



**SOAS**  
University of London



## D3.2 Models for linguistic-cultural documentation and revitalization via community engagement and transdisciplinary research

Engaged humanities in Europe: Capacity building for participatory research in linguistic-cultural heritage

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 692199.



## Credits

Editing and compilation:

Ebany Dohle  
Joanna Maryniak  
Justyna Olko

Text:

Ebany Dohle  
Bartłomiej Chromik  
Elwira Sobkowiak  
Justyna Olko

Presentations in order of appearance:

Peter Austin  
Ebany Dohle  
Marta Ostajewska  
Language Landscape  
Candide Simard  
Plants. Animals. Words.  
Sophie Mu  
Maciej Mętrak  
Andrzej Żak  
Martin Grabowski  
Samantha Goodchild  
Sara Selleri

## **Acknowledgements**

The work represented in this document is a collective result of the hours spent working alongside minoritized and indigenous communities all around the world. It is thanks to the strength of community ties and those who have taken the time to engage with visitors and academics that this type of research is possible. Many thanks go to representatives and hosts who welcomed our research events in Wilamowice, Poland and Tlaxcala, Mexico.

This deliverable is a result of the joint efforts of the EngHum team and its representatives at collaborating universities, SOAS, University of London and Leiden University. Many thanks to Justyna Olko, John Sullivan, Stanisław Kordasiewicz, Julia Sallabank, Peter Austin, Maarten Jansen, and Willem Adelaar.

## Introduction

This document does not wish to claim that there is only one model for language documentation and revitalization which can be superimposed on any given linguistic situation. If anything, the work carried out over the past two years has demonstrated the importance of taking into account local, historical and environmental contexts when working with minority and indigenous languages. It is best therefore to view this collection of work as a set of different examples and case studies from different parts of the world.

The first section outlines the aims and practices of language documentation and revitalization over the past 100 years and opens the question of where the future of this field of study will take us next (Austin). Section 2 outlines practical strategies for language documentation and revitalization by presenting guides for the creation of audio-visual material (Dohle & Ostajewska), and the software and metadata recording practices which need to be employed in order to be able to make use of the raw data (Austin). The section continues with the proposal of transdisciplinary models via de research group *Plants. Animals. Words.* (Simard), and ends with a tool which supports collaborative community driven efforts in today's digital world (Language Landscape). Finally, Section 3 presents case studies from around the world, as they were presented during the two EngHum field school events at Wilamowice, Poland and Tlaxcala, Mexico (Chromik, Mu, Dohle, Goodchild & Selleri). It also provides an example of the type of research that collective and transdisciplinary practices can yield. The presentation includes exemplary results of sociolinguistic fieldwork carried out by one of the working groups during the EngHum Field school in Tlaxcala, Mexico.

Two more texts describing the two fieldwork events which have taken place to date throughout the duration of this project follow. The first provides some historical context of the linguistic situation found in Poland; the second an overview of the collaborative research activities, which took place during the field school in Mexico.

## Report 1: Poland

### *Developing innovative models for fieldwork and linguistic documentation: ENGHUM experience in Hałcnów, Poland*

*Bartłomiej Chromik, Scientific Coordinator of the ENGHUM Project*

Hałcnów (ger. *Alzen*, alz. *Alza*), formerly a separate village, currently belongs to the north-eastern district of the city of Bielsko-Biała (a large town close to Wilamowice where the 2016 ENGHUM field school was carried out). Until the end of World War II, the place was predominantly German, however its inhabitants spoke *Alznerish*, a variety which is hardly mutually intelligible with High German. Despite the fact that most of the Halcnovians were not politically connected to Nazism, after the end of the war they suffered severe persecution. The vast majority of the population of the settlement was either killed, banished to the Soviet Union or resettled to Germany. The village was repopulated by Poles migrating to new industrial plants in Bielsko-Biała. The communist regime tried to erase all “signs of Germaness” from public and private spaces.

As a consequence, *Alznerish* also became invisible. When the political situation in Poland changed and post-war anti-German sentiment declined, most of the scholars supposed that it is too late to find (at least in Hałcnów) any native-speakers of the language. The fieldwork conducted in 2013 by scholars from Adam Mickiewicz in Poznań proved that there are at least 8 native-speakers of the language, all of them elderly. These pilot studies marked the beginning of academic interest (rather modest) in the language.

During the 2016 ENGHUM field school in the town of Wilamowice (a place where other endangered language is spoken and where some activities of the ENGHUM project have been focused) the major task of one of the working groups was to document a linguistic and cultural heritage of Hałcnów. A multi-ethnic group consisting of seven people developed an innovative methodological approach to the problem. In the first phase of the fieldwork they focused on tracing the (hidden) elements of the linguistic landscape of Hałcnów. These actions were an attempt to recreate the materiality of ecology of *Alznerish*, but they also attempted to establish whether the German past of the village is seen as an integral part of heritage of the place by new inhabitants of Hałcnów. The positive externality of this work was gaining knowledge in local topography.

In the second part of the fieldwork the group was divided into two parts. The first group met with some of the native speakers (unfortunately it appeared that some of the people recognized in 2013 had already passed away) and conducted either less structured conversations in *Alznerish*, German and Polish, or some elicitation in *Alznerish*. The second group attempted to reach the most crucial social actors of Hałcnów: the priest, teachers, local historians and activists. Except for the obvious and overt aim of this work – gaining knowledge on current ideologies and attitudes toward the language, and asking about some other people that may know *Alznerish*, there was also another essential purpose to the fieldwork. In Poland researchers enjoy high respect of the society. Science indeed is perceived as a privileged discourse. Moreover, as a result of a long-lasting isolation of Poland in a communist period, foreigners from behind the Iron Curtain, especially outside big urban centers, are treated with esteem. Taking this into account, the interest of foreign scholars in *Alznerish*, inevitably led to increasing prestige of the local linguistic variety. This visit, was an indirect

and non-intrusive way to change linguistic ideologies. The work of this group led however, to some unexpected discoveries. In a local school, teachers were told about the 'dictionary of Halcnovian language' created by school students. The dictionary was found. It appeared to be a short glossary of the Polish variety used in Hałcnów, which is a testimony of emergence of a new linguistic community. Moreover, thanks to work of the subgroup, a previously unknown fluent speaker of *Alznerish* was identified.

In the third stage, the group acted together again. The meeting of all *Alznerish*-speaking participants of the research was organized. It was quite striking, but it appeared that despite the fact that most of them are neighbours, in some cases they did not know about their skills in their mother tongue. Their joy from this discovery was noticeable. It has to be admitted that the scholars taking part in the fieldwork did not know *Alznerish* (which should be a prerequisite for this kind of research; however, taking into account the short period and quite accidental character of the study, it was not demanded from them) could communicate in German or Polish. Very soon it turned out that using the latter language was more beneficial for the research. Halcnovians asked in German, replied in German. The 'distance' between Polish and *Alznerish* was big enough to prevent constant code-switching. The conversation concerned the pre-war time in the village and its "ethnography". Currently, it is perhaps the only domain where language can be used.

The last phase of research activities took place in Wilamowice. Halcnovians were asked to participate in an event summarizing the field school. They were treated as special guests and received an opportunity to speak publicly in their language. It was perhaps the first time after the end of World War II, that *Alznerish* was used publicly without fear and with pride. It is worth mentioning that this fact was noticed by a journalist of the regional newspaper, who devoted an article to the case of Hałcnów (<http://www.dziennikzachodni.pl/wiadomosci/bielsko-biala/a/jak-sie-mowi-po-halcnowsku-w-bielskubialej-kto-jeszcze-zna-ten-jezyk-zdjecia-wideo,10688031/2/> ). Later, he said, that he had received positive feedback from the readers about that publication.

The described pilot study is an innovative methodological proposal for a short-term study. It was focused on documentation of the language, networking of its users and either external or internal promotion of *Alznerish*. Combination of three factors may give some hope that the effects of the study will be extended in time. It is however highly recommended to immediately conduct more systematic transdisciplinary research in Hałcnów. Another essential result has been the creation of substantial linguistic-cultural documentation that can be used for multiple purposes, including research, teaching and revitalization.

## Report 2: Mexico

### *Lessons learnt from a hands-on practical approach to community based language documentation and revitalization*

*Elwira Sobkowiak, Ebany Dohle & Justyna Olko*

The 2017 Field school in San Miguel Xaltipan (Tlaxcala, Mexico), an event organized as part of the EngHum project with the participation of 'Artes Liberales' Faculty at the University of Warsaw, School of Oriental and African Studies (SOAS) at the University of London, Leiden University and representatives of San Miguel Xaltipan, took place between the 19th and 29th of August 2017.

Among the participants of the school that was organized in a Nahuatl-speaking zone were native speakers of various variants of Nahuatl itself and speakers of other indigenous languages of Mexico including Yucatec Maya, Ayuuc and Mixtec. Numerous scholars working on language documentation and language revitalization as well as language activists from Mexico, El Salvador, the U.S., Poland, United Kingdom, Italy and Catalonia also gathered in San Miguel Xaltipan. Activities scheduled for the duration of the School included workshops on language documentation techniques (audio and video recordings), data and metadata management, linguistic software used in language documentation and revitalization, design of pedagogical materials in language revitalization, etc. The School was also an opportunity for the exchange of experiences and making valuable contacts with fellow language activists working in language revitalization. Participants of the School had the chance to practice newly acquired skills during numerous workshops in a series of tasks assigned to them and involving documenting Nahuatl spoken in San Miguel Xaltipan, San Pedro Tlalcuapan and San Francisco Tetlanohcan. The programme of the School also included a number of academic and non-academic presentations on a range of topics in language documentation and revitalization strategies involving work with minority languages in Mexico, Central America, Africa and Europe. On the 25th and the 26th of August another edition of the Interdialectal Nahuatl Encounter took place. Several dozens of people gathered in Xaltipan in order to discuss a range of topics concerning the Nahuatl language and to do so in Nahuatl, which was the language of communication during the encounter. Nahuatl was also one of the working languages during the duration of the school beside Spanish.

The event's programme included trips to two Nahuatl-speaking towns: San Miguel Tenango in the state of Puebla and San Pedro Tlalcuapan in the state of Tlaxcala. An excursion to the archaeological sites of Cacaxtla and Xochitecatl were made as well. During the school several hundred books in Nahuatl published by the University of Warsaw (the 'Totlahtol' series) were distributed free of charge to the Nahua people in San Miguel Xaltipan and in other Nahua communities located nearby.

Field school activities were organized in such a manner that collaborative participation was encouraged. Attendees formed working groups of 8-10 people and were asked to carry out documentation efforts within the local community. Participants were expected to learn core language documentation techniques such as the creation of audio and visual materials, in addition to being aware of data management, ethical consent and team management issues. The event provided a useful opportunity for participants of different backgrounds to learn from each other directly, gaining practical hands on experience.

Once data had been collected via language documentation techniques, the participants were shown how to work with raw data. Workshops on how to utilize transcription and video editing software were provided, and the participants were asked to create revitalization materials using the raw documentation materials that had been previously collected. The final presentation within this collection shows an example of the types of results that these efforts resulted in. It is a presentation that was created by members of Group 3 (Gladiola Villegas, Martín Tonalmeyotl, Abel Hernández, Samantha Goodchild, Tomasz Wicherkiewicz, Gregory Haimovich, Refugio Nava Nava, and Genner Llanes).

The strategy for conducting fieldwork was based on close collaboration between indigenous participants (mainly Nahuatl-speaking students, researchers, teachers and activists coming from many different communities located in several distinct states of Mexico) and non-indigenous persons taking part in the event, including new speakers of the language and representatives of other minority groups from Mexico and Europe. This made it possible to conduct interviews in Nahuatl rather than in Spanish (the dominant language) and create spaces for spontaneous conversations, being mindful of Nahua communication conventions and rules of respect. This has been of special importance given that Nahuatl is a highly endangered, low-prestige/strongly discriminated language in the communities where the fieldwork was carried out, and has been gradually eliminated from most public and also private spaces over the last several decades to make way for Spanish. Out of six working groups of the ENGHUM field school, five were engaged in direct fieldwork in three indigenous communities and in creating linguistic-cultural documentation for the subsequent use by these communities. The themes of fieldwork embraced local linguistic and cultural heritage, customs, language attitudes and ideologies, local concepts of psychological and physical well-being and health - also including the role of the heritage language -, traditional healing, local craftsmanship, artistic and economic activities. Also, conversations envisioned as teaching materials were created by the group working on teaching strategies and materials. As a result, multi-purpose documentation has been created and handed in to the representatives of participating local communities (San Miguel Xaltipan, San Francisco Tetlanohcan and San Pedro Tlalcuapan). It is hoped that these materials will be used not only for research purposes, but also, become part of incipient communitarian archives of linguistic and cultural heritage which can be further used in revitalization and teaching activities.



## **Presentations**

### **Section 1**

Austin, Peter, "Language Documentation: Past, Present and Future", EngHum Summer School 2016, 14/06/2016, London, UK

### **Section 2**

Dohle, Ebany, "Creando grabaciones de audio", EngHum Field school 2017, 21/08/2017, Xaltipan, Mexico

Ostajewska, Marta, "The Basic Principles of Creating Audio-visual Material", EngHum Field school 2017, 21/08/2017, Xaltipan, Mexico

Austin, Peter, "Software Tools", EngHum Summer School 2016, 14/06/2016, London, UK

Austin, Peter, "Data management, metadata and mobilisation", EngHum Summer School 2016, 14/06/2016, London, UK

Goodchild, Samantha & Ebany, Dohle, "Language Landscape: Un modelo y una plataforma digital para la autodocumentación de idiomas", EngHum Field school 2017, 28/08/2017, Xaltipan, Mexico

Simard, Candide, "Plants. Animals. Words.", EngHum Summer School 2017, Leiden, Holland

Dohle, Ebany & Mu, Sophie, "PAW – Conclusions for the Day: Impact of Transdisciplinary Data Collection on Revitalization Efforts", Summer School 2017, Leiden, Holland

### **Section 3**

Mętrak, Maciej & Żak, Andrzej, "Documenting Halcnovian", EngHum Field school 2016, Wilamowice, Poland

Grabowski, Martin & Mu, Sophie, "A Tale of Two Countries: Language ideologies and identities in Mengguang, China and the Basque Country, Spain", EngHum Field school 2016, Wilamowice, Poland

Dohle, Ebany, "El código de las plantas: Motivos, estrategias y recomendaciones", EngHum Field school 2017, 21/08/2017, Xaltipan, Mexico

Goodchild, Samantha, "El uso de idioma(s) en Essyl, Senegal", EngHum Field school 2017, 24/08/2017, Xaltipan, Mexico

Selleri, Sara, "Un dialogo sobre la colaboración interdisciplinaria entre etnomusicólogos y sociolingüistas para la revitalización de la lengua y de la música", EngHum Field school 2017, 21/08/2017, Xaltipan, Mexico

Gladiola Villegas, Martín Tonalmeyotl, Abel Hernández, Samantha Goodchild, Tomasz Wicherkiewicz, Gregory Haimovich, Refugio Nava Nava, Genner Llanes, "Actitudes Lingüísticas", EngHum Field school 2017, 29/08/2017, Xaltipan, Mexico

# Section 1

---

ENGHUM Summer School 2016

# Language documentation: past, present and future

Peter K. Austin

Department of Linguistics, SOAS, University of London

2016-06-14

---

---

# Outline

- Language documentation – past
  - Language documentation – present
    - ‘multipurpose record’
    - interfaces and accessibility
    - output gaps
  - Language documentation – future?
    - diversity
    - transdisciplinarity
    - metadocumentation
    - relationships with revitalisation
    - beliefs and ideologies
-

---

# Language Documentation – past

---

---

# Language documentation 1

- Term used in late 19<sup>th</sup> and early 20<sup>th</sup> 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” (Himmelman 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
-

---

# Evidence

## Student interest

- 180 students graduated from SOAS *MA in Language Documentation and Description* 2004-16 – currently 20 are enrolled
- 15 graduates in PhD in Field Linguistics – 20 currently enrolled
- other documentation programmes, eg. UTAustin have similar experience
- Summer schools (3L, InField/Colang), training courses (DocLing, Mahidol etc.)

## Publications

- Gippert et al 2006 *Essentials of Language Documentation*. Mouton
  - *Cambridge Handbook of Endangered Languages* 2011
  - *Routledge Essential Readings* 2011
  - *Oxford Bibliography Online* 2012
  
  - *Language Documentation and Description* – 13 issues (paper and online)
  - *Language Documentation and Conservation* – 6 issues (online only)
-

# Big money – DoBeS projects



# Even bigger money – ELDP projects

**ENDANGERED LANGUAGES DOCUMENTATION PROGRAMME**

WHAT WE DO ABOUT US APPLY OUR GRANTS - **OUR PROJECTS -** NEWS & EVENTS

**ARCADIA** **SOAS**  
University of London

**TUYUKA NATHALIE PIRES VLCEK**

**DOCUMENTATION OF BRAZILIAN TUYUKA**  
Tuyuka is an Eastern Tukano language spoken by approximately one thousand people distributed in some twenty communities in Brazilian and Colombian territories in northwestern Amazonia. The language as it is used in communities served by the Tuyuka indigenous school, and to aid the school's efforts toward language maintenance. The school serves five communities in Brazilian and Colombian territories: Sao Pedro, Cachoeira Comprida, Fronteira, Papunha and Bella Vista, and fieldwork will be conducted primarily in the first two, on the Brazilian side of the upper Tiqui river.

<http://elar.soas.ac.uk/deposit/0300>

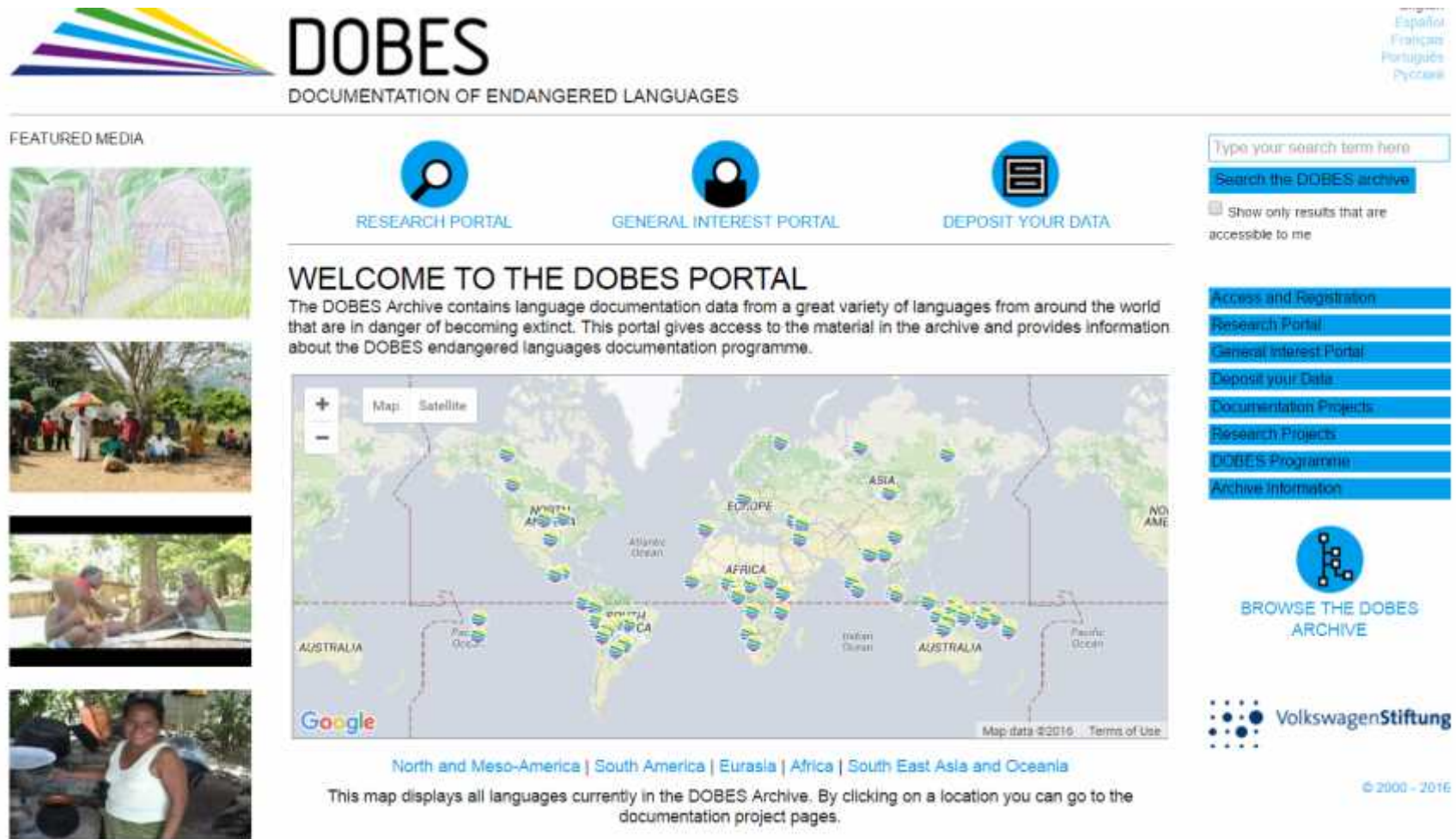
**PROJECTS WE HAVE FUNDED**

The image shows a world map with numerous orange location pins indicating funded projects. The pins are densely clustered in South America, particularly in the Amazon region, and are also scattered across Africa, Asia, and Europe. A sidebar on the left provides details for a specific project: 'TUYUKA NATHALIE PIRES VLCEK'. The text describes the documentation of Brazilian Tuyuka, an Eastern Tukano language spoken by about 1,000 people in 20 communities in northwestern Amazonia. It mentions a school serving five communities (Sao Pedro, Cachoeira Comprida, Fronteira, Papunha, and Bella Vista) and notes that fieldwork will be conducted primarily in the first two on the Brazilian side of the upper Tiqui river. A URL is provided: <http://elar.soas.ac.uk/deposit/0300>. At the bottom of the map, the text 'PROJECTS WE HAVE FUNDED' is displayed.

# Big archives – ELAR at SOAS



# Big archives – DoBeS at MPI Nijmegen



The screenshot shows the homepage of the DOBES (Documentation of Endangered Languages) website. At the top left is the DOBES logo, a stylized rainbow fan, followed by the text "DOBES DOCUMENTATION OF ENDANGERED LANGUAGES". In the top right corner, there are language selection options: Español, Français, Português, and Pycckий.

Below the header, there are three main navigation buttons: "RESEARCH PORTAL" (with a magnifying glass icon), "GENERAL INTEREST PORTAL" (with a person icon), and "DEPOSIT YOUR DATA" (with a document icon). To the right of these buttons is a search bar with the placeholder text "Type your search term here" and a "Search the DOBES archive" button. Below the search bar is a checkbox labeled "Show only results that are accessible to me".

On the left side, there is a "FEATURED MEDIA" section with four small images: a drawing of a person and a hut, a group of people in a field, two people sitting and talking, and a woman in a white tank top.

The main content area features a "WELCOME TO THE DOBES PORTAL" heading, followed by a paragraph: "The DOBES Archive contains language documentation data from a great variety of languages from around the world that are in danger of becoming extinct. This portal gives access to the material in the archive and provides information about the DOBES endangered languages documentation programme." Below this is a world map showing the distribution of languages in the DOBES Archive, with pins indicating specific locations. The map is labeled with continents: NORTH AMERICA, SOUTH AMERICA, AFRICA, EUROPE, ASIA, and AUSTRALIA. It also shows the Atlantic Ocean, Indian Ocean, and Pacific Ocean. The Google logo is in the bottom left of the map, and "Map data ©2016 Terms of Use" is in the bottom right.

Below the map, there is a navigation bar with links: "North and Meso-America | South America | Eurasia | Africa | South East Asia and Oceania". Below this bar is a caption: "This map displays all languages currently in the DOBES Archive. By clicking on a location you can go to the documentation project pages." In the bottom right corner of the page, there is a "BROWSE THE DOBES ARCHIVE" button and the VolkswagenStiftung logo. At the very bottom right, the copyright notice "© 2000 - 2016" is visible.

---

# 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
-

---

# A 2010 example – Stuart McGill

- 4 year PhD project at SOAS, plus 2 year post-doc
  - documentation of Cicipu (Niger-Congo, north-west Nigeria) in collaboration with native speaker researchers
  - outcomes:
    - a corpus of texts (video, ELAN, Toolbox)
    - 2,000 item lexicon
    - archive (956 files, 50Gbytes)
    - overview grammar (134 pages)
    - analysis of agreement (158 pages)
    - website, cassette tapes, books, orthography proposal and workshop
-



# McGill Cicipu corpus

The screenshot displays the svtmg001.caf application interface. On the left, a video player shows two men in traditional attire. The main window is divided into several sections:

- Subtitle List:** A table with columns for 'Nr', 'Annotation', 'Begin Time', 'End Time', and 'Duration'. The current selection is row 69: 'Hausa has swept them away'.
- Timeline:** A horizontal axis at the bottom showing time markers from 00:00:00.000 to 00:03:31.000.
- Subtitle Editor:** A detailed view of the selected subtitle (69) showing the text 'Hausa has swept them away' and its alignment with the video frame.

Nr	Annotation	Begin Time	End Time	Duration
68	yes	00:03:27.030	00:03:27.790	00:00:00.760
69	Hausa has swept them away	00:03:28.260	00:03:29.450	00:00:01.190
70	yes	00:03:30.260	00:03:30.960	00:00:00.700
71	if they want proper Cicipu, it's only the former people	00:03:33.910	00:03:37.830	00:00:03.920
72	m-hm, not the children of now	00:03:38.350	00:03:40.310	00:00:01.960
73	Cicipu only the former people	00:03:40.310	00:03:42.310	00:00:02.000
74	yes	00:03:42.310	00:03:44.310	00:00:02.000
75	yes	00:03:46.550	00:03:47.570	00:00:01.020
76	yes	00:03:58.230	00:03:59.010	00:00:00.780
77	one place, doing their work in one place	00:04:00.550	00:04:03.790	00:00:03.240
78	they would do it together	00:04:06.070	00:04:07.870	00:00:01.800

# Cicipu annotations

Toolbox - Dictionary.tst

File Edit Database Project Tools Checks View Window Help

[no filter]

Text.tst

Reference: svimg001.099  
Start Time: 208.260  
End Time: 209.150  
Speaker: Tami

Text: **tikógó**    **tihúcò**                    **rè**

Morphemes: tì kógá tì    húcò l. H l. rè

Gloss: NC6- Islam AGR6- sweep -RLSp 3PP.PRO

Part of Speech: nc- n    agv- v    -vtonep pro

Free Translation: Hausa has swept them away

Reference: svimg001.100  
Start Time: 210.260  
End Time: 210.260  
Speaker: Tami

Text: **óó**

Morphemes: ò'íí

Gloss: yes

Part of Speech: inj

Dictionary.tst

Lexeme	Citation form	English Gloss	Part of speech	Gloss (1)	Paradigm
kógó	*empty*	Hausa person	n	*empty*	3/2
káká	(m) káká	small_drum_k.o.	n	*empty*	4/5?
kákóp	*empty*	drum_k.o.	n	kuge?	8/3?
kámá	(k) kámá	drum?	n	*empty*	4/?
komo	*empty*	cover	v	*empty*	*empty*
kámá	(ù)-kámá	salt	n	*empty*	7/8
káo	*empty*	carry	idco	da wuri	*empty*
káo	(k) káo	egg	n	*empty*	1/2
káo	(ù)-káo	floor	n	tartajya	7
kóhúu	*empty*	lung	n	huhu ; kuhu	8
káré	*empty*	indeed	adv	kwarai	*empty*
kóti	(m) kóti	stump	n	*empty*	4/5
koto	*empty*	finish	v	karke	*empty*
késá	(k) késá	back of head	n	*empty*	1/?

Metasab.tst

id	file	con contributor	audio file	recording place
svgd001	Discussion of chieftanc	(H)M ; JN ; King	..\..\audio_visual\svgd001.wav	King's guest house, Korisino
svmk001	Greeting the Mallu	MM ; MK	..\..\audio_visual\svmk001.wav	Mallu's guest hut in Kacalada
svmy001	Norman Biggs' grave	MY	..\..\audio_visual\svmy001.wav	At the grave of Norman Biggs, Sakaba
svsdt001	Ukula mountain	MoMu ; SDT (Sani the	..\..\audio_visual\svsdt001.wav	On the Ukula (Maburya) mountain
svimg001	Interview about the old	MoMa ; TMG ; Yaaki	..\..\audio_visual\svimg001.wav	IMG's compound in Ka'ingawa KaGaladima

ik kógó      835/195      Cicipu.ppt

# Cicipu archival deposit

## Cicipu documentation

[Home](#) [Resources](#)

---

### Cicipu documentation

Language: Cicipu [awc]

Depositor: Stuart McGill

Location: Nigeria

---

#### Summary of deposit

This corpus contains folktales, riddles, historical narratives, casual conversation, commentaries on festival videos, interviews, songs, prayers, and sermons. Nine Pear Film narratives are also included. In total there are approximately six hours of interlinearised time-aligned texts are provided in Toolbox/ELAN format. The corpus also contains an accompanying lexicon in Toolbox format, collected from the texts as well as from the SIL Africa Area 1700-item wordlist. A large number of elicitation sessions are also provided (conducted in either Hausa or Cicipu). GPS data of the Cicipu area is included.

#### Group represented

The Acipu of Kebbi and Niger State, Nigeria

#### Language information

Called Acipançi in Hausa. Called 'Western Acipa' in Ethnologue 15th edition.

#### Special Characteristics

The deposit includes A. B. Mathews' 'Historical and anthropological report on the Achifawa', an unpublished typewritten manuscript from 1926. There is a physical copy in the National Archives, Kaduna (K2, 068), from which the electronic copy in this corpus was photographed.

#### Search this deposit

[Reset keywords](#)

#### Access protocol

URLS (407)

#### Language

[more](#) ▾

- Cicipu (259)
- Damakawa (2)
- Duka (2)
- English (1)
- Hausa (8)
- [more ...](#)

#### Type

- Audio (287)
- Document (20)
- ELAN (60)
- Image (141)
- Text (37)
- Transcriber (32)
- Video (53)
- XML (2)
- Zippped collection (7)

#### Tags

- Kezzeme (5)
- Photo (1)
- Photos (8)

#### Genre

[more](#) ▾


- Christian (2)

#### Deposit status


✓ **Curated:**  
Resources online and curated

#### Depositor

Stuart McGill

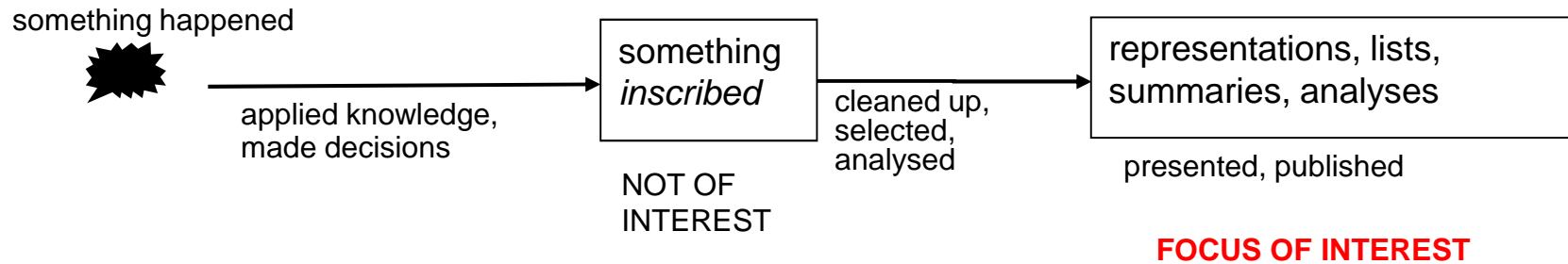


**Nationality:** UK  
**Affiliation:** School of Oriental and African Studies

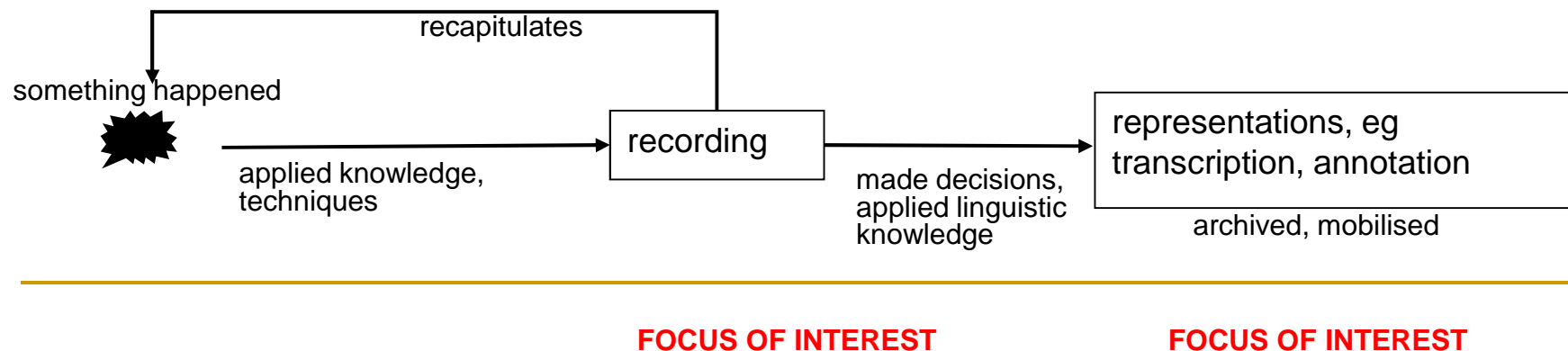


# Workflow

## Description



## Documentation



---

# As a consequence

- Documentation 2 needs an epistemology for media capture – audio and video recording (see Nathan)
  - 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
-

---

## Woodbury (2003:46-47) – a good corpus

- *diverse* — containing samples of language use across a range of genres and socio-cultural contexts, including elicited data
  - *large* — given the storage and manipulation capabilities of modern information and communications technology (ICT), a digital corpus can be extensive and incorporate both media and text
  - *ongoing, distributed, and opportunistic* — data can be added to the corpus from whatever sources that are available and be expanded when new materials become available
-

- 
- *transparent* — the corpus should be structured in such a way as to be useable by people other than the researcher(s) who compiled it, including future researchers
  - *preservable*, and *portable* — prepared in a way that enables it to be archived for long-term preservation and not restricted to use in particular ICT environments
  - *ethical* — collected and analysed with due attention to ethical principles and recording all relevant protocols for access and use.
-

---

# Recording

- *audio* – basic and familiar in modern linguistic work. Important considerations: environment, equipment choice, microphones, monitoring, file type (wav not mp3 generally recommended)
  - *video* – immediate, rich in authenticity, multi-dimensional in context, great interest to communities, can be produced independently by community members BUT more difficult to produce, process, access without time-aligned annotation, transfer, store and preserve
  - *text* – compact, stable, easy to store, access and index, can express hypertextual links to other text and media BUT relies on literacy and is less rich than audio/video
-



- 
- *metadata* – data about the data: needed to identify, manage, retrieve data. Provides context and understanding of data to oneself and others. Types:
    - Cataloguing — identifying and locating data, eg. language code, file id, recorder, speaker, place of recording, date of recording etc
    - Descriptive — kind of data found in a file, eg. abstract/summary of file contents, knowledge domain represented
    - Structural — specification of file organisation, eg. textfile is a bilingual dictionary
    - Technical — file format, kind of software needed to view, preservation data
    - Administrative — work log, intellectual property rights, moral rights, access and distribution restrictions
-

- 
- *meta-documentation* – documentation of language documentation models, processes and outcomes, goals, methods and conditions (linguistic, social, physical, technical, historical, biographical) under which the data and analysis was produced (should be *as rich and appropriate* as the documentary materials themselves)
-

---

# Adding value

- requires decision making (selection, editing, choice of method and theory) and is very time consuming (eg. annotation can be 100:1 in terms of time required)
  - linguistic value adding ('thick' meta-data):
    - *transcription* – textual representation of audio signal (orthographic, phonemic, phonetic) typically time-aligned to media
    - *annotation* – overview, code, morphological, grammatical, semantic ('gloss'), syntactic, pragmatic, discourse. Fixation among documenters on 'interlinear glossing', cf. overview annotation/summary
    - *translation* – levels: word, sentence, paragraph, text. Types: literal, running, parallel, literary (Woodbury 2005, Evans and Sasse 2005)
-

---

# Tools for value adding

- application programs, components, fonts, style sheets, and document type definitions (DTD).
  - application programs:
    - *general purpose* software – user must design data structures and manipulation routines, eg. LibreOffice, MS Office (Word, Excel, Access)
    - *specific purpose* software – designed for particular tasks, eg. Transcriber, ELAN, Arbil, Flex, Toolbox
  - Important: design and use a work flow that enables data transfer (export, import) without loss/corruption of encoded knowledge
-

---

# Archiving

A digital language archive:

- is a trusted repository created and maintained by an institution with a commitment to the long-term preservation of archived material
  - has policies and processes for acquiring, cataloguing, preserving, disseminating, and format/content migration
  - is a platform for building and supporting relationships between data providers and data users
-

---

# And *endangered* languages archiving?

- extremely diverse context – languages, cultures, communities, individuals, projects
  - typical source - fieldworkers
  - typical materials - documentation
  - difficult for archive staff to manage
  - sensitivities and restrictions
-

# Example – deposit home page

Preservation of the Jewish Iraqi spoken language - Windows Internet Explorer

http://elar.soas.ac.uk/deposit/timan2008jewishiraqi

Endangered Languages Archive

Logged in as [egvahoo](#) Logout | [Feedback](#) | [ELAR catalogue](#)

## Preservation of the Jewish Iraqi spoken language

Home Metadata Resources

Find resources by ...

