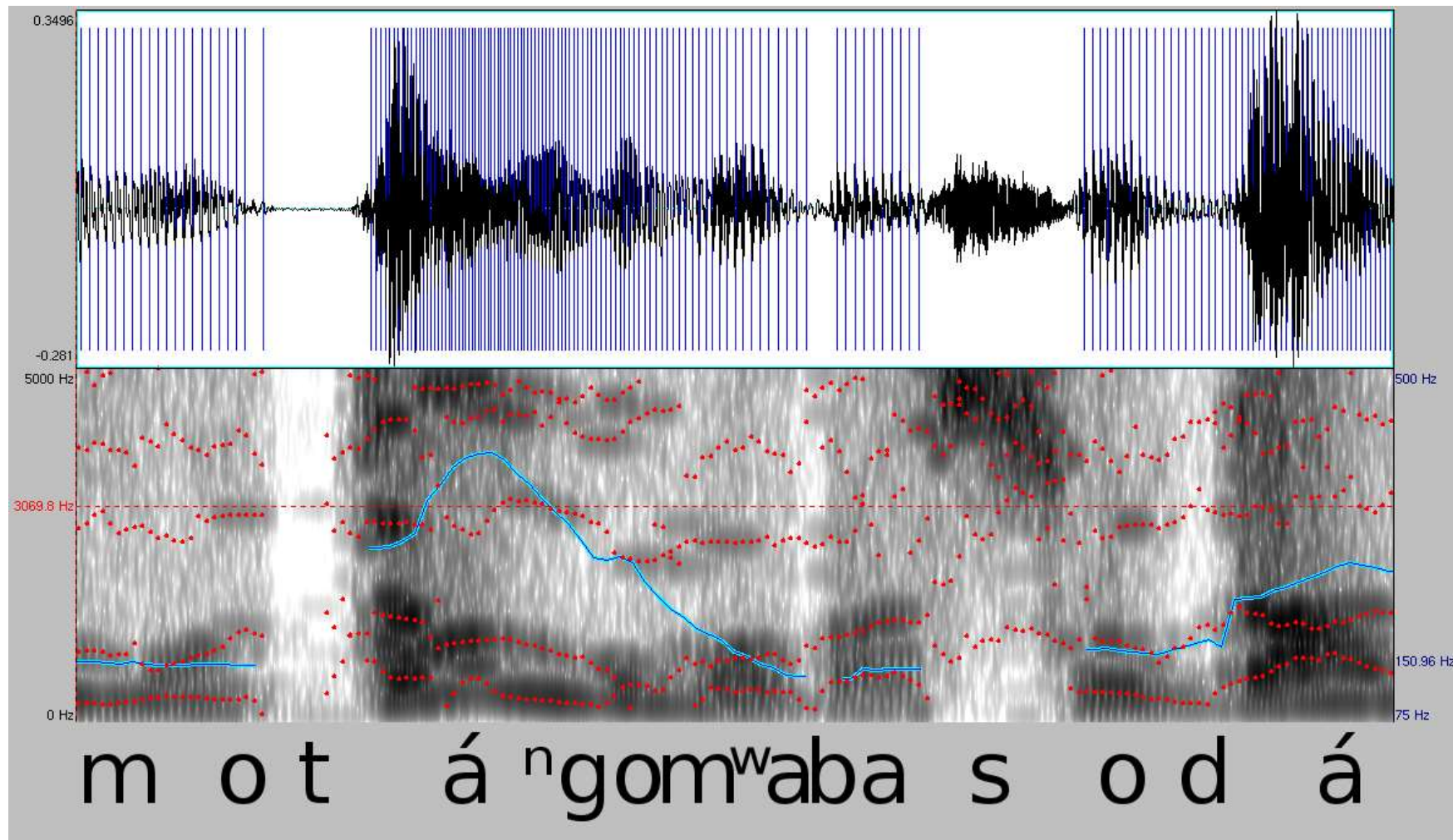0   5   10   15   20

Cursor : 0

Transcriber plus and minus

- Relatively easy to set up and use
- XML format for easy file exchange
- Handles most audio formats
- Multi-platform, open source

- Lacks video support
- Overlapping speech tricky to handle when exporting to Toolbox
- Not (really) designed for linguists – unlikely to integrate with linguistic analysis tools in the future

Praat

- Praat is a tool for assisting with phonetic (especially acoustic) analysis, synthesis, and manipulation, and for creating high-quality pictures for your articles and thesis
- Can label intervals and time points in multiple tiers for transcribing
- Can create spectral analysis and pitch (fundamental frquency) analysis
- Can manipulate pitch and duration, and synthesise sounds
- Statistics analysis and programming language (scripts)
- http://www.fon.hum.uva.nl/praat/

m o t á ⁿgomʷaba s o d á

Praat plus and minus

- Powerful acoustic analysis tool
- Communicates with ELAN
- Exports for easy file exchange
- Handles most audio formats
- Multi-platform, open source

- Lacks video support
- Complex and difficult to learn, not (really) designed for fieldworking linguists – unlikely to integrate with linguistic analysis tools, eg. FLEX, in the future

You've transcribed. Now what?

- Grammar analysis
- Lexicon building
- Cultural/ethnographic notes
- ???

Tools that help you do some of these things:
Toolbox

Fieldworks Language Explorer (FLEx)

– both from SIL

Toolbox

- Toolbox is a data management and analysis tool for field linguists.

- It is especially useful for maintaining lexical data, and for parsing and interlinearizing text, but it can be used to manage virtually any kind of data.

Toolbox plus and minus

- Tried and tested
- (Relatively) easy to use after some initial study
- Large and helpful user community
- Interoperability with ELAN
- Can produce printed or online dictionaries with MDF or Lexique Pro

- Standard Format (backslash codes) not really well-structured
- 'End of life'? It is very old, not being developed actively
- Limited interaction with media files
- Mac only under emulation

Fieldworks Language Explorer

- "FieldWorks is a set of software tools that help manage cultural and linguistic data from initial collection through submission for publication"
- It can be used to record lexical information and develop dictionaries.
- It can interlinearize text.
- The morphological parser provides the user with a way to check the grammatical rules they have recorded against real language data.
- The grammar information can also be compiled in an automatically generated grammar sketch.

