Language documentation for language revitalisation and support -- what would it look like?

Peter K. Austin
Department of Linguistics
SOAS, University of London
Outline

- Language documentation and language revitalisation
- Models of fieldwork and data collection
- Relations between the two
- Some current challenges and developments
- A case study
- Conclusions
Language documentation

“concerned with the methods, tools, and theoretical underpinnings for compiling a representative and lasting multipurpose record of a natural language or one of its varieties” (Himmelmann 1998)

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

Outcome is annotated and translated corpus of archived representative materials on a language, cf. DoBeS, ELAR
Language revitalisation

- efforts to increase language vitality by taking action to:
  - increase the domains of use of a language and/or
  - Increase the number of speakers (often in the context of reversing language shift)

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

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

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


• Hinton and Hale 2001 ‘The green book of language revitalisation’

• Hinton 2002 ‘How to keep your language alive’

• Hinton 2013 ‘Bringing our languages home: revitalisation for families’
What is language documentation?

• “concerned with the methods, tools, and theoretical underpinnings for compiling a representative and lasting multipurpose record of a natural language or one of its varieties” (Himmelmann 1998)

• has developed over the last 20 years in response to the urgent need 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
What it is not

• it's not about collecting stuff to preserve it without analysing it

• it's not = description + technology

• it's not necessarily about endangered languages *per se*

• it's not a passing trend
Evidence

**Student interest**
- 140 students graduated from SOAS MA in Language Documentation and Description 2004-14 – currently 20 are enrolled
- 10 graduates in PhD in Field Linguistics – 20 currently enrolled
- other documentation programmes, eg. UTAustin have similar experience
- Summer schools (3L, InField/Collang), training courses

**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* – 12 issues (paper and online)
- *Language Documentation and Conservation* – 6 issues (online only)
Big money – DoBeS projects
Big archives – ELAR at SOAS
Main features (Himmelmann 2006: 15)

- **Primary data** – collection and analysis of an array of primary language data to be made available for a wide range of users;

- **Accountability** – access to primary data and representations of it makes evaluation of linguistic analyses possible and expected;

- **Long-term storage and preservation of primary data** – includes a focus on archiving in order to ensure that documentary materials are made available to potential users now and into the distant future;
Main features (cont.)

- *Interdisciplinary teams* – documentation requires input and expertise from a range of disciplines and is not restricted to mainstream ("core") linguistics alone

- *Cooperation with and direct involvement of the speech community* – active and collaborative work with community members both as producers of language materials and as co-researchers

- Outcome is **annotated and translated corpus** of archived representative materials on a language
McGill Cicipu corpus
Cicipu annotations

Toolbox - Dictionary.txt

svtmg001.099
208.260
209.450
Tcni

annotations

Hausa has swept them away

svtmg001.100
210.260
210.960
Tcni

Metadata.txt

svgd001
Discussion of chieftaincy
GDM ; JN ; King
MM ; MK
MY

svnk001
Greeting the Mall
MoMu ; SDT (Sani the
MoMa ; TMG ; Yaaki

svmy001
Norman Biggs' grave
MoMu ; SDT (Sani the
MoMa ; TMG ; Yaaki

svsd001
Ukula mountain
MoMu ; SDT (Sani the
MoMa ; TMG ; Yaaki

svtg001
Interview about the old
MoMu ; SDT (Sani the
MoMa ; TMG ; Yaaki

Recording

King's guest house, Korisino
Malli's guest hut in Kadaanda
At the grave of Norman Biggs, Sakaba
On the Ukula (Maburya) mountain

TMG's compound in Ka'ingawa KaGaladima
Components of documentation

- **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
Issues in language documentation

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

- Arguably, we should document language ecologies, not just individual languages
  - Multilingual repertoires, mixed codes, translanguaging, contact effects (Mühlhäusler 2003, Grenoble 2011)

- Ideals are often not lived up to
  - Lack of collaboration, corpora are not always representative, outputs are not always accessible – especially to communities
Fieldwork frameworks

- Grinevald (2003:58-60), Cameron 1992
  - fieldwork ON a language – traditional and (later) ‘ethical research’
  - fieldwork FOR the language community – 1960’s ‘advocacy research’
  - fieldwork WITH the language speakers – 1980’s ‘action research’, ‘negotiated fieldwork’
  - fieldwork BY trained language speakers – ‘empowerment research’
- Add in 2015: fieldwork BY anyone – crowd-sourced language documentation (Youtube, Facebook, Aikuma)
Aikuma

- Collection of naturalistic materials by older speakers, respoken as ‘oral documentation’ by younger speakers
- Problematic in terms of quality of recording and analysis
- Relies on external ‘expert’ to collect and process materials
Towards a typology of projects

- researchers and projects are still overwhelmingly ‘lone wolf’ and research ON, or FOR
- projects have a range of working arrangements, but perhaps we can develop a **typology** of these
- using a metaphor from human land use practices:
  - Hunting & gathering
  - Slash & burn swidden
  - Sedentary intensive cultivation
  - Plantation
  - Sustainable land use
1. Hunting and gathering
Hunter-gatherers