Search

Reset search

Access protocol  
U R C S (23)

Language  
Jewish Iraqi (23)  
English (18)  
Hebrew (11)  
Classical/Standard Arabic (8)  
Muslim Iraqi (5)

Type  
Audio (23)  
ELAN (23)  
Image (14)


Topic  
Apples (2)  
Drummers (2)  
Henna (2)  
House (2)  
Jars (2)  
Painting (2)  
Persian (2)  
Prayer (2)  
Wall (2)  
Writing (2)  
Alembic (1)  
Amascus (1)  
Arabesque (1)  
Arak (1)  
Bamboo (1)  
Berlin (1)

Preservation of the Jewish Iraqi spoken language

Language: Jewish Iraqi

Depositor: Eli Timan

Location: Iraq



**Summary of deposit**  
This deposit contains personal stories, songs and descriptions of life in Iraq in the first half of the twentieth century, as told by Jewish Iraqi informants in London, Canada and Israel in 2007.

The image on the right is a painting by Eli Sawdayee, showing a man and his donkey with the Alembic, distilling rose water. To access the story and picture, search for or click on [alembic](#) in the left panel.

**Deposit contents**  
The stories, songs and descriptions have been transcribed and translated to English, time-aligned with the audio recordings, using ELAN software. There are also original paintings on which some of the stories are based.


**Special Characteristics**  
Contains original paintings by informants, which are the focus of discussions.

**Depositor**  
Eli Timan

**Nationality:** British  
**Affiliation:** School of Oriental and African Studies

**Your access**  
Default access protocol: U R C S  
Your access roles: U R C S

**Deposit**  
**Group represented:** Iraqi Jews  
**Location:** London, Canada, Israel



Map Terrain

Done, but with errors on page.

Internet | Protected Mode: Off

100%

# Example – file deposit pages

The screenshot shows a web browser window displaying the Endangered Languages Archive (ELAR) website. The page title is "Preservation of the Jewish Iraqi spoken language". The search results show 23 resource bundles. The selected bundle is "Titipampa" with a file view showing three files: "Titipampa.eaf", "titipampa.wav", and "TITIPAMPA.jpg". The metadata for the selected bundle is as follows:

ID:	Titipampa
Title:	Refurbishing mattresses
Linguistic topic:	----
Date created:	23/12/2006
Location:	Israel
Language:	English - Hebrew - Jewish Iraqi
Participants:	Eli Meir Sawdayee
Contributor:	Eli Timan
Devices:	M660

The page also includes a sidebar with filters for "Find resources by ...", "Access protocol", "Language", "Type", and "Topic". The "Topic" filter shows various categories like "Apples (2)", "Drummers (2)", etc. The "Deposit" information on the right indicates the depositor is Eli Timan, with a nationality of British and affiliation at the School of Oriental and African Studies. A map of the Middle East is also visible, with a red pin indicating the location in Israel.



# Mobilisation

- Creation of usable outputs for a range of different audiences, eg. multimedia websites, sub-titled video, apps
- There are tools to help with this (LexiquePro, CuPed) and people working on app development who can help, e.g. Ma! Iwaidja



---

# 1. “Research ON a language”

- Usual in first half of 20<sup>th</sup> century
    - “Salvage linguistics”
  - Who is language documentation for?
  - Still continues: ‘lone wolf’ linguist encouraged by some funding models
  - “Community members report sometimes feeling that the linguist comes in, reifies the language, turns it into a commodity, and then takes it away.” (Bower 2011: 468)
-

---

## 2. “Research FOR the community”

- Developed in 1960s
    - period of civil rights movements in USA
  - Fieldworkers ‘give something back to the community’
    - e.g. educational materials,
    - advocacy: Labov 1982
  - Endangered language speakers are not just sources of data
    - often economic and social problems contribute to language shift
  - Not all linguists have other needed skills (e.g. social work, medical expertise)
-

---

### 3. “Research WITH the community”

- Developed in 1980s
    - “Action Research”
    - “Negotiated fieldwork”
  - Equal say and partnership to speakers of the language
    - Full participation, from planning to outputs
  - Now dominant model
    - at least in rhetoric!
  - May be difficult to find funding for
-

---

## 4. “Research BY a community”

- The project is community-driven
  - May include maintenance/revitalisation measures, creating language teaching programmes, etc.
    - e.g. Dieri Aboriginal Corporation
  - Multidisciplinary approach
  - Role of external linguist:
    - Training, teaching, mentoring native speakers ...
-

- 
- Language might not be their major concern
  - Or it may be part of a larger set of interrelated concerns.  
e.g. economic development or health or environmental issues
  - If concerned about language vitality, communities may be keener on revitalisation than description
  - Growing interest in many indigenous communities in traditional knowledge and indigenous paradigms of teaching, learning, and research:
    - may clash with Western scientific models which seek to quantify measurable entities and results
    - but don't make essentialist assumptions (e.g. Grenoble & Whitecloud 2014)
-

---

# 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.’ (Bowerman 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.
-



# ILG blindness

Texts.txt	
Reference	svtmg001.099
Start Time	208.260
End Time	209.450
Speaker	Tenii
Text	tíkógó tihúfò rè
Morphemes	tì- kógó ti- hudò -L -H -L rè
Gloss	NC6- Islam AGR6- sweep -RLSp 3PP.PRO
Part of Speech	nc- n agv- v -vtonep pro
Free Translation	Hausa has swept them away
Reference	svtmg001.100
Start Time	210.260
End Time	210.960
Speaker	Tenii
Text	óó
Morphemes	ò'íi
Gloss	yes
Part of Speech	intj

many documenters believed that interlinear glossing is the 'gold standard' of annotation but it is very time-consuming and illegible to non-linguists – **overview annotations** may be a preferable as a primary goal: 'roadmap' or index of a recording – approximately time-aligned information about what is in the recording, who is participating, and other interesting phenomena

---

# Holton 2014

**Item 408:** Oral Literature Collection, Tape 343, Side B. Robert Zuboff (*Kak'weidí* clan, *Kaakáakw Hít*) and Susie James (*Chookaneidi* clan, *T'akdeintaan yádi*), July 27, 1972; interviewed by Nora Marks Dauenhauer, migrated from reel to CD. Length 60:14. **Content by DK:** story of how the Sea Otter came to be is told, 0-4:15; raven sounds are given by Zuboff, and their meaning/use, 4:16-11:10; Zuboff tells a story about a man who became an invisible man (*tlékanáa*) (13:24); 11:11-13:24; story of a man named Naawan that bit the tongue off of a raven, 13:25-16:09; general conversation and questions about Tlingit phrases, 16:10-19:57; story of a man named *Gáneix*, 19:58-21:40; discussion about language and storytelling, mention of the Salmon Boy story, 21:41-24:12; Zuboff tells the story about the Woman that Raised the Wood Worm, attributes the story's people, 24:13-27:34; Susie and Nora talk, Susie speaks about the Man Who Commanded the Tides (*Yookis kookeik*) and his sister and raven. She then tells the story of bringing in the house that was way out on the ocean and how raven got the octopus tentacle to bring in the house. She then talks about the type of resources that were in the house but not in detail. She mentions the whale, cod etc. She then goes back to the man who commanded the tide and rescues his mother by placing her in the skin of a black duck, 27:35 to the end of the recording. Notes on file.

---

---

# Language Documentation – present

---

---

# Language documentation – present

- ‘multipurpose record’
  - interfaces and accessibility
  - output gaps
  - documentation and revitalisation
-

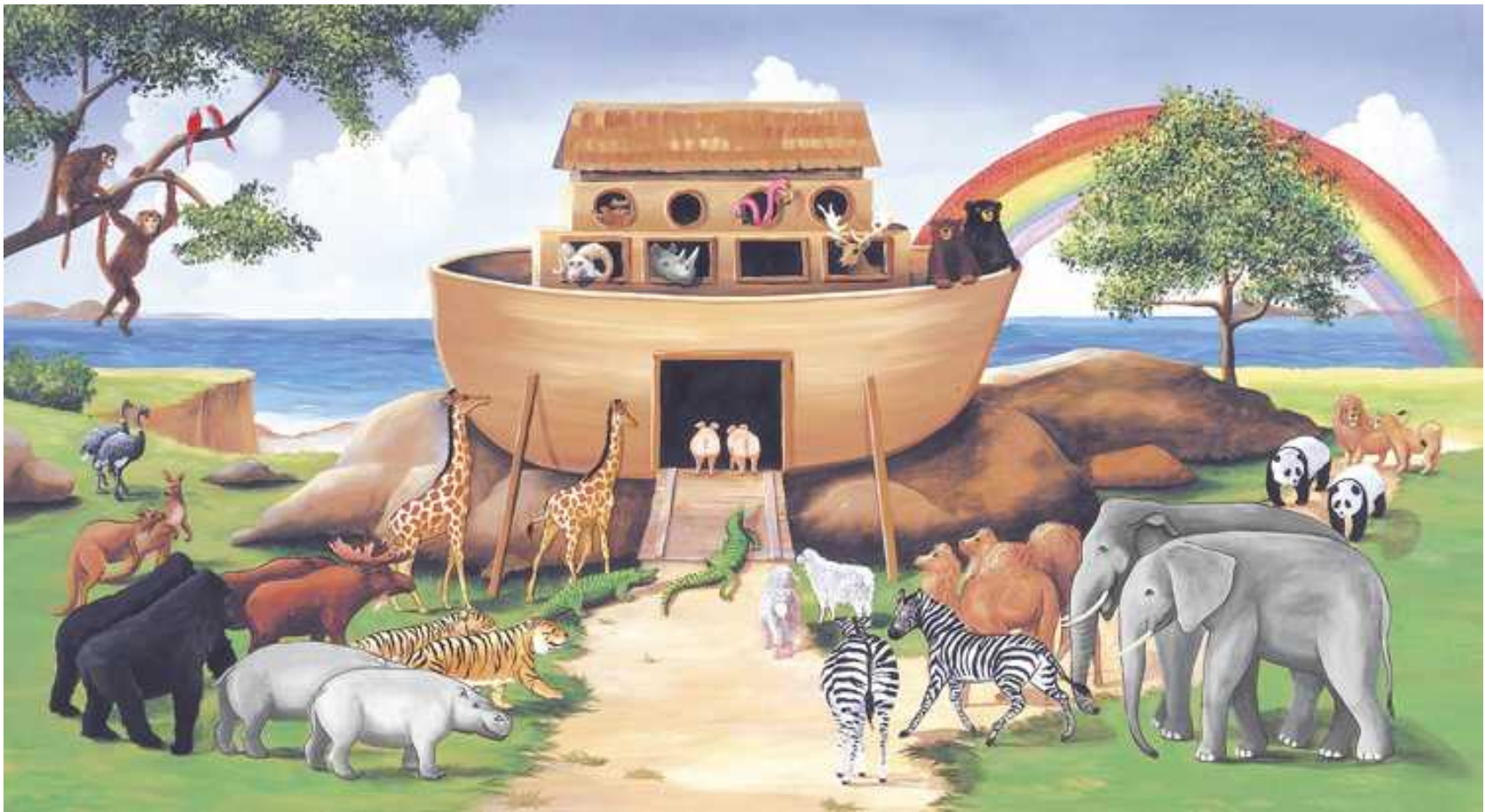
---

# 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.)
-

# The documentation model 1995-2010




Noah's arc(hive) – saving the morphemes 2-by-2

---

# ‘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

 Endangered Languages Archive Not logged in. [Login](#) | [New user](#) | [Search](#) | [Home](#)

## Search

Found 46 resources in the archive (page 1 of 6)

[1](#) [2](#) [3](#) [4](#) [5](#) [6](#) [next >](#) [last >>](#)

### Search ELAR

[Reset keywords](#)

#### Language

- Avatime (Dominant) (5)
- Chinese (1)
- Cicipu (5)
- English (6)
- Gurindji Kriol (7)
- Gurindji (2)
- Kriol (2)
- Nar (2)
- Nepali (4)
- Northern Arapaho (1)
- Phu (1)
- Sadu (1)
- Spanish (4)
- Yakkha (4)
- Zapotec (4)

#### Type

- Audio (31)
- Deposit (3)
- Document (4)
- ELAN (17)
- Image (2)
- Text (4)
- Video (10)
- Zipped collection (1)

- **Text documentation of N|uu**  
*Tom Güldemann*  
... 22 hours of audio and 10 hours of video material of the last 10 known speakers of the Nuu language. Texts include folktales, personal stories, and conversations.

---

- **Cicipu documentation**  
*Stuart McGill*  
... folktales, riddles, historical narratives, interviews, songs, prayers, and sermons; 9 Pear Film narratives and six hours of interlinearised time-aligned texts.

---

- **Documentation and grammatical description of Yakkha, Nepal**  
*Diana Schackow*  
... documenting Yakkha, an endangered language of Nepal

---

- ▶ **FM09\_a115 (Documenting Gurindji Kriol, an Australian mixed language)**

#### 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:





## Search

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

1 2 3 4 5 6 7 8 9 ... [next >](#) [last »](#)

### Search ELAR

[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)  
 Koyl 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, Watungene 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:

# Corpus accessibility – I found it, what now?

**Cicipu documentation**

Home Resources


Found 60 bundles in this deposit with keyword **ELAN x** (page 1 of 8)

1 2 3 4 5 6 7 8 next > last »

▼ **Discussion of chieftancy**

svgd001.eaf Access protocol: [URCS](#)  
Download

svgd001.001.mpg Access protocol: [URCS](#)



00:05 00:17

Download

**Search this deposit**

Search

[Reset keywords](#)

**Access protocol**

[URCS](#) (60)

**Language** *more* ▼

Cicipu (58)  
Tidipo (5)  
Tikula (3)  
Tirisino (6)  
Damakawa  
*more* ...

**Type**

**ELAN x**  
Audio (60)  
Image (8)  
Transcriber (3)  
Video (10)  
Document  
Text  
XML  
Zipped collection

**Tags**

Kezzeme (2)  
Photo  
Photos


**Genre** *more* ▼

**Deposit status**

✓ **Curated:**  
Resources online and curated

**Depositor**

Stuart McGill




**Nationality:** UK  
**Affiliation:** School of Oriental and African Studies

**Your access**

Your roles: [URCS](#)

**Tools**

[Download metadata](#)  
[Add to My Bookmarks](#)



# Corpus accessibility – I can't even find it

The screenshot displays the interface of 'The Language Archive'. On the left is a hierarchical tree view of corpora, including 'DoBeS archive' and 'Bainouk Gubeeher'. The tree view shows a path: DoBeS archive > Plants > Elicitations > Tree list > DJI10312CDD. The main content area on the right shows the details for the selected item 'DJI10312CDD'. The details are organized into sections: Session, Description, Location, Project, Content, Languages, Actors, and Mediafile. The 'Session' section shows Name: DJI10312CDD, Title: Tree list, and Date: 2012-03-01. The 'Description' section contains the text: 'Verification of the pronunciation and agreement patterns of all tree names.' The 'Content' section lists metadata: Genre: Elicitation, SubGenre: lexical elicitation, Task, Modalities: speech, Subject, Interactivity: interactive, PlanningType: planned, Involvement, SocialContext, EventStructure, and Channel. The 'Languages' section lists 'Language: Bainouk Gubeeher (c)' and 'Language: French (c)'. The 'Actors' section lists 'Actor: Alexander Cobbinah' and 'Actor: Jean Marie Sagna'. The 'Mediafile' section shows Type: audio, Format: audio/x-wav, Size: 460 MB, and Quality: Unspecified. The 'RecordingConditions' and 'TimePosition' sections are also visible, with 'TimePosition' showing Start and End as 'Unspecified'.

---

# Shifting the sociology of the academy?

- The development of language documentation from 1995 looked like a possible avenue to legitimise data collection and analysis and shift the sociological power balance between ‘theoretical linguists’ and ‘fieldworkers’ (or ‘butterfly collectors’) as it developed its own theoretical and analytical machinery
  - This is the context that led in 2010 to the LSA Resolution Recognizing the Scholarly Merit of Language Documentation:
    - “the products of language documentation and work supporting linguistic vitality are of significant importance to the preservation of linguistic diversity, are fundamental and permanent contributions to the foundation of linguistics, and are intellectual achievements which require sophisticated analytical skills, deep theoretical knowledge, and broad linguistic expertise”
-

---

There is an output gap



---

# The output gap

Outputs from language documentation projects have bifurcated into:

- ❑ **Published** grammars, (bilingual) dictionaries and (glossed) texts – ‘revival’ of familiar genres linguists have been comfortable with for 100+ years
  - ❑ **Archive** deposits – hundreds or thousands of files, professionally curated by archivists, but often poorly organised or structured, with little if any contextualisation
-

---

# What is missing?

- Meta-documentation – the documentation of documentation projects, goals, methods, IP contributions, outcomes
  - New (unfamiliar) genres that link and contextualise analytical outputs and the archival corpus:
    - ethnographies of documentation project designs
    - accounts of data collection (cf. archaeology ‘field report’)
    - finding-aids to corpus collections
    - ‘exhibitions’ or ‘guided tours’ of archival deposits
  - Evaluation measures that enable properly-based peer assessment of documentations, equivalent to the way traditional outputs are judged
-

---

# 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
  - beliefs and ideologies
-

---

# Diversity

- of goals, contexts, people, data, corpora, outcomes
    - move away from Noah's Arc(hive) 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)
-

---

# Meta-documentation

- identity of **stakeholders** involved and their roles in the project
  - **attitudes** of language consultants, both towards their languages and towards the documenter and documentation project
  - **relationships** with consultants and community
  - **goals** and **methodology** of researcher, including research methods and tools (see Lüpke 2010), corpus theorisation (Woodbury 2011), theoretical assumptions embedded in annotation (abbreviations, glosses), potential for revitalisation
  - **biography** of the project, including background knowledge and experience of the researcher and main consultants (eg. how much fieldwork the researcher had done at the beginning of the project and under what conditions, what training the researcher and consultants had received)
  - for funded projects, includes original grant application and any amendments, reports to the funder, email communications with the funder and/or any discussions with an archive
-

---

# New genres

- Woodbury (2015) ‘Archives and audiences: Toward making endangered language documentations people can read, use, understand, and admire’:

“I urge documenters to take **authorial control** of their work, as they would if each archived collection were a book of language materials

- make a guide to your own documentary corpus
- include meta-documentation: describe the design of activities or projects from which the corpus arose, offer a theorization of the corpus (or several, from different perspectives), and describe the appraisal process used to select and assemble the corpus
- write narratives, logs, and journals
- think of your corpus as belonging to a **genre**.

To some extent, all this means documenters taking on some of the work traditionally done by archivists.”

---



[www.elpublishing.org](http://www.elpublishing.org)

A new free open access online venture launched in July 2015 which:

- has the infrastructure of familiar models of publication (editorial board, peer assessment, etc.)
  - publishes familiar outputs: LDD journal, books, multimedia, but without the limitations of traditional publishing, eg. length, hypertext
  - provides a platform to encourage experiments in new genres of output, eg. “Language Contexts”
  - will provide a space and an interface to move towards evaluations of these new outputs so that the underlying desire of the LSA statement might be realised
-



---

# 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?
-

---

# Transdisciplinarity

- Siebert (2016) ‘Documentary linguistics: a language philology of the 20<sup>th</sup> century’:

“documentary linguistics’ focus on ‘direct representation of discourse’ requires a broader conceptualization of the field that moves **beyond purely linguistic concerns**. This article recasts documentary linguistics as a philology, broadly understood as the inquiry into ‘the multifaceted study of texts, languages, and the phenomenon of language itself’ ... The reconceptualization of documentary linguistics described in this article opens documentary linguistics to non-core linguistic types of language documentation efforts and situates the documentary activities more broadly in the humanistic enterprise of communicating, discussing, studying, and understanding human achievements of other times and places.”

---

---

# 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
-

---

# 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!

I would like to acknowledge the input of Lise Dobrin, Anthony Jukes, Yan Marquis, David Nathan, Julia Sallabank, Candide Simard and other colleagues in discussions which informed the development of this presentation.

But address all criticisms to me!

Email: [pa2@soas.ac.uk](mailto:pa2@soas.ac.uk)

---

---

# References

Austin, Peter K. 2014. Language documentation in the 21<sup>st</sup> 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 Gruyter.

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.

---

## Section 2



A thick black L-shaped frame surrounds the text. It starts at the top left, goes right, then down, then right again at the bottom right.

# CREANDO GRABACIONES DE AUDIO

Ebany Dohle  
SOAS, Universidad de Londres  
21 de Agosto, 2017  
San Miguel Tlaxcala

[e\\_dohle@soas.ac.uk](mailto:e_dohle@soas.ac.uk)

# Outline

- 5 min - Introducción
- 20 min – Audio, Ética y Responsabilidad Social y Manejo de Datos
- 20 min - Video
- 30 min – Trabajo en grupo (grupos de 5), familiarizandose con el equipo
- 45 min – Recopilación de Data
- 15 min – Traspaso de datos
- 45 min – Presentaciones de grupos

# ÉTICA Y RESPONSABILIDAD SOCIAL



# LSA – Sociedad Lingüística de la Americas

- Responsabilidad al consultor individual
  - Responsabilidad a la comunidad
  - Responsabilidad a los estudiantes y colegas
  - Responsabilidad a la beca
  - Responsabilidad a la universidad
  - Responsabilidad al público
- 
- Mayo 2009, LSA  
[http://www.linguisticsociety.org/sites/default/files/Ethics\\_Statement.pdf](http://www.linguisticsociety.org/sites/default/files/Ethics_Statement.pdf)

# Responsabilidad al Consultor

- Respetar los deseos del consultor: Son colegas y colaboradores
- Adquirir permisos para grabar audio, video y fotografiar
- Compensación por el tiempo
- Acceso a los datos recopilados



# Responsabilidad a la Comunidad

- Qué impacto tendrá mi investigación en la comunidad?
- Quién representa la comunidad?
- A quién le pertenece el conocimiento y la cultura?
- Cómo presento mi proyecto a la comunidad?
- Propiedad intelectual



# Responsabilidad a los Colegas

- Reconocer las contribuciones de estudiantes y colegas
- No discriminar
- Asegurar que los estudiantes/colegas que participen estén enterados de la política ética del proyecto de investigación
- Conducta profesional

# Responsabilidad a la Institución

- Respetar los estandares de la instución académica o financiera
- Respetar las oportunidades de futuros colegas e investigadores
- Preservar datos originales y materiales de documentación
- Cumplir con sus compromisos





# Responsabilidad al Público

- Los resultados y las metodologías de las investigaciones deben estar abiertas al público
- Tomar en consideración el impacto de la investigación al público
  - *Daños, malinterpretación, impacto social etc.*

# Proyectos

- Cronolectos – grabar conversaciones entre nahuahablantes en diferentes contextos, luego traducir.
- Rescate de la comidas – acercamiento con familias, recopilación de verbos y sustantivos.
- Morfología de la lengua nahuat - las raíces, sufijos, prefijos etc.
- Censo nahuat – formación de instrument para recolección de datos

# Metodología

- Acercamiento a la comunidad – líderes y lideresas comunales
- Presentar el proposito del Proyecto/visita/conversatorio
- Informar el cuerpo de seguridad de la zona
- Dialogo de respeto mutuo
- Aprender nahuat e incorporar un nahuahablante

# METADATOS Y MANEJO DE DATOS



# Manejo de Datos

**Pictures library**  
Includes: 2 locations

Name	Date	Tags	Size	Rating
102D5000	02/09/2015 08:39			☆☆☆☆☆
20140823_MNL	11/02/2015 02:56			☆☆☆☆☆
20140913_mujerindi...	15/09/2014 11:21			☆☆☆☆☆
20140921_PAW	10/02/2015 05:00			☆☆☆☆☆
20150101-MNL	11/02/2015 02:53			☆☆☆☆☆
20150205_SOAS	11/02/2015 02:52			☆☆☆☆☆
20150210_BSC-BKK	10/02/2015 05:18			☆☆☆☆☆
20150218-JKT	19/03/2015 20:34			☆☆☆☆☆
20150307_SDG	09/03/2015 17:45			☆☆☆☆☆
AF	18/02/2016 18:29			☆☆☆☆☆
Calendar 2015	10/09/2014 16:57			☆☆☆☆☆
Calendar2016	04/11/2015 11:31			☆☆☆☆☆

- Video
- Audio
- Fotografía
- Metadata
- Documentos



# Protocol de Acceso

- ELAR – Archivo de Idiomas en Peligro de Extinción, SOAS, Universidad de Londres
- URCS(X)
  - *Universal*
  - *Investigadores/Academia*
  - *Comunidad Indígena*
  - *Subscripción*
  - *Secreto*

[www.elararchive.org](http://www.elararchive.org)

# CARPETAS Y ARCHIVOS



# Nombres de Archivos

- El nombre de un archivo tiene tres partes
  - *Nombre, punto, y extensión*
  - *tlax20170821 . WAV*
- Crear el nombre del archivo usando un código que represente el proyecto, ejemplo: TLAX
- Agregar número de secuencia
  - *tlax001, tlax002, tlax003*
  - *tlax001-1a, tlax001-1b, tlax001-2a..*



# Nombres de Archivos

- NO utilizar: !@#\$%^&\*()~` ,. ?/ \ | { } []
  - [eb@ny.WAV](#)
  - *tlaxnáhuatl.WAV*
- NO dejar espacio en el nombre
  - *Tlaxcala Nahuatl.WAV*
  - *tlaxcala\_nahuatl.wav*
- Sí usar letras, números, guion y guion bajo
  - *TLAX-001.WAV*
  - *TLAX001\_20160304.mp3*

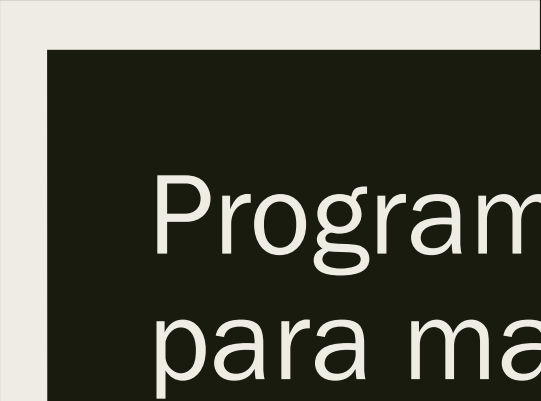
La creación de códigos para nombrar los archivos digitales nos permite agrupar archivos de sesión

- *tlax001.wav, tlax001.jpg, tlax001.pdf*
- *tlax001-a.wav, tlax001-b.wav, tlax001-c.wav*

Manejo de archivos de múltiples investigadores

- *dohle001.wav; gonzalez001.wav*
- *Dohle: nht-tlax001.wav; Gonzalez: nht-tlax501.wav*

## Carpetas y Archivos



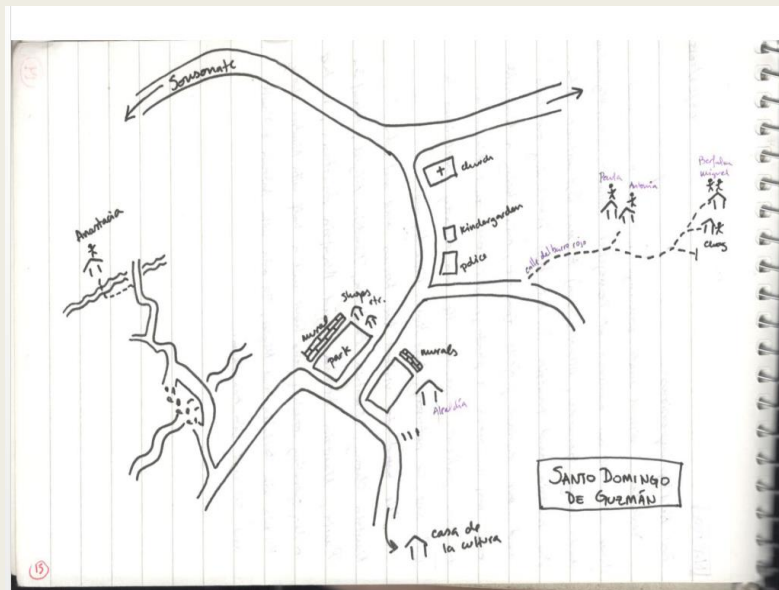
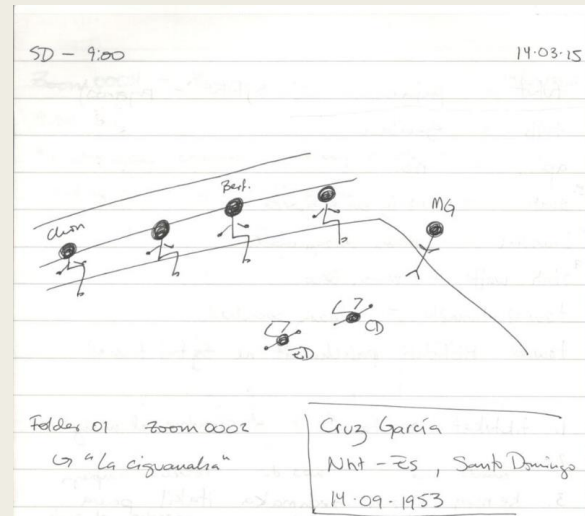
# Programas para manejar datos

- [Karen's Directory printer](#) – crea una lista completa de todos los datos
- [Beyond Compare](#) – para comparar datos de diferentes carpetas/disco duros/computadoras
- [VLC](#) – programa para escuchar audio y video
- [Audacity](#) – edición de audio
- [Arbil](#) – creación y edición de metadata
- [Bulk Rename Utility](#) – renombrar archivos

# Programas de Analisis Lingüístico

- [FLEX](#) – creación de diccionarios léxicos
- [Transcriber](#) – programa de transcripción
- [ELAN](#) – programa de transcripción (difícil)

# Cuadernos de Investigación



paterna - wama  
 pepeto - pepet  
 pepeto peludo -  
 guayaba -  
 cujín - kujé  
 ujushte - ujushte  
 alberja - perume (t)  
 algodón - algodo  
 chichipince - chichipince  
 ↳ huevo para el dolor de estomago

mora → se come con huevo  
 alcapate - n/a  
 zigapate - zigapate [las mujeres lo traen después de dar luz].  
 jilote -  
 mamon - mamon  
 coyol - coyol  
 cojon - teksis kuyamet  
 tempate - tempate  
 maizjillo - 'alboroto'  
 maizy - tauiyal  
 mangana rosa - n/a  
 ulzapote - ulzaput  
 epapote - n/a  
 hesa - n/a  
 rabawo - n/a  
 papa malanga - papa malanga  
 yuca - pauj  
 yame - yamej  
 cacao - chwkwlat  
 paste - paste  
 patashite - n/a  
 nance - nance  
 jicama - shikama  
 amaranto - n/a

[00:13:00]  
 9:45  
 24-05-2015  
 Melba Lopez Ramirez

ajfisa  
 ajwa - ladrar (in)  
 -ajwa - resñar (trans).

pitahaya - n/a  
 tomate de gato - chiripiyu (tumat)

Zoom 0038

tenquique -  
 queso - kesu  
 uva - uwa  
 morro - wajkal / morro  
 jicasta - shikal  
 camote - kesu kamuj  
 pito - pito  
 duffe - chujte  
 flor de jamaica -  
 café - café  
 muggo - n/a  
 sacate - sabat  
 guía de cochinos -  
 cochinito - kuyametchin  
 cincaya - cincaya  
 mejorana - n/a  
 sandía de culbra - No se come  
 ↳ sandía de culbra

# Cuadernos de Investigación

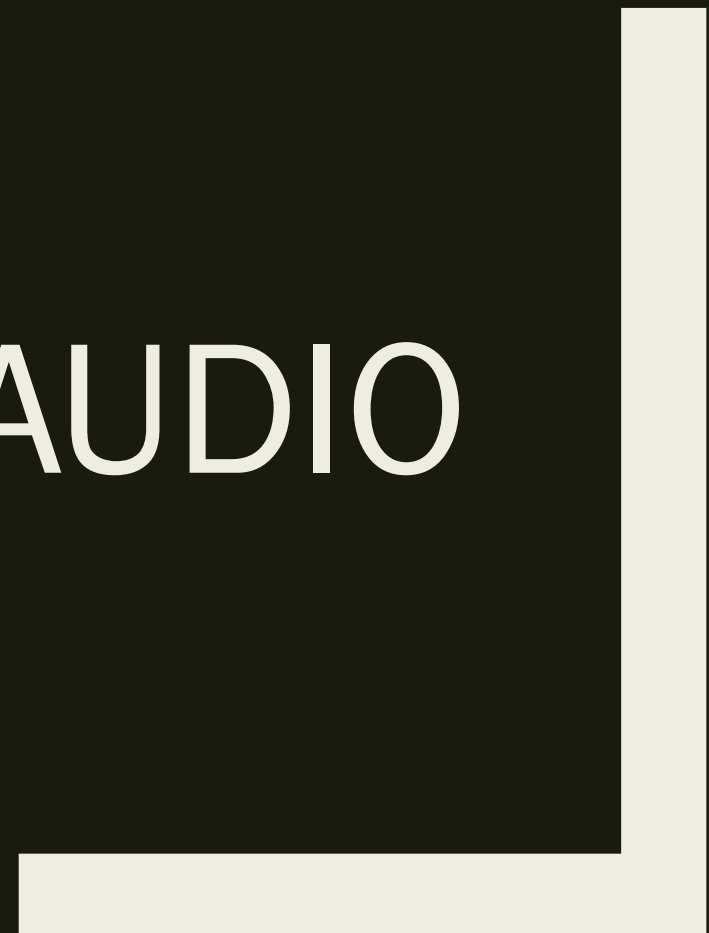
El instrumento mas importante

- *Diário de campo*
- *Recurso primario de datos*

Necesita:

- *Tapa dura o resistente*
- *Tinta indeleble*
- *Números de página*
- *Números de línea*
- *Versión digital*

GRABANDO AUDIO





# Grabadoras

- Marcas de Grabadoras de Audio
  - *Zoom, Edirol, Marantz, telefonos, tablets*
- Marcas de Cámaras de Video
  - *Canon, JVC, Sony, telefonos, tablets*
- Entradas para los micrófonos y audífonos: Mini Jack vs. XLRs
- Formatos: .wav, .mp3, .mp4, formatos propietarios



Mono vs. Stereo

Lavalier

Cardioid

Shotgun/Pistola

Microfonos internos

Tripodes y accesorios

- *Esponja, gato muerto, baterías*

# Micrófonos



# Ambiente

- Viento
- Ruidos ambientales
  - *Pajaros, maquinas, refrigeradoras, radios y música, Iglesias etc.*
- Luz de fondo y reflejo de sonido
- Personas
  - *Uno mismo (comportamiento, manejo de equipo y cables, respuestas a consultores)*
  - *El consultor*



# Objetivos

- Monologo/Dialogo/Conversación/Elicitación
- Posicionamiento
- Nivel de analisis lingüístico
  - *Fonológico vs. social vs. analisis de discurso*
- Modalidad
  - *Hablado vs. dibujado vs. señalado*
- Metadata

# Stereo vs. Mono

- Stereo



- Mono



# Dispositivos vs. Grabadora

- Telefono



- Grabadora



# Viento y Esponja



# Reflejo de Sonido

Superficie sin cobertura



Superficie con cobertura





# MATERIALES AUDIOVISUALES

Principios Básicos

Marta Ostajewska, Artes Liberales, UW  
Tlaxcala, Mexico  
21/08/2017





# Revisar equipo audiovisual

- Crear una lista de equipo requerido y revisar el equipo
- Revisar baterías y tarjetas de memoria
- Asegurar que la cámara y el microfono funcionan
- Revisar y familiarisarse con el equipo creando una grabación de prueba
- Revisar el audio mientras el equipo esté grabando. Revisar posicionamiento de micrófono y el canal de audio.

# ERRORES COMUNES



# Material no grabado

- Siempre es necesario revisar si el boton rojo que empieza a grabar a sido presionado, si la batería ha sido cagada y llevar una tarjeta de memoria extra.