Sena 3 - FieldWorks Language Explorer

File Edit View Data Insert Format Tools Parser Window Help

Sena

Lexicon

- Lexicon Edit
- Browse
- Dictionary
- Categorized Entr
- Classified Diction
- Bulk Edit Entries
- Bulk Edit Senses
- Reversal Indexe
- Bulk Edit Revers

Lexicon

Texts

Words

Grammar

Entries

Headword △	Gloss
Show All	Show
_	dividir
gaya	vaidad
	ser vai
goa	vale
goce	farelo
godama	joelha
godamira	joelha
gogodo	osso
gogoma	ajoelh
gombe	marge
gona	dormi
gopa	ter me
gubudza	sacudi
gula	compr
	vende

Entry

Show Hidden Fie

godamira *V.* Por joelhar a favor de algém Eng kneel for someone

Lexeme Form   **godamir**

   Morph Type   root

Citation Form   godamira

Entry Type   Main Entry

Sense 1

Gloss   Por **joelhar**

          Eng **kneel**

Definition   Por joelhar a favor de algém

          Eng kneel for someone

Grammatical Info. Verbo

Example

Translation

Reference

Semantic Domains

Lexical Relations

Parsing Note

File   Edit   View   Data   Insert   Format   Tools   Parser   Window   Help

English

Texts

Texts

Text

Edit
Document

Title

Show All

My Green Mat

My Green Mat

This is a simple sentence about seeing a green mat.

Word	pus	yalola		nihimbilira			
Morphemes	pus	yalo	-la	ni-	him-	*bili	-ra
Lex. Entries	pus_1	yalo	-la	ni-	hiN-	*bili	-ra
Lex. Gloss	green	mat	1SgPoss	1SgSubj	3SgObj	to.see	Pres
Lex. Gram. Info.	adj	N (I)	N:(Possessor)	V:(Subject)	V:Object	trans (1)	sta:Tense
Word Gloss	green	mat		I see			
Word Cat.	mod	N		V			

.

Free: I see my green mat.

Lexicon

Texts

Words

Gram...

Lists

Queue: [-/-/-] No Parser Loaded

1/1

FLEX plus and minus

- Better data structure than Toolbox - XML
- Very powerful parsing and grammatical analysis tools
- Designed to hold all your linguistic and cultural data and notes

- Poor handling of media
- Large application, memory hog
- Windows only
- Poor integration with Toolbox

Another dictionary tool – WeSay

- WeSay helps non-linguists build a dictionary in their own language.

- It has various ways to help native speakers to think of words in their language and enter some basic data about them (no backslash codes, just forms to fill in).

- Designed for teamwork – one 'advanced' user does the complicated set-up work, very simple interface for other users

Home | Dictionary Browse & Edit | Actions | Collect Words By Semantic Domain

bth

abit ab'it <u>water carrying basket</u>

abit
abo-abo
aboh
abur
abus
abŭs
abut
abŭt
abŭt kupong
adat
addi
adis
adoh
adoi
adu
adu-adu
adud
adŭd

Word

bth abit

ipa ab'it

Meaning 1

en water carrying basket

Picture

Remove Image...

POS

noun

Example

bth

Meaning 2

en

New Word   Delete This Word   Show Uncommon Fields

We Say plus and minus

- Very simple to use
- Will run on netbooks and other low-powered machines
- Good data structure
- Easy export via Lexique Pro for print/web

- No tools for interlinearising or analysis
- Limited media support
- Windows only

Comparison of programs

	Transcriber	ELAN	Toolbox	FLEx	WeSay
Audio time-alignment	✓	✓	✗	✗	✗
Video time-alignment	✗	✓	✗	✗	✗
Multi-tier annotation	✗	✓	✓	✓	✗
Interlinear support	✗	✗	✓	✓	✗
Lexicography	✗	✗	✓	✓	✓✗
Word collection	✗	✗	✓	✓	✓
Simple to learn	✓	✗	✗	✓✗	✓
Special char. input	✗	✓	✓	✓	✓
XML data	✓	✓	✗	✓	✓

Managing metadata

- There are a few programs that can be used to manage metadata

- Arbil (from MPI Nijmegen) can be used online or stand alone for IMDI metadata

- CIMDI Maker for offline CIMDI metadata http://cmdi-maker.uni-koeln.de/

- SayMore (from SIL) can be used to harvest metadata from files and then say more about it (also transcription or translation)

- Being developed but now very solid

SayMore

- Pull the day's recordings directly off your camera or audio device, creating Session folders for each one, or record directly into SayMore.

- Enter meta-data about the session as a whole, and also for individual files.

- Add any kind of file related to the session - audio and video recordings, photographs, & transcriptions.

- Enter information about session participants.

- Add evidence of informed consent, be it audio, video, or photographs or a signed form.

SayMore

- Can auto-segment your media into bite-size chunks, or do it by hand, or import from ELAN .

- Transcribe and translate, then export to FLEX for interlinearization, or do further work in ELAN.

- Enlist a native speaker to easily record Careful Speech annotations and Oral Translations.