# Rotación de nivel horizontal a vertical

- No filmar con la cámara en posición vertical. Mejor, en posición horizontal – el montaje de los videos será más fácil.

# Imágenes torcidas

- Mira las líneas de composición y buscar nivelarlas, manteniendo la cámara estable utilizando un trípode. El trípode debe ser nivelado antes de empezar a filmar.

# Perspectiva

- Al filmar un edificio, evitar levantar la cámara de video. Es mejor alejarse lo tanto posible o elevarse sobre una silla, mesa, colina etc.



# Luz

- No filmar a una persona bajo la luz directa del sol, enfrente de una Ventana o en un lugar con luz variable. Mejor filmar con un fondo neutro.
- Filmar con luz de día en un sitio bien iluminado
- No combinar luz natural con luz artificial

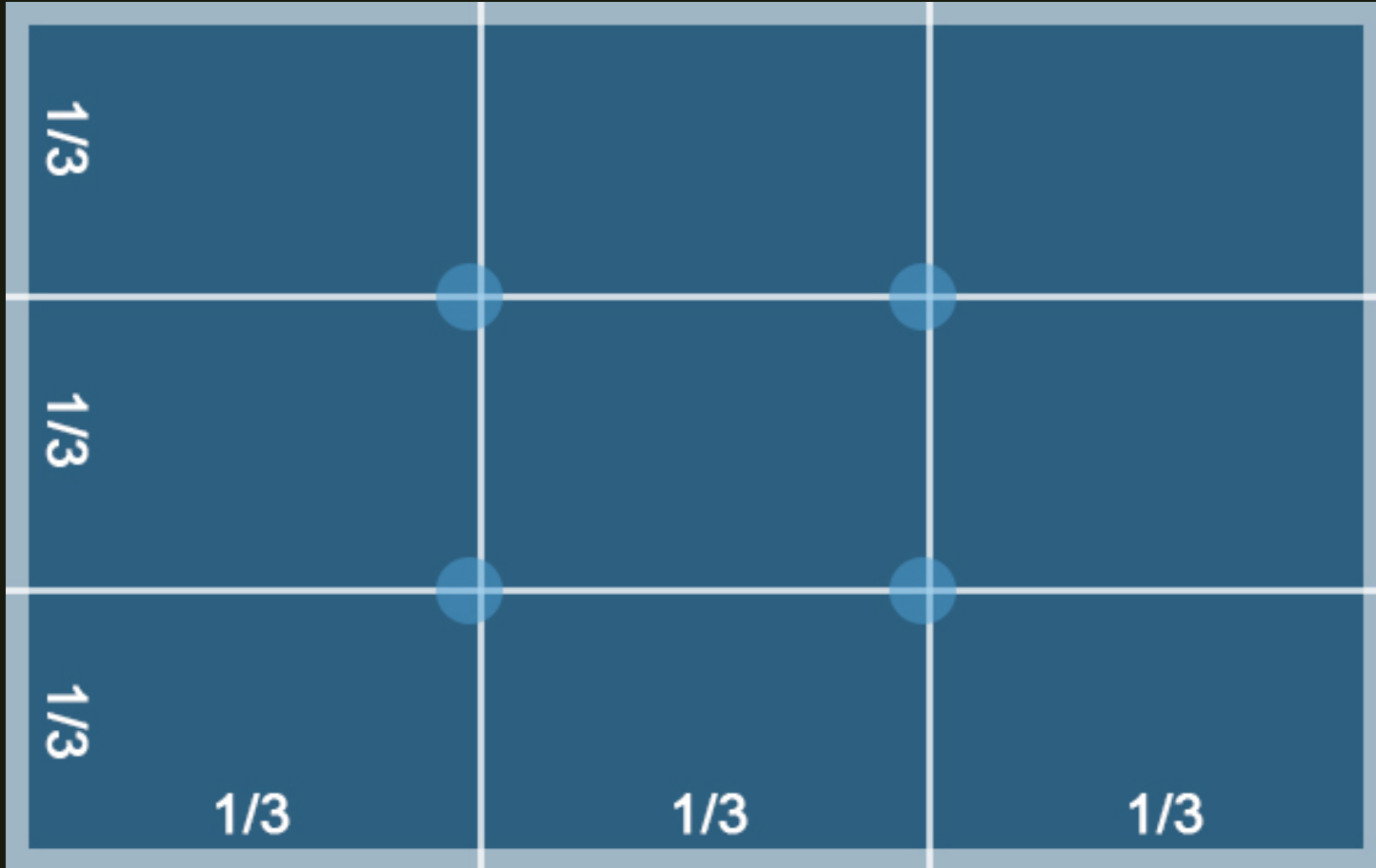
# Movimiento

- Es importante permitir que un escenario se estabilice antes de mover la cámara. Mantener calma y no hacer movimientos bruscos.
- No cambiar el zoom demansiado
- En vez de utilizar el zoom, es mejor acercarse físicamente a la persona objeto siendo filmado.
- Si es necesario hacer zoom, utilizar un trípode y permitir suficiente tiempo para que la imagen se ajuste



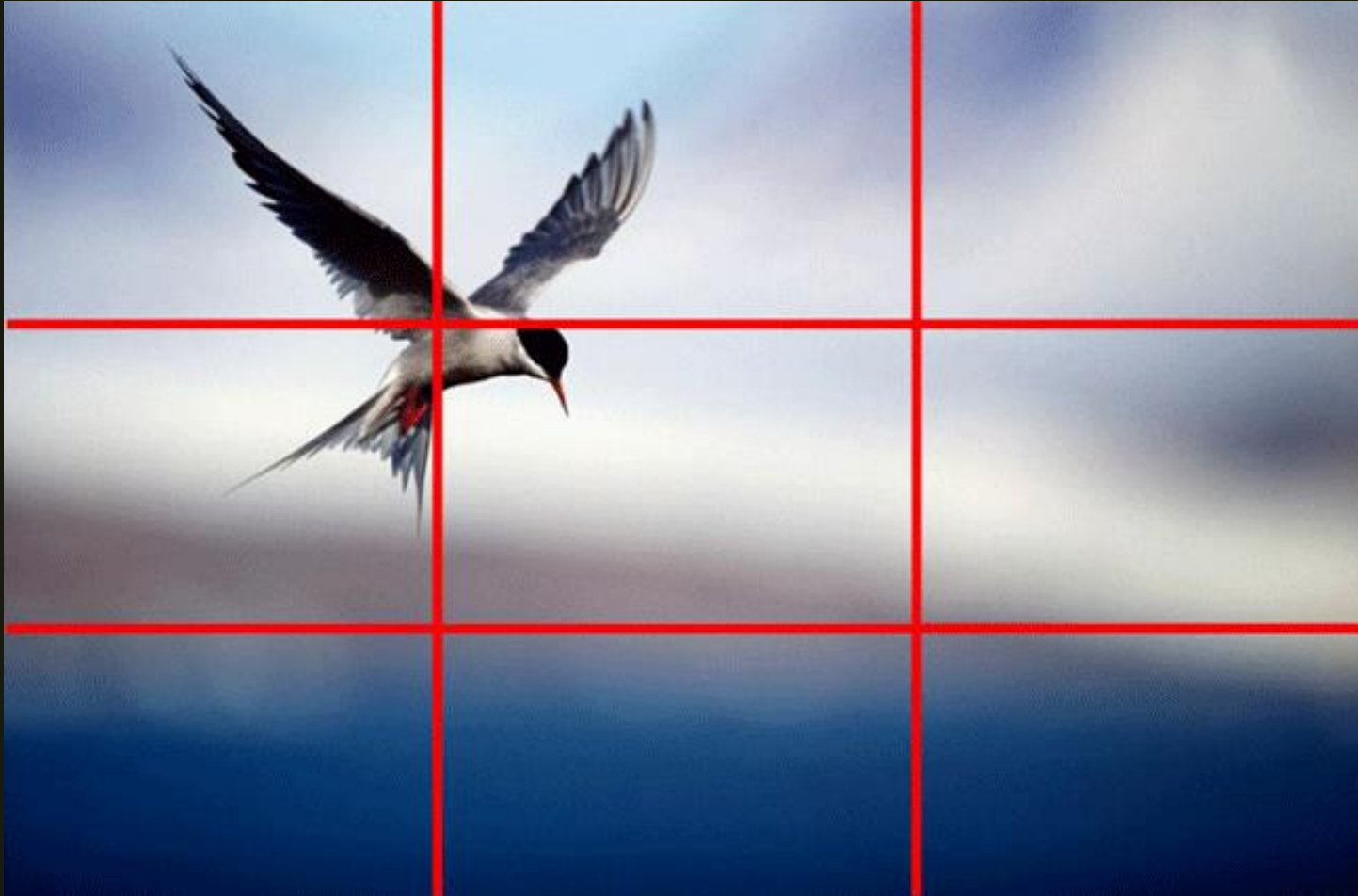
# Enfoque

- Practicar utilizar el enfoque manual, hacienda zoom, enfocando manualmente y desenfocando
- Lo positivo de utilizar el Auto-enfoque
  - *Facil de usar*
  - *Siempre trabaja aún en situaciones inesperadas*
  - *Punto de enfoque particular*
  - *Permite que uno se enfoque en otros aspectos de filmar*
- Lo negativo de utilizar el auto-enfoque
  - *Se pierde el enfoque facilmente al moverse*
  - *Falta de control sobre el*
  - *Problemas con cambios de luz*
  - *Ruido interno*



## Composición

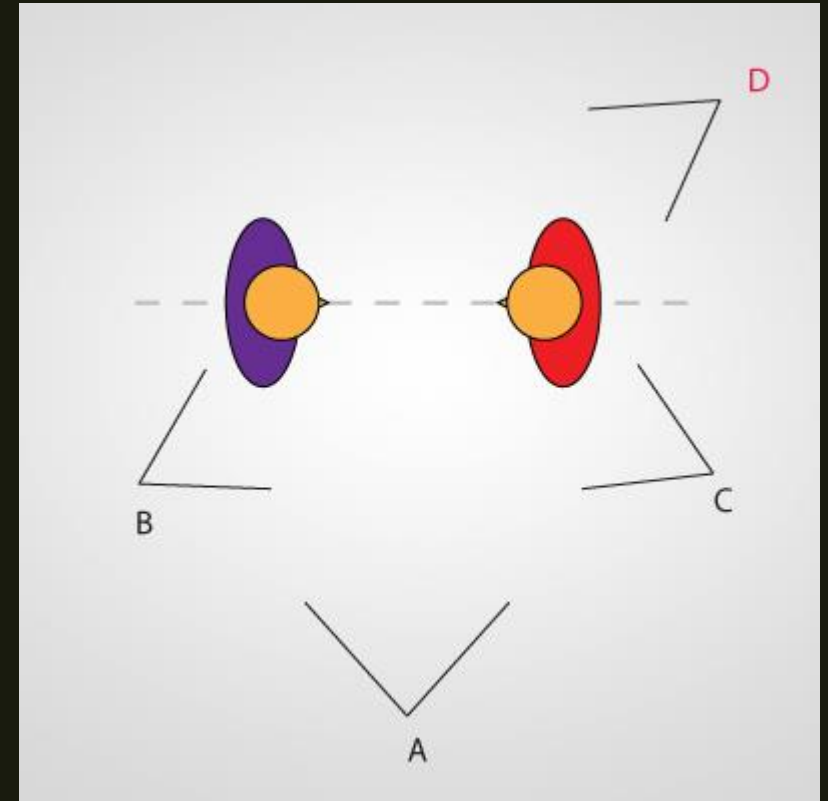
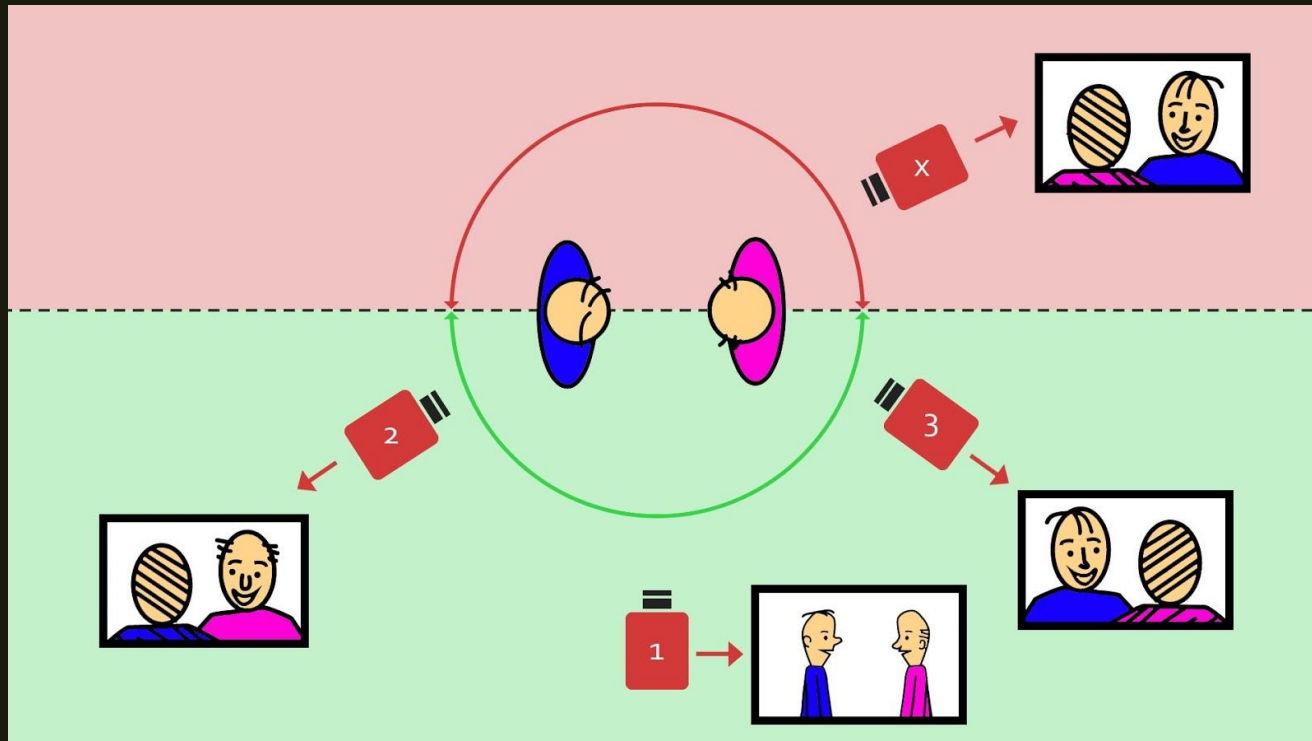
- La regla de tercios

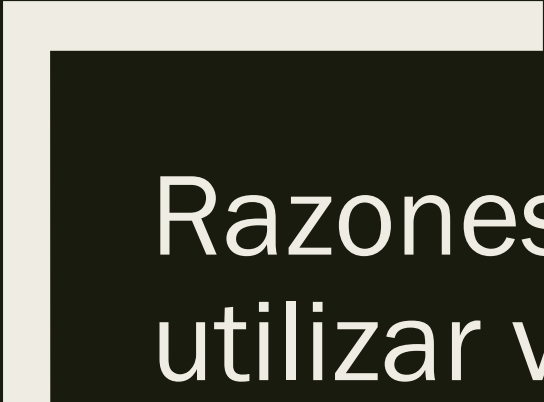


## La regla de los tercios

- Una imagen debe ser dividida en 9 partes iguales por dos líneas horizontales y dos líneas verticales. El enfoque de la imagen debe ser colocada en los cruces de estas líneas.

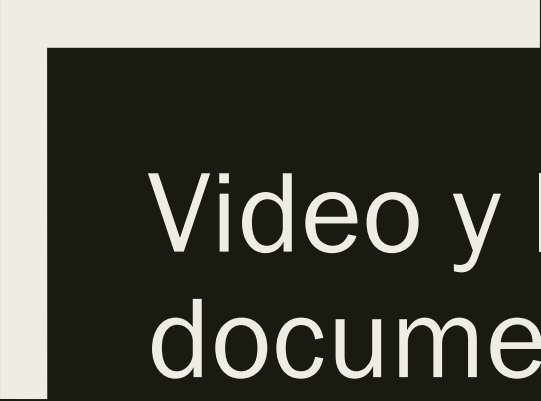
# Regla de 180 grados





# Razones por utilizar video en la documentación de los idiomas

- Los gestos son un elemento importante para la comunicación
- Son importantes para las funciones cognitivas
- Son un element importante de la cultura
- Tener acceso a video y audio facilita la transcripción y asegura que el context del discurso sea colectado



# Video y la documentación de idiomas

- Es importante considerar los permisos para grabar y compartir lo grabado además de la información individual de las personas que participant.
- Es importante conocer su equipo.
- Es importante sentirse bien para que la personas que participant también se sientan bien en el ambiente para grabar. Para lograr esto, practiquen con amigos y familia hasta que logres conocer bien al equipo.

En la video documentación, es importante asegurar que los gestos caben en la toma.



Enmarcación  
mala: no  
decapitar a  
las personas!





Toma amplia



# Acercamiento medio



# Acercamiento y acercamiento extremo

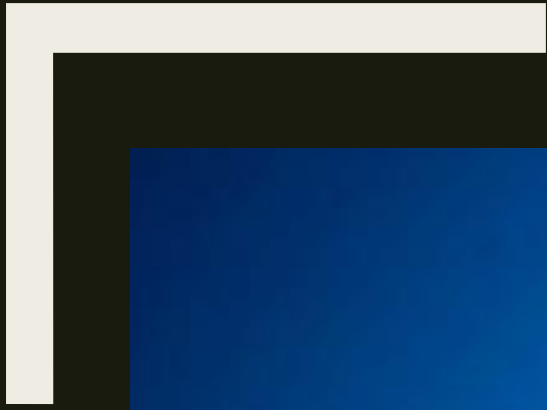


# Contando cuentos – 3 planes de trabajo

- Si vas a filmar un evento, es importante hacer un plan antes de empezar a filmar
- Considerar el area cubierta, los angulos y los detalles









# Imágenes intercortables

- Un imagen que ayuda la transición entre escenas
- Ejemplos: un actor o una actriz hablando, un reloj etc.



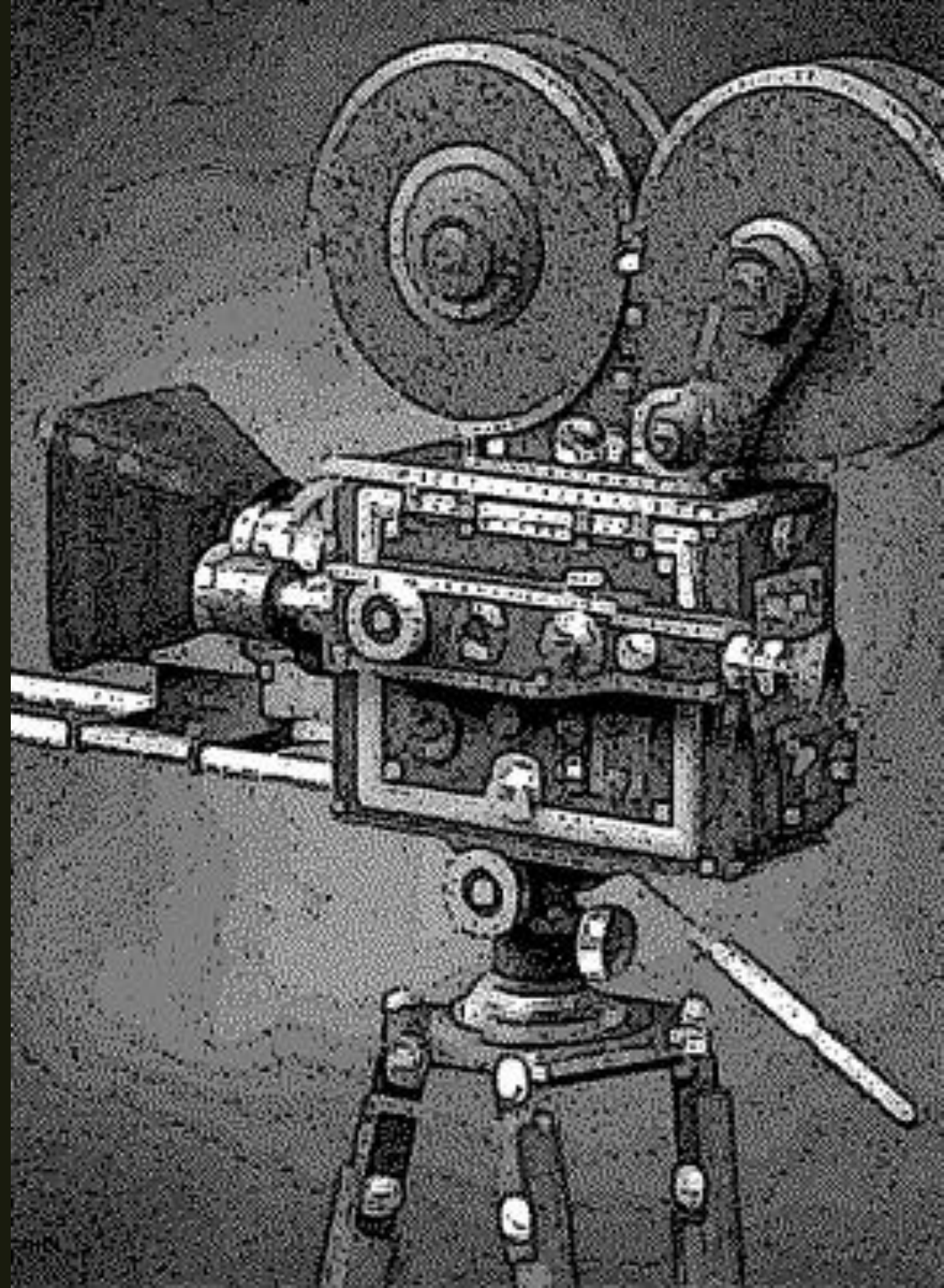


# Problemas de Sonido

- No grabar unicamente con el microfono de la camara de video. Utilizar microfono externo
- No grabar con viento o Lluvia fuerte
- Vestir el microfono para salir
- Evitar sobrecarga



DISFRUTEN!





# The Basic Principles of Creating Audiovisual Material

Marta Ostajewska, Artes Liberales, UW

---

# Check audiovisual equipment

---

- ❖ Make a list of the things that you need and check if you have everything
- ❖ Do you have spare batteries and memory cards?
- ❖ Does the camera and the microphone work?
- ❖ Test an equipment by making a trial recording
- ❖ Check audio during trial recording: is the microphone firmly taped and positioned on the right channel?



# Common mistakes

---

# The material is not recorded

Always check whether the recording button is pressed, whether the battery is charged and if you have a spare memory card

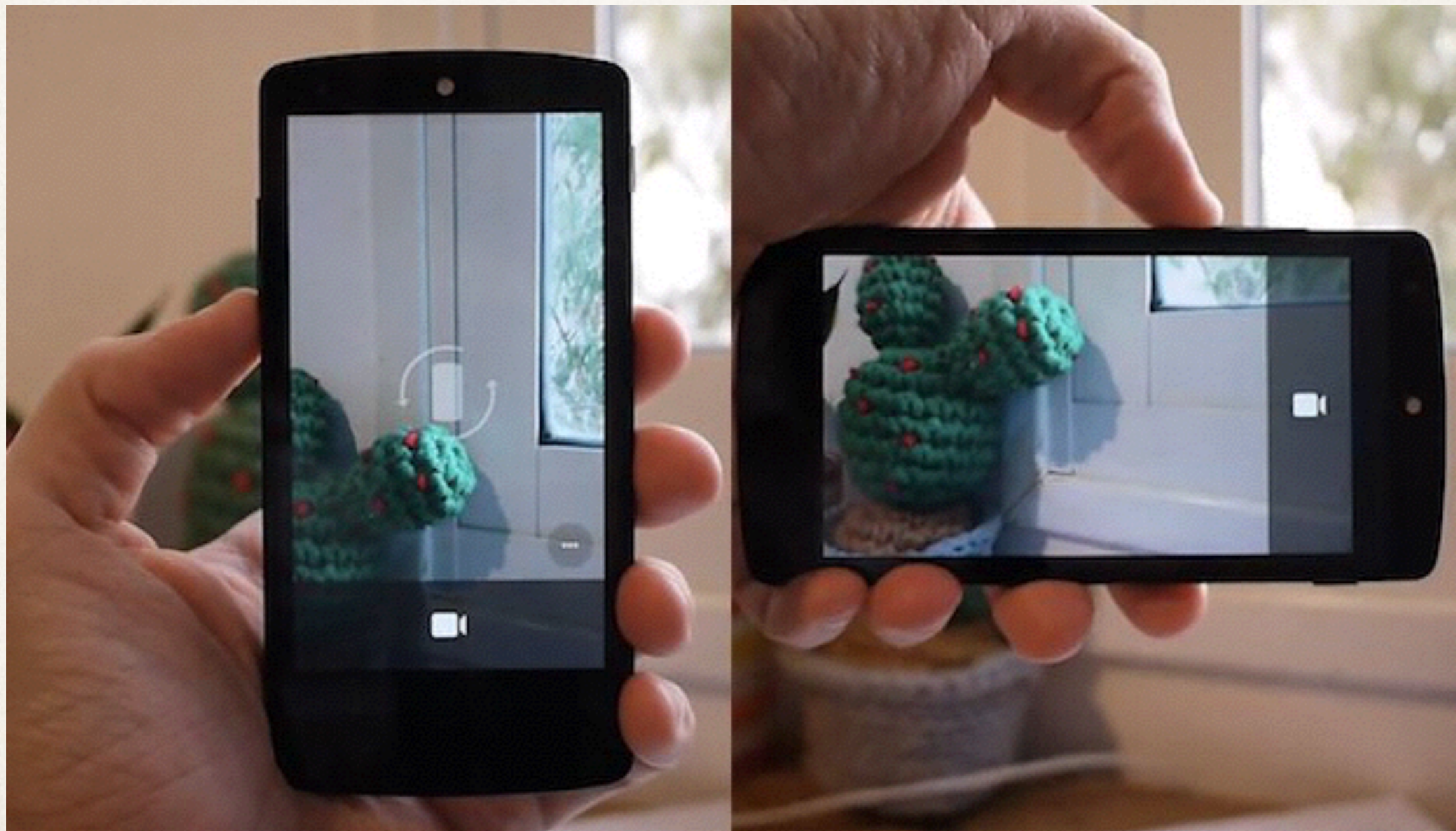
---



# Rotation from horizontal level to vertical level

Do not turn the camera, shoot horizontally, it makes your montage easier

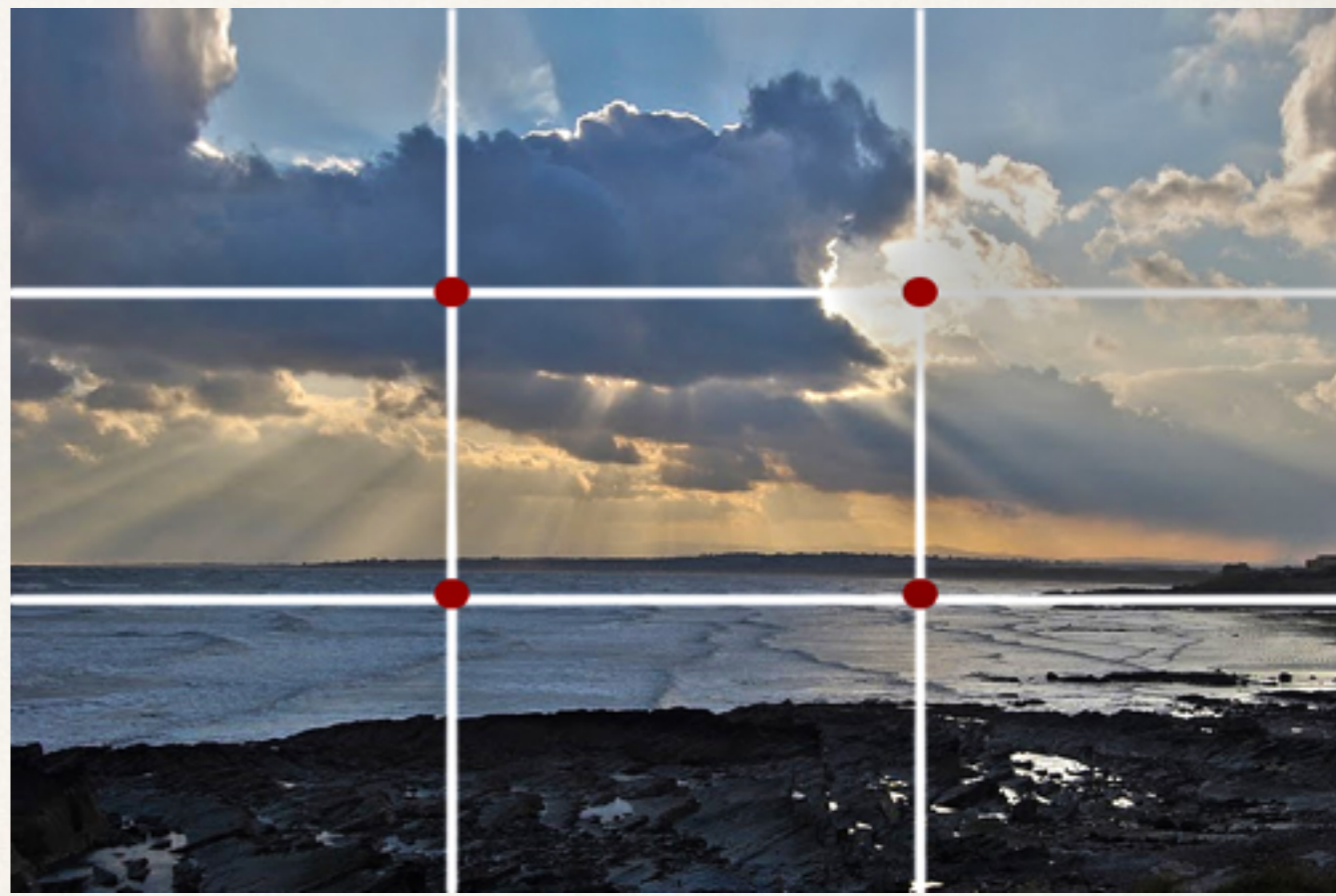
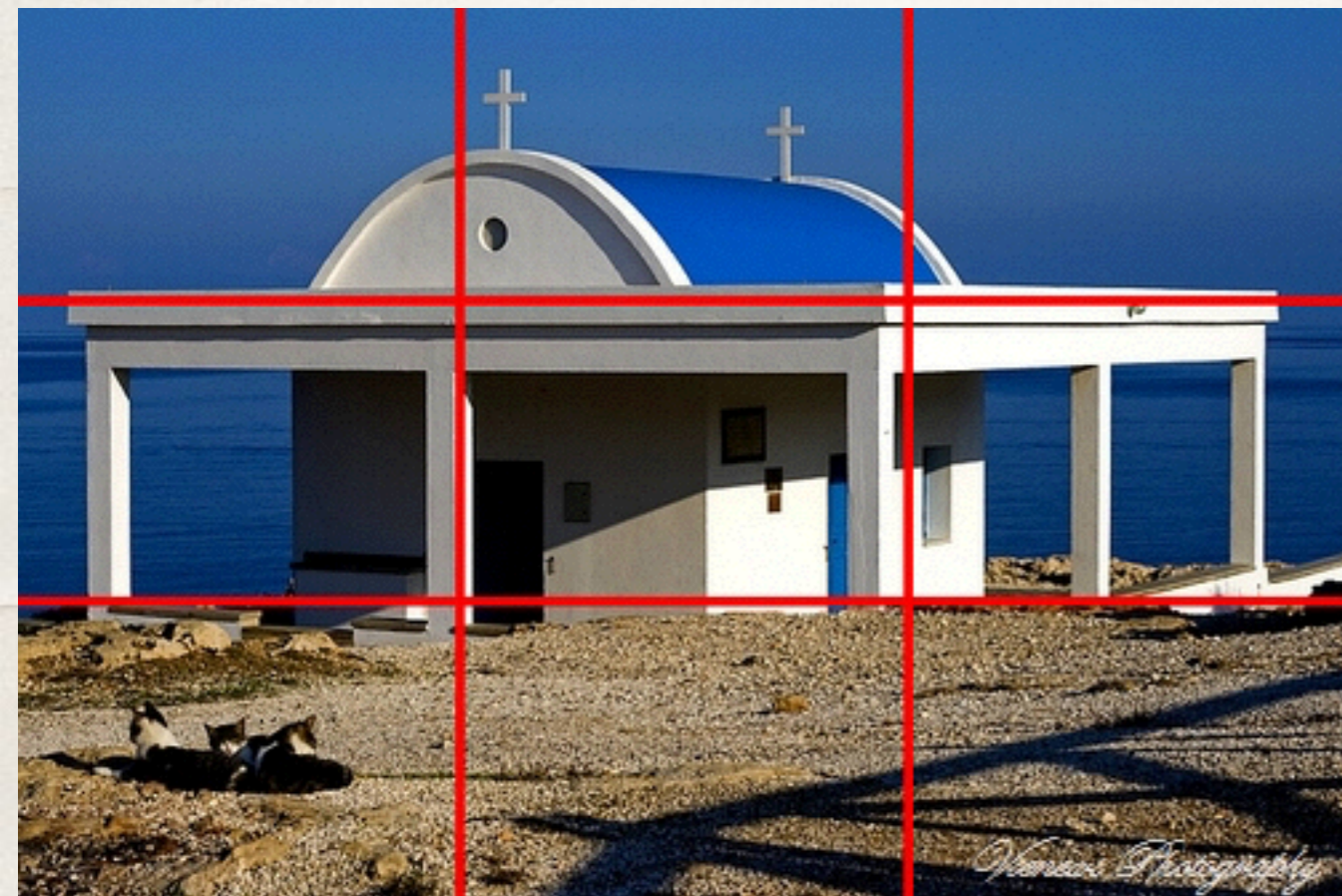
---



# Crooked images

Look at the lines and try to keep the camera even, level to the horizontal lines – frame of the buildings. Always level a tripod before recording.

---



# Shortcuts of the perspective

Do not raise up the camera. Move as far away from the building as possible and zoom in (if you have the opportunity) or – preferably – stand on some elevation (hill, bench, stairs, etc.).

---





# Bad lighting

---

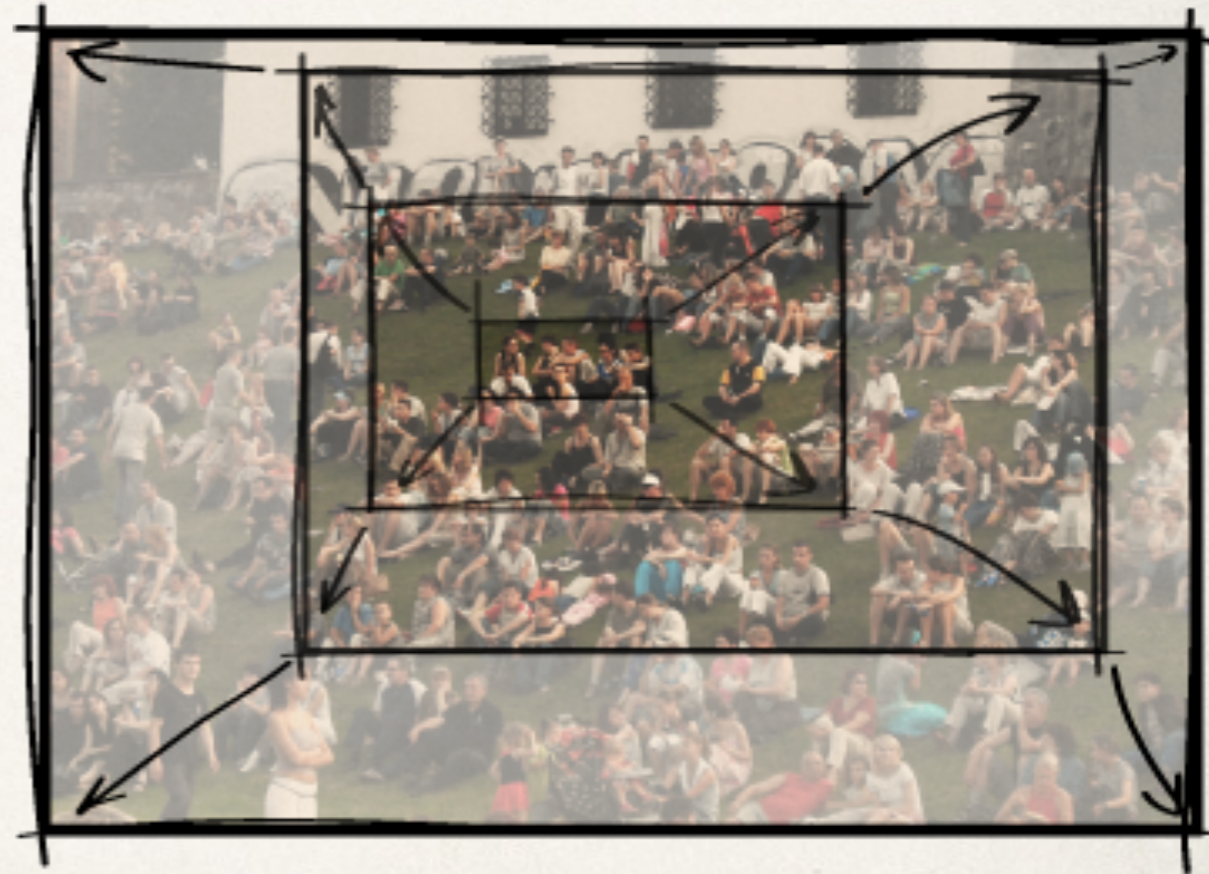
- ❖ Do not film a caller under the sun and against the window, in a place with variable light. Shot a scene preferably on a neutral background
- ❖ Record under daylight in a well lit room
- ❖ Do not mix artificial light with natural light



# Moving to much

---

- ❖ Give a scene time to resound (do not change the position of the camera constantly). Try to be calm during filming
- ❖ **Do not make jumpy zoom in – zoom out**
- ❖ Get closer to the subject with the camera instead of zooming
- ❖ If you zoom in with a camera on the tripod, set a good frame and give the scene a time



# Out of focus

Turn on manual focus, zoom in, sharp the image, zoom out

---

## Positives of autofocus

- ❖ For beginners
- ❖ It works in incalculable conditions
- ❖ It sharpens on a particular element from the frame
- ❖ It allows to focus on other aspects of film - composition, movement, etc.

## Negatives of autofocus

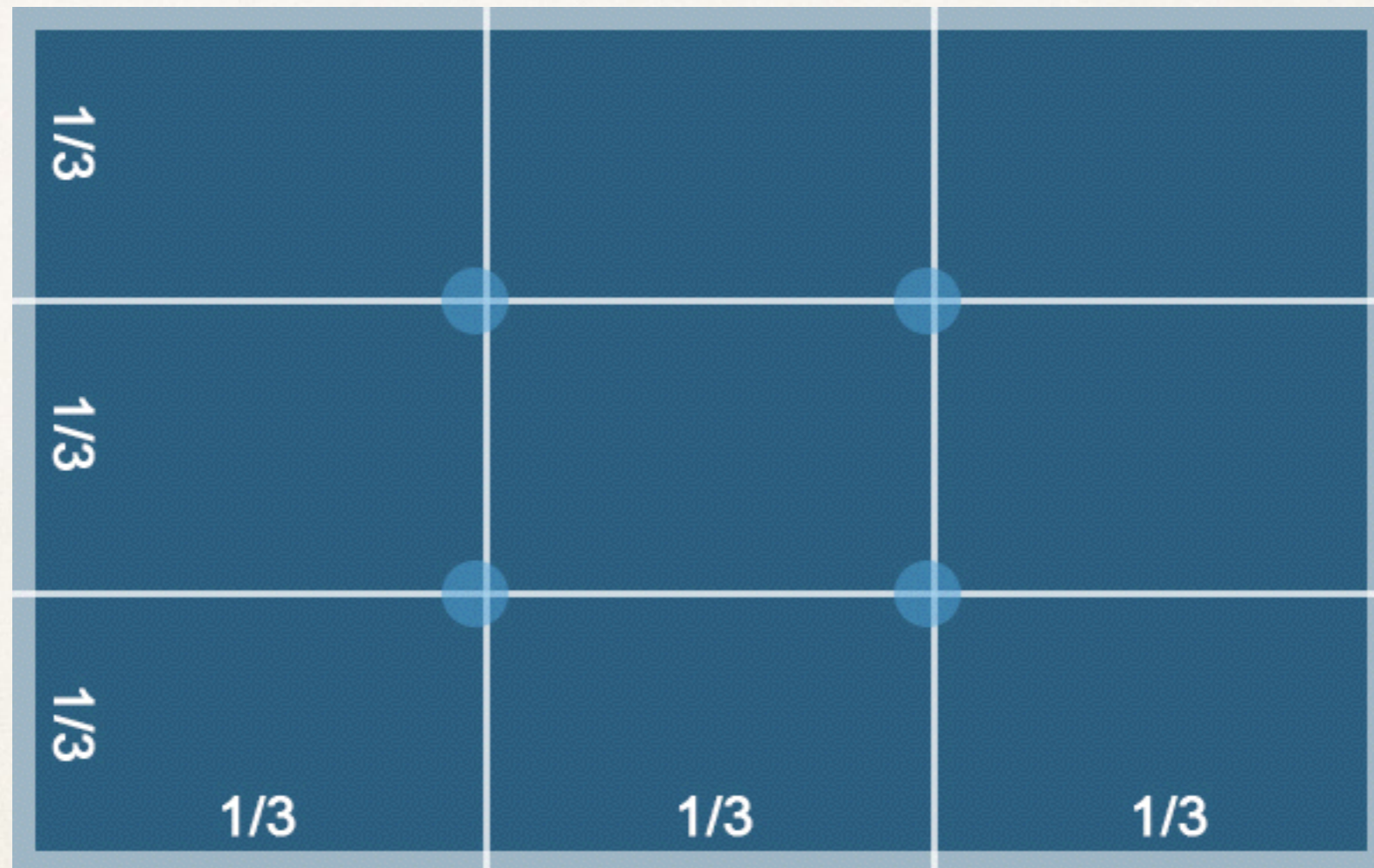
- ❖ Loss of focus when moving in the frame
- ❖ Lack of full control over depth of field
- ❖ Problem with changing lighting, blurring focus
- ❖ Engine noise



# The composition of the picture

Lookroom, Headroom, The Rule of Third

---

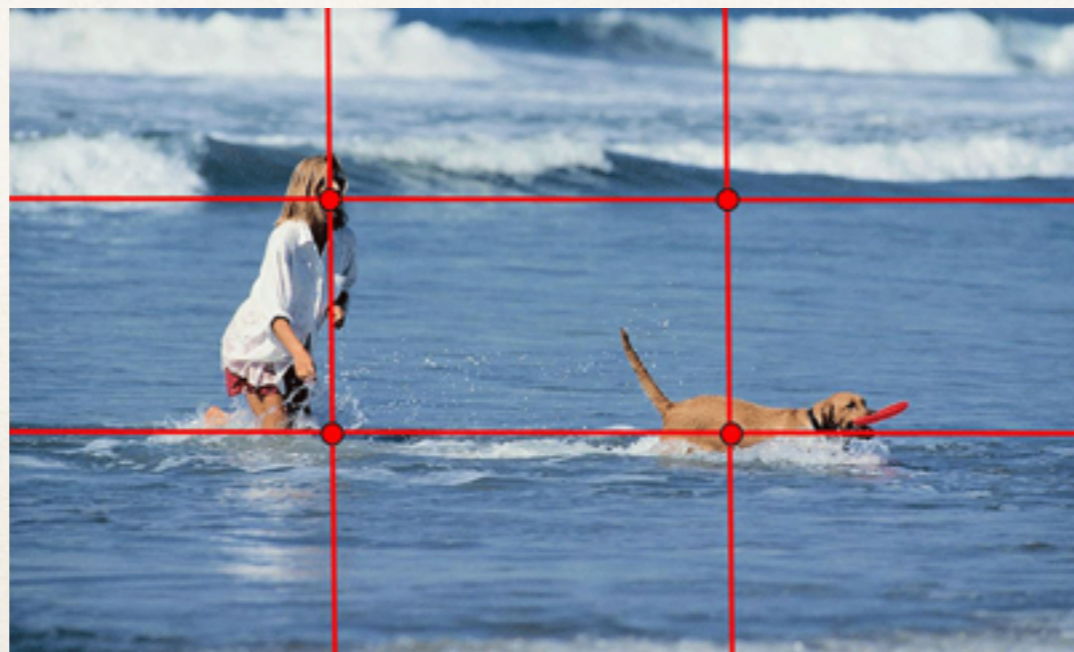
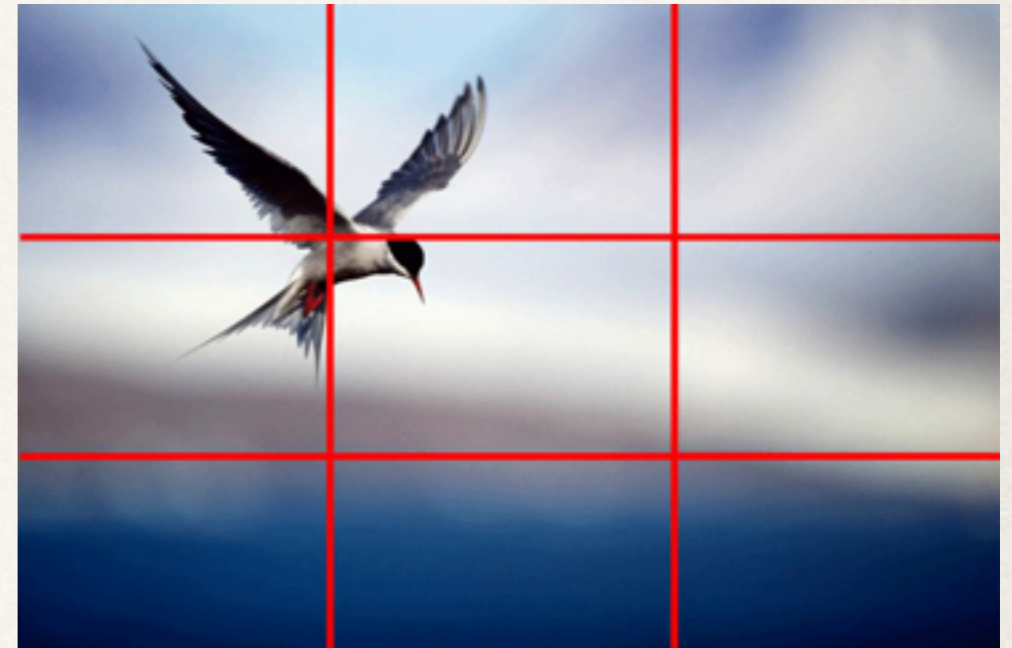
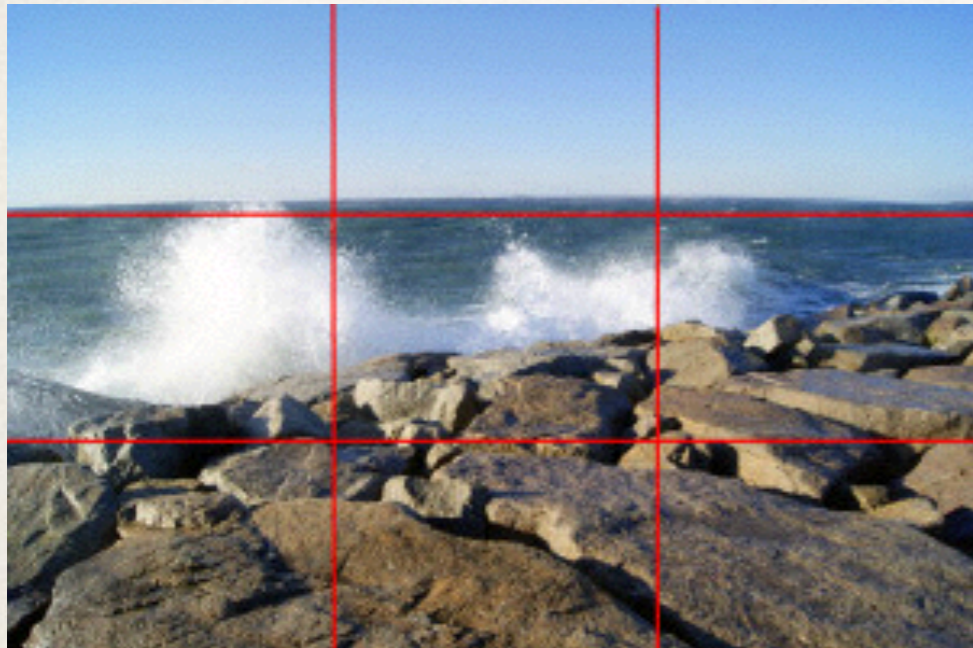


# The Rule of Thirds

An image should be imagined as divided into nine equal parts by two equally spaced horizontal lines and two equally spaced vertical lines, and that important compositional elements should be placed along these lines or their intersections

---

---



# The Rule of Thirds

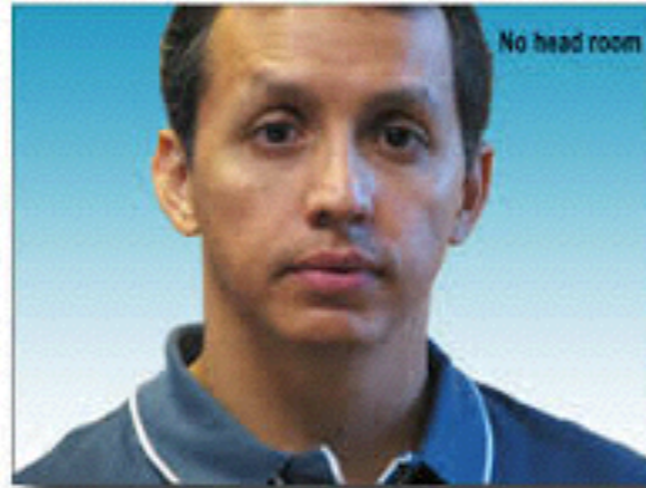
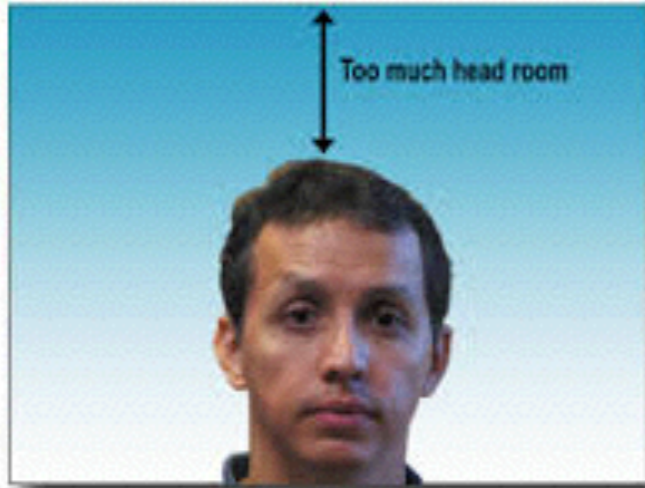
---

## The Rule of Thirds

When shooting a close-up of the face, the eyes are the center of attention : divide the screen into thirds and try to compose your shot so they're at one of the points where the lines intersect.



# Lookroom and Headroom

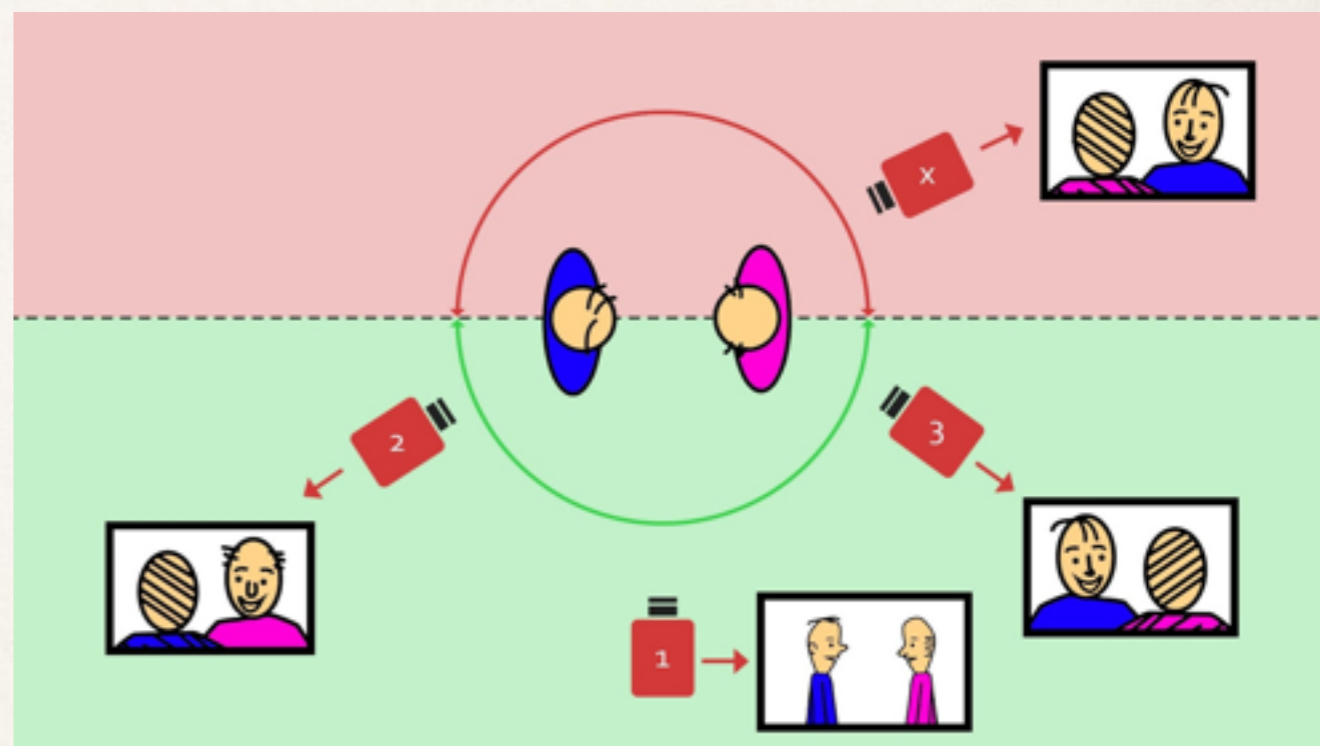
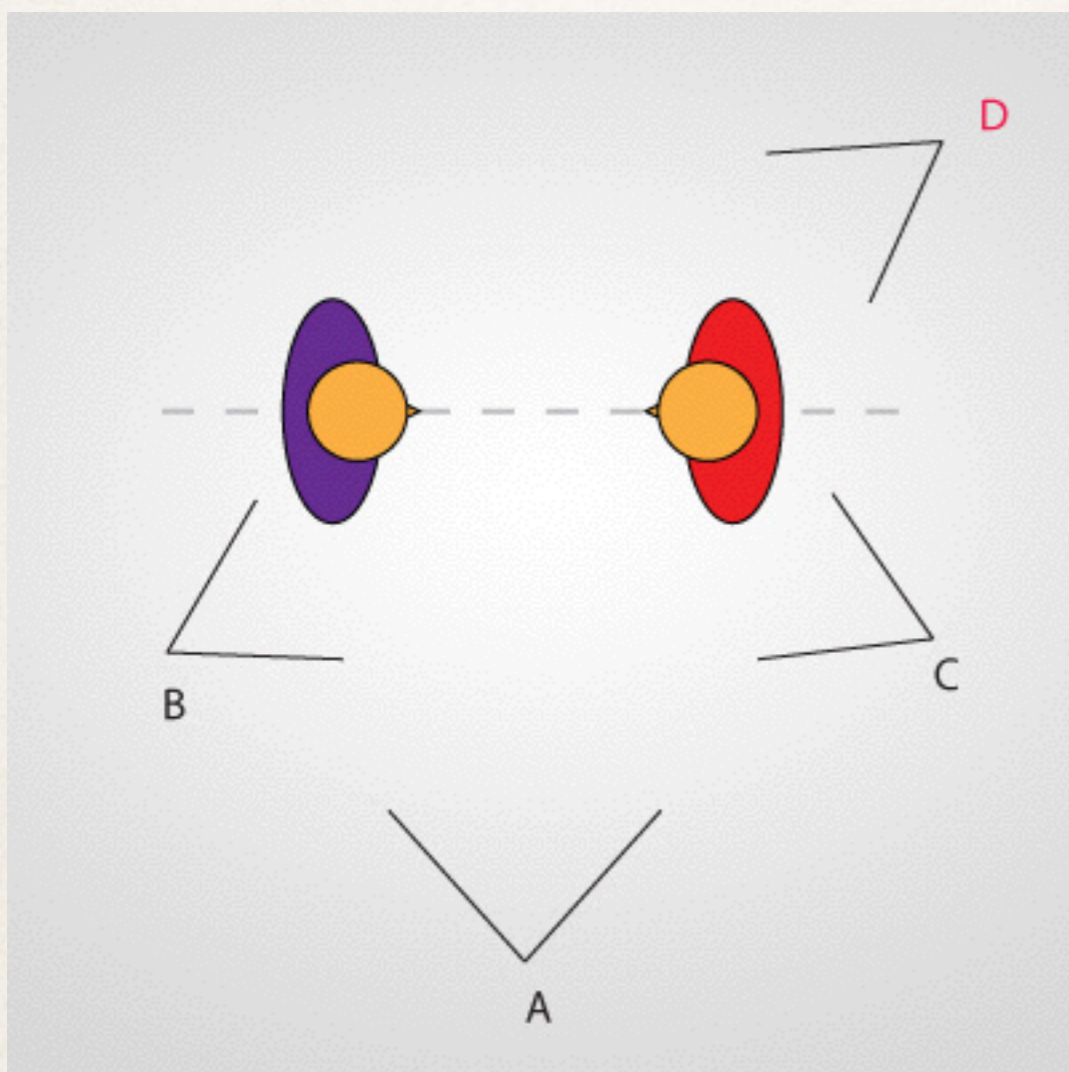


The Lead Room is the space in front of where the actors are looking. In this shot from *Butch Cassidy and the Sundance Kid* (1969), the lead room is on the left side of the frame.

# 180 degree rule

It allows you to record dialogue between characters correctly. According to this principle, we have to lead an invisible line crossing the two recorded callers and simply do not cross it during recording. (A, B, C – correct, D – incorrect)

---





# Reasons you should use video in language documentation

---

- ❖ Gesture is an important part of communication
- ❖ Gesture is an important part of cognition
- ❖ Gesture is an important part of culture
- ❖ Having video as well as audio makes transcription easier, and ensures that the elements of discourse that are not in the spoken channel are still collected

# Video and language documentation

---

- ❖ You have to discuss sharing and permissions with the community, and the individuals you are recording with and have the permission to record
- ❖ Setting up any recording situation is obtrusive
- ❖ Being comfortable with your equipment and making people feel comfortable with your presence mitigates many of those problems. Practice setting up as many times as you can before you begin the project. Record your friends and family. Know your equipment well enough now to continue chatting throughout the setup.

In video documentation of the language it is important that the gestures fit in the frame, do not go beyond the frame, are not cut

---



# Bad framing: cutting people's legs and heads

---

- ❖ Try to be consistent in framing: if you change plan from far to near, from near to far pay attention to where do you stop framing
- ❖ Try not to "cut" the body at the height of the joints (ankles, knees, hips, etc.)
- ❖ When you are filming humans use: Wide shot, Medium shot, Medium close up, Close up, Extreme close up



# Wide shot



# Medium shot



# Medium close up



Close up



Extreme close up



# Story telling (add three plans)

---

- ❖ If you shot an event, except for the shooting the action itself, make an additional three plans
- ❖ Broad plan: A broad plan covering the area
- ❖ Changing angle: Place from another angle, e.g. a view of people watching events
- ❖ Detail: Closeup on smaller objects, details





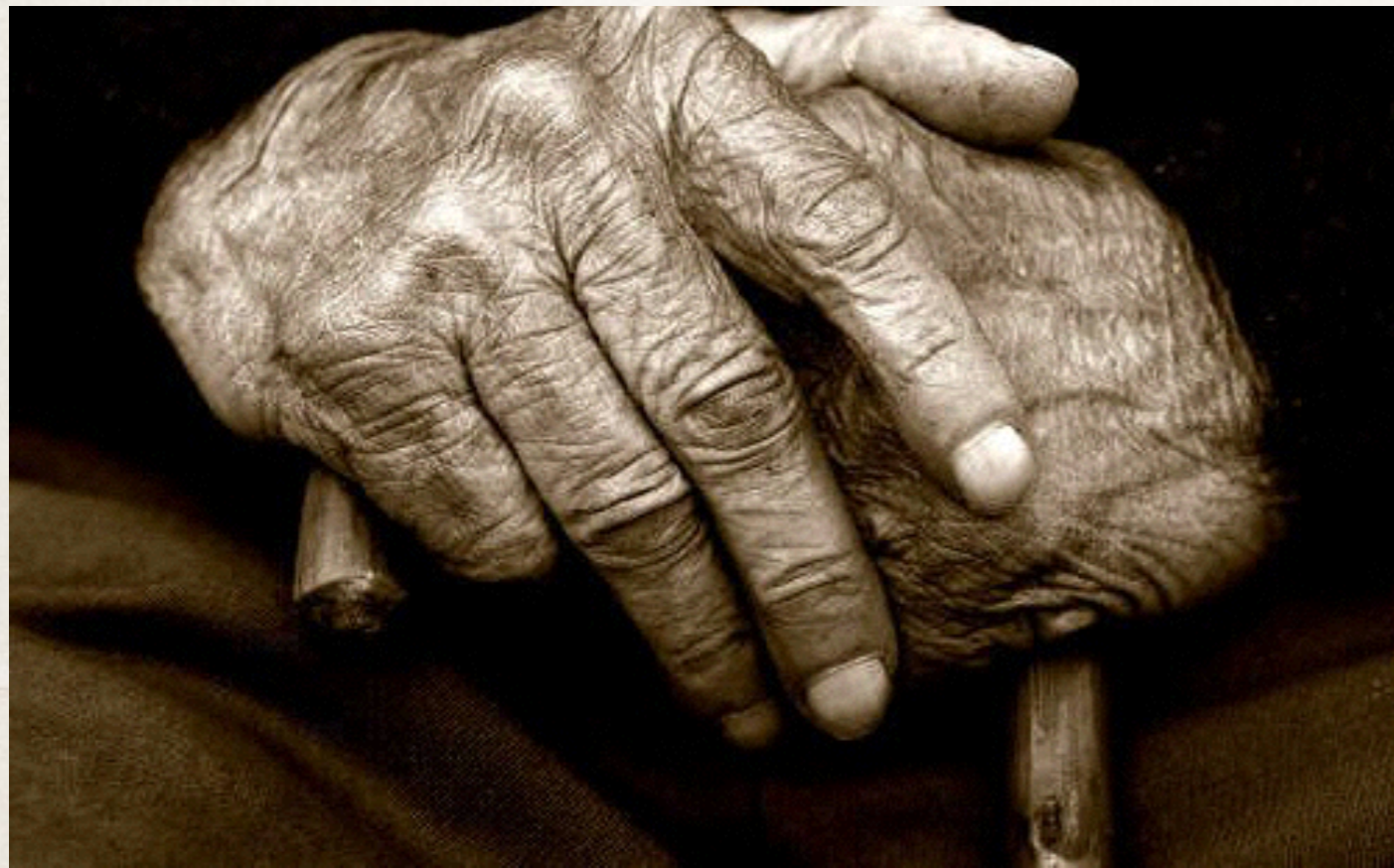




# Inter-cutttable shots

A kind of shot helping to smoothly integrate scenes that do not take part in the narrative mainstream. For example, "a talking actor," "a ticking clock (inter-cutttable shots)," "a talking actress".

---

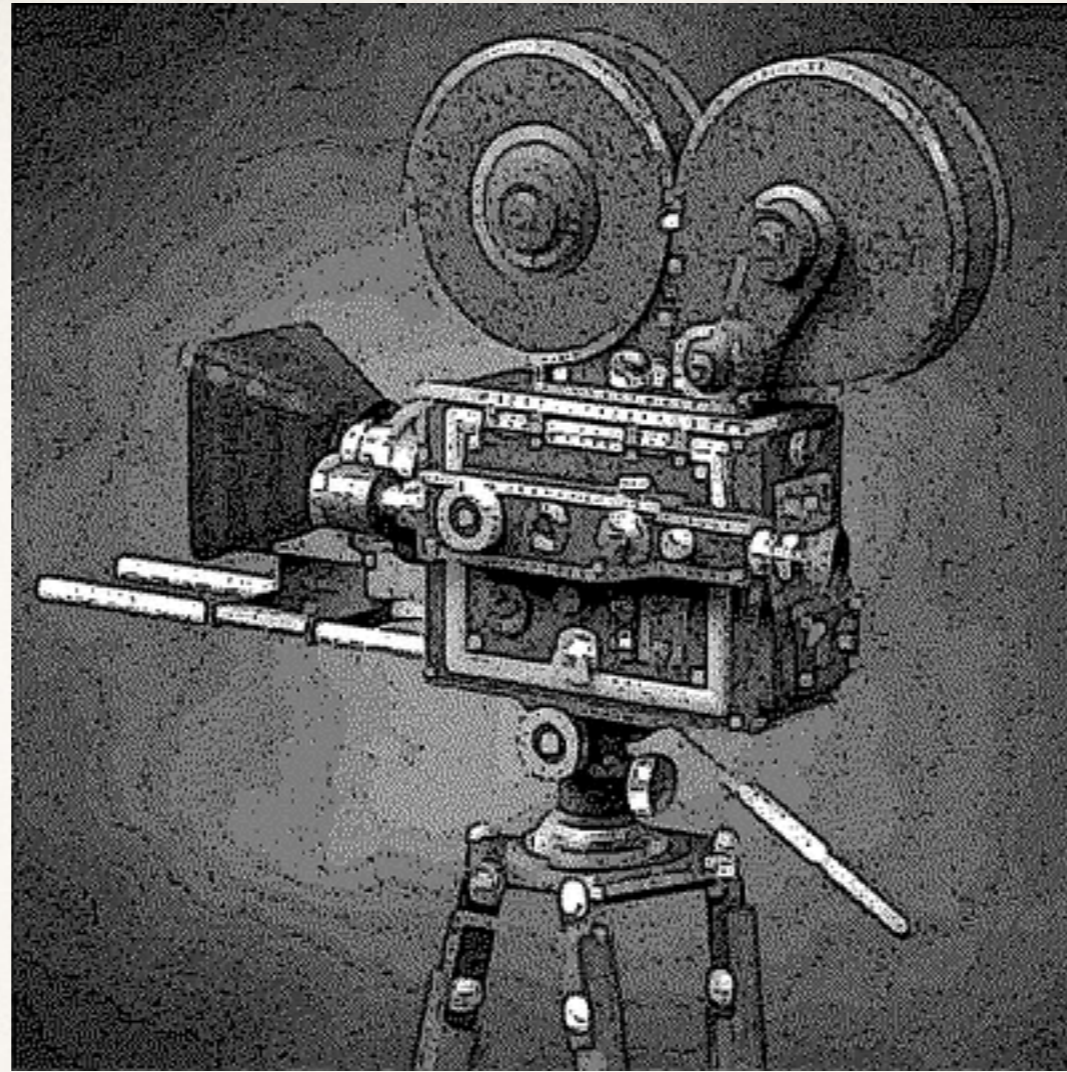


# Sound problems

---

- ❖ Do not record sound only on the camera, have an external microphone
- ❖ Try not to record in the strong wind
- ❖ Dress up your microphone outdoor
- ❖ Avoid overdrive





**ENJOY!**

## Ejercicios

### Grupo 1

Pregunta: Describe tu vestimenta de hoy. ¿Por qué lo elegiste?

Dividir responsabilidades entre el grupo: Entrevistador/a, anotador/a de permisos, metadatos y respuestas, persona responsable por monitorear y grabar audio, persona responsable por monitorear y grabar video, coordinador/a de proyecto, persona responsable por los datos y por crear la presentación, persona entrevistada.

1. Crear dos grabaciones de vídeo, la primera: bajo la luz sin regular el nivel de la cámara. Segundo: con buena luz, nivelada. Ambas grabaciones se producen sin cortes.
2. Anoten la metadatos y permisos de cada video y escriban una descripción corta por cada video.
3. En grupo, discutan lo más fácil y más difícil de este ejercicio.
4. Crear una presentación.

### Grupo 2: Video (2 grabaciones)

Pregunta: ¿Qué te gusta de la comunidad y por qué?

Crea un comercial corto para la comunidad Nahuatl de Tlaxcala.

5. Crear dos videos cortos. El primer video debe ser grabado con plan amplio, de cerca de ancho. Debe tener buenos niveles de luz, buen sonido y tomado sobre un trípode. Ojo a la perspectiva y la composición de la imagen. La segunda grabación de video debe ser grabada con mala luz, mal sonido y mala composición de imagen.
6. Anoten la metadatos y permisos de cada video y escriban una descripción corta por cada video.
7. En grupo, discutan lo más fácil y más difícil de este ejercicio.
8. Crear una presentación.

### Grupo 3: Audio (4 grabaciones)

Pregunta: ¿Cuáles plantas crecen aquí?

Dividir responsabilidades entre el grupo: Entrevistador/a, anotador/a de permisos, metadatos y respuestas, persona responsable por monitorear y grabar, persona responsable por el micrófono, coordinador/a de proyecto, persona responsable por los datos y por crear la presentación, persona entrevistada.

1. Crear dos grabaciones de audio de la mejor posible calidad. Una grabación de audio debe ser grabada adentro de un edificio, la otra afuera en un lugar público.
2. Crear dos grabaciones de audio de la peor calidad posible. Una grabación de audio debe ser creada adentro y la otra afuera en un lugar público.
3. Anotar metadatos, permisos y descripción de cada grabación.
4. En grupo, discutan que encontraron más fácil y más difícil de este ejercicio.
5. Crear una presentación.

#### Grupo 4: Audio y Video (3 grabaciones)

Pregunta: ¿Qué te parecen los perros?

Dividir responsabilidades entre el grupo: Entrevistador/a, anotador/a de permisos, metadatos y respuestas, persona responsable por monitorear y grabar audio, persona responsable por monitorear y grabar video, coordinador/a de proyecto, persona responsable por los datos y por crear la presentación, persona entrevistada.

1. Grabar un video sosteniendo la cámara con la mano. Practicar mantener el enfoque mientras cambiando el zoom.
2. Grabar un video utilizando un trípode. Mantener el video enfocado y no jugar con el zoom.
3. Grabar todos los animales posibles (perros, gatos, pajaros etc.) en la mejor calidad de audio y video posible.
4. Anotar metadatos, permisos y descripción de cada grabación.
5. En grupo, discutan que encontraron más fácil y más difícil de este ejercicio.
6. Crear una presentación.

#### Grupo 5: Audio y Video (3 grabaciones)

Pregunta: ¿Cuántos idiomas aprenden los niños en la escuela y por qué?

Dividir responsabilidades entre el grupo: Entrevistador/a, anotador/a de permisos, metadatos y respuestas, persona responsable por monitorear y grabar audio, persona responsable por monitorear y grabar video, coordinador/a de proyecto, persona responsable por los datos y por crear la presentación, persona entrevistada.

1. Encuentra un lugar perfecto para la grabación, fondo neutro, buena luz del día sin cambio de luz fuerte y crear el primer video. Hacer por lo menos dos tiros inter-cortables (los objetos en el ambiente, registrando la mano del llamador) que se pueden utilizar en el montaje.
2. Encuentra sitio con mala iluminación y fondo neutro y crear el segundo video. Hacer por lo menos dos tiros inter-cortables (los objetos en el ambiente, registrando la mano del llamador) que se pueden utilizar en el montaje.
3. Anotar metadatos, permisos y descripción de cada grabación.
4. En grupo, discutan que encontraron más fácil y más difícil de este ejercicio.
5. Crear una presentación.

#### Grupo 6: Audio y Video (2 grabaciones)

Pregunta: ¿Cuál es tu lengua materna? ¿Cuántas personas hablan este idioma?

Dividir responsabilidades entre el grupo: Entrevistador/a, anotador/a de permisos, metadatos y respuestas, persona responsable por monitorear y grabar audio, persona responsable por monitorear y grabar video, coordinador/a de proyecto, persona responsable por los datos y por crear la presentación, persona entrevistada.

1. Grabar la mejor posible grabación de audio y video utilizando un teléfono móvil.
2. Anotar metadatos, permisos y descripción de cada grabación.
3. En grupo, discutan que encontraron más fácil y más difícil de este ejercicio.
4. Crear una presentación.

### Grupo 7: Audio (12 grabaciones)

Pregunta: ¿Cuál fue tu juguete o juego favorito cuando eras niño/a?

Dividir responsabilidades entre el grupo: Entrevistador/a, anotador/a de permisos, metadatos y respuestas, persona responsable por monitorear y grabar, persona responsable por el micrófono, coordinador/a de proyecto, persona responsable por los datos y por crear la presentación, persona entrevistada.