- Get 3-track audio combining original, careful speech, and oral translations

- View charts and other statistics of your progress

- http://www.sil.org/resources/software_fonts/saymore

- Video: https://youtu.be/xmjxg6H1cyQ
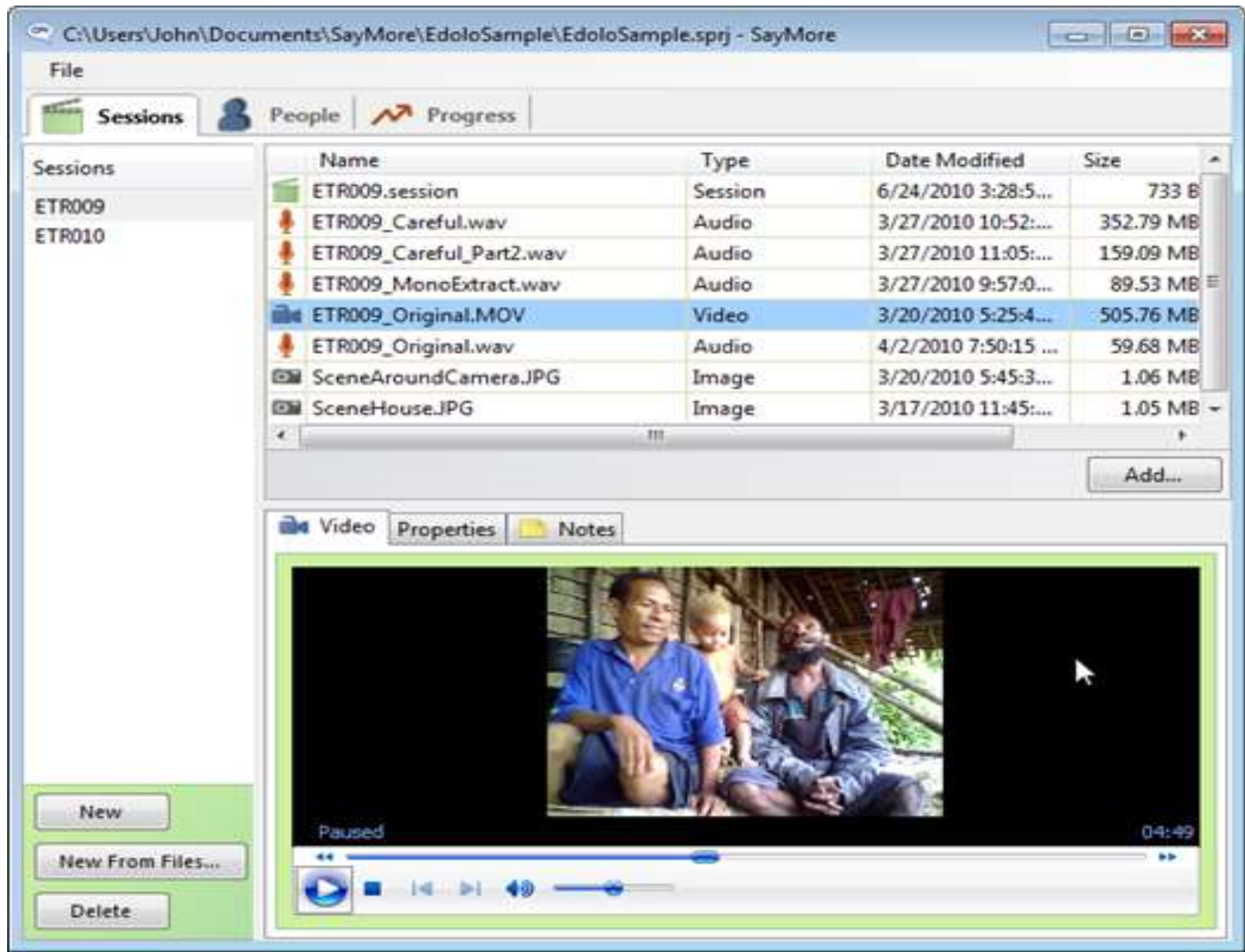
Language revitalisation and support

Language revitalisation

- efforts to increase **language vitality** by taking action to:

  - increase the domains of use of a language and/or

  - Increase the number of speakers (often in the context of reversing language shift)

- older than language documentation (serious work began in 1970s and 1980s among Maori, Native American groups and others)

- Speech/language community members are often more interested in revitalisation than documentation

- Often assumed revitalisation = formal language learning (school lessons, immersion)

What does revitalisation involve?

- Usually driven by 'the community' (who are they?)
- 'Speakers create opportunities to use the language, and address the social attitudes that triggered the abandonment of the language.' (Anderson & Harrison 2007)
- Usually involves second language learning
- and 'the ability to shift the language into new domains of language activity' (Williams 1992)
  - These are ideologically-based assumptions - may be challenged!
- Is it bottom-up or top-down?

Examples of revitalisation

1. "Language nests"

- Translation of Māori 'Te Kōhanga Reo'
  - originated in New Zealand in 1980s
- Pre-schools in endangered languages
- "Totally immerses children in Māori language and culture in an effort to promote learning within a context/situation that is relevant to the children and which draws on Māori styles of learning and teaching" (http://www.kohanga-reo.co.nz/)
- A replacement for family transmission?
- Tries to engage entire community (*whānau*)
  - e.g. cultural events, adult learning

2. Master (mentor)-Apprentice programmes

- Pioneered by Native American groups in California (Hinton 1997)
- Fluent speakers are paired with learners or latent speakers
- 'Learning through doing': activity-based
- Useful practice for learners
  - may have passive exposure but little productive competence
- Helps older users stay fluent
  - language with a useful social purpose
- Can be combined with documentation
  - learners record sessions
- Simple in principle
  - requires little funding or bureaucracy
  - in America funded programs recompense participants for time
  - training needed for both partners
- Emissaries now visit other groups to help set up programs

Language maintenance

- 'Whereas the goal of revitalization is to increase the relative number of speakers of a language and extend the domains where it is employed, maintenance serves to protect current levels and domains of use' (Grenoble and Whaley 2006: 13)

- Communities may not (want to) realise their language is endangered

- Is language shift in evidence?
  - e.g. loss of domains, less use among the young

Language maintenance involves:

- Increasing status / prestige of a language
- Promote additive bi/multilingualism rather than shift
    - E.g. cognitive benefits of bilingualism
    - May involve language mixing
- Literacy – if desired
    - the language may not have a writing system
- 'Mother-tongue' education

Language support

- Encouragement, assistance, funding of any activities which promote the use of a language, including *revitalisation, maintenance, preservation* and *development*
  - Language Support and Revitalisation is an under-theorised field
- Aims are often not articulated or evaluated
- Needs to take into account language attitudes, ideologies, discourses

Documentation vs. description

- Documentation 2 needs an epistemology for media capture – audio and video recording

- Need to pay attention for good practices in recording – eg. microphone choice and spatiality in audio, framing-lighting-editing for video ("recording arts")

- Some concern for socio-cultural context ("ethnography of speaking")

- Concern for data structuring and data management – eg. 'portability', relational modelling, XML

- Concern for 'standards' and cross-project comparability, especially typology and data mining

- Concern for ethics of research – documentation collects language use in 'intimate' personal contexts, impacts on potential users and uses of documented speech events

- Changing models of research and relationships with people

Issues in language documentation 2

- Objectification and commodification of languages
  - 'Community members report sometimes feeling that the linguist comes in, reifies the language, turns it into a commodity, and then takes it away.' (Bowern 2011: 468)
  - 'Technical parameters such as bit rates and file formats are now often foregrounded to the point that they eclipse discussions of documentation methods' (Dobrin, Austin & Nathan 2009: 42)

- Arguably, we should document language ecologies, not just individual languages

  - Multilingual repertoires, mixed codes, translanguaging, contact effects (Mühlhäusler 2003, Grenoble 2011)

- **lack of audio skills:** little or no knowledge about recording arts and microphone types, properties and placement (microphone choice and handling is the single greatest determiner of recording quality)

- **video madness**: video recordings made without reference to hypotheses, goals, or methodology, simply because the technology is available, portable and relatively inexpensive

- **corpus taming**: little ability at corpus and metadata management, file naming and bundle organisation – ELAR spent huge amounts of time and energy simply cleaning up deposits before they could be archived.