1. Grabar audio en los siguientes sitios: (1) Adentro (2) afuera (3) lugar ventoso (4) lugar público con tráfico, (5) lugar con superficies duras como vidrio o concreto (6) sitio con superficies suaves y absorbentes como textiles y muebles. En cada una de estas instancias, crear una grabación monitoreada y otra sin.
2. Anotar metadatos, permisos y descripción de cada grabación.
3. En grupo, discutan que encontraron más fácil y más difícil de este ejercicio.
4. Crear una presentación.

### Grupo 8: Audio (4 grabaciones)

Pregunta: ¿Cuál es tu lugar favorito en Tlaxcala? ¿Por qué?

Dividir responsabilidades entre el grupo: Entrevistador/a, anotador/a de permisos, metadatos y respuestas, persona responsable por monitorear y grabar, persona responsable por el micrófono, coordinador/a de proyecto, persona responsable por los datos y por crear la presentación, persona entrevistada.

1. Grabar entrevistas individuales con un niño/a, mujer, hombre y una persona mayor. Audio debe ser grabado afuera en un lugar público.
2. Anotar metadatos, permisos y descripción de cada grabación.
3. En grupo, discutan que encontraron más fácil y más difícil de este ejercicio.
4. Crear una presentación.



## TASKS

### One:

Question: **What are you wearing today and why?**

One person from the team makes an interview, another shots video, third creates an audio, fourth creates a metadata sheet, the fifth coordinates the whole project and makes a presentation, or, if necessary, is an informant

1. Create two video recordings, the first: **under the light without** catching the **camera level**.

Second: with **good light, leveled**. Both recordings are produced without cuts.

2. Create a metadata sheet and movies descriptions.

3. Together with the team think about what was the easiest for you in this task, what was the most difficult.

4. Make a presentation.

### Two:

Question: **Do you love this place and for what?**

Create a short commercial of the settlement.

One person from the team makes an interview, another shots video, third creates an audio, fourth creates a metadata sheet, the fifth coordinates the whole project and makes a presentation, or, if necessary, is an informant.

1. Create two video recordings, the first one: **wide plan, wide close up, close up**. The video should be leveled, with **good light** and **good sound**, on a **tripod**. Watch out for perspective and composition.

The second recording has to be in each of these plans with **poor light** and poor picture and **poorly recorded sound**.

2. Create a metadata sheet and movies descriptions.

3. Together with the team think what was the easiest for you in this task, what was the most difficult.

4. Make a presentation.

### Three (just audio, 4 tracks):

Question: **What kind of plants are growing here?**

One person from the team makes an interview, another holds a microphone, third creates an audio, fourth creates a metadata sheet, the fifth coordinates the whole project and makes a presentation, or, if necessary, is an informant.

1. Create a short audio clip of **the best quality**, in which the informant talks about the types of plants. You record in the middle of **the building and outside**.

2. Create a short audio recording, as **poor quality** as the informant tells about the types of plants. You record in the middle of the building and outside.

3. Create a metadata sheet and movies descriptions.

4. Together with the team think what was the easiest for you in this task, what was the most difficult.

5. Make a presentation.

### Four:

Question: **What do you think about dogs?**

One person from the team makes an interview, another shots video, third creates an audio, fourth creates a metadata sheet, the fifth coordinates the whole project and makes a presentation, or, if necessary, is an informant.

1. Record a video from your hand asking the informants what they think about dogs. **Losing** and catching **focus** by **zoom in and zoom out**

2. Second video **recorded on a tripod, sharp, no zoom in and out, fix pose**.

3. Try to **record as many animals as possible** (primarily dogs, but also birds and cats and other animals) in the best quality.

4. Create a metadata sheet and movies descriptions.
5. Together with the team think what was the easiest for you in this task, what was the most difficult.
6. Make a presentation.

#### **Five:**

Question: **How many languages do children learn in school and why?**

One person from the team makes an interview, another shots video, third creates an audio, fourth creates a metadata sheet, the fifth coordinates the whole project and makes a presentation, or, if necessary, is an informant.

1. Find perfect spot for the recording, **neutral background, good daylight** without strong light changing and create the first video. Make at least **two inter-cutttable shots** (the objects in the environment, recording the caller's hand) that can be used in montage.
2. Find spot with **bad lighting and not neutral background** and create the second video. Make at least **two inter-cutttable shots** (the objects in the environment, recording the caller's hand) that can be used in montage.
3. Create a metadata sheet and movies descriptions.
4. Together with the team think what was the easiest for you in this task, what was the most difficult.
5. Make a presentation.

#### **Six:**

Question: **What is your native language? How many people use this language?**

One person from the team makes an interview, the second and the third shot video and create an audio, fourth creates a metadata sheet, the fifth coordinates the whole project and makes a presentation, or, if necessary, is an informant.

1. Create the best possible audio and video recording using a mobile device e.g. phone.
2. Create a metadata sheet and movies descriptions.
3. Together with the team think what was the easiest for you in this task, what was the most difficult.
4. Make a presentation.

#### **Seven (just audio, 6 tracks):**

Question: **What was your favorite game/toy when you were a kid?**

One person from the team makes an interview, another holds a microphone, third creates an audio, fourth creates a metadata sheet, the fifth coordinates the whole project and makes a presentation, or, if necessary, is an informant.

1. Create audio recordings in the following locations: (1) indoor, (2) outdoor, (3) windy, (4) crowded areas, (5) locations with hard surfaces like concrete or glass, (6) locations with absorbent surfaces like textiles and furniture. In each of these create a monitored recording and one with no monitoring.
2. Create a metadata sheet.
3. Together with the team think what was the easiest for you in this task, what was the most difficult.
4. Make a presentation.

#### **Eight (just audio, 4 tracks)**

Question: **What is the most beautiful in the area and what we definitely should see?**

One person from the team makes an interview, another holds a microphone, third creates an audio, fourth creates a metadata sheet, the fifth coordinates the whole project and makes a presentation, or, if necessary, is an informant.

1. Conduct a one to one interview with three different people (child, woman, man, senior), outdoors in a crowded area.

One person from the team makes an interview, another holds a microphone, third creates an audio, fourth creates a metadata sheet, the fifth coordinates the whole project and makes a presentation, or, if necessary, is an informant.

2. Create a metadata sheet.

3. Together with the team think what was the easiest for you in this task, what was the most difficult.
4. Make a presentation.

---

# ENGHUM Summer School

## Software Tools

Peter K. Austin

Department of Linguistics  
SOAS, University of London

2016-06-14

---

---

© 2016 Peter K. Austin

Creative commons licence

Attribution-NonCommercial-NoDerivs

CC BY-NC-ND

---

---

With thanks to ...

Stuart McGill, Anthony Jukes and Candide Simard who all contributed to the development of these materials for various training courses

---

---

## 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 SayMore are three that linguists are using these days (NB. Praat can be used as a transcription tool also; eXMaraLda is also being adopted in Germany)

---

---

# 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 Toolbox and FLEX for interlinearisation, then reimported)
  - <https://tla.mpi.nl/tools/tla-tools/elan/>
-



---

# 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 untrained non-researchers
  - 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

Node	Wioi014_0001	Wioi014_0002
ref@BH [11]		
t@BH [11]	jadi, ore' to a:...	makawengke, oras makawengke.
fe@BH [11]	so, if the one who	will get married, when they're going to
fn@BH [11]	jadi, kalau yang	mokaweng, diwaktu mokaweng
w@BH [33]	jadi,   ore'   to   aa	makawengke,   oras   makawengke.
m@BH [49]	jadi   ore'   to   aa	ma- kaw Ce   oras ma- kaw Ce
g@BH [49]	so   or   NR   um	AV- marr CPL time AV- marr CPL
p@BH [49]	conj   conj   prt   int	pref- v   prt   n   pref v   prt

---

# 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

## PLUS

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

## MINUS

- 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>
-

Transcriber 1.5.1

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

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

## PLUS

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

## MINUS

- 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
-



---

# 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 International

---

---

# 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

## PLUS

- 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 LexiquePro

## MINUS

- Standard Format (backslash codes) not really well-structured
  - ‘End of life’? It is very old, not being developed actively
  - Limited interaction with media files
  - Macintosh 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.
-



**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
-	dividir
<b>gaya</b>	vaidac ser vai
<b>goa</b>	vale
<b>goce</b>	farelo
<b>godama</b>	joelha
<b>godamira</b>	joelha
<b>gogodo</b>	osso
<b>gogoma</b>	ajoelh
<b>gombe</b>	marge
<b>gona</b>	dormi
<b>gopa</b>	ter me
<b>gubudza</b>	sacudi
<b>gula</b>	comp
	vanda

**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

Kalaba - Ls-hovland - FieldWorks Language Explorer

File Edit View Data Insert Format Tools Parser Window Help

English

**Texts**

Edit  
Document

**Texts**

Title  
Show All  
My Green Mat

**Text**

My Green Mat

This is a simple sentence about seeing a green mat.

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

.

Free: I see my green mat.

Queue: [-/~/] No Parser Loaded

1/1

---

# FLEx plus and minus

## PLUS

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

## MINUS

- Poor handling of media
  - Large application, memory hog
  - Windows only
  - Poor integration with Toolbox
  - All data and analysis in one file so corruption risk
-

---

## 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
-





bth



**abit ab'it** water carrying basket

abit

abo-abo

aboh

abur

abus

abūs

abut

abūt

abūt kupong

adat

addi

adis

adoh

adoi

adu

adu-adu

adud

adūd

**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

## PLUS

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

## MINUS

- 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 (C)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 starting to look solid
-

---

# Demonstration of software

- ELAN
  - Toolbox
  - Lexique Pro
  - Fieldworks Language Explorer (FLEEx)
  - WeSay
  - SayMore
-

---

# ENGHUM Summer School

## Data management, metadata and mobilisation

Peter K. Austin

Department of Linguistics  
SOAS, University of London

2016-06-16

---

---

© 2016 Peter K. Austin

Creative commons licence

Attribution-NonCommercial-NoDerivs

CC BY-NC-ND

---

---

With thanks to ...

David Nathan and Anthony Jukes who  
contributed to the development of these  
materials for various training courses

---



---

# 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” (Himmelman 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*
-

---

# 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

 Endangered Languages Archive Not logged in. [Login](#) | [New user](#) | [Search](#) | [Home](#)

## Search

Found 46 resources in the archive (page 1 of 6)

[1](#) [2](#) [3](#) [4](#) [5](#) [6](#) [next >](#) [last >>](#)

### Search ELAR

[Reset keywords](#)

#### Language

- Avatime (Dominant) (5)
- Chinese (1)
- Cicipu (5)
- English (6)
- Gurindji Kriol (7)
- Gurindji (2)
- Kriol (2)
- Nar (2)
- Nepali (4)
- Northern Arapaho (1)
- Phu (1)
- Sadu (1)
- Spanish (4)
- Yakkha (4)
- Zapotec (4)

#### Type

- Audio (31)
- Deposit (3)
- Document (4)
- ELAN (17)
- Image (2)
- Text (4)
- Video (10)
- Zipped collection (1)

- **Text documentation of N|uu**  
*Tom Güldemann*  
... 22 hours of audio and 10 hours of video material of the last 10 known speakers of the Nuu language. Texts include folktales, personal stories, and conversations.

---

- **Cicipu documentation**  
*Stuart McGill*  
... folktales, riddles, historical narratives, interviews, songs, prayers, and sermons; 9 Pear Film narratives and six hours of interlinearised time-aligned texts.

---

- **Documentation and grammatical description of Yakkha, Nepal**  
*Diana Schackow*  
... documenting Yakkha, an endangered language of Nepal

---

- ▶ **FM09\_a115 (Documenting Gurindji Kriol, an Australian mixed language)**

#### 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:

- Deposit
- Bundle
- Person



## Search

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

[1](#) [2](#) [3](#) [4](#) [5](#) [6](#) [7](#) [8](#) [9](#) ... [next >](#) [last »](#)

### Search ELAR

[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)  
 Koyl 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, Watungene 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:

# Corpus accessibility – I found it, what now?

**Cicipu documentation**

Home Resources

Found 60 bundles in this deposit with keyword **ELAN x** (page 1 of 8)


1 2 3 4 5 6 7 8 next > last »

▼ Discussion of chieftancy

svgd001.eaf Access protocol: [URCS](#)

Download

svgd001.001.mpg Access protocol: [URCS](#)



00:05 00:17

Download

**Search this deposit**

Search

[Reset keywords](#)

**Access protocol**

[URCS](#) (60)

**Language** *more* ▼

Cicipu (58)  
Tidipo (5)  
Tikula (3)  
Tirisino (6)  
Damakawa  
*more* ...

**Type**

**ELAN x**  
Audio (60)  
Image (8)  
Transcriber (3)  
Video (10)  
Document  
Text  
XML  
Zipped collection

**Tags**

Kezzeme (2)  
Photo  
Photos


**Genre** *more* ▼

**Deposit status**

✓ **Curated:**  
Resources online and curated

**Depositor**

Stuart McGill




**Nationality:** UK  
**Affiliation:** School of Oriental and African Studies

**Your access**

Your roles: [URCS](#)

**Tools**

[Download metadata](#)  
[Add to My Bookmarks](#)



Map

Soain

Greece

Tunisia

Morocco

Algeria

Libya

estern Sahara

auritania

# Corpus accessibility – I can't even find it

The screenshot displays the interface of 'The Language Archive'. On the left is a hierarchical tree view of corpora. The main area on the right shows search results for a specific session.

**Search Interface:** The top navigation bar includes 'The Language Archive' logo, a home icon, and links for 'about', 'manual', 'register', 'user: anonymous', and 'Log In'. Below this are search and action buttons: 'METADATA SEARCH', 'CONTENT SEARCH', 'MANAGE ACCESS', 'REQUEST ACCESS', 'CITATION', 'DOWNLOAD ALL', and 'VERSION INFO'. A 'BOOKMARK' button is also present.

**Left Panel (Tree View):** A list of corpora with expandable folders. The 'DoBeS archive' is expanded, showing sub-folders like 'Aché', 'Akie', 'Aru languages', 'Aweti', 'Bainouk', 'Bainouk Gubeeher', 'Bainouk Gubeeher', 'People', 'Plants', 'Elicitations', 'Tree list', 'Interviews', 'Photos', and 'Pots'. Under 'Tree list', three files are listed: 'DJI10312CDD', 'DJI151112AC', and 'DJI200212CDD1-2'. The file 'DJI10312CDD' is highlighted.

**Main Panel (Session Details):** The details for session 'DJI10312CDD' are shown. The 'Name' is 'DJI10312CDD', the 'Title' is 'Tree list', and the 'Date' is '2012-03-01'. The 'Description' is 'Verification of the pronunciation and agreement patterns of all tree names'. The 'Location' is 'Project DoBeS-0P'. The 'Content' section lists metadata: Genre (Elicitation), SubGenre (lexical elicitation), Task (speech), Modalities (speech), Subject (interactive), Interactivity (interactive), PlanningType (planned), Involvement (planned), SocialContext (planned), EventStructure (planned), and Channel (planned). The 'Languages' section lists 'Bainouk Gubeeher (c)' and 'French (c)'. The 'Actors' section lists 'Alexander Cobbinah' and 'Jean Marie Sagna'. The 'MediaFile' section lists 'Type audio', 'Format audio/x-wav', 'Size 460 MB', and 'Quality Unspecified'. The 'RecordingConditions' section is empty. The 'TimePosition' section lists 'Start Unspecified' and 'End Unspecified'.

---

There is an output gap





---

# The output gap

Outputs from language documentation projects have bifurcated into:

- ❑ **Published** grammars, (bilingual) dictionaries and (glossed) texts – ‘revival’ of familiar genres linguists have been comfortable with for 100+ years
  - ❑ **Archive** deposits – hundreds or thousands of files, professionally curated by archivists, but often poorly organised or structured, with little if any contextualisation
-

---

# What is missing?

- Meta-documentation – the documentation of documentation projects, goals, methods, IP contributions, outcomes
  - New (unfamiliar) genres that link and contextualise analytical outputs and the archival corpus:
    - ethnographies of documentation project designs
    - accounts of data collection (cf. archaeology ‘field report’)
    - finding-aids to corpus collections
    - ‘exhibitions’ or ‘guided tours’ of archival deposits
  - Evaluation measures that enable properly-based peer assessment of documentations, equivalent to the way traditional outputs are judged
-

---

# New genres

- Woodbury (2015) ‘Archives and audiences: Toward making endangered language documentations people can read, use, understand, and admire’:

“I urge documenters to take **authorial control** of their work, as they would if each archived collection were a book of language materials

- make a guide to your own documentary corpus
- include meta-documentation: describe the design of activities or projects from which the corpus arose, offer a theorization of the corpus (or several, from different perspectives), and describe the appraisal process used to select and assemble the corpus
- write narratives, logs, and journals
- think of your corpus as belonging to a **genre**.

To some extent, all this means documenters taking on some of the work traditionally done by archivists.”

---



[www.elpublishing.org](http://www.elpublishing.org)

A new free open access online venture launched in July 2015 which:

- has the infrastructure of familiar models of publication (editorial board, peer assessment, etc.)
  - publishes familiar outputs: LDD journal, books, multimedia, but without the limitations of traditional publishing, eg. length, hypertext
  - provides a platform to encourage experiments in new genres of output, eg. “Language Contexts”
  - will provide a space and an interface to move towards evaluations of these new outputs so that the underlying desire of the LSA statement might be realised
-

---

# 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?
-

---

# Transdisciplinarity

- Siebert (2016) ‘Documentary linguistics: a language philology of the 20<sup>th</sup> century’:

“documentary linguistics’ focus on ‘direct representation of discourse’ requires a broader conceptualization of the field that moves **beyond purely linguistic concerns**. This article recasts documentary linguistics as a philology, broadly understood as the inquiry into ‘the multifaceted study of texts, languages, and the phenomenon of language itself’ ... The reconceptualization of documentary linguistics described in this article opens documentary linguistics to non-core linguistic types of language documentation efforts and situates the documentary activities more broadly in the humanistic enterprise of communicating, discussing, studying, and understanding human achievements of other times and places.”

---

---

# 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
-



---

# 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
-

---

# Managing your data

- design and use a filename system
  - work out ('model') your basic units of documentation and the relationships between them
    - if you get these right, it will do the “heavy lifting” of your data management strategy
    - data and metadata are intertwined, points in a spectrum rather than different things
-

---

# Three important qualities

- consistency
- documentation of conventions, structures, methods
- machine readability
  - “computer programs can act on data in terms of its proper structures and categories”



---

# Data management

- understand and model the data (units, relationships)
  - use appropriate data structure methods – in both file *contents* and *organisation*
  - use appropriate and conventional data encoding methods (e.g. Unicode)
  - be explicit and consistent
  - plan for flow of data, working with others, across different systems
  - document steps, decisions, conventions, structures
  - think ahead to archiving
-

---

# Managing data in your computer

- design a well-organised system of folders so that you can always find your stuff according to *what it is*, **not**:
    - ❑ where the software decided to put it
    - ❑ what the software decided to call it
    - ❑ when/where you last used it
    - ❑ what someone else called it
-

---

# File structures and names

- design folder structure as a logical hierarchy that suits your goals, content and work style
    - have documentary materials within one overall directory (e.g. for backup)
    - make directories for **relevant categories**, e.g. sessions, media types, dates
    - design it so that you will *always* be able to find things
  - you may need to restructure at different points in your project, e.g. move from date-based to session-based structures
-

---

# Designing a file/folder structure

- it should relate to reality
  - locations should make sense, so you (and others) will know where to look for things (where do you keep your passport; favourite cup?)
  - the best location is “the place that one would naturally look to find it”
-

---

# File naming

- filenames that are unsystematic or are non-standard will cause problems, eventually
  - unsystematic file naming *might* be (*just*) OK if
    - you already have many files
    - you have a working method that already does everything you need to do
    - your “system” will do everything you need to do in the future
-



---

# Manage file names from the start

- a new file:
    - ❑ don't just accept the default filename or location suggested by the application when you first save the file
    - ❑ put it where it belongs, *immediately*. If necessary, create the place (directory/path) where it belongs
    - ❑ name it according to your naming system!
    - ❑ if you have an inventory/index of files, add an entry for the new file
-

---

## Filename rules

- all filenames should have correct extensions
  - each filename should have only one ".", before the extension
  - use only ASCII characters (US keyboard)
  - use only letters, numbers, hyphens (-) and underscores (\_)
  - keep filenames short, just long enough to contain the necessary identifier - don't fill them up with lots of information about the content (that is metadata!)
  - (advised) use only lower case letters
-

# How about these file names?

1. ready.audio.wav
2. ReALLyhArDtOReAd.txt
3. éclair.jpg
4. e'clair.jpg
5. french-cake.jpeg
6. french-cake.jaypeg
7. -2011.psd
8. lexicon-master
9. I .eaf
10. ice cream.doc
11. Obama.TXT
12. オバマ.txt

---

# Make filenames sortable

- make filenames usefully sortable:
  - ❑ 20100119lecture.doc
  - ❑ 20100203lecture.doc

gr_transcription_1.txt	gr_transcription_001.txt
gr_transcription_12.txt	gr_transcription_005.txt
gr_transcription_5.txt	gr_transcription_009.txt
gr_transcription_9.txt	gr_transcription_012.txt

---

# Associating files

- you can make resources sortable together by giving them the same filename root (the part before the extension), or part of the root:

gr_reefs.wav	paaka_photo001.jpg
gr_reefs.eaf	paaka_photo002.jpg
gr_reefs.txt	paaka_txt_conv203.wav
	paaka_txt_conv203.eaf
	paaka_txt_lex.doc

- **document your conventions and system if you do this**
-

---

# Avoid metadata in filenames

- avoid putting *metadata* into filenames. *A filename is an identifier, not a data container*
  - better to  
use a simple (semantic) filename  
or  
a key (i.e. meaningless) filename, and then create a metadata table to contain all the relevant information
  - a table can properly express all the information, contain links etc, and is extensible for further metadata
-

- e.g.  
Paaka\_Reefs\_Dan\_BH\_3Oct97.wav

- better:

- paaka\_063.wav

plus

- paaka\_063.txt

paaka\_063.txt

language	topic	speaker	location	date
Paakantyi	Reefs at Mutawintyi	Dan Herbert	Broken Hill	1997- 10-03

---

## A file naming *system*

- carefully design a filename **system** for your data and *document the system so that somebody else can understand it*
- one documenter's new system:

aaa\_bb\_cc\_yyyy-mm-dd\_nnn.wav

---



---

## A filenaming *system*

- `aaa_bb_cc_yyyy-mm-dd_nnn.wav`

aaa = village **code**

bb = (main) speaker **code**

cc = genre/event **code**

yyyy-mm-dd = date (why this order?)

nnn = optional number (e.g. 001)

`.wav` = correct extension for file content type

---

---

# Different types of metadata

- there are many types of metadata
  - different types of materials may have different metadata
    - eg metadata for photos and videos may have technical parameters, lists of people appearing
    - e.g. metadata for transcriptions may have date, version, who transcribed, notes on progress
-

---

# Meta-documentation

- you should keep an updated description of the methods, conventions, abbreviations you use
- .. so somebody could fully understand (and use) your data and methods *in your absence*



---

# Data/file versions

- need to distinguish or keep versions depends on purposes
  - by adding a suffix to the filename, eg
    - fugu1.txt  
fugu2.txt
    - or
    - fugu\_1.txt  
fugu\_2.txt
  - which of the above methods is better?
-

- 
- fugu\_14022013.txt  
fugu\_20130214.txt  
14022013\_fugu.txt  
20130214\_fugu.txt
  - which of the above would be best?
-

---

# Managing data/file versions

- do you need to keep every version?
    - it may be OK to keep “original” plus current
  - if information is regularly updated, corrected, you can keep 1 filename and put dates in the document itself, or record dates in a catalogue/metadata file
  - however, a series of files may have inherent value, e.g. your transcriptions/annotations, as your understanding and analysis changes, so
    - date and keep files
- 
- use different tiers in ELAN?

---

# Metadata

- metadata is *data about data*
    - for *identification, management, retrieval* of data
    - provides the *context* and *understanding* of that data
  - carries those understandings into the *future*, and to *others* (and hence is important for archiving and preservation)
  - reflects *knowledge* and *practices* of data providers
  - *defines* and *constrains* audiences and usages for the data
  - *all* value-adding to inscriptions involves the creation of metadata – all annotations (transcriptions, translations, glosses, pos tagging, etc.) are metadata (Nathan and Austin 2004)
-

---

# Metadata

- recommendations for creating metadata for language documentation have been primarily influenced by library concepts (eg. Dublin Core), and key metadata notions have been *interoperability*, *standardisation*, *discovery*, and *access* (OLAC, EMELD, Farrar & Langendoen 2003).
  - the goals of language documentation mean this is not powerful enough and we need a *theory of metadata*, largely lacking until now
-



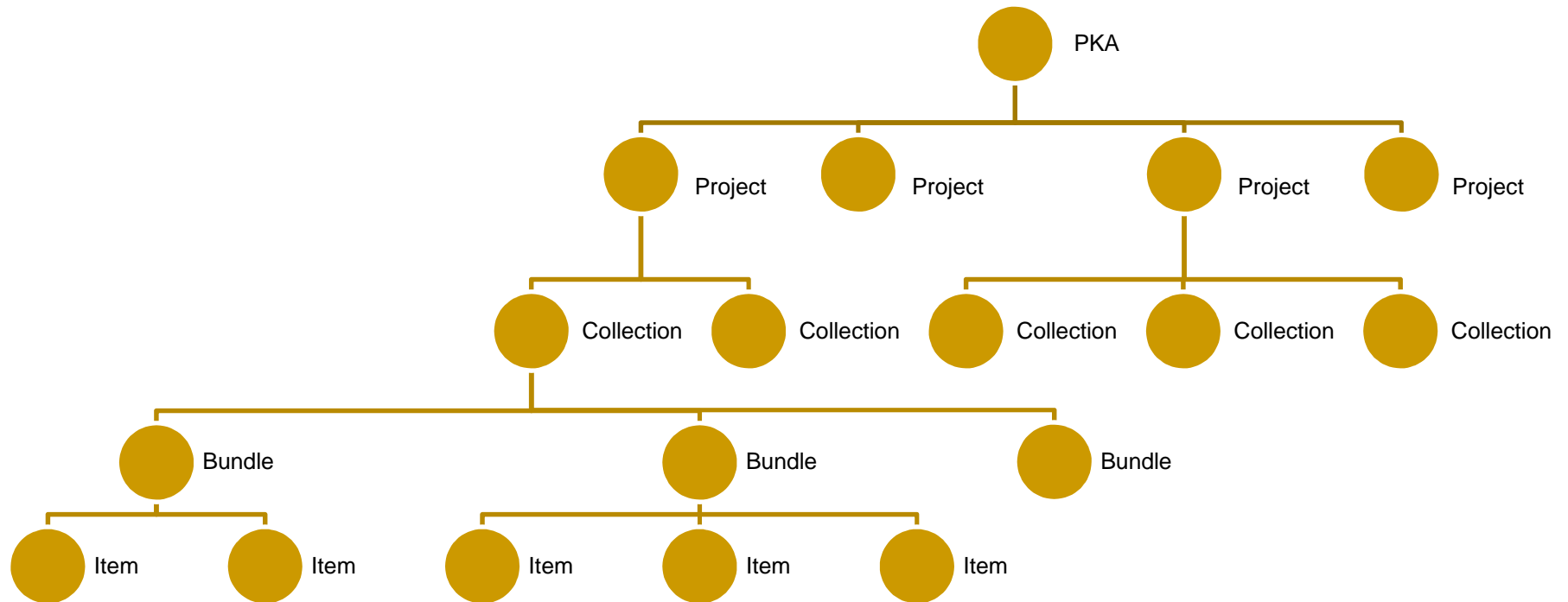
---

# Types of metadata

- *people* metadata – creator's / delegate's details
  - *descriptive* metadata – content of data
  - *administrative* metadata – eg. date of last edit, relation to other data
  - *preservation* metadata – character encoding, file format
  - *access and usage* protocols – eg. URCS
  - metadata for *individual* files or *bundles*
  - Metadata can apply at various levels
-

---

# A possible model



---

# How do we store and manage metadata?

- *in our heads* – problem: degrades rapidly and not preservable or portable
  - *on paper* – problem: not easily searchable or extensible
  - *within files* (headers) – problem: not easily searchable or extensible
  - *in file/folder names* (eg. SasJBpka09-12\_int03.wav – problem: difficult to maintain, breaks easily, not all semantics can be expressed
  - *in a metadata system*
-

---

# Metadata systems

- free text
  - structured text (eg. Word tables, XML, Toolbox)
  - spreadsheet (eg. Excel) – see ELAR sample
  - database (eg. Filemaker Pro, Access, MySQL)
  - metadata manager (CIMDI maker, Arbil, SayMore)
  
  - or some combination of these that is *usable, flexible* and sufficiently *expressive*
-

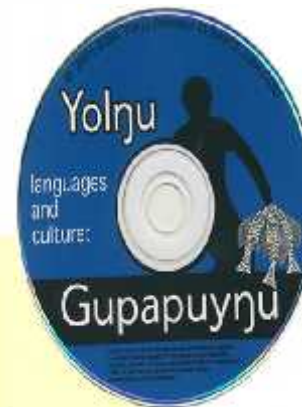
# Mobilisation: outcomes, audience research & design



[29] [Play](#)

gamu	wanyja-ngumay	yurra	juga-l
water+ABS	where-ABL+ABS	2PL.NOM(A)	dip-PP2

Lit. 'You dip (a billy can and get) water from where?', i.e. 'Where do you get water from?'



# Links and access

- In this PPT document, there are many [links](#)
- Links to items available on the internet are marked\* with an asterisk
  - You can access this item [link](#)\*
  - You can't access this item [link](#)

# Woodbury adds

- “language documentarians take their material ... as a starting point ... for **engaging diverse audiences: community members interested in family, neighbors, community identity, verbal art, education, reclamation, or nostalgia; scientists interested in philology, ethnohistory, human ecology, language typology, or linguistic theory; humanists interested in linguistic expression and its products; or general publics** with any of these interests, and more”

- Woodbury 2014:2

## So we need more than this ...

*Pulhapayara-lu kanya-nyja pirru ngunha.*  
 [name]-erg carry-past meat.acc that.acc  
 'Pulhapayara carried that meat.' [T45s3]

(Peter Austin, Jiwari)

(Carina Jahani,  
Galesh)

ej ze n Eng permission FAR  
 el yde adv Eng separate FAR علی  
 هده، جداگانه  
 em adv Eng also FAR هم  
 -em v Eng 1SG FAR  
 =em pro.clit. Eng my FAR  
 em pers Eng we FAR  
 ems 1 adv Eng this year FAR  
 -en n Eng are,COPUL. FAR هستند  
 end 1) vt Eng throw FAR 2 )  
 vt  
 end t v Eng throw FAR  
 enj m n Eng accomplishment  
 es vi Eng stop FAR ایستاد  
 eš2 vt Eng leave FAR  
 eš3 vt Eng put FAR می گذاشتیم  
 =eš1 nomprt Eng her or his .PRON  
 FAR  
 es adv Eng now FAR  
 esbe adj Eng white FAR سفید  
 esf n nprop Eng proper name FAR



# Outcomes – what influences them?

- goals
  - contribute to knowledge - academic outputs
  - contribute to language maintenance - pedagogical materials, activities
  - personal and social rewards
  - behave ethically and encourage ethical practices
  - improve career and financial status
  - ... ?
  
- how would YOU prioritise (rank) these?

# Outcomes – what influences them?

- constraints
  - time
  - ethics
  - funding
  - equipment
  - skills and people
  - funder's requirements
  - contacts and relationships
  - community needs and contexts

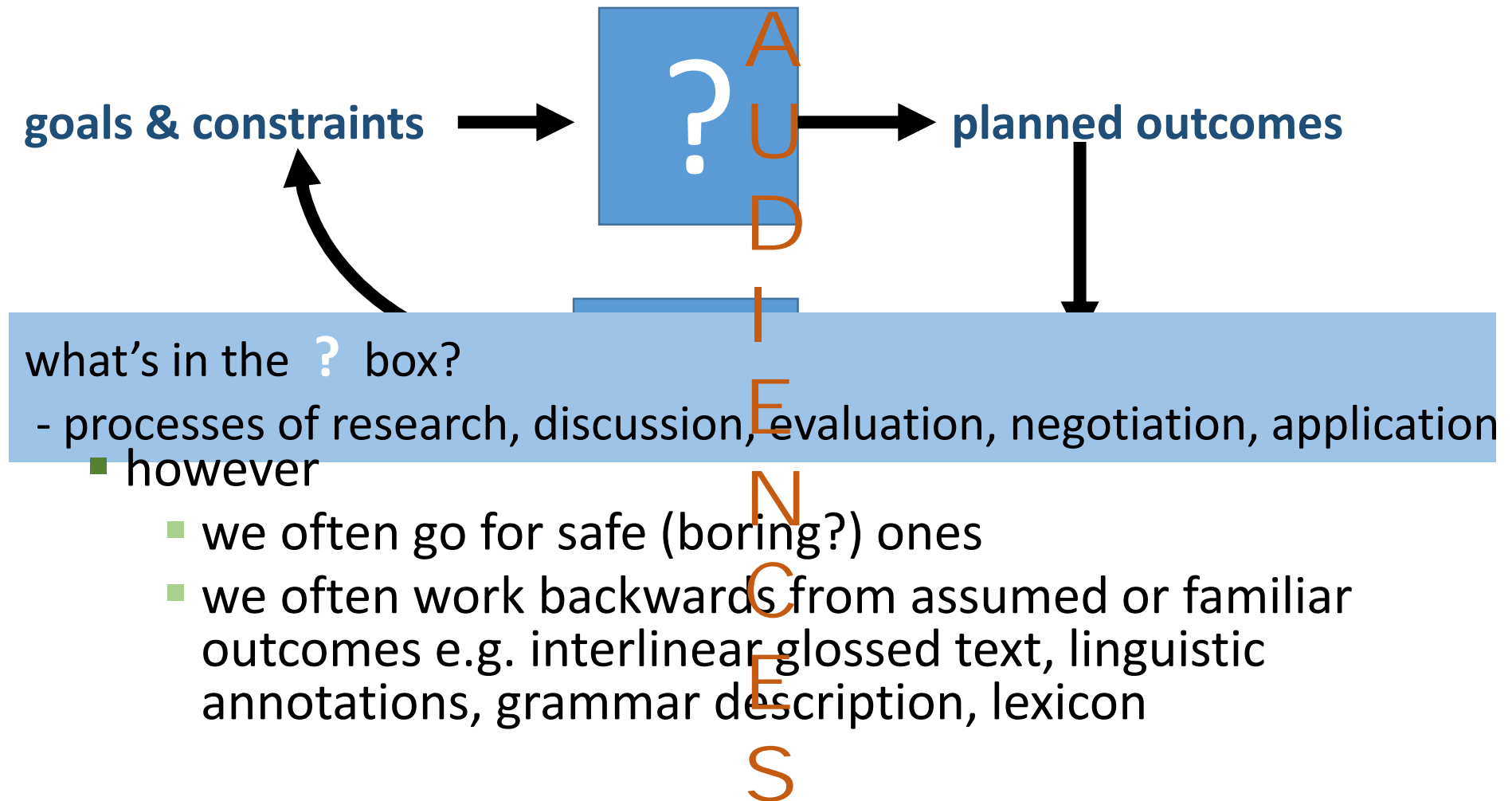
- how would YOU prioritise (rank) these?

# Audiences

- for *effective* outcomes, we need to understand our *audiences*
- audiences?
  - may have been defined, e.g. a school or community that employs you
  - may be who you intend/imagine/hope will use your outcomes
  - may be some category of people, e.g. linguists interested in the use of nasalisation in managing conversation turn-taking
- how can we understand them?
  - research!
    - current situation and resources, skills, literacies, infrastructure: do they have internet? electricity? schools and teachers?
    - their values and goals
  - also evaluate and iterate

# If we were scientific ...

- **goals** and **constraints** lead to planning outcomes



# Some outcome types

- qualitative descriptions
- descriptive (linguistic) materials
- sociolinguistic materials
- pedagogical materials
- community self-documentation
- other unprocessed resources
- new media apps
- media publication
- interactive multimedia

# Some outcome types

- qualitative descriptions
  - e.g. describe methods, activities, histories of documentation events and contexts

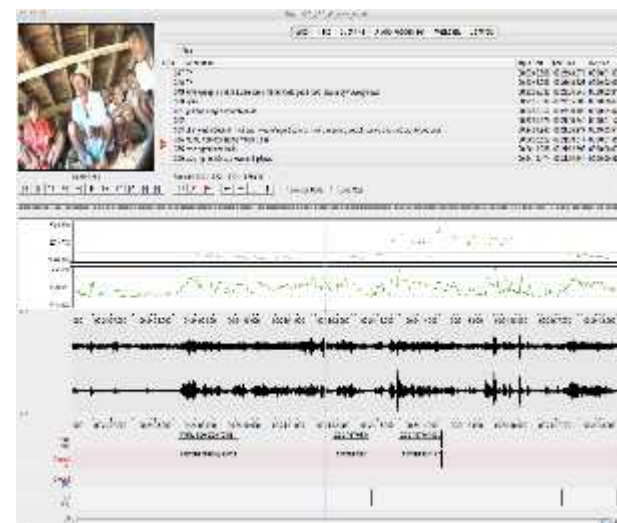
... this resource documents the ceremony of the changing of the local traditional authorities in San Juan Quiahije. [It] began at night on December 31, 2009 and culminated at noon on January 1, 2010.

These events take place at the city hall in San Juan which is located in the main square ... There is a conversation inside the city hall with higher ranking authorities about the ritual places where they place the candles when they go pray, and the last prayer by two head elders conduct in the main altar of city hall. They are Wenceslao Cruz Cortés and Evencio Cruz Apolonio. Evencio was the head judge and Wenceslao the head elder. This is the last post that Wenceslao will ever serve at city hall. Wenceslao will now transition to be part of Consejo of Elders in the community ... There are prayers inside the church, a recording of us preparing to record the ceremony ... ambient noise, and recordings outside the city hall.

# Some outcome types

- descriptive (linguistic) materials
  - e.g . interlinear and annotated material, grammar, lexicon

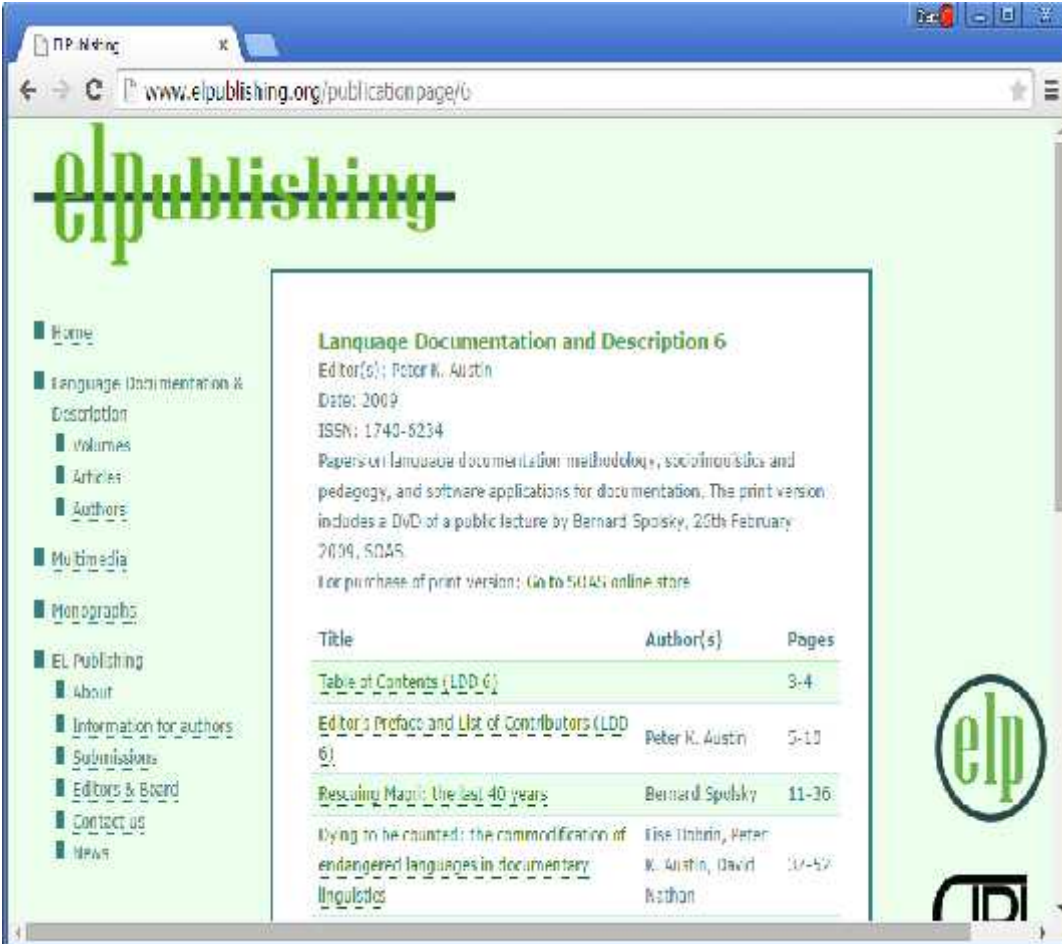
*ngaygu* / *warrungu* /  
1SG.GEN Warrungu-ABS  
[I answered,] My [language is]  
Warrungu.  
(Alf Palmer, Tsunoda Tasaku)





# Some outcome types

- sociolinguistic
  - e.g. papers, surveys, interviews, statistics

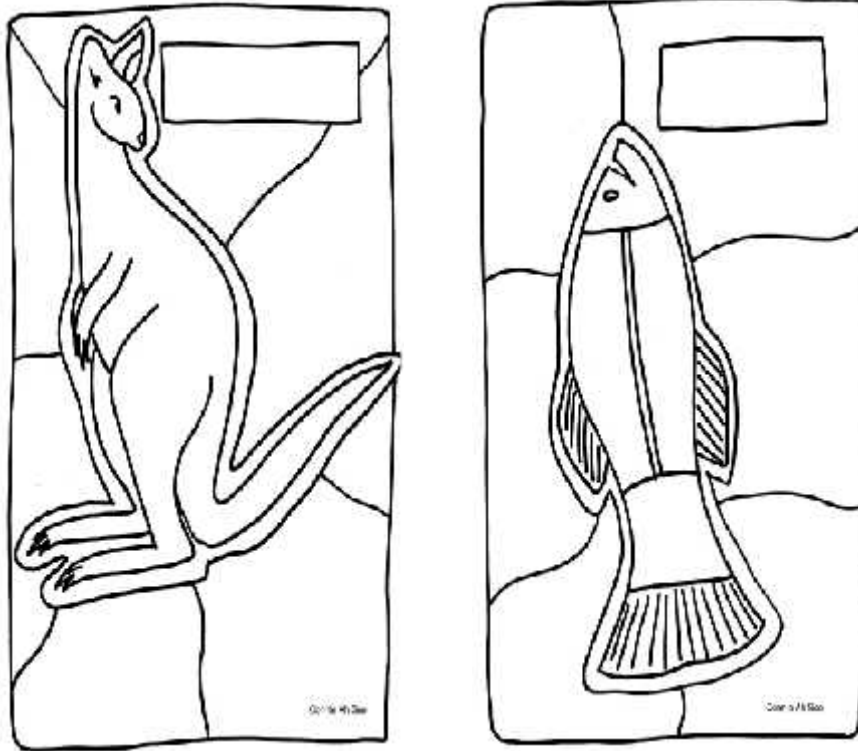


The screenshot shows a web browser window displaying the website [www.e-publishing.org/publication/page/6](http://www.e-publishing.org/publication/page/6). The page features the 'elpublishing' logo at the top left and a navigation menu on the left side. The main content area displays information for the publication 'Language Documentation and Description 6', edited by Peter K. Austin in 2009. The ISSN is 1740-5234. The description mentions papers on language documentation methodology, sociolinguistics, and pedagogy, along with a DVD of a public lecture by Bernard Spolsky. A table of contents is provided below the description, listing the title, author(s), and page numbers for each section.

Title	Author(s)	Pages
<a href="#">Table of Contents (LDD 6)</a>		3-4
<a href="#">Editor's Preface and List of Contributors (LDD 6)</a>	Peter K. Austin	5-10
<a href="#">Rescuing Manin: The last 40 years</a>	Bernard Spolsky	11-36
<a href="#">Dying to be counted: the commodification of endangered languages in documentary linguistics</a>	Eise Uthman, Peter K. Austin, David Nathan	37-49

# Some outcome types

- pedagogical materials
  - instructional – exercises etc



(Wiradjuri: Connie Ah See)

# Some outcome types

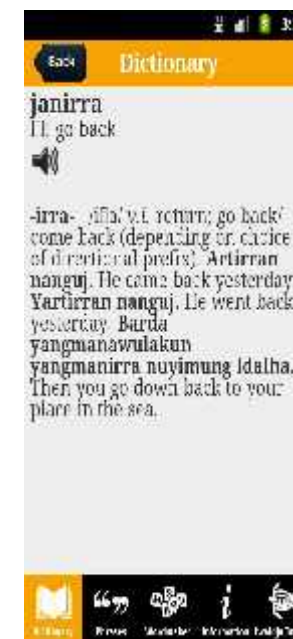
- pedagogical materials
  - resources – for teachers and creative learning (be clear about learners, learning goals, levels, learning contexts, equipment etc)

(Chatino: Hilaria Cruz and Anthony Woodbury)

		COMPLETIVE			PROGRESSIVE			POTENTIAL		
ENGLISH	JG	SJQ			SJQ			SJQ		
eat sth.	A	yku	4	ʔj 4	nt <sup>y</sup> ku	32	ʔj 4	ku	4	ʔj 4
					nskw					
pluck sth.	A	skwə	4	ʔj 4	ə	32	ʔj 4	xkwə	4	ʔj 4
heat sth.	B	ntsɔ	14	ʔj 0	ntsɔ	1	ʔj 0	ktsɔ	14	ʔj 0
sprinkle sth.	B	sne	14	ʔj 0	nsne	1	ʔj 0	sne	14	ʔj 0
grind sth.	C	yo	2	ʔj 1	nd <sup>y</sup> o	2	ʔj 1	ko	14	ʔj 0
					nt <sup>y</sup> kw					
vomit	C	ykwɛ	2	ʔj 1	ɛ	2	ʔj 1	kwɛ	14	ʔj 0

# More outcome types

- community self-documentation
  - e.g. [websites](#)\*, [YouTube](#)\*, [Indigenous Tweets](#)\*
- other unprocessed resources
  - e.g. recordings, notes, radio shows - [ABC news](#)\*
- new media apps
  - e.g. [phone apps](#)\*
- media publication
  - produced, edited and published video
- interactive multimedia



GUNDHI  
HOMEGAAY/GARAY  
WORDSMANDI  
SENTENCESGUWAABAL  
STORIESYUGAL  
SONGSYULUGI  
GAMESMAYABIL  
LINKS

HELP X

Language: Yuwaalaraay

## Burrulaa birralii (Lots of children)

- Kelsey Strasek-Barker

**Yinayu**                      **nhama yilamaldanha.**

yinarr                      -gu    nhama yilama-li -lda-y                      -nha  
[Aboriginal]woman -doer.to that,the cook                      -continuous -Present

**Dhayndu**                      **nhama bamba guwaaldanha.**

dhayn                      -gu    nhama bamba guwaa-i -da-y                      -nha  
Aboriginal man -doer.to that,the hard talk                      -continuous -Present

**Birraliigal**                      **yulugilanha.**

birralii -gal    yulu-gi -gi-la-y                      -nha  
child                      -plural dance                      -continuous -Present

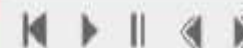
**Burrulaa mari**                      **wiidja.**

burrulaa mari                      wi -ga  
many                      Aboriginal person fire -in,at,on

The women are cooking.  
The men are talking loudly.  
The children are dancing.  
Lots of maris [murris] at the fire.

Verse 1

Page 2 of 5

 Play to end[Go back to song menu](#)

# Examples - links

- targeted/restricted distribution
  - [Spoken Karaim](#) – given to Karaim in Lithuania, Poland, Russia, Ukraine
  - community desire to restrict or control distribution - [Paakantyi CD](#) Barbara Glowczeski “Dream Trackers” anecdote [app](#)
  - Paakantyi [app](#)
- web
  - Ed Garrett’s [speech bubble player](#)\*
  - Notebooks of William Dawes [website](#)\* [XML data](#)
  - eg Kamilaroi [dictionary](#)\* and [launch](#)\*
  - Jewish Iraqi (Arabic) [website](#)\*
  - collection of social media <http://indigenoustweets.com/>\*
  - Tsunoda Warrungu [text](#) <http://www.dnathan.com/language/warrungu>\*

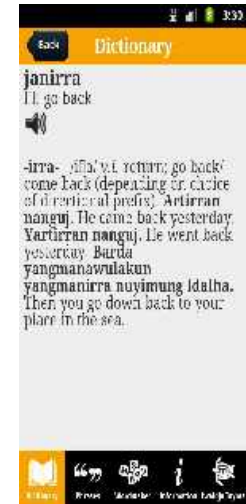
# Examples - links

- [Gayarragi, Winangali](#) ([get it online](#)\*)
- [Sago grubs – Tombulu](#)\* (Anthony Jukes/YouTube)
- [Kezzeme festival](#)\* with post commentary (Stuart McGill)
- [Sorosoro](#)\*
- [Living Tongues Institute](#)\*
- [Endangered Languages Project](#)\*
- Online archives, eg [California Language Archive](#)\*
- Community-based, eg [First People's Cultural Council](#)\*

# Examples - phone apps

- phone apps taking off – more buzz\* than language?

“The Ma! Iwaidja free mobile phone app aims to prevent the extinction of the Iwaidja language -- one of Australia's 100 endangered languages.”





## Conclusion: the next big challenge ...

“ The next big challenge facing language documentation is the discovery and widespread use of software interfaces that make aesthetically appealing documentation materials that can be easily and flexibly used by a wide range of users, including the speech communities whose endangered languages we want to document. ”

- Gerrit Dimmendaal

in *Language Documentation and Conservation*, vol 4

# Links and references


- Woodbury 2014 Archives and audiences: Toward making endangered language documentations people can read, use, understand, and admire <http://www.elpublishing.org/PID/135>\*
- Eli Timan Jewish Iraqi website  
<http://www.jewsofiraq.com/texts.html>\*
- Tsunoda Warrungu  
<http://www.dnathan.com/language/warrungu>\*
- Ed Garrett's Speech Bubble Player  
<http://lah.soas.ac.uk/projects/dev/bubble-player/wilbur.html>\*
- Nathan et al, Notebooks of William Dawes  
<http://www.williamdawes.org/>\*
- Giacom & Nathan, Gayarragi, Winangali  
<http://www.dnathan.com/projects/gw>\*

# Links and references (continued)

- McGill Kezzeme festival  
[http://www.cicipu.org/video\\_kezzeme.html](http://www.cicipu.org/video_kezzeme.html)\*
- Sorosoro Foundation <http://www.sorosoro.org/en/>\*
- First People's Cultural Council <http://www.fpcc.ca/>\*
- Living Tongues Institute <http://www.livingtongues.org/>\*
- Endangered Languages Archive <http://www.elar-archive.org>\*
- California Language Archive <http://cla.berkeley.edu/>\*
- Endangered Languages .com  
<http://www.endangeredlanguages.com/>\*

# Links and references (continued)

- Indigenous tweets (and blogs)  
<http://indigenoustweets.com/>\*
- Austin & Grenoble 2007, Current trends ...  
<http://www.e-publishing.org/PID/047>\*
- Dimmendaal Language Description and “The New Paradigm” ...  
<http://scholarspace.manoa.hawaii.edu/bitstream/10125/4475/7/dimmendaal.pdf>\*
- Nathan 2013, Access and accessibility  
[http://www.dnathan.com/eprints/dnathan\\_accessibility.pdf](http://www.dnathan.com/eprints/dnathan_accessibility.pdf)\* 2012
- Eric Michaels 1986, The Aboriginal Invention of Television  
<http://astro.temple.edu/~ruby/wava/eric/>\*
- Linn 2014, Living archives  
<http://www.e-publishing.org/PID/137>\*



# Language Landscape: Un modelo y una plataforma digital para la autodocumentación de idiomas

Samantha Goodchild  
Ebany Dohle

28/08/2017  
San Miguel Xaltipan



# Introducción

- Quienes somos
- La plataforma digital
- Por qué agregar tus grabaciones
- Muestras de proyectos

# Quienes somos



## Grétar sings the colours of the rainbow

Added by [ebany\\_d](#)

ADD TRANSLATION

ADD TRANSCRIPTION



DATE: 08/11/2013  
 SPEAKERS: EBANY, GRÉTAR, SIG BORGARNES  
 PLACE: BORGARNES  
 LANGUAGES: ENGLISH - ICELANDIC



NAME: EBANY  
 DOB: EL SALVADOR  
 BIRTHPLACE: EL SALVADOR  
 LANGUAGES: ENGLISH - GERMAN - PORTUGUESE, BRAZILIAN, FORTALEZA - SPANISH, SALVADORAN

This is a song about the colours of the rainbow sung by young Grétar in Icelandic

**Metadata**

**Transcriptions (0)**

**Translations (0)**



NAME: GRÉTAR  
 DOB: 01/01/2007  
 BIRTHPLACE: BORGARNES  
 LANGUAGES: ICELANDIC



NAME: SIG  
 DOB: ICELANDIC - ENGLISH  
 BIRTHPLACE: ICELANDIC - ENGLISH  
 LANGUAGES: ICELANDIC - ENGLISH







## ebany\_d

[PREVIEW PROFILE PAGE](#)[LOG OUT](#)

EMAIL:

ebany@languagelandscape.org

DESCRIPTION:

Ebany is currently doing a PhD in linguistics at SOAS and is interested in the languages of Central America. She has worked with communities in El Salvador, Guatemala, Nepal and the Philippines in researching policy for environmental sustainability, bilingual education and cultural preservation. She has also created several podcasts as part of SOAS Radio's Endangered Languages Series.

As part of the student-initiated Language Landscape team, Ebany is involved with coordinating outreach projects with schools and communities and setting up events.

4 PROJECTS

[ADD PROJECT](#)

12 RECORDINGS

[ADD RECORDING](#)[UPDATE PROFILE](#)

## Add Recording

HELP

TITLE:

AUDIO FILE:

 No file chosen

YOUTUBE URL:

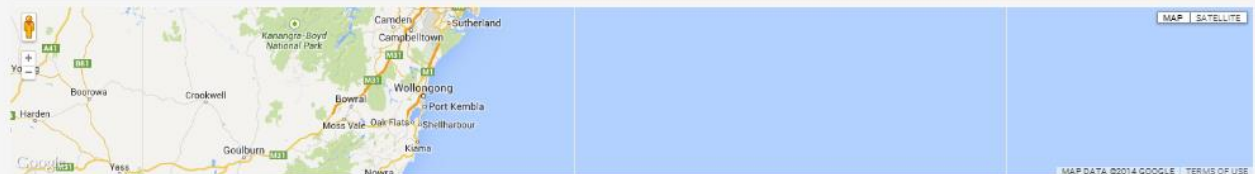
PHOTO:

 No file chosen

TIME:

PLACE:

Enter a location



MAP | SATELLITE

MAP DATA ©2014 GOOGLE | TERMS OF USE

[ADVANCED LOCATION DATA](#)

SPEAKERS:

 Type some text to search in this autocomplete

LANGUAGES:

 Type some text to search in this autocomplete

TOPICS:

 Type some text to search in this autocomplete

GENRES:

 Type some text to search in this autocomplete



### Add Recording

HELP

TOPICS:



GENRES:



PROJECTS:

KEYWORDS:

DESCRIPTION:

» ADVANCED METADATA

OTHER PEOPLE CAN EDIT THIS RECORDING:




ADD RECORDING



# Usos de la página

- Búsqueda avanzada
- Comparación de distribución de idiomas en relación a género, edad y fecha
- Familias lingüísticas
- Identificar y nombrar idiomas propios
- Mapa y proyecto personalizado



# Por qué?

- Compartir y demostrar grabaciones en página propia
- Proyectos temáticos
- Comunidad internacional
- Visualización geográfica de los idiomas del mundo
- Acceso a redes sociales
- Audio, video, y fotografías



# Proyectos

- [World Oral Literature Project](#)
- Eleanor's Project: [Vatlongos](#)
- [Inyujyulu Takwikat](#)
- Lingüistas de SOAS y su trabajo de campo:  
[SOAS in the Field](#)

A large red graphic on the left side of the slide, consisting of several thick, angular lines that form a stylized, abstract shape, possibly resembling a large letter 'L' or a similar geometric pattern.

# #mapyourlanguage

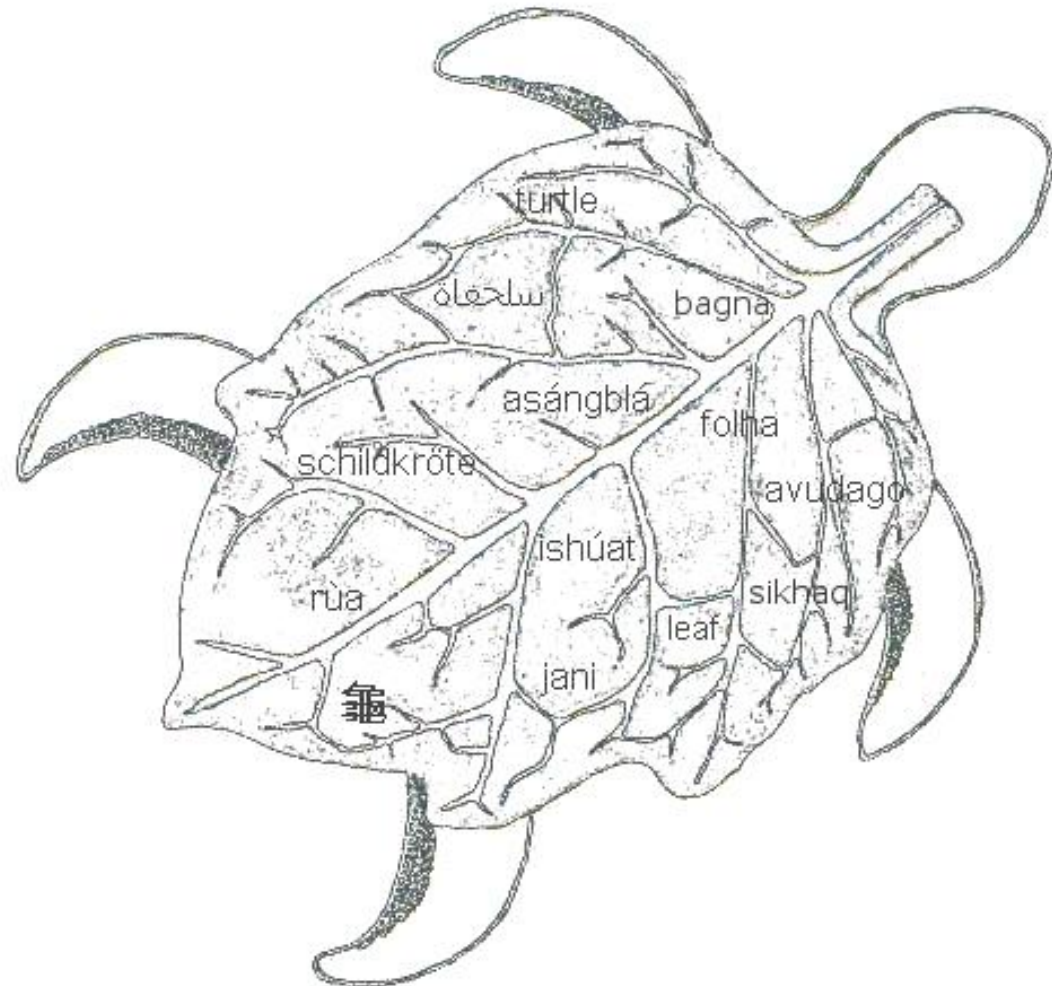
- [www.languagelandscape.org](http://www.languagelandscape.org)
- Fb: /languagelandscape
- Twitter: @langlandscape
- Blog: [www.languagelandscapeper.org/blog](http://www.languagelandscapeper.org/blog)

[samantha@languagelandscape.org](mailto:samantha@languagelandscape.org)

[ebany@languagelandscape.org](mailto:ebany@languagelandscape.org)



**SOAS**  
University of London



plants. animals. words.



# Exploring peoples, their languages and the natural world



**PAW** is a network of researchers focused on bridging the disciplinary boundaries that have typically separated studies about people, animals, and their environment.

Our members include academic and non-academic individuals and organisations, from anthropology, to linguistics or ethnoornithology.

It is a platform for exchange for practitioners from different fields, to learn each other's tools and methods, and develop new ones, and question our theories. We also suggest that academics producing primary research need to consider its impact on the social agenda.

This is about engagement and collaboration and how we can make research relevant for the world in which we live.

Our aims:

- Create a space for dialogue across the usual discipline boundaries
- Nurture basic skills for ethical research methodologies
- Foster an awareness of what is required by each disciplines to create an understanding of our own limitations, enabling us to move beyond them



## Outcomes

- Workshops (2013, 2014, 2015, 2017)
- Practical resources and guidelines for interdisciplinary documentation
- Network of fieldworkers
- Website and mailing list
- PAW book
- Radio podcast on BBC World Service



Join in for a hands-on one day workshop where you will learn how to make good quality audio recordings, collect botanical samples, and create a map recording the location of your collection.

# Plants. Animals. Words: Mapping a day in the Square!

14 December 2015 | 9.30am-5pm  
To register: [www.soas.ac.uk/linguistics/events/](http://www.soas.ac.uk/linguistics/events/)

Organised by PAW / SOAS, Department of Linguistics

CENTRES & PROGRAMMES OFFICE  
SOAS, University of London  
Russell Square, London WC1H 0XG  
[www.soas.ac.uk](http://www.soas.ac.uk)

Design: SOAS Centre & Programmes Office | Printed by SOAS Print Shop



View the slides online:  
[bit.ly/PAWmapping](http://bit.ly/PAWmapping)

181 members all over the world!



<http://paw.plantsanimalswords.org/page/paw-participants>

Resources from linguistics, botany, archiving etc website:

<http://paw.plantsanimalswords.org/index>

Meet and exchange : jiscmail and Facebook

<https://www.facebook.com/groups/paw2013/>

PAW on the BBC World Service! (about 19 minutes in)

<http://www.bbc.co.uk/programmes/p045103p>

E-mail: [plantsanimalswords@gmail.com](mailto:plantsanimalswords@gmail.com)



# PAW

CONCLUSIONS FOR THE DAY: IMPACT OF TRANSDISCIPLINARY DATA  
COLLECTION ON REVITALIZATION EFFORTS

# Pela: Case Study

- ▶ Outline of project
- ▶ Motivations
- ▶ Lessons learnt
- ▶ Measuring Impact





# Pela: Project Outline

- ▶ To document language use of the Pela in a multilingual setting. Audio, video and photographic records pay special attention to endangered genres such as religious ritual and folk stories.
- ▶ Approximately 400 speakers

# Pela: Motivations

- ▶ Family connection
- ▶ Academic value
- ▶ Desire of the community for documenting their own language

# Pela: Lessons Learnt

- ▶ Adopt appropriate methodology according to the real situation in the community, not just follow books...and the internet
- ▶ Insider v.s Outsider
- ▶ Learn from other disciplines



# Pela: Measuring impact

- ▶ Is the capacity building successful?
- ▶ Is there a continued work on the documentation project?
- ▶ What did people do with the documentation materials created during this project?
- ▶ Is there increasing interests in using Pela/creating language materials?

# Náhuat-Pipil Case Study

- ▶ Outline of project
- ▶ Motivations
- ▶ Lessons learned
- ▶ Measuring Impact



# Náhuat-Pipil: Project Outline

- ▶ To document the manner in which knowledge of the natural world is encoded in the Náhuat-Pipil language of El Salvador
  - ▶ Naming conventions, categorization of plants, noun-verb interaction
  - ▶ Understanding local taxonomies and strategies for perceiving and interacting with the natural world.
  - ▶ Recording uses, cultivation practices and the relationship between TEK and identity
- ▶ Links with anthropologists, botanists, activists and artists
- ▶ Approximately 500 speakers
- ▶ Age: 50+



# Náhuat-Pipil: Motivations

- ▶ Community driven: speakers identified an aspect of their daily lives which they wanted to preserve and 'rescue'.
- ▶ Personal interest: home country, interest in plants, reconnection with roots
- ▶ Improving access and appreciation of local knowledge



# Náhuat-Pipil: Lessons Learnt

- ▶ Establishing local connections across existing disciplines can provide much needed support
- ▶ Support from a varied group of people provides more validation for endangered or minority language communities
- ▶ Incorporating community interests into documentation and revitalization efforts is key e.g. agriculture and clay work



# Náhuat-Pipil: Measuring Impact

- ▶ Has the group of people interested in the outcomes of the research project grown or expanded?
- ▶ Is there a continued interest in the research?
- ▶ Are new ideas being proposed within the community members?
- ▶ Are new initiatives (personal and community based) being pushed forward? E.g. song writing, creative solutions to overcoming obstacles
- ▶ Are there any conflicts?

# Documentation of the botanical knowledge of Northern Khmer speakers

- ▶ Outline of project
- ▶ Motivations
- ▶ Lessons learnt
- ▶ Measuring Impact



# Northern Khmer: Outline

- ▶ document the botanical knowledge of the speakers of the Northern Khmer language in Surin Province, Thailand
- ▶ Interdisciplinary: speakers, botanists, psychologist/CLA, linguists
  - ▶ spoken by 1, 400, 000 people
- ▶ study how this knowledge is passed on to the younger generation in order to design better maintenance and revitalisation programmes that fit the context, by assessing the learning styles traditionally privileged by the Northern Khmer

# Northern Khmer: Motivations

Community-driven: solicited by speakers who have already participated in language documentation project

- ▶ close spiritual-cultural links to their natural environment
- ▶ concern that loss of language leads to loss of knowledge

Language documentation and revitalisation efforts has not paid much attention to the experience of young learners

- ▶ Their well-being, now and in the future: exposure to language and culture influences young children's developing sense of self and community (e.g., Chandler & Lalonde, 2008)
- ▶ Success of LR itself

# Lessons Learnt

- ▶ Speakers involved in documentation of language motivated to continue with 'something practical'
- ▶ Knowledge about the forest: important for identity
- ▶ Interdisciplinarity planned from the start, not as an 'afterthought'

# Northern Khmer: Measuring impact

- ▶ Have the community members continued to be motivated in all steps of the project?
- ▶ How do the people in the villages/diaspora evaluate the findings? How does this compare to the evaluation of the non-local participants? What can we learn from that?
- ▶ Is language revitalization more successful as a component of broader cultural revitalization rather than language-only initiative?

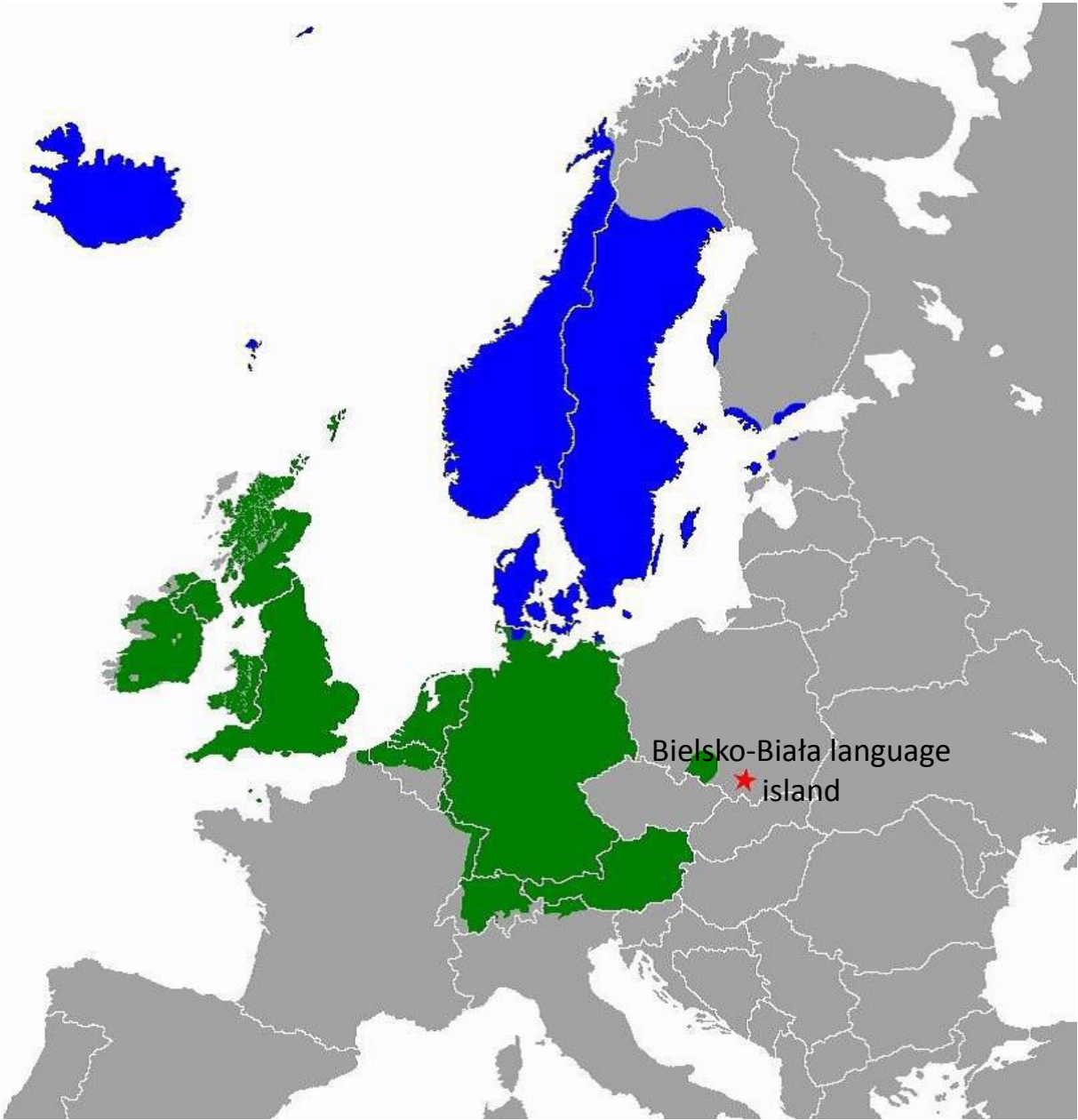
## Section 3

# Documenting Halcnovian





# Germanic languages in Europe



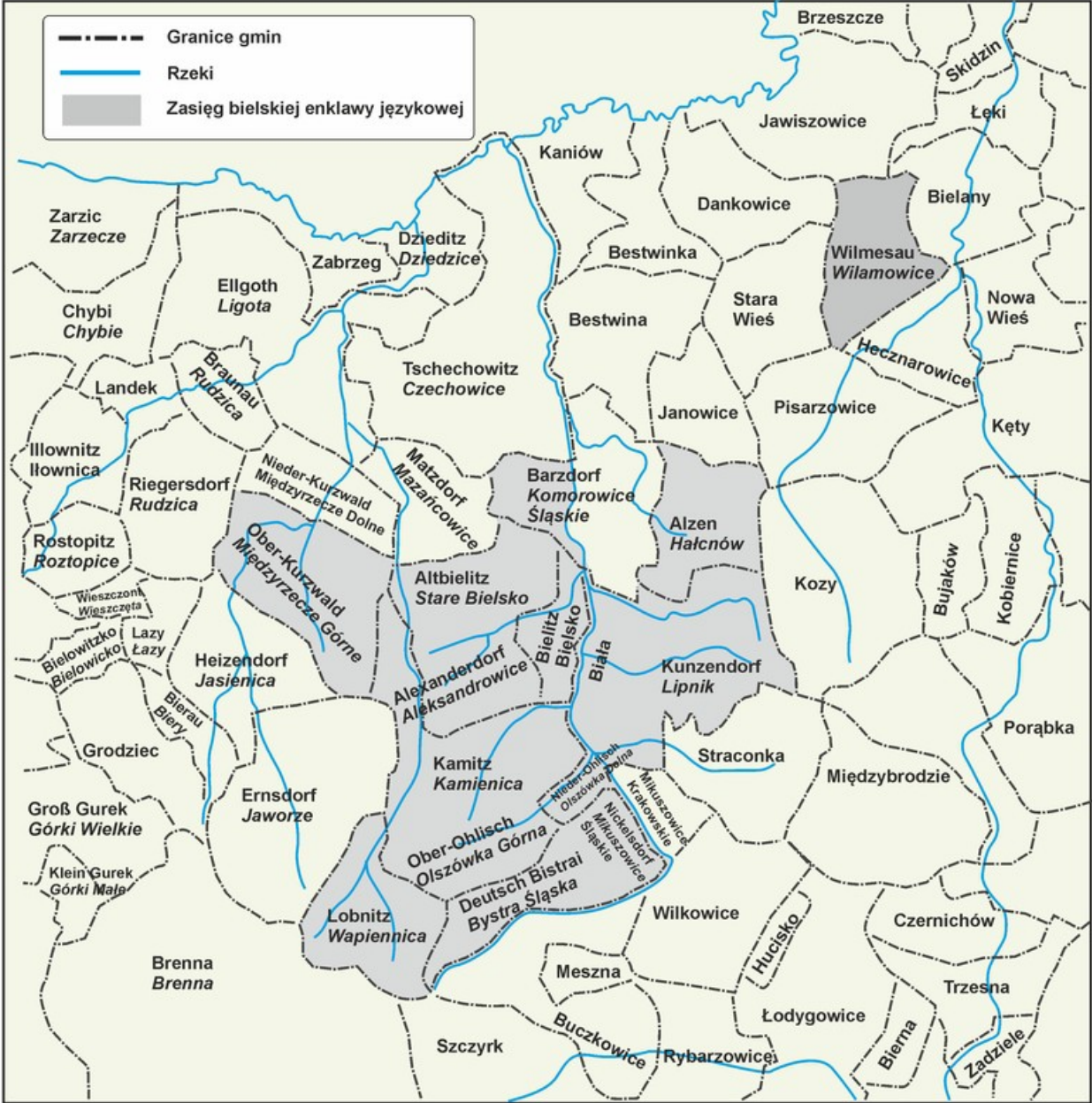
Hałcnów / Alzen / Alza



Hałcnów / Alzen / Alza



# Bielsko-Biała language island pre-WWII



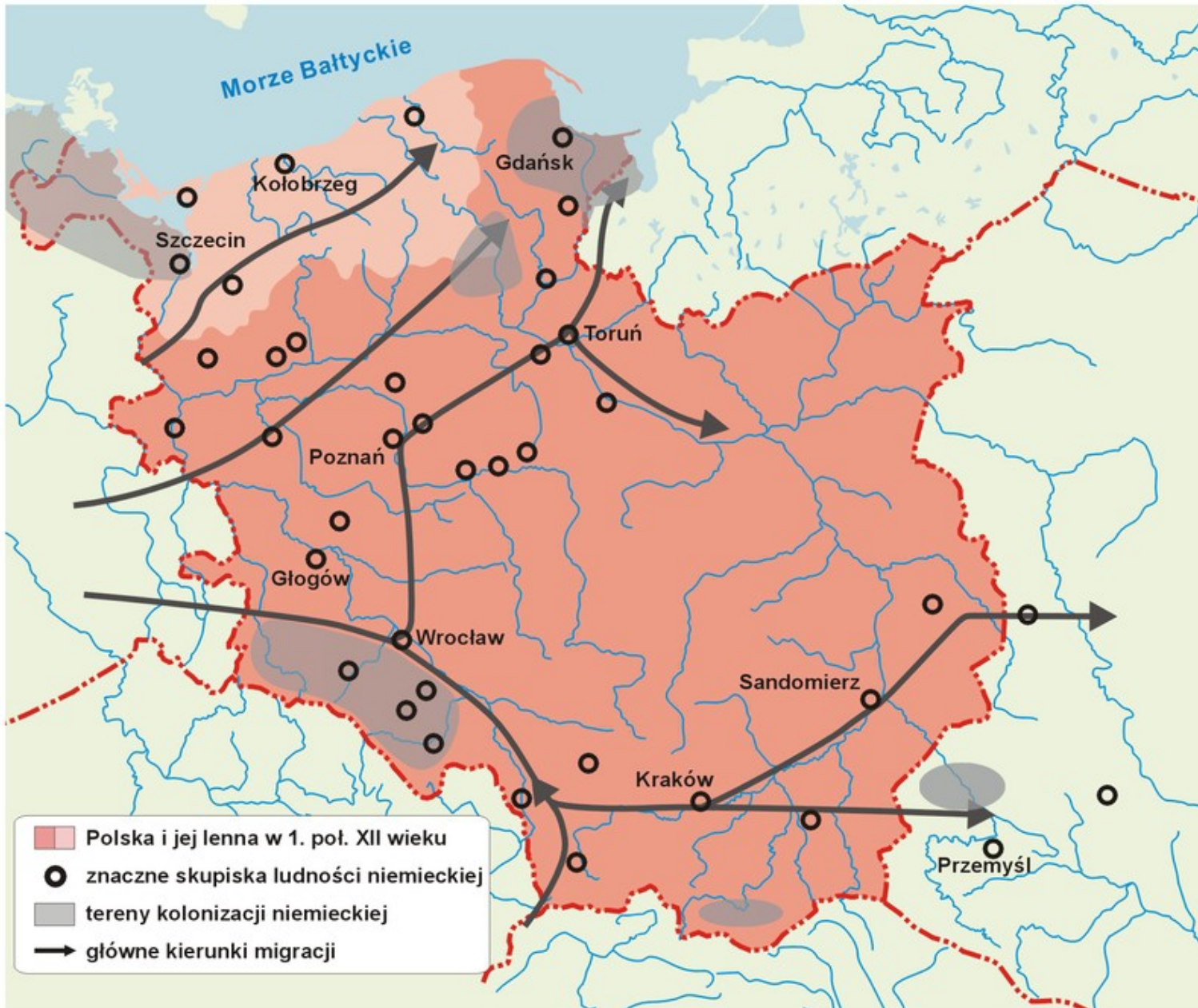
red. Jacek Ciešlewicz, <http://inne-jezyki.amu.edu.pl/>