Language Documentation – present

Despite the rhetoric

- lone wolf linguists primarily focussed on language

- little real interdisciplinary interest

- the linguist decides what to deliver to academia and communities and produces familiar and traditional outcomes (dictionaries, orthographies, story collections, etc.)

'multipurpose record'

- Ideals are often not lived up to
  - lack of collaboration
  - corpora are not always representative, and may be heavily biased towards easy-to-hand 'stimuli' (Pear Story, Frog Story) – see following slides for ELAR
  - outputs are not always accessible – especially to communities – interfaces are problematic, even for 'apps' or websites that claim to be 'user friendly'

Unrepresentative corpora

Search

Found 196 resources in the archive (page 1 of 25)

1 2 3 4 5 6 7 8 9 ... next › last »

Search ELAR

Frog story | Search

Reset keywords

Language

!Xo (1)
Adelaide dialect (10)
Avatime (Dominant) (9)
Avatime (1)
Brisbane dialect (10)
Cashibo-Cacataibo (1)
Chinese (1)
Choguita Rarámuri (4)
Ecuadorian Siona (1)
English (Dominant) (1)
English (9)
French (5)
Gija (6)
Guari-Guari (8)
Gurindji Kriol (10)
Hupa (1)
Ju|'hoan (1)
Kabardian (2)
Kibena (Kisovi) (2)
Kibena (Nga?veta) (1)
Kibena (8)
Koyi rai (1)
Kubokota (4)
Melbourne dialect (10)
Nalu (6)
Northern dialect (20)

■ **Documentation and Analysis of Kabardian as Spoken in Turkey**

Ayla Applebaum Bozkurt

... documentation of Kabardian, a typologically rare and threatened Northwest Caucasian language as spoken by the Turkish Kabardian community.

■ **Documentation of Rongga**

I Wayan Arka

... Rongga has 4,000 speakers in the villages Tanarata, Bamo, Watunggene and Waelengga, on Flores Island, Indonesia. Data collected includes audio and visual recordings of interviews and observations and linguistic descriptions ...

■ **The painter's eye, the painter's voice: language, art and landscape in the Gija world**

Frances Kofod

The painter's eye, the painter's voice: language, art and landscape in the Gija world

How to use search

You can search in two ways:

- enter text in the search box and press 'Search'. Search is not case sensitive, and variations of words are found, e.g. 'Village' finds 'villages' and 'Indian' finds 'India'; or

- click a keyword in the left panel to find a set of resources. Click another keyword to refine the results (a black keyword) or to find a new set (a brown keyword)

To refine your search:

- enter two or more words for results containing all those words; e.g. entering 'nigeria' and 'audio' finds the deposit *Damakawa wordlist* which includes recordings made in northern Nigeria.

- use the keywords in the left panel to browse and select further categories; e.g. if you search for 'nigeria' and 'audio', a list (under 'Tags') includes place and language names: Akoko, Ikaann, Damakawa and Sakaba. Click one to find a resource pertaining to that name

To reset search and display all keywords, press 'Reset keywords'.

Colour coding of results

Search results can include deposits, bundles (file groups within deposits) and people. These are colour coded:

■ A deposit: click on its title to view the

Corpus accessibility – I found it, what now?

Revitalisation and documentation – not a simple relationship

❑ 'Revitalisation' has been seen by some documentary linguists as a waste of time, or a simple technical add-on

= orthography, dictionaries, videos, primers, multimedia

❑ But documentary methods and outputs are not always useful for revitalisation

"Work on language documentation to this point has tended to focus on what sorts of records are required to facilitate the creation of grammars, dictionaries, and texts, rather than, for instance, considering what kinds of records are required to adequately document patterns of variation in a community or to provide sufficient context to inform community efforts at language standardization." (Childs, Good & Mitchell 2014)

Documentation and revitalisation

- most language documentation outputs are **unsuitable** for revitalisation:
  - inappropriate genres or topics
  - primarily speech of older fluent speakers (reflects linguists' ideology of "saving the language" or "getting the best language") – may be difficult for learners to process
  - no learner-directed speech (cf. Slow Italian website)
- Observed language practices may not match perceived/stated ones
- Some speakers/language activists may prefer 'folk linguistics' or purism to documentary evidence
  - Documentation which demonstrates low vitality, attrition, 'decline', variation and change may be unwelcome

Language Documentation – future?

Documentation – future?

- diversity
- metadocumentation
- transdisciplinarity
- relationships with revitalisation

Diversity

- of goals, contexts, people, data, corpora, outcomes
  - move away from generic to more focused documentation, eg. ELDP 2012 grant list: bark cloth making, libation rituals, fishing practices, child language, interactive speech, and ethnobotany
  - diverse **inputs** – field interviews, experiments and observations (traditionally the bread and butter of documentation and description) but also Youtube uploads, Twitter feeds, Facebook, blogs, email, chat, Skype, local pedagogy in revitalisation
  - diverse **outputs** – books, papers and archive deposits (the bread and butter of 1990's documentation) but also Youtube uploads, Twitter posts, Facebook, blogs, email, chat, Skype, local pedagogy in revitalisation, mobile apps, Kindle readers

Meta-documentation

- meta-documentation = documentation of language documentation models, processes and outcomes
- the goals, methods and conditions (linguistic, social, physical, technical, historical, biographical) under which the data and analysis was produced
- meta-documentation should be **as rich and appropriate** as the documentary materials themselves, for:
  - developing good ways of **presenting** and **using** language documentations
  - future **preservation** of the outcomes of current documentation projects
  - **sustainability** of field
  - helping future researchers **learn** from the successes and failed experiments of those presently grappling with issues in language documentation (Austin 2010)
  - documenting **IP contributions** and career trajectories (Conathan 2011)

Transdisciplinarity

- Is language documentation a new sub-field of linguistics? (as per Himmelmann, Austin) or

- Is it a new transdisciplinary approach that: "must draw on concepts and techniques from linguistics, ethnography, psychology, computer science, recording arts and more" (Woodbury 2011), where "more" includes history, archiving, museum studies, project management, creative writing, social media, ornithology, biology (cf. PAW project at SOAS), political science, development studies?