# Bielsko-Biała language island on current administrative map



red. Jacek Cieślewicz, <http://inne-jezyki.amu.edu.pl/>

# Medieval Germanic colonisation

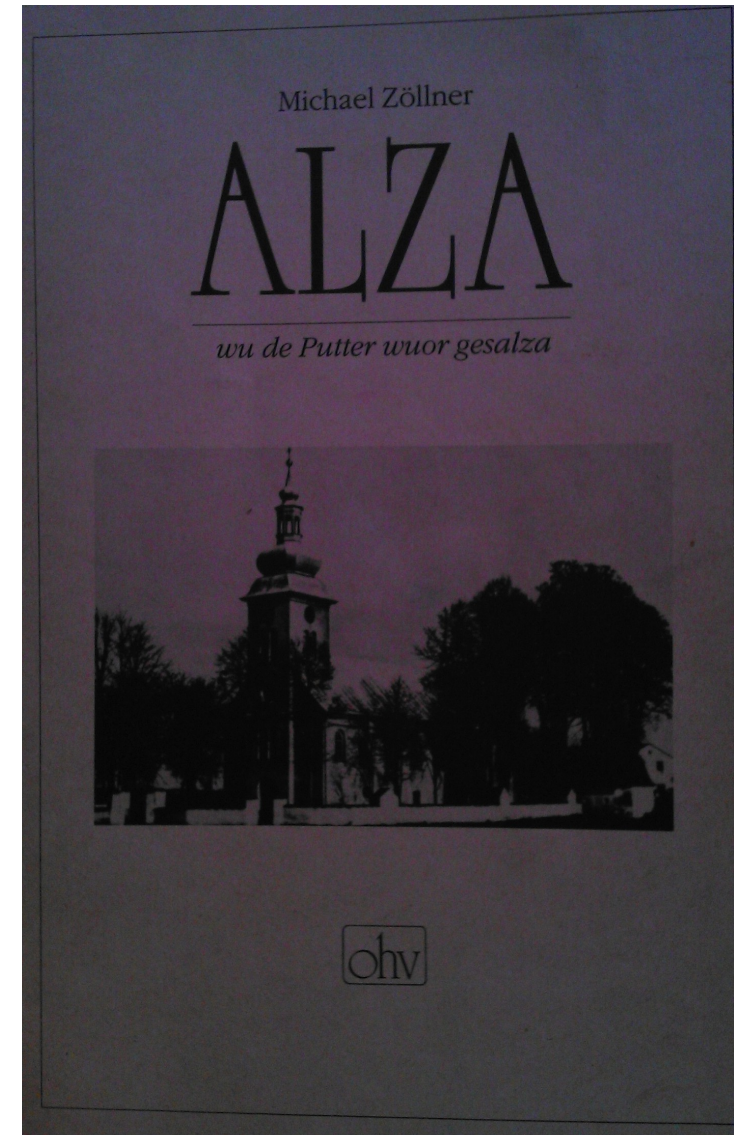


***De jeschta Ajlzner***

*Vu Alzenaa  
em Fränkischa  
met Bow an Kejnd,  
met Fad an Rejnd  
de Schles ganz düöchj  
wie Schnepf an Schtüöchj  
sän se gezünj,  
bi se verlüönj  
em Bajgsgehelz  
ne wäjt vu Beltz  
de Wanderlost...*

**The first peple of Hałcnów**

from Alzenau  
in Franconia,  
with an infant and with a child,  
with a horse and with a calf,  
through the whole of Silesia  
like a woodcock and a stork  
did travel  
until they lost  
their will to walk further  
in a mountain forest  
near Bielsko...



Remnants of bilingual landscape – house numbers, crossroads shrine



Property of W. Kominiak



## Remnants of bilingual landscape – Memorial linden tree (planted 2010)





Jon i Gertruda Olejdz z d. Rozmer





Speakers of Hałcnovian

Karolina Olma



Karolina Owczarz



# List of Hałcnów inhabitants cca. 1944

Haus nr	Korn	
73	Soblich	Franz jekel Hanslik Johann (zu Gals) ...
74	Solich	Josef (Kojn fuski) ...
75	Guralczyk	Stephanie ...
76	Rosner	Johann (zu Paluch) ...
77	Crimina	Franz ...
78	Sakamrat	Johann Lorenz ...
79	Jaroszek	Karl ...
80	Peterko	Franz ...
81	Jantscha	Josef ...
82	Pokladnik	Josef ...
83	Bogner	Johann ...
84	Kitsch	Anna ...
85	Jantscha	Franz ...
86	Jantscha	Johann ...
87	Piesch	Anton ...
88	Kaba	Theresie ...
89	Wiesner	Josef ...
90	Jantscha	Josef Jantscha

Bartel  
Mikolaj Dankowski

Property of Hanusz family

# List of hałcnovian words written down by Józef Jancza

POLSKI	NIEMIETZ	HAŁCNOVIAK
- KOBIECE		<del>KIECH</del> KIECH
- DOM		HOJS
- ULICA		STR OSS
- NIEŚ		DYŃF
- CHŁEB		Brut
- KON		Fad
- MAMA		Müttero
- TATA		Vötys
- BRAT		Brädys
- SIOSTRA		Schwästy
- WÓZ		Wion
- DRZEWO		Hule
- KWIATY		Blüma

# Trilingual signage in Wilamowice (2016)





## Selected references

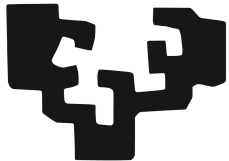
Dolatowski, Marek 2015. *Pochodzenie etnolektu hańcnowskiego w świetle fonetyki i fonologii historycznej*. [In:] D. Lipiński (ed.), *Badania diachroniczne w Polsce* .

Ritchie, Carlo 2012. *Some Considerations on the Origins of Wymysorys*. Thesis for the degree of Bachelor of Arts. Sydney: The University of Sydney.

<http://www.inne-jezyki.amu.edu.pl> [access: 18.12.2016]

Thank you!

eman ta zabal zazu



Universidad  
del País Vasco

Euskal Herriko  
Unibertsitatea



**SOAS**  
University of London

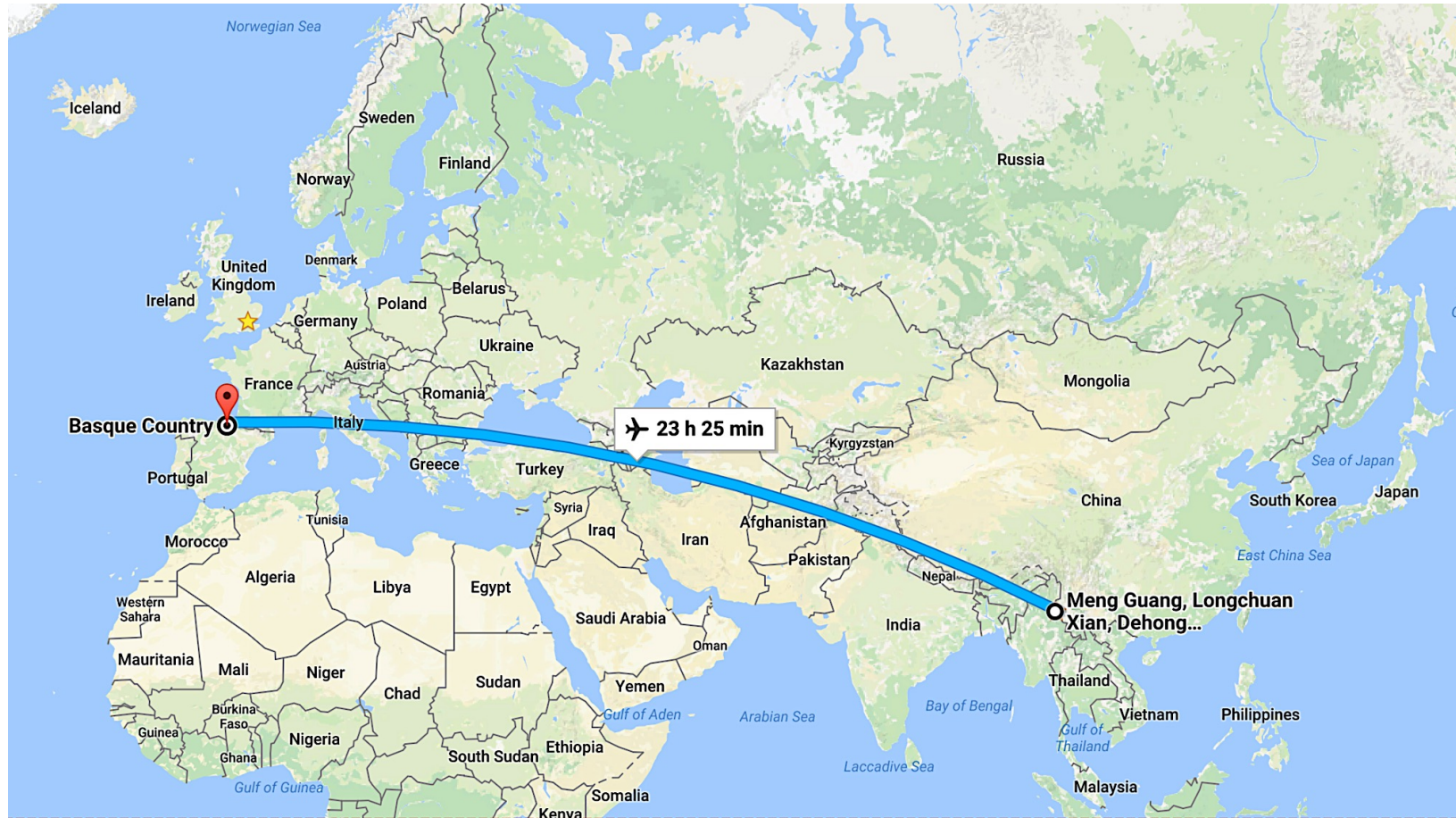
# A tale of two countries

Language ideologies and identities in Mengguang, China and the Basque Country, Spain

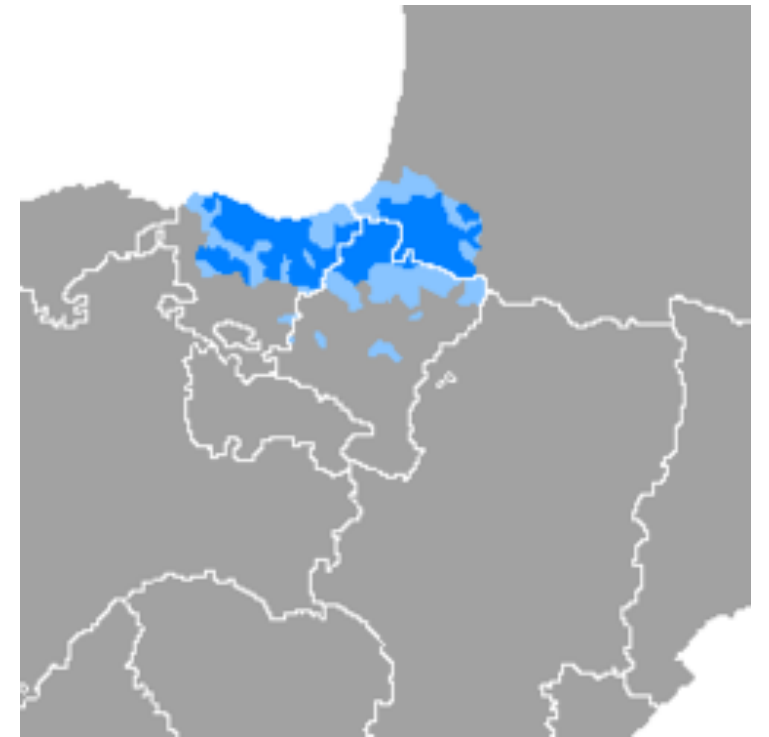
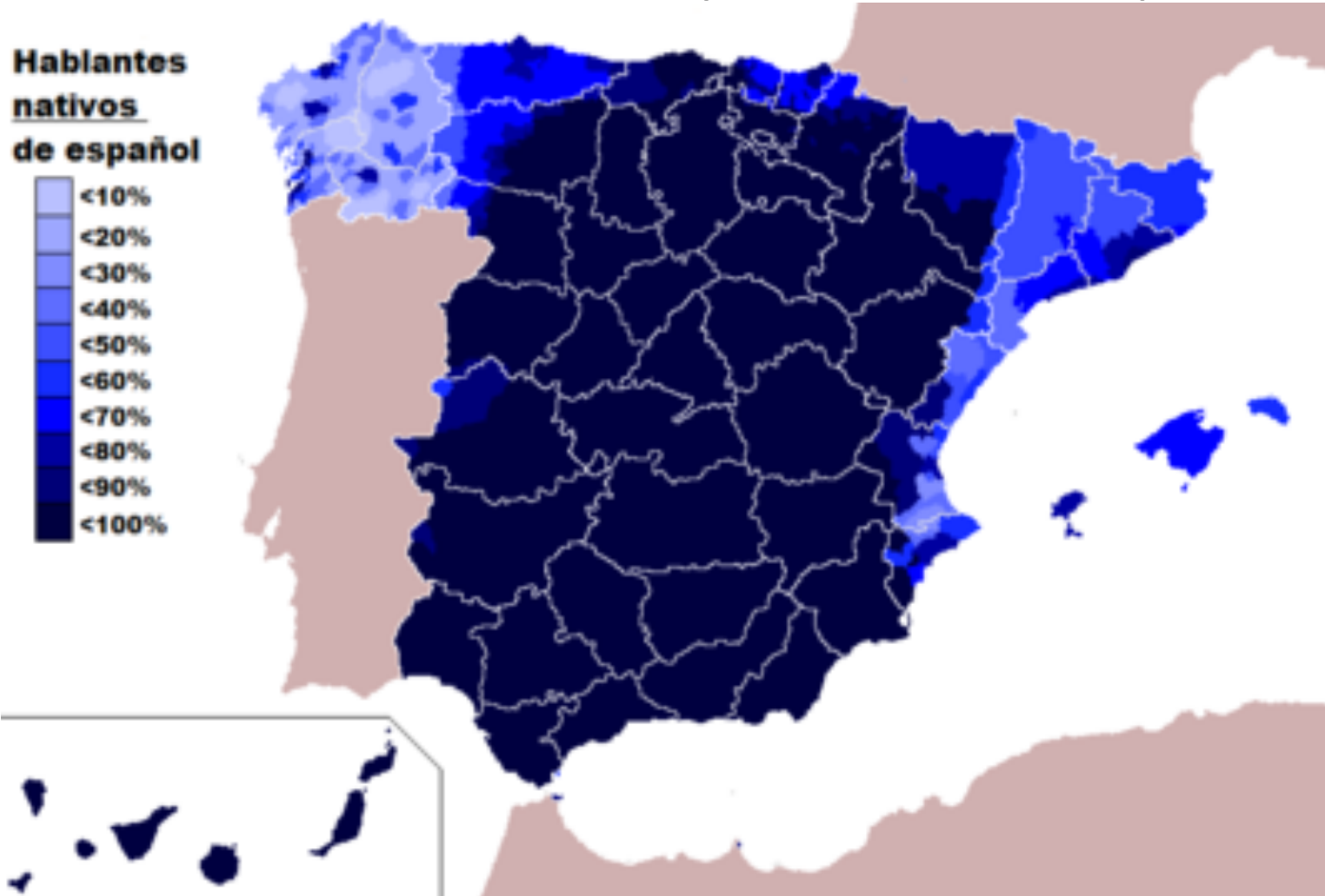
Martin Grabowski (University of the Basque Country)

Yingying Mu (SOAS, University of London)

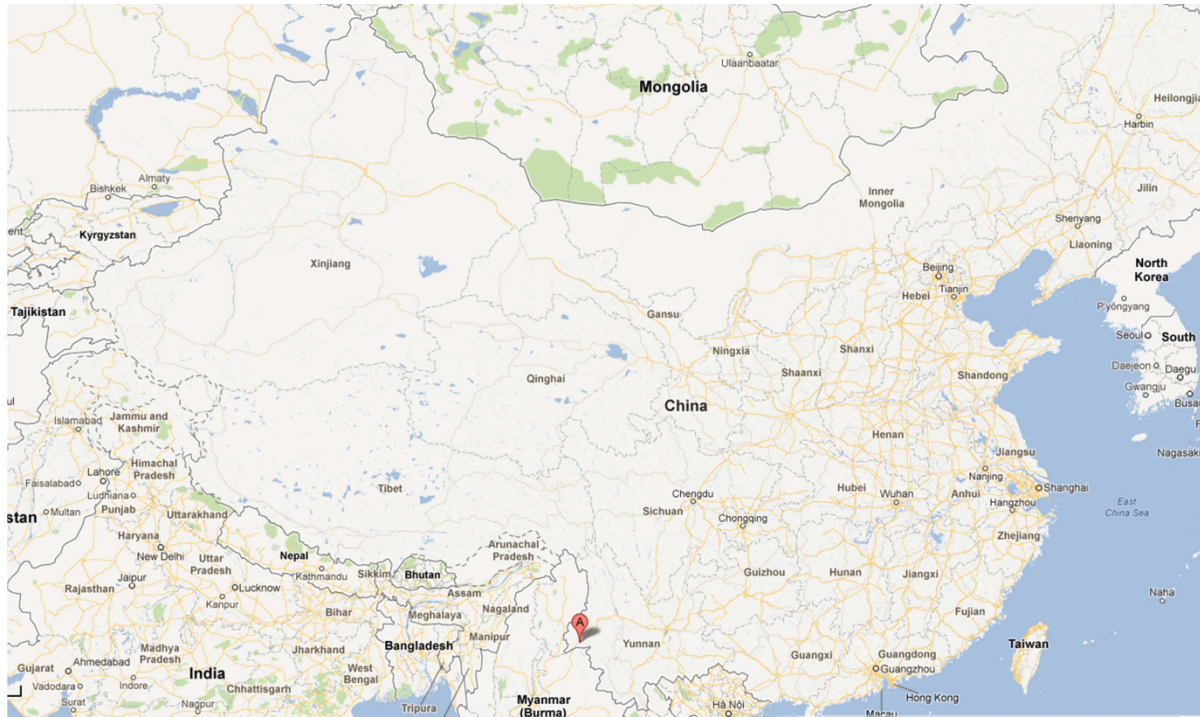
# Where are these two communities?



# Where is Basque Country



# Mengguang and its people



# Languages used in Mengguang



Langwo

Pela

Zaiwa

Han (Southwestern Mandarin)

# Standardization of Basque Language - Euskera

Before 1968, the basques borrowed writing system from nearer country, for example: one of the most popular basque surnames is *Etxeberri*. It was *Etcheberry* for French basques and *Echeverri* for Spanish basques .

“Before the standardization movement, they were not able to understand each other.” –Xabier Kintana

# Academy of Basque Language - Euskaltzaindia

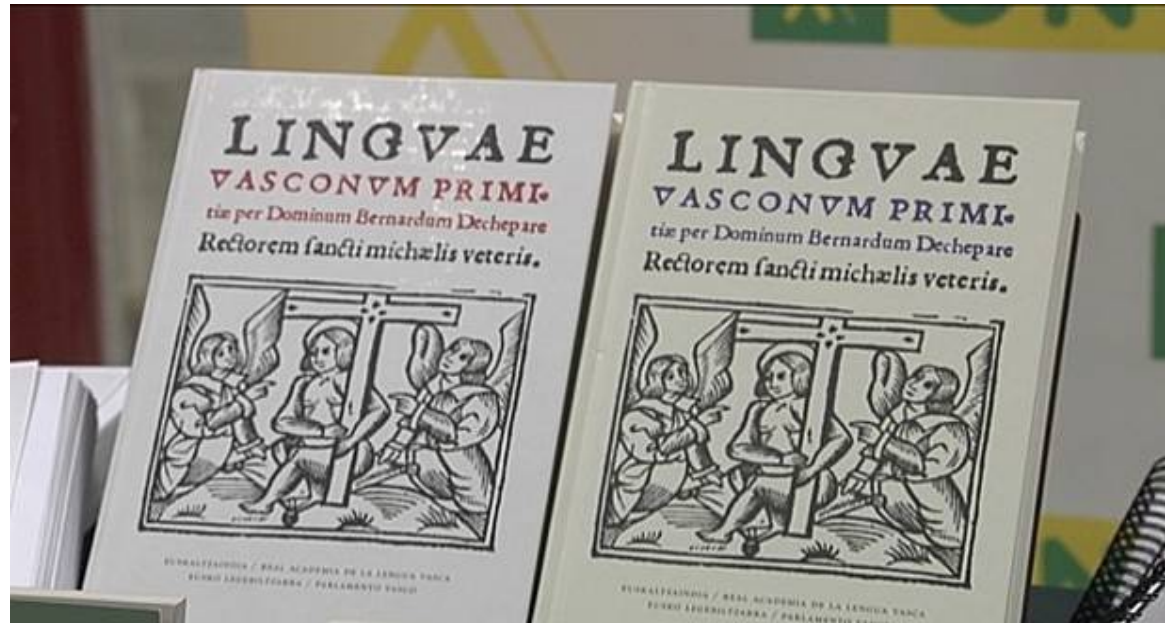
- 1919
- 1968
- 1981



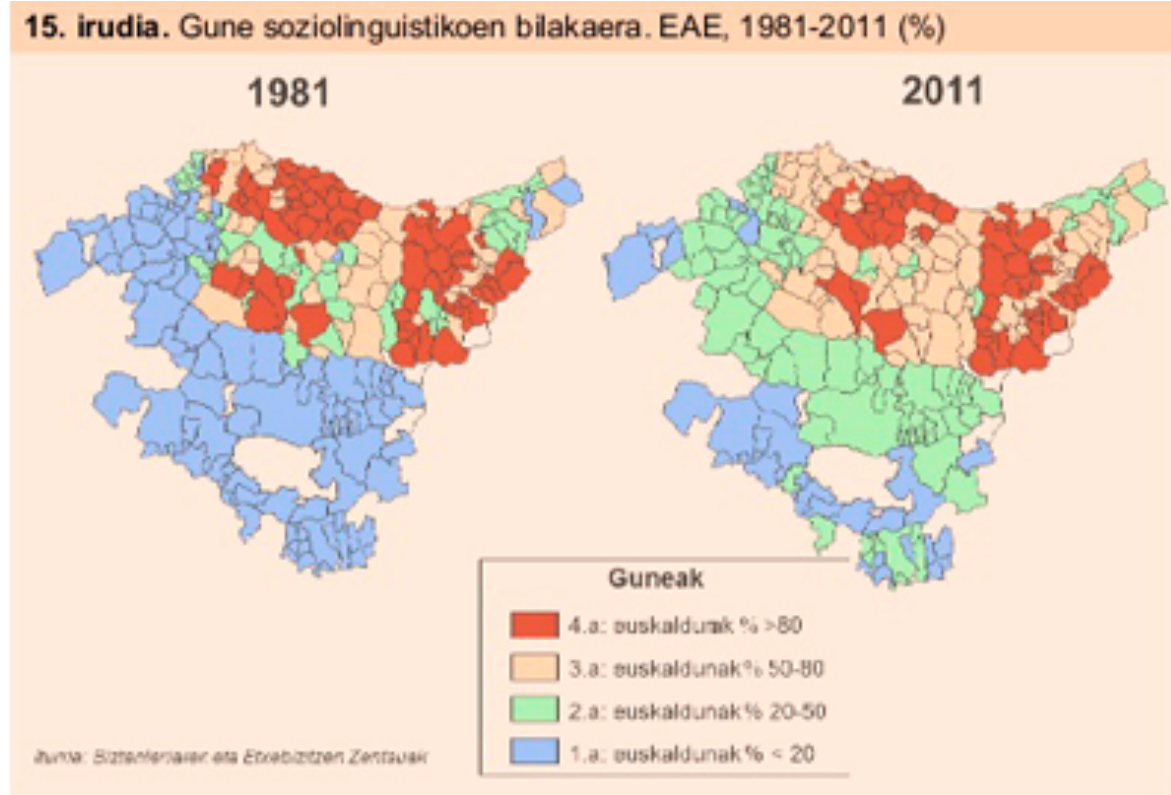


# How did they standardize Euskera

- They were based on:
  - Texts from XVI centuries
  - Central dialects



# Basque at school



Ikastola

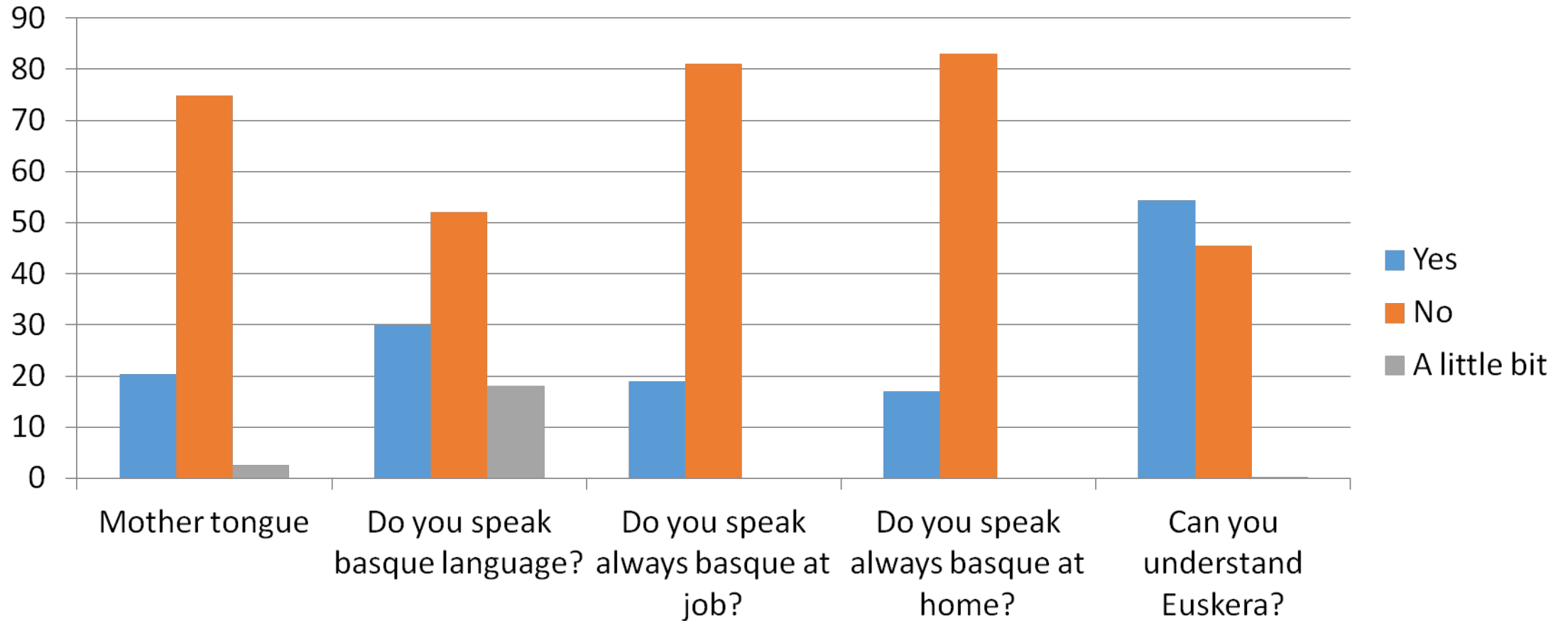
Curso académico 2011/2012<sup>2</sup>

	Modelo A	Modelo B	Modelo D	Modelo X
<b>Educación infantil</b>	5,5%	24,7%	69,1%	0,7%
<b>Educación primaria</b>	5,7%	24,6%	69%	0,7%
<b>Educación secundaria</b>	11,4%	26,7%	61,2%	0,7%
<b>Bachillerato</b>	38,7%	1,6%	59,0%	0,6%
<b>Total</b>	<b>15,8%</b>	<b>19,6%</b>	<b>64%</b>	<b>0,6%</b>

# Others activities



# Language situation in the Basque country now





“Every one can speak all four languages used in the village, it’s not something special” Pela man, 31

It is very common to hear more than one language in one conversation

# Language Practices in Mengguang

“well, yes we should always speak our own language with everyone in the village and I do always speak Langwo to people who are older than me. But, with younger ones... hmm, whichever is appropriate at that moment.” Langwo man 34

It is expected that one always speak your own language at all time, however, in most cases men follow this rule mainly with older members of the community.

# Language Practices Mengguang

“He used to speak Zaiwa to me when we were dating and once we got married, he started speaking half Pela and half Zaiwa, and now we have children, he only speaks his language (Pela) at home. I’ve always been speaking Zaiwa”.

Zaiwa woman, 32

Men change languages they use with their wife before and after the marriage, it is rarer for women to do so

# Language Practices in Mengguang

“I speak Zaiwa because everyone else does the same.” Pela girl 13

It is rarer to hear more than one language used in a group of children/teenagers.



# Language ideologies concerning language choice

“Don’t steal my language, speak your own!” Zaiwa woman 41

Very clear idea of which language is/is not yours

# Language ideologies concerning language choice

“Showing your *Kongga* is the most important part of an interaction.”  
Pela man, 76

What is *Kongga*?