Documentation **for** revitalisation

- what would language documentation look like if it was done with a goal of producing outputs for revitalisation?
- *different genres*: conversation, not just narratives or rituals (Sugita, 2007; Amery, 2009), interactions (greetings, leave takings, 'phatic communication')
- *'chunks' of language*, from fixed/formulaic expressions to whole discourses (eg. 'Welcome to Country')
  - Dorian 1980 'semi-speaker' – "a speaker of an endangered language who has a partial linguistic competence" but can sometimes appear 'more competent' because they can interact appropriately
  - research suggests proficient language users know a large number of formulaic sequences (e.g. Fillmore, Kay, and O'Connor, 1988; Wray, 2002), with Pawley and Syder (1983) suggesting that speakers know several hundred thousand of these sequences

From post-vernacular to revitalisation – via documentation

- Documentation needs to be accessible and useful for revitalisation and teaching
    - e.g. conversation, child-directed language, functions
- Documentation and theorisation of revitalisation needs to be developed
    - and made accessible to activists and practitioners
- More genuine collaboration is needed
    - including applied linguists
- Community and disciplinary ideologies, and power relations, need to be explored and taken into account
- See Leonard and De Korne 2016 issue of LDD

Colonialism in language documentation and revitalisation

Overview

- **Defining colonialism**

- **Colonialism and Linguistics – past**

- **Colonialism and Linguistics – present**
  - Language documentation and revitalisation mantras
  - Linguistic heroes?
  - Models of training
  - British Museum syndrome
  - Here's the grammar of your language – pity you can't read it
  - Decolonialising spaces for languages

- **Colonialism and Linguistics – future?**

Defining colonialism

- policy of a foreign polity seeking to extend or retain its authority over other people or territories, generally with the aim of opening trade opportunities. The colonizing country seeks to benefit whilst the colonized country or land mass ... modernizes in terms defined by the colonizers, especially in economics, religion, and health.

- domination of an indigenous majority by a minority of foreign invaders who rule in pursuit of the coloniser's interests.

- *Types*: settler colonialism, exploitation colonialism, surrogate colonialism, and internal colonialism

Defining colonialism

- *Exploitation colonialism* – few colonists, exploitation of natural resources or population as (indentured or slave) labour, typically to the benefit of colonising centre

- *Surrogate colonialism* – settlement supported by a colonial power where most of settlers do not come from the same ethnic group as the ruling power.

- *Internal colonialism* – uneven structural power between areas of a state, originating from within the state, e.g. control and exploitation passing from whites in colonizing country to white (or surrogate) immigrant population within a newly independent country.

Impact of colonialism

- spread of diseases, **unequal social relations**, **exploitation**, **enslavement**, medical advances, the **creation of new institutions**, abolitionism, improved infrastructure, and technological progress.

- spread of colonist languages, literature and cultural institutions, **endangering or obliterating** those of local peoples.

- Universalism – concepts developed in the colonial centre must apply everywhere; the centre is presented as the source of ostensibly enlightened policies imposed on the colonies

Post-colonial approaches

- Attempt to deconstruct ideologies, social roles, and political power embedded in rhetorical stances and narratives of a discipline

- Address politics of knowledge in terms of epistemology, ethics and political theory

- Aim at making clear and challenging the theories (intellectual and linguistic, social, and economic) used by colonialists to 'perceive', 'understand', and 'know' the world.

- Establish intellectual spaces for subordinated peoples to speak for themselves, in their own voices, and produce cultural discourses of philosophy, language, society and economy, balancing the imbalanced us-and-them binary power-relationship between the colonist and the colonial subjects

Example of post-colonialist critique

- Political geographers suggest that colonial behaviour was reinforced by physical mapping of the world, creating a visual separation between 'them' and 'us'. Geographers look at how space was materially and symbolically appropriated to enable colonialism

- cartography was often manipulated during the colonial era as map-makers used **rhetoric** in their formation of boundaries and in artistic representations, favouring the views of the conquering Europeans, making any map created by a non-European **inaccurate**. Harley (1989): 'The steps in making a map – selection, omission, simplification, classification, the creation of hierarchies, and 'symbolization' – are all inherently rhetorical.'

- *Key concepts*: objectification, boundedness of entities, quantification, commodification, hierarchy of values (cf. Dobrin, Austin & Nathan 2007)

Colonialism and Linguistics – past

Errington (2001: 1) "European technologies of literacy enabled missionary and non-missionary linguistic work that resulted in representations of languages [as objects abstractable from texts] as powerful **icons of spiritual, territorial, and historical hierarchies** that emerged in colonial societies. As descriptions of languages travelled from exotic colonial peripheries to European metropoles, they came under the purview of comparative philology. This disciplinary precursor to modem linguistics helped to legitimize colonial linguistic projects and legislate colonial difference on a global scale." (emphasis added)

p2. "language difference could become a resource -- like gender, race, and class -- for figuring and naturalizing inequality in the colonial milieu"

Colonial linguistics includes mapping monolingual blocks to control uncertainties of multilingual flux and reductively standardising them for literacy and religion (thereby ranking 'degenerate' variants below them) – cf. Luepke 2015

Colonialism and Linguistics – past

Irvine (2008: 1) "European ideologies of language, and the conditions in which linguists' carried out their research, influenced the resulting descriptions of African linguistic structures and the delimitation of linguistic boundaries."

Irvine (2008:16) "colonization … turns cultural traditions and genealogies into 'ethnicity,' turns linguistic practices into named 'languages' corresponding (supposedly) to ethnic groups, and interprets multilingualism as a secondary effect."

Dobrin (2018: 1) "The missionaries' linguistic interpretations naturalized social hierarchies and reshaped languages on a European nationalist model that takes there to be an essential isomorphism between social groups (nation states) and languages." In addition, there is one 'heart language' through which Truth can be expressed (and communicated to God), cf. multilingualism

Colonialism and Linguistics – present

- Resurgence of interest in linguistic diversity since 1992, funding since 2002

- Ideology emerges of 'endangered languages' being replaced by larger, more powerful languages, including lingua francas (which had been ignored in the colonial era) through 'language shift'

- Many aspects of colonialist conceptualisations continue to be taken for granted: languages as bounded objects, hierarchical ranking of ways of speaking, one 'mother tongue', multilingualism and 'code mixing/switching' as abberations, simple linguistic ecologies (and pragmatics and metapragmatics), need for literacy to transcribe and translate (and gloss interlinearly), GIDS to measure 'vitality'

- 'scientific' (Western) research vs. 'unscientific' ethno-linguistics – the former achieved only by study and training vs. the latter from 'tradition'

- Reproduction of hierarchical positions of 'researcher' and 'researched'

Language documentation & revitalisation mantras

- "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)
- Features:
  - *Focus on primary data*
  - *Accountability*
  - *Long-term storage and preservation of primary data*
  - *Interdisciplinary teams*
  - *Cooperation with and direct involvement of the speech community*

- *Narrow view*: outcome is **annotated and translated corpus** of archived representative materials on use of a language, cf. DoBeS/TLA, ELAR – separate from **description** (language as system)
- *Broad view*: outcome is transparent records of a language with description and theorisation dependent (Woodbury)

Linguistic heroes?

Sarvasy & Forker (2018: 1) *Word hunters: Unsung heroes of linguistics* – "It reflects poorly in our societies that the contributors to this volume are not household names. In fact, these career-long linguistic fieldworkers are true heroes. … There is no public reward for heroism. … Fieldwork – taken here as *in situ* language research – is surely the ultimate all-around challenge that a linguist can voluntarily undergo."

Note: all the contributors are white Europeans doing research in ex-colonies

Frameworks for ethical language research

Ethical research

Advocacy research

Empowering research

Ethical research

"… there is a wholly proper concern to minimize damage and offset inconvenience to the researched, and to acknowledge their contributions. … But the underlying model is one of 'research on' social subjects. Human subjects deserve special ethical consideration, but they no more set the researcher's agenda than the bottle of sulphuric acid sets the chemist's agenda."

(Cameron, Frazer, Harvey, Rampton, and Richardson 1992, p. 14-15)

Advocacy research

"… characterized by a commitment on the part of the researcher not just to do research on subjects but research on and for subjects. Such a commitment formalizes what is actually a rather common development in field situations, where a researcher is asked to use her skills or her authority as an 'expert' to defend subjects' interests, getting involved in their campaigns for healthcare or education, cultural autonomy or political and land rights, and speaking on their behalf."

(Cameron, Frazer, Harvey, Rampton, and Richardson 1992, p. 15)

Empowering research

"… as research **on**, **for** and **with**. One of the things we take that additional 'with' to imply is the use of interactive or dialogic research methods, as opposed to the distancing or objectifying strategies positivists are constrained to use. It is the centrality of interaction 'with' the researched that enables research to be empowering in our sense; though we understand this as a necessary rather than a sufficient condition … we [propose three] programmatic statement[s] and then pose various questions: 4 (a) 'Persons are not objects and should not be treated as objects.' (b) 'Subjects have their own agendas and research should try to address them' (c) 'If knowledge is worth having, it is worth sharing.'"

(Cameron, Frazer, Harvey, Rampton, and Richardson 1992, p. 22-24)

Models of collaboration and training

Extensive discussion of "collaborative models" (Czaykowska-Higgins 2009, Glenn 2009, Leonard & Haynes 2010) however most in documentation projects "training" involves teaching software tools like ELAN, Toolbox, FLEx or SayMore to "native speakers"

Narrow set of computer skills that are reliant on consistent literacy (for searching) not transferable outside the tasks of transcription, translation and annotation

Result: creation of "plantations" or "sweat-shops" of local workers who are given basic tasks to create standardized template computer files (using local lingua francas) then sent to non-local researchers for value-adding (using academic lingua francas and specified analytical currency, e.g. Leipzig glossing rules)

British Museum syndrome

Bringing home the linguistic bacon

Creation of large-scale digital archives (TLA, ELAR, to a lesser extent Paradisec, AILLA) to collect the outputs from grants that researchers are required to deposit and self-curate (mediated by standardized and difficult to use software tools)

Result: terabytes of things (files) only accessible to those with the necessary keys (fast internet access, knowledge of software tools to open and access files, expert knowledge to understand and manipulate the analytical metalanguage (overwhelmingly English))

Unlike (colonial) museums there are only limited catalogues, and no exhibitions, finding aids or available guides for the user to interrogate or ask for help

Pressure for "open access", subjugation of speaker control over their instances of language performance and use

Here's the grammar of your language – pity you can't read it

Most academic products for endangered languages (e.g. grammars, dictionaries, articles, teaching materials, apps) are written in a metalanguage that the speaker communities cannot access, replicating colonial hierarchies of language and power

Creation of monolingual materials and metalanguage has a long history (going back to Ancient Greeks, Romans, Sanskrit grammarians) and exists for state languages, both large (e.g. English, Malay, Swahili) and small (e.g. Samoan school grammar, Mari reference grammar, Somali, Bislama grammar)

Rare to find monolingual materials for non-state languages, but cf. Poerwadarminta 1939 "Baoesastra Djawa", Q'anjob'al and K'ichee' monolingual grammars (Guatemala), Chatino (Mexico) monolingual dictionary, Luqa (Solomon Islands) monolingual teaching books