- Humbleness
- Admiration
- Respect

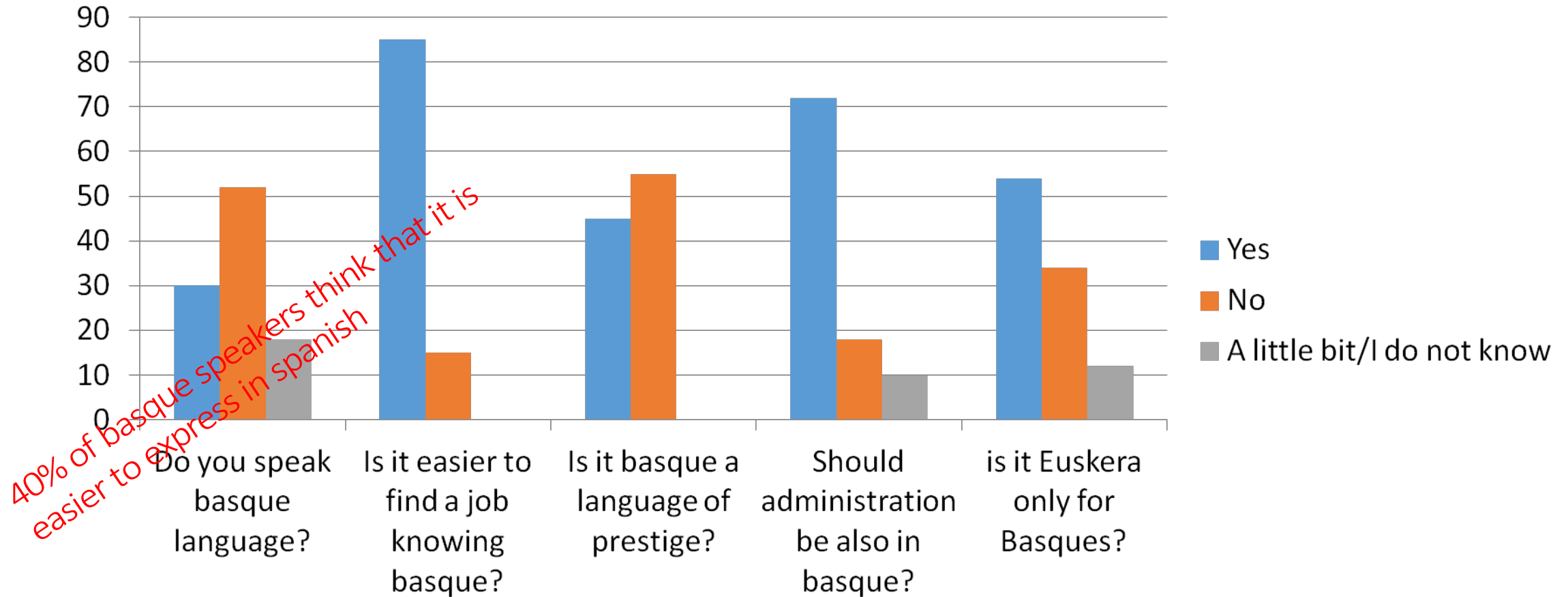
# Language ideologies concerning language choice

“Han, especially standard Mandarin, is the future.” Zaiwa woman, 26

National Language policy

The language used in education

# Ideologies concerning Euskera



# Factors for creating a new language ideology

“I will have a clean hands and I will not speak a basque”

- “Stereotypes”
- Spanish as “better” language (folklore stories for Euskera)
- New language policy
  - Spanish and Basque place names

# What is an identity and how does it appear

- Identity Isolation vs. Interaction
- Identity and Identification
- Epoch of state nation
- Factors
  - Internal
  - Mixed
  - External

Autonomous  
Culture

Alien Culture

Appropriate  
Culture

Imposed  
Culture

# Identities constructed in language choice

- Ethnicity
- Gender
- Group member
- Status

# We have a dream.....

- International interdisciplinary communication and cooperation





Thank you



Eskarikasko

谢谢

Gvetzhu Gvobe

Dziynkujã



# El Código de las plantas: Motivos, Estrategias y Recomendaciones

Ebany Dohle  
SOAS, Universidad de Londres

[e\\_dohle@soas.ac.uk](mailto:e_dohle@soas.ac.uk)

24/08/2017  
San Miguel Xaltipan

# Mesoamerica

Centro de diversidad biológica,  
lingüística y cultural (Nettle and  
Romaine 2000; Gorenflo et al 2012;  
Stepp 2015)

300+ idiomas

Mapa de biodiversidad  
Mesoamericana (CEPF 2016)



	Taxonomic Group	Nr. Species	Nr. Endemic Species	Endemism %
1	Plants	17,000	2,941	17.3
2	Mammals	440	66	15.0
3	Birds	1,113	208	18.7
4	Reptiles	692	240	34.7
5	Amphibians	555	358	64.5
6	Freshwater Fish	509	340	66.8

# Idiomas de Mesoamerica



# El Salvador



# El Salvador – El pulgarcito

- Territorio: 21, 041 km<sup>2</sup>
- Costa Pacífica – el único territorio en Mesoamérica que no tiene costa caribeña
- Área volcánica: 60+ (Sanchez Ruiz 2006)
- Elevación de 600m+
- Un país ‘homogéneo’ con 5 idiomas: castellano (oficial), lenguaje de señas, Potón, Pisbi y Náhuat-Pipil
- La Matanza de 1930 – 5,000-30,000 muertos
- Guerra Civil 1980-1992

# Náhuat, Nawat, Pipil, Náhuat-Pipil

- Código ISO [ppl]
- ‘Críticamente en peligro de extinción’ (UNESCO: Moseley 2010)
- ‘Casi extinto’ (Ethnologue: Lewis, Simons and Fennig 2016)
- Censo Nacional 2008: 200 hablantes
- Censo Nacional 2013: 300 hablantes
- Edad: 50+
- Bilingües en español y náhuat-pipil

# Clasificación yuto-azteca

## Yuto-Azteca: Campbell 1985

- I. Numic (Plateau Shoshone)
  - I. Mono (Monachi), Paviotso
  - II. Panamint, Shoshone
  - III. Kawaaisu, Ute
- II. Tübatulabal (Kern River)
- III. Takic (Southern Californian Shoshone)
  - I. Serrano, +Kitanemuk, +Vanyume, +Alliklik
  - II. +Fernandeño, +Gabrieleño, +Nicoleño
  - III. Cahuilla, Luiseño, Cupeño, +Juaneño
- IV. Hopi
- V. Piman
  - I. Pima, Papago
  - II. Northern Tepehuan, Southern Tepehuan
- VI. Taracahitic
  - I. Tarahumara
  - II. Cahita, Varihio (Guarihio)
  - III. Opata
- VII. Cora-Huichol
  - I. Cora
  - II. Huichol
- VIII. VIII. Nahuat (Aztecan, Nahuatlan)
  - I. +Pochutec
  - II. Pipil
  - III. Core Nahua (All other Nahua varieties)

Note: + = extinct.

## Yuto-Azteca: Miller 1983, 1984

- I. Numic
  - I. Western Numic: Mono, Paviotso (or Northern Paiute) (Bannock)
  - II. Central Numic: Panamint, Shoshoni, Comanche
  - III. Southern Numic: Kawaaisu, Ute (Chemehuevi, Southern Paiute, Ute)
- II. Tubatulabal
- III. Takic
  - I. Serrano-Gabrielino
    - I. Serran: Serrano, Kitanemuk
    - II. Gabrielino (Gabrielino, Fernandino)
  - II. Cupan
    - I. Cupeño, Cahuilla
    - II. Luiseño
- IV. Hopi
- V. Southern Uto-Aztecan
  - I. Sonoran
    - I. Tepiman: Upper Piman (Papago, Pima, Nevome), Lower Piman, Northern Tepehuan, Souther Tepehuan, (Southern Tepehuan, Tepecano)
    - II. Taracahitian
      - I. Tarahumaran: Tarahumara (Eastern Tarahumara, Western Tarahumara, Souther Tarahumara), Guarijio (Upland Guarijio, Lowland Guarijio)
      - II. Opatan: Opata, Eudeve
      - III. Cahita (Mayo, Yaqui)
    - III. Tubar
    - IV. Corachol: Cora, Huichol
  - II. Aztecan
    - I. General Aztec: Pipil, Aztec (Classical Aztec, Tetelcingo, Zacapoaxtla, and others)
    - II. Pochutec



# Motivos





# Estrategias

- Acercamiento a la comunidad
- Medios sociales y Tzunhejekat
- Investigación interdisciplinaria
  - Antropología, etnobotánica, artistas, amigos y colegas
- Conocimiento de las plantas autóctonas
- Apoyo institucional
  - MINED – Ministerio de Educación
  - MUNA – Museo de Antropología
  - UTEC – Universidad Tecnológica
  - SOAS – Universidad de Londres

# Analizando identidad

## IWGIA – Grupo Internacional de Trabajo de Asuntos Indígenas

- Religión y Sistema de valores propios
- Idioma
- Vestimenta
- Ancestros
- Auto-identificación
- Acceso a territorios ancestrales

## Náhuat-Pipil

- Católico and Evangelico
- Español y náhuat-pipil
- Pocos utilizan refajos
- No existen registros
- Poco auto-identificación
- Genocidio 1930s

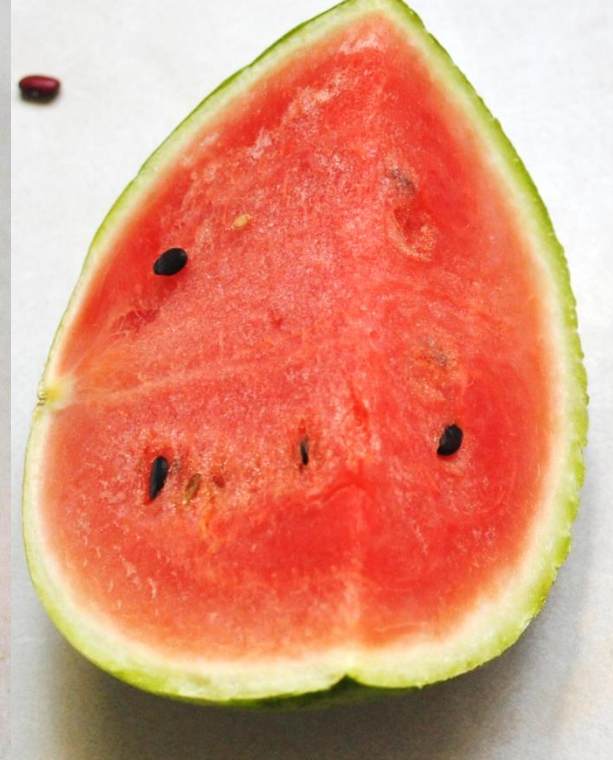


# Identidad náhuat-pipil

- Orgullo por sus tierras
  - Agricultura, alfarería, parteras, curanderos
- Respeto por los 'lugares vivos'
  - Minas de barro, ríos, cuevas y cavernas, cascada
- Historia Oral
  - Cuyancúa, siguanaba, cadejos
  - Desaparecen con el evangelismo
- Idioma
  - 'niweli chupi'

# Documentación Lingüística

- Semántica y la codificación del conocimiento de las plantas y la naturaleza en el idioma náhuat-pipil
  - Categorización de plantas y ‘frutas’
  - Usos rituales o medicinales
  - Usos prácticos – construcción y comida
  - Importancia cultural – puchut (ceiba); wajkal
  - Símbolos y metáforos – se, ume, makwil; nushuchiww
- Lingüística cognitiva – universalismo
  - Existe un modelo universal para nombrar y categorizar las plantas?
  - Existe una cosmovisión cultural náhuat-pipil: conocimiento de grupo vs. Conocimiento individual
  - Cuáles son las categorías léxicas y cuáles son sus prototipos?









# Categorías

- Shukut – Jocote, mamon
- Ayut – Ayote, chayote, chilacayote
- Kilit – Chipilin, loroco, güisquil, añil, berro
- Tzaput – Sapote, nispero, anona
- Chil – Chiles
- Et – Frijoles, chilipucas, arvejas, lentejas
- Kamuj – Yuca, zanahoria, camotillo, camote, batata

# Ne shukut wan ne tzaput

- Acido
- Sostenido por las puntas de los dedos
- Puede ser consumido crudo
- Dulce
- Cabe en la mano
- Semilla dura y brillante
- Partir antes de consumer



# Ne kamuj wan ne kilit

- Raíz
- Textura masosa
- Debe ser cocido antes de consumir
- Hierba o fruta con tayo quebradizo
- Verde
- Debe ser cocido antes de consumer



# Retos

- Cambio climático: Las plantas igual como los idiomas están desapareciendo
- Medicina occidental
- “Estatus”
- Percepción personal
- Desconfianza y miedo a los no conocidos
- Falta de apoyo institucional

# Recomendaciones

- Trabajar en grupo, buscar contrapartes y apoyo
- Considerar el impacto de la geografía, el contexto histórico, el clima
- Considerar y respetar los motivos e intereses de las contrapartes indígenas
- Considerar y respetar los motivos e intereses de

Padiush!

# El uso de idioma(s) en Essyl, Senegal.

Reuniendo actitudes lingüísticas  
hacia el monolingüismo y  
situaciones naturales del  
multilingüismo.



**24 Agosto 2017**  
**Samantha Goodchild**  
**sg76@soas.ac.uk**



**The Leverhulme Trust**

# El índice de la ponencia

2

- Introducción al contexto
- Presentación del proyecto doctoral
- Repertorios lingüísticos diversos
- Ejemplos del multilingüismo
- ¿Cómo reunir el uso de idiomas naturales con actitudes lingüísticas?
- ¿Se puede describir Essyl como una aldea monolingüe?
- Conclusión
- Bibliografía



# Introducción al contexto Senegal – África del oeste



# Introducción al contexto Essil en un reino de 10 aldeas



# Introducción al contexto

## Idiomas oficiales y nacionales

- Idioma oficial
  - francés
  
- Idiomas nacionales
  - wolof
  - joola
  - soninke
  - sereer
  - pulaar
  - mandinka

# Introducción al contexto

## Algunas de las idiomas – Familia Atlántica

(Podzniakov & Segerer en prensa)

6

### ▪ **Grupo Joola**

Joola Eegimaa/Joola

Banjal

Joola Hitou

Joola Buluf

Joola Youtou

Joola Fogny

Joola Kaasa

Joola Kujireray

Arame

Bayot

### ▪ **Grupo Nyun**

Bainounk Gubëeher

### ▪ **Grupo Fula-Sereer**

Pulaar

Sereer

### ▪ **Grupo Wolof**

Wolof

### ▪ **Grupo Manjaku**

Mankanya

# Introducción al contexto

## Algunas idiomas más...

7

### ▪ Familia Mande

Mandinka

### ▪ Familia Indo-Europea

francés

inglés

castellano

alemán

italiano

### Criollos portugueses

Kriolu Casamance

Kriolu Guinea-Bissau

# Presentación del proyecto doctoral

- Sagna (2008; 2016) y Sagna & Bassène (2016: 44) describe Essil como una aldea monolingüe – Joola Eegimaa
- Tesis sobre la movilidad y su conexión con el repertorio lingüístico
- Además utilizando grabaciones de conversaciones “naturales” para establecer si es una aldea monolingüe dentro del multilingüismo extensiva de la región

# Repertorios lingüísticos diversos

9

- Calvet & Dreyfus (1990) “*séjours linguistiques*” ‘estancias lingüísticas’

- **VB**

- Joola Eegimaa
- Francés

Essil en Mof Avvi

Ziguinchor (escuela)

- **RM1**

- Joola Eegimaa
- Francés
- Mandinka
- Kriolu
- Wolof

Enampor en Mof Avvi

Dakar

Brikama en Gambia

Banjul en Gambia

Ziguinchor

Essil

- Ejemplos del multilingüismo
  - Ver clip ELAN ESS080217SG2a\_CUT1 – 68
  - Ejemplo de multilingüismo joola “intra-familial”
  - Ver clip ELAN ESS090316SG1b –
  - Ejemplo de multilingüismo “inter-familial”





## ¿Cómo reunir el uso de idiomas naturales con actitudes lingüísticas?

- “*et moi qui ne parle que le joola*” ‘y yo solamente hablo el joola’ VB [ESS040317SG1].
- RM1: joola eegimaa – pero también joola fogny, joola kujireray – multilingüismo receptivo dentro de “un idioma”

# ¿Se puede describir Essyl como una aldea monolingüe?

12

- Percepciones de monolingüismo – ayuda a mantener el idioma dentro de una situación multilingüe
- Quiere decir que la gente que viene a vivir en Essyl sea esperado a adquirir el Joola Eegimaa
- Joola – al mismo tiempo una práctica lingüística mono- y multilingüe

# Conclusión

- Importante incluir actitudes lingüísticas para ayudar a un idioma en peligro.
- Cómo podríamos crear un lugar monolingüe dentro de las ciudades para la gente en la diáspora?
- Essyl es al mismo tiempo una monolingüe y multilingüe.

- Calvet, Louis-Jean, & Martine Dreyfus. 1990. “La Famille Dans l’espace Urbain: Trois Modèles de Plurilinguisme.” *Plurilinguismes* 3 (1990): 29–54.
- Pozdniakov, Konstantin, & Guillaume Segerer. En prensa. “A Genealogical Classification of Atlantic Languages.” In *The Oxford Guide to the Atlantic Languages of West Africa*, edited by Friederike Lüpke. Oxford: Oxford University Press.
- Sagna, Serge. 2008. “Formal and Semantic Properties of the Gújjolaay Eegimaa: (A.k.a Banjal) Nominal Classification System.” PhD, School of Oriental and African Studies, University of London.
- Sagna, Serge. 2016 “‘Research Impact’ and How It Can Help Endangered Languages.” Edited by Christopher Moseley. *OGMIOS Newsletter* 59, no. 30 April 2016.

# Bibliografía

- Sagna, Serge, & Emmanuel Bassène. 2016. “Why Are They Named after Death? Name Giving, Name Changing and Death Prevention Names in Gújjolaay Eegimaa (Banjal).” Edited by Mandana Seyfeddinipur. *African Language Documentation: New Data, Methods and Approaches*, Language Documentation and Conservation, no. Special publication 10: 40–70.

# Gracias

- Thank you
- Tlazokamantli
- Yoo
- Merci

**UN DIÁLOGO SOBRE LA  
COLABORACIÓN  
INTERDISCIPLINARIA ENTRE  
ETNOMUSICÓLOGOS Y  
SOCIOLINGÜISTAS EN LA  
REVITALIZACIÓN DE LA LENGUA Y  
DE LA MÚSICA**

**ENGHUM  
ESCUELA XALTIPAN  
Agosto 2017**

**Sara Selleri**

**SOAS  
London  
(UK)**

# QUÉ ES LA ETNOMUSICOLOGÍA?

- La Etnomusicología es el estudio de la música desde los aspectos culturales y sociales de las personas que lo hacen. También ha sido definida como “música en cultura”, “música y cultura” y “música como cultura”.
- En términos generales, la etnomusicología puede describirse como una investigación holística de la música en sus contextos culturales. Combinando aspectos del folclore, la psicología, la antropología cultural, la lingüística, la musicología comparada, la teoría musical y la historia, la etnomusicología ha adoptado perspectivas desde una multitud de disciplinas.
- Entre los libros fundamentales de la disciplina, tienen:
  - *The Anthropology of Music*, Alan P. Merriam (1964)
  - *How Musical is Man?*, John Blacking (1974)



# VISIÓN GENERAL/ESTUDIOS ACTUALES EN LA COLABORACIÓN ENTRE EL CAMPO DE LA ETNOMUSICOLOGÍA Y DE LA SOCIOLINGÜÍSTICA

- Existe una larga relación entre los campos cuyos fundamentos fueron consolidados por el antropólogo lingüístico Feld (1974, 1990, Feld & Fox 1994)
- Ejemplos recientes incluyen trabajo de etnomusicólogos Linda Barwick y Allan Marett con la lingüista Lysbeth Ford en la región *Daly River* del territorio del Norte de Australia (2013) y de la etnomusicóloga Monika Stern con el lingüista Alexandre François en las regiones del Norte de Vanuatu (2015).
- Varias publicaciones y conferencias académicas recientes han explorado el trabajo interdisciplinario en temas de *endangerment* y/o revitalización incluyendo *Language Documentation and Description* en lenguaje, prosodia y música (Svantesson et al, 2012) y la conferencia en 2015 de la *Foundation for Endangered Languages* en la temática de “música de las lenguas en peligro de extinción” (FEL, 2015).

# VISIÓN GENERAL/ESTUDIOS ACTUALES EN LA COLABORACIÓN ENTRE EL CAMPO DE LA ETNOMUSICOLOGÍA Y DE LA SOCIOLINGÜÍSTICA

Otros trabajos interdisciplinarios han investigado las interconexiones directas entre la vitalidad y la viabilidad de las lenguas y de los géneros musicales, como:

- Faudree (2013) explora los papeles de palabra y canción dentro del renacimiento cultural de los indígenas en México
- Johnson (2011, 2015) examina las conexiones entre la canción y la revitalización de lenguas en peligro en la isla de Jersey
- Minks (2013) traza el papel de la música en la revitalización de la lengua en la costa atlántica de Nicaragua
- Además, varios proyectos aplicados han utilizado la música como un vehículo para ayudar a fortalecer las lenguas en peligro de extinción, como la que usó la canción como una forma de reclamar el *Dharug* de Australia (descrito por Green, 2010)

CATHERINE GRANT (ENERO 2017)

**A CASE FOR GREATER INTERDISCIPLINARY COLLABORATION IN LANGUAGE  
AND MUSIC REVITALIZATION**

**UNA PROPUESTA PARA UNA MAYOR COLABORACIÓN INTERDISCIPLINARIA  
EN LA REVITALIZACIÓN DE LA LENGUA Y LA MÚSICA**

Ella expone tres razones específicas para una mayor colaboración entre las dos disciplinas en temas de revitalización y sostenibilidad:

- 1) En relación con el valor de un lenguaje compartido – referenciando el marco conceptual *intangible cultural heritage* que abarca tradiciones y expresiones orales, incluida la lengua como vehículo del patrimonio cultural inmaterial, así como las artes escénicas y la música (UNESCO, 2003a, Art 2)
- 2) Para desarrollar tanto la teoría como la práctica – en relación con la colaboración entre sociolingüistas y etnomusicólogos que podría conducir al avance de marcos teóricos versátiles y herramientas para la investigación aplicada
- 3) Para reforzar la defensa y el activismo en torno a la salvaguardia y el apoyo a los idiomas, la música y otras expresiones intangibles de la cultura – en 3 niveles: *advocacy* interna, académica y pública.

DAVID SAMUELS (DEPARTMENT OF MUSIC, NYU)  
*MUSIC'S ROLE IN LANGUAGE REVITALIZATION -  
SOME QUESTIONS FROM RECENT LITERATURE*  
PAPEL DE LA MÚSICA EN LA REVITALIZACIÓN DE LA LENGUA -  
ALGUNAS PREGUNTAS DE LA LITERATURA RECIENTE

- Faudree, Paja. 2013. *Singing for the Dead: The Politics of Indigenous Revival in Mexico*. (trad. Cantando para los Muertos: La Política del Renacimiento Indígena en México). Durham: Duke University Press.
- Grant, Catherine. 2014. *Music Endangerment: How Language Maintenance Can Help*. (trad. Music Endangerment: Cómo el mantenimiento del idioma puede ayudar.) New York: Oxford University Press.
- Minks, Amanda. 2013. *Voices of Play: Miskitu Children's Speech and Song on the Atlantic Coast of Nicaragua*. (trad. Voces de Juego: Discurso y canto de miskitu en la costa atlántica de Nicaragua). Tucson: University of Arizona Press.

THE NEW YORK TIMES (MARCH 30, 2016)  
*VANISHING LANGUAGES, REINCARNATED AS MUSIC*  
IDIOMAS QUE ESTÁN DESAPARECIENDO,  
REENCARNADAS COMO MÚSICA

- Un número cada vez mayor de compositores está fijando su atención en idiomas extintos, en peligro de extinción o distintivos de pequeños grupos de hablantes en lugares lejanos, con el objetivo de tejerlas en obras musicales que celebran, conmemoran o lamentan las lenguas y las culturas que les dieron a luz.
- Ejemplos: Opera de Cologne (Alemania) donde la compositora Australiana Liza Lim incluye una lengua de silbido turco de un pequeño pueblo de montaña en su composición; la Filarmónica de Nueva York interpretó la sinfonía multimedia "Nu Shu" de Tan Dun, el resultado de la investigación del compositor sobre un sistema de lenguaje y escritura transmitido entre las mujeres habitantes de una pequeña aldea en la provincia de Hunan en China durante 700 años.
- El Sr. James, fundador del *Vanishing Languages Project*, en una reciente entrevista dijo que el objetivo era "no poner las lenguas a la música, sino ponerlas como música", en referencia a la gama de sonidos a los que diferentes lenguajes pueden abrir el cuerpo humano y la voz. Entre sus preocupaciones están también cuestiones éticas de extraños que obtienen beneficios económicos o el prestigio del uso de estas lenguas, o el uso de las voces registradas de los muertos en las culturas donde eso es tabú.

# TRABAJO DE CAMPO – PUERTO RICO

## CONTEXTO SOCIAL, POLITICO Y CULTURAL

- **CONTEXTO SOCIAL y CULTURAL** – La realidad socio-cultural de Puerto Rico es una singular mezcla de influencias Tainos-amerindias, españolas, europeas, africanas y norteamericanas, dando lugar a múltiples capas de culturas, grupos étnicos, tradiciones y creencias.
- **CONTEXTO POLITICO** – La Carta Autonómica que concedió la autonomía política y administrativa a la isla se aprobó en España en 1897 después 4 siglos de dominio colonial, concediendo a Puerto Rico un gobierno autónomo; sin embargo, los EE.UU. se apoderaron de la isla quitándole el control a España en 1898 durante la Guerra Española-Americana, a pesar de que España ya no tenía posesión de la isla desde 1897. Desde entonces, Puerto Rico y los EE.UU. han sufrido una relación política y económica complicada y controvertida. Entre las implicaciones de este fenómeno está el último referéndum de junio de 2017, que ha mostrado una sociedad polarizada entre *estatalistas* (que están a favor de convertirse en el estado número 51 de los EE.UU.) e *independentistas* (que están a favor de separar del EE.UU.).

# TRABAJO DE CAMPO – PUERTO RICO

## CONTEXTO MUSICAL Y EDUCATIVO

- **CONTEXTO MUSICAL Y EDUCATIVO** – Esta realidad socio-política tiene conexiones directas y repercusiones significativas sobre la realidad musical y educativa. De hecho, la música y la educación han sido explotadas como instrumentos para limpiar o disminuir diversidades culturales autóctonas en favor de los modos de vida de los colonizadores, como: la importación de formas y contenidos de enseñanza de música clásica europea bajo los españoles y luego Estados Unidos; la imposición del inglés como lengua obligatoria en las escuelas en 1902; la música Jazz en los currículos musicales; la noticia de principio de Agosto que da cuenta como la consultora que funge como Secretaria de Educación Julia Keleher (estadounidense) eliminó la *Semana de la Puertorriqueñidad* de las escuelas, demostrando cómo el fenómeno ocurre todavía.
- Por el contrario, la música y la transmisión de músicas autóctonas también han sido utilizadas por los locales para reafirmar su identidad, sus raíces y conexiones con su pasado afro descendiente, que a menudo se minimizan y desprecian en la sociedad y dentro de las prácticas educativas. De hecho, en los últimos años, Puerto Rico ha visto resurgir la enseñanza de músicas y danzas como la *Bomba*, encarnando el anhelo de los grupos históricamente marginados de avanzar su visibilidad y representación en la sociedad puertorriqueña a través y junto a sus músicas. Los grupos más discriminados tratados como ciudadanos de “segunda clase” son generalmente los que se encuentran en la intersección entre clase social pobre, de ascendencia africana, con discriminaciones adicionales hacía las mujeres.

# BOMBA PUERTORRIQUEÑA





# TRABAJO DE CAMPO – PUERTO RICO

## LA BOMBA PARA FORTALECER COMUNIDADES

Artículo de *El Nuevo Día* de 1 Agosto 2017 titulado “La Bomba para fortalecer comunidades; mediante el baile, el Taller Tambuyé ayuda a identificar metas y valores”

- **Marién Torres López fundó en el 2003 la organización Tambuyé para ayudar a su comunidad a fomentar la autogestión y formar parte del quehacer cultural a través de la Bomba. Ella explica como empezó a desarrollar espectáculos en comunidades vulnerables poniendo en escena las problemáticas de esas comunidades como el desplazamiento, la expropiación y la violencia, donde los/las que actúan son personas de esas mismas comunidades. A través del baile y la música de la Bomba consiguen transmitir valores y cultura en las comunidades, así que se puedan re-apoderarse de su historia y su cultura. “En la medida que sabemos quiénes somos, podemos tomar decisiones de hacia dónde vamos” ella comenta.**
- **Por ejemplo, este año el Taller Tambuyé ha presentado su espectáculo anual de Bomba, bajo el título “Perspectivas”. Fue un homenaje a seis puertorriqueñas: mujeres que, en su época, rompieron esquemas y estereotipos que se esperaban de una mujer en la sociedad. “Ellas con su lucha desde diferentes disciplinas, nos abrieron el camino a muchas mujeres” (Torres).**

# TRABAJO DE CAMPO – PUERTO RICO

## LA BOMBA PARA FORTALECER COMUNIDADES



# TRABAJO DE CAMPO – PUERTO RICO

## DESIGUALDADES EN LA SOCIEDAD, EDUCACIÓN Y ENSEÑANZA DE LA MÚSICA

- **DESIGUALDADES EN LA SOCIEDAD** – La percepción de ciudadanos de "primera clase" vs "segunda clase" basada en la clase social, la etnia, el género y los orígenes nacionales también se muestran a nivel educativo. Para especificar, "primera clase" abarca en su mayoría a los descendientes "blancos" de América del Norte o de Europa, por lo general de alto nivel educativo y medios económicos, mientras que "segunda clase" comprende a personas pobres y de clase baja, a menudo afro descendientes o mixtos *Latinos*.
- **DESIGUALDADES EN LA EDUCACION Y ENSEÑANZA DE LA MUSICA** – Estas discriminaciones socioculturales se reflejan en las prácticas musicales y en los entornos educativos, ya que se reconocen diferentes *status* a las expresiones musicales de los dos grupos. La música clásica de Europa y el Jazz estadounidense son vistos como formas "superiores", mientras que la música tradicional local autóctona incluyendo Salsa, Bomba y Plena son consideradas "inferiores". Esto se refleja en su papel en la instrucción formal: todos los grados oficiales de música universitaria en Puerto Rico están en clásica, jazz o música popular. La Bomba, Plena y Salsa se enseñan principalmente por la tarde y después de la escuela en clases privadas y actividades extra-curriculares.

# TRABAJO DE CAMPO – PUERTO RICO

## IMPORTANCIA DE LA ENSEÑANZA Y TRANSMISIÓN DE FORMAS AUTÓCTONAS EN RELACIÓN A LA PRESERVACIÓN DE EXPRESIONES ARTÍSTICAS, RAÍCES ÉTNICAS Y LINGÜÍSTICAS

- A pesar de esto, la enseñanza de estas formas autóctonas de música y baile tiene un papel fundamental en el redescubrimiento de las raíces étnicas y lingüísticas.
- Por ejemplo, la Bomba permite encontrar enlaces directos a las raíces africanas de la población puertorriqueña a través de sus ritmos y pasos de danza, mientras que las letras de las canciones, también de acuerdo a las diferentes zonas de la isla, permite recuperar terminologías y formas lingüísticas que incorporan la historia del país y las diversas influencias que ha sufrido durante siglos.
- La mayoría de los/as académicos/as y músicos/as locales que entrevisté creen que el hecho de que no hay grado o título oficial en las músicas puertorriqueñas y del Caribe es causado exclusivamente por el estatus de segunda clase que se otorga a las formas musicales autóctonas y no está relacionado con su valor musical intrínseco; de hecho, es causado por ser históricamente expresiones artísticas de ciudadanos de “segunda clase”, emergiendo así una discriminación integrada en el intercambio entre las prácticas socioculturales, musicales y educativas.

# TRABAJO DE CAMPO – PUERTO RICO

## CONCLUSIONES SOBRE LA IMPORTANCIA DEL DESARROLLO DE INSTITUCIONES Y PROGRAMAS EDUCATIVOS INCLUSIVOS

- Para contrarrestar la perpetuación de las dinámicas históricamente heredadas de discriminación crónica y políticas abusivas, la comunidad académica, los órganos políticos y la sociedad civil tienen que trabajar juntos para crear medidas que pueden tener un impacto.
- La educación juega un papel fundamental, ya que es clave para dar forma al pensamiento de las generaciones actuales y futuras. Por lo tanto son cruciales aplicaciones no etnocéntricas o centradas en los países ‘desarrollados’ del *Norte*, la introducción de una perspectiva de género y garantizar la representación equitativa de las expresiones de grupos históricamente excluidos como los indígenas y las comunidades vulnerables en las instituciones educativas y en los programas escolares. Es igualmente imperativo traer a la superficie dinámicas de exclusión e invisibilización al fin de contrarrestar la reiteración perpetua, la asimilación pasiva y la aceptación implícita de la injusticia y la intolerancia en la sociedad y en la escuela.

# **MÚSICA, EDUCACIÓN Y INCLUSIÓN SOCIAL**

## **ICTM (INTERNATIONAL COUNCIL OF TRADITIONAL MUSIC)**

- **El foco temático del simposio giró en torno a grupos históricamente excluidos como las mujeres y las niñas, las minorías étnicas, las personas con vulnerabilidades (incluidos los discapacitados y otros grupos marginados) que han estado sistemáticamente no representados o están infrarrepresentados en la educación musical.**
- **La exclusión refleja prácticas más amplias socialmente discriminatorias que se perpetran y transmiten a través de la escolarización, en formas que dan forma a la sociedad en general más allá de las escuelas y las instituciones académicas.**
- **Los trabajos se centraron en explorar prácticas educativas multifacéticas en relación con un espectro más amplio de temas y pensamientos, incluyendo:**
  - **Educación y Representación**
  - **Cuestiones de Identidad en la Educación**
  - **Inclusión social y educación**
  - **Educación y Desarrollo Internacional**
  - **Etnomusicología, Prácticas de Transmisión (enseñanza / aprendizaje) e Inclusión Social**

# CONSIDERACIONES

- En un contexto de: revitalización de una cultura en peligro; de (re)-construcción y redescubrimiento de identidad; de cuestionar patrones culturales y formas de construir conocimientos; de derechos de pueblos indígenas y de empoderamiento – las expresiones musicales autóctonas también tendrían que ser consideradas, investigadas y puestas en el debate.
- La Música es cultura y expresión de identidad. El manejo de la música de manera utilitarista o en estilo “copia-pegar” donde se ponen textos en Nahuatl a patrones musicales de tradición hispánico/europea puede ser peligroso y cuestionable. Usar la música solo como un medio y de manera exclusivamente utilitarista para la revitalización lingüística podría ser ofensivo desde una perspectiva socio-cultural dentro de las comunidades indígenas.
- Un trabajo de investigación, de búsqueda y que podría visibilizar expresiones musicales indígenas pre-coloniales y contemporáneas es recomendable.
- Un trabajo de documentación de expresiones musicales como canciones de cuna permitiría desarrollar conocimientos sobre elementos musicales indígenas pre-coloniales como: melodía, armonía, estructura musical, ritmos, timbres, sonidos, instrumentaciones, etc. y podría abrir las puertas para una revitalización músico-lingüística más cercana a las raíces musicales y culturales de las poblaciones indígenas.

Escuela de campo de Xaltipan  
19 – 29 de agosto de 2017  
Proyecto "Engaged Humanities"

## ACTITUDES LINGÜÍSTICAS

GRUPO 3: Gladiola Villegas, Martín Tonalmeyotl,  
Abel Hernández, Samantha Goodchild, Tomasz  
Wicherkiewicz, Gregory Haimovich, Refugio Nava  
Nava, Genner Llanes



# Interlocutores Comunitarios

- Marcelina Cahuantzin (78)
- Florentino Rosete (70)
- Josefina Rivera (32)
- Constantina Bautista (71)
- Nieves Flores (70)
- Angel Acoltzi (57)
- Rosa Villalba (53)
- Benito Muñoz Xelhuantzin (57)
- Maximina Xochitemol (78)

## Marcadores socio-lingüísticos:

- Edad o generación.
- Clase social o nivel socioeconómico.
- Nivel de formación.
- Competencia lingüística (primera o segunda lengua).
- Actitudes lingüísticas / ideologías.

# Edad o generación

- Gente mayor de 70-80 años quienes crecieron escuchando y/o hablando náhuatl.
- Gente mayor de 40 años quienes crecieron en una época en la que la lengua náhuatl estaba siendo marginada o en la que se discriminaba a los hablantes.
- Gente menor de 40 años que se han socializado en un espacio ambiguo de revaloración lingüística y de permanencia de racismo contra modos de vida y personas indígenas.

# Clase o nivel socioeconómico

- Antiguos campesinos.
- Trabajadoras del hogar (tortillera).
- Profesores de primaria / secundaria (inglés).
- Artesanas tejedoras. Obreros tejedores.
- -----
- Clase media (relativamente acomodada).
- Clase trabajadora de bajos ingresos.

# Nivel de escolarización

- Personas adultas mayores que no saben leer ni escribir, hablan más y mejor la lengua náhuatl.
- Personas más jóvenes que no tuvieron educación escolarizada, también hablan náhuatl con confianza y orgullo.
- Personas jóvenes con mayor nivel de escolarización, tienen una actitud más ambigua, más ideológica que práctica, respecto a la lengua náhuatl (cuando no la rechazan por completo).

# Competencia lingüística I

- Personas que crecieron hablando la lengua náhuatl cuando eran niños; escuchaban el idioma y lo aprendieron en la práctica. Aprendieron castellano en una etapa posterior, cuando eran adolescentes o adultos.
- Personas que crecieron escuchando y entendiendo náhuatl, pero no lo hablaban porque no era considerado importante. Su lengua de uso cotidiano fue el castellano. Posteriormente, comenzaron a hablar náhuatl.

## Competencia lingüística II

- Personas que crecieron escuchando solamente el castellano y que no entendían el náhuatl, pero que aprendieron más tarde y que se muestran interesados en seguir aprendiendo.

# Actitudes lingüísticas

- Personas para quienes el náhuatl es un vehículo de comunicación cotidiano “predeterminado”. No les afecta tanto la discriminación.
- Personas para quienes el náhuatl es un idioma de uso “alternativo” (sólo lo usan para ciertos temas, pero no para otros). *Son más vulnerables a la discriminación.*
- Personas que muestran una ideología de valoración del idioma náhuatl pero que raramente lo emplean. Representan una oportunidad para la revitalización.



# Conclusiones I

- Los hablantes nativos de náhuatl tienen actitudes diferentes, dependiendo del nivel de escolarización y socioeconómico. A más escolaridad, menos uso de la lengua.
- Hay una diferencia entre los hablantes más viejos y los más jóvenes que también tiene relación con el nivel económico. Personas de clase trabajadora, mayores de 40, hablan más que los menores de 40, escolarizados, tanto de clase media como clase trabajadora.

## Conclusiones II

- La identidad también influye en las actitudes lingüísticas. Los que se identifican con orgullo como “macehualmeh” muestran una actitud y práctica más asertiva que los que se auto-definen como “mestizos”.
- Ninguno de los interlocutores ha tenido una actitud consciente de enseñanza activa del náhuatl a sus hijos. Apenas comienzan a tenerla.