Public lecture 27 November 2018

Alpheaus Graham Zobule, Founder and Director of *Kulu Language Institute* of the Solomon Islands, recounts a decades-long project that has allowed speakers of a vernacular tongue (Luqa) to study their own language in that language – an inspirational story of teaching literacy to strengthen an indigenous language.
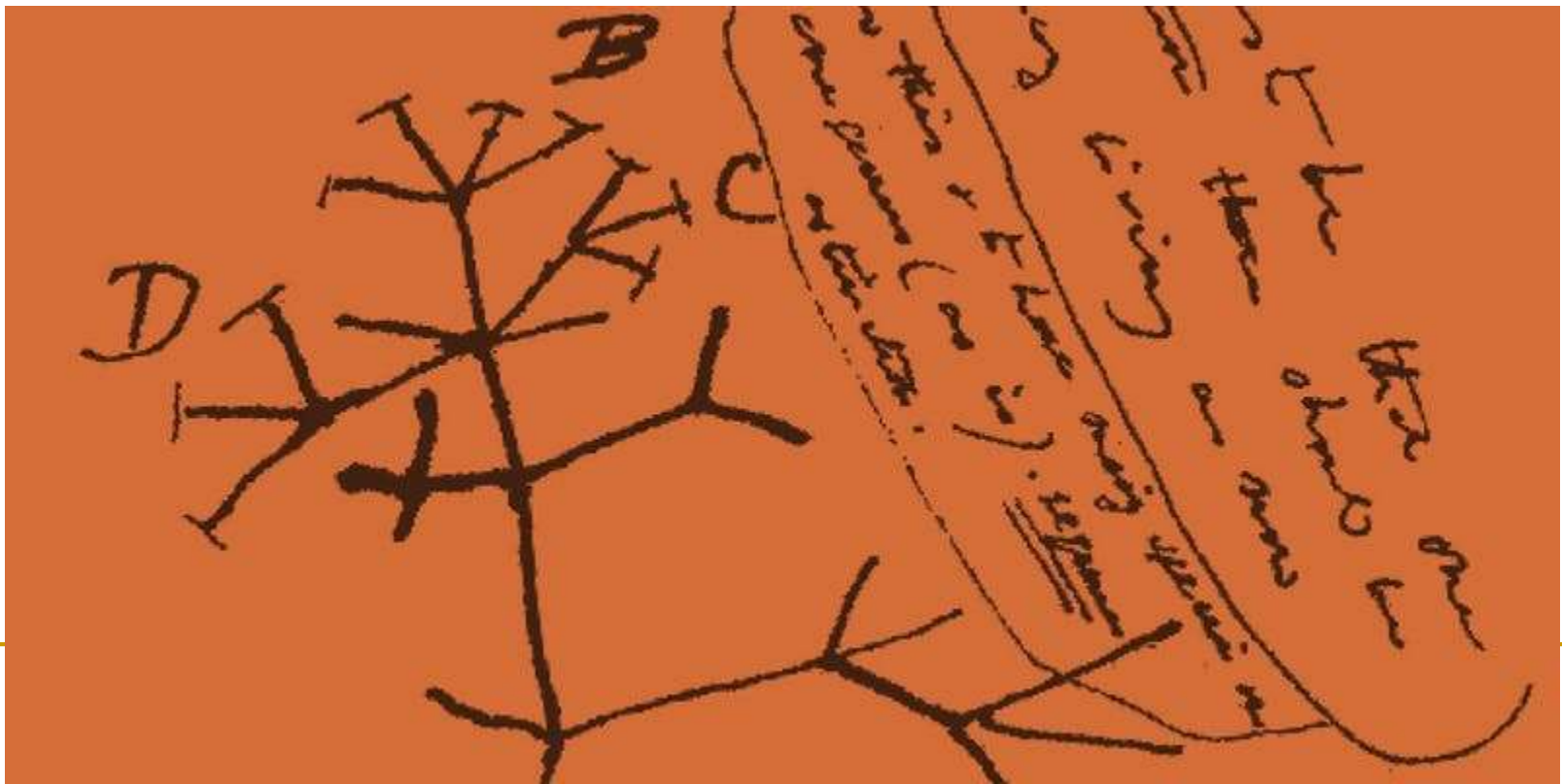
Decolonialising spaces for languages

Some interesting recent examples of creating academic and research spaces for endangered and minority languages.

Juan Carlos Reyes delivered a lecture at ENGHUM summer school in Leiden 2017 on "Community strategies for the strengthening and development of the Ayuuk language" in Ayuuk (Mexico)

Justyna Olko (Warsaw) organized session at 2018 American Society for Ethnohistory: *Protecting Land and Water, Cherishing Language,* where researchers presented talks in Nahuatl, Mixtec, and Zápotec with Powerpoint showing English and Spanish translations

Justyna Olko (2018) describes cultural documentation project in Mexico led by indigenous researchers and carried out entirely in Nahuatl

Non-academic responses

Communities and individuals have responded by:

- ❑ Language revitalisation initiatives to increase domains of use and/or increase number of speakers, often through education or grass-roots activities (master-apprentice, language nests, language camps). This has often resulted in development of educational materials, mostly "grey literature" with limited distribution

- ❑ Engagement in language exchange through social media, especially Twitter, Facebook and WhatsApp or WeChat, involving hundreds of languages. Most material is siloed within closed groups and within the platforms, and not accessible outside but cf. Indigenous Tweets and Indigenous Blogs

Language	Users	Tweets	Top User	Tweets	First Tweet
Euskara	17052	10664821	berria	131815	eastigarraga
Kiswahili	1296	8413349	radiomaisha	169950	issamichuzi
Cymraeg	14273	5698636	newyddcymraeg	100353	meigwilym
Kreyòl Ayisyen	14270	5604681	amour109	80701	tichrist
Gaeilge	13719	2596781	aonghusoha	105259	imeall
Kapampangan	1379	2183819	keeyttguevarra	21764	desperada
Frysk	2712	1439821	petrakramer	464477	eetweetje
Setswana	314	817857	sesutho	55484	WameDre
Asturianu	779	637236	iyangc	38880	Pingarates
Hausa	1331	461283	bbchausa	49726	mojaam
Ikinyarwanda	289	451935	TweetRwanda	60099	kwitob
Gàidhlig	1842	368405	sconewt	27303	Seumas
Soomaaliga	558	309362	BBCSomali	19451	HaPpYMaXaMeD
Yorùbá	2239	297132	jag_bros	9001	kojere
Boarisch	5	229809	bayernuhr	223452	steffkellerband
Igbo	1180	217592	IgboBible	19299	Igbopeople
Aragonés	187	185220	MTrujilloBerges	16756	jesmar
Māori	619	151594	maonewt	29810	DigitalMaori
Gaelg	288	110444	glvold	40479	greinneyder
Brezhoneg	553	91286	YKastell	13319	strangeland
Lingua Latina	72	89840	Vallensis	18220	henduu14
Davvisámegiella	234	70713	YleSapmi	15324	kirstiguvsam
Chamoru	65	68101	chanewtest	25699	Aufhebung

Language	Blogs	Posts	Words	Top Blog
Cymraeg	278	28286	6741085	BlogMenai.com
Gaeilge	188	17651	4358526	Smaointe Fánacha Aonghusa
བོད་སྐད	78	6104	1063911	བོད་ཀྱི་རྩོམ་པ་པོ་
Gàidhlig	74	3274	571331	Gàidhlig – cha bhithinn às a h–aonais!
Occitan	69	8626	2541720	Democracia per Occitània
ייִדיש	51	2651	832754	קטלא קניא
Føroyskt	45	10818	3049278	birk
Aragonés	45	5854	840940	Tierra de barrenaus
Hausa	42	3097	2089253	Kimiyya da Fasahar Sadarwa!
Kreyòl Ayisyen	41	2594	653575	..
Afaan Oromoo	40	2509	906516	Oromo Free Speech
Furlan	35	8493	1332595	UNE VÔS ZOVINE, UNE VÔS LIBAREII blog di Christian Romanini
Ikinyarwanda	35	3843	1474499	Rwandan News
Tetun	33	17581	8600742	TIMOR AGORA
Brezhoneg	29	1635	688535	BREZHONEG DIGOR
Kapampangan	27	1077	1218643	Kapampangan English Bible
Удмурт	23	1134	217000	мынам малпанъёсы
Davvisámegiella	22	3578	466985	villle.se
Bicol	22	820	686553	An Libro Ni Mormon (An Saro Pang Tipan Ni Jesucristo)
Papiamentu	21	4357	1097799	Curacao Softball
Frysk	21	5142	1117462	Stikelstekker
vakaViti	19	141	101507	Nai Vola vanua kei Yacata
Māori	17	254	64175	Los Fairhalls en Extranjero

Citizen science initiatives

There are two projects that focus on non-academics providing language material that is freely and openly accessible via a website:

- The Endangered Languages Project
- Language Landscape

The Endangered Languages Project

The Endangered Languages Project

- Established 2012 by Google.org, then assigned to *Alliance for Linguistic Diversity* (First Peoples' Cultural Council, UH Manoa). Has a Governing Committee.

- "through this website, users can not only access the most up to date and comprehensive information on endangered languages as well as language resources being provided by partners, but also play an active role in putting their languages online by submitting information or samples in the form of text, audio or video files. In addition, users will be able to share best practices and case studies through a knowledge sharing section and through joining relevant Google Groups."

The Endangered Languages Project

- Most content uploaded to the Endangered Languages Project is hosted on several associated Google products or services, including YouTube, Picasa and Google Docs.

- Each service has its own product policies and content guidelines. In addition, all content submitted through other Google products or services must be in accordance with their associated terms. These include but are not limited to: a prohibition on content containing pornography, obscenity, pedophilia, bestiality or other sexually explicit material; hateful or violent content; harassing content or content that infringes another's privacy.

Language Landscape

Language Landscape

- Language Landscape grew out of *London's Language Landscape* project by staff and students which featured at SOAS Endangered Languages Week in May 2011

- Now a not for profit run by a group of current and former SOAS students, funded by grants and donations

- Runs a website – users can upload recordings of language events, are tagged for geolocation and metadata

- Organises outreach activities at London schools and communities; completed a pilot educational programme in east London, providing students with practical training in recording techniques and helping them to learn about issues such as multilingualism and language endangerment, using the website as a starting point for discussion and activities.

Colonialism and Linguistics – the future?

- Despite the rhetoric of "empowerment research" creating multifunctional outputs for multiple audiences through collaboration, much of the work in language documentation and revitalisation continues colonial ideologies and histories

- Recently there have been some initiatives that suggest decolonialism of research spaces and research outcomes may be possible; hegemonies of multilingualism have been queried

- However, the audit culture of current academia that values certain kinds of products written in a limited number of large languages creates institutional barriers to change

- Also, the dominant culture of academic conferences and meetings and existing structural barriers to entry for women, minorities and early career researchers also mean that it will be difficult to create post-colonial real world spaces for minority languages and their speakers to fully participate, which does not mean we should not try

The times they are a changing'?

- Perhaps, a little, in recent times

- But we need to challenge and question current rhetorics and behaviours that perpetuate world views and hegemonies that originate in colonial histories of the distant and more recent past

- Even if this makes some people uncomfortable

Conclusions

- Some researchers have presented language documentation as a return to the Boasian past while others see it as a new approach to the study of human language that pays better attention to data collection and analysis, and to communities, contexts and roles

- it appeared to be an opportunity to shift the socio-political academic balance between "fieldworkers" and "armchair linguists" (typologists, theoreticians) by providing a foundation (theory, best practices) for documentation, in contrast to language description

- Over the past 20 years, and especially the last 10 years, we have seen shifts in the goals, methods, foci and contexts of Language Documentation to make it more pluralistic, open, and socially networked and responsive

- However challenges remain, including encouraging new genres that bridge the output gap, more reflexivity, and better engagement with transdisciplinarity and the ethnography of our research and its contexts

Thank you!

We would like to acknowledge the input of Lise Dobrin, Anthony Jukes, Friederike Luepke, Yan Marquis, David Nathan, Justyna Olko, Candide Simard and other colleagues in discussions which informed the development of this presentation.

References

Austin, Peter K. 2014. Language documentation in the 21st century. LIPP 3, 57-71.

Grenoble, Lenore. 2010. Language documentation and field linguistics: The state of the field. In Grenoble, Lenore A. and N. Louanna Furbee (eds.) *Language Documentation: Practice and values*, 289-309. Amsterdam: John Benjamins Publishing Company.

Good, Jeff. 2010. Valuing technology: Finding the linguist's place in a new technological universe. In Lenore A. Grenoble and N. Louanna Furbee (eds.) *Language Documentation: Practice and values*, 111-131. Amsterdam: John Benjamins Publishing Company.

Berge, Anna. 2010. Adequacy in documentation. In Grenoble, Lenore A. and N. Louanna Furbee (eds.) *Language Documentation: Practice and values*, 51-66. Amsterdam: John Benjamins Publishing Company, 2010.

Schultze-Bernd, Eva. 2006. Linguistic annotation. In Jost Gippert, Nikolaus P. Himmelmann and Ulrike Mosel (eds.) *Essentials of Language Documentation*, 213-251. Berlin: Mouton de Gruyer.

Woodbury, Anthony C. 2007. On thick translation in language documentation. In Peter K. Austin (ed.) *Language Documentation and Description*, vol. 4, 120-135. London: SOAS, University of London.

Woodbury, Anthony C. 2014. Archives and audiences: Toward making endangered language documentations people can read, use, understand, and admire. In David Nathan and Peter K. Austin (eds.) *Language Documentation and Description*, vol. 12. London: SOAS, University of London.