- “primary subsistence method that involves the direct procurement of edible plants and animals from the wild, by foraging and hunting, without significant recourse to the domestication of either.” [Wikipedia]

- **in linguistics**: rapid surveys relying on questionnaires (primarily lexical) for language identification and classification, and collecting basic typological data, never staying in one place for any length of time (Jane Simpson’s “fifo linguistics”)}
“Greg and David travel around the world to collect recordings of some of the last speakers of dying languages”
2. Slash and burn swidden
Slash and burn

- “cutting and burning of forests or woodlands to create fields for agriculture ... or for a variety of other purposes. It is sometimes part of shifting cultivation agriculture, and of transhumance livestock herding ... and operates on a cyclic basis” [Wikipedia]

- **in linguistics**: 3-5 year projects aimed at creation of Boasian trilogy (grammar, texts, dictionary) then moving on to next language
Slash and burn

- R.M.W. (Bob) Dixon
- 1963-1984 Cairns rainforest
- 1985-1990 Fiji
- 1991-2008 Amazonia
- 2009- Cairns rainforest
3. Sedentary intensive cultivation
Sedentary agriculture

- ranges from feudal to communal, with employment of local serfs and artisans in temporary or specialist roles (and necessary application of fertiliser and pesticides, or crop rotation)
- **in linguistics**: long-term projects in a single site, often with Christian missionary connections, eg. SIL
4. Plantation
Plantation

- use 3rd world local residents to grow consumable products in the correct form and extract them to refine and add expensive value in 1st world, with or without “fair trade” label

**in linguistics**: train native speakers to transcribe and gloss using ELAN and Toolbox – linguist takes files “home” to further process and publish academic papers (note skills acquired have no local application)
5. Sustainable use

ecology-driven wholistic approach, including reforestation and recuperation of damaged land
Reduce nutrient losses by:
- Erosion
- Leaching
- Volatilization

Nutrient management:
- Fertilizer use
- BNF
- Recycling

Land clearing:
- Manual
- Shear blade
- In situ burning

Sustainable land-use system

Production systems:
- Food crops
- Perennials
- Livestock
- Mixed systems

Erosion control:
- Mulch farming
- Conservation tillage
- Cover crops
- Vegetative hedges
- Engineering structures

Conversion

Resource assessment

Yes

Increasing trend in per capita productivity
High environmental quality

TRF
Sustainability

- we do not have good models for sustainable documentation projects
- we understand sustainability of archived data but how do we sustain projects and relationships beyond the typical 3-5 year academic life cycle?
- how can documentation contribute to sustaining endangered languages and the communities who want to maintain and develop them?
• we need to move beyond our current models and involve different stakeholders and different models and methods

• one partial answer to these questions may come from considering what documentation would look like if the goal was revitalisation and language maintenance and support
‘Revitalisation’ has been seen by some documentary linguists as 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)
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
‘Elders’ may traditionally be seen as language authorities

- May not be willing to lose this role to linguists or activists

Activists may not want to wait for description and analysis

- Sense of urgency; enthusiasm rather than planning
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
Nathan and Fang 2008

- Meta-data for language pedagogy:
  - Identification and description of socially/culturally relevant events such as songs, which are of great interest to community members and which provide invaluable teaching materials
  - Phenomena that provide learning domains, e.g. numbers, kinship, greetings
  - Socially important phenomena such as register and code switching
  - Notes on learner levels, and links to associated materials that have explanations and examples
  - Notes on previous selections and usages of material for teaching and how to use material for teaching
  - Notes and warnings about restricted materials or materials which are inappropriate for young or certain groups of people (e.g. profane, archaic etc)

- Sociolinguistic survey for education authorities or community bodies to identify potential learner groups and their abilities, needs, and motivations, as well as potential teachers and consultants and their particular skills
Documentation of revitalisation

- Language teaching and associated activities can provide useful language data and insights about structure and use.
- Nathan and Fang (2008) state that language classes provide a unique locus for uncovering language attitudes, paths of acquisition, language change, literacy, language in use, new types of language usage, or identifying new consultants (or roles for people with a range of language skills, e.g., Dieri project ‘errors’ by semi-speakers became materials for language learning games).
- Gomez (2007: 101) argues that a language teaching needs to precede more linguistic documentation in a community for community members to be fully informed and empowered in any participation, and to make their contributions richer.
Case study: Dieri
A bit of history

- In 1974 (yes, that’s 40 years ago!) I did the 4\textsuperscript{th} (honours) year of my undergraduate studies at ANU in Canberra in Linguistics – as part of this I did fieldwork in northern SA on Diyari, which then had about 12 fluent speakers who had learnt it as a 1\textsuperscript{st} or 2\textsuperscript{nd} language as children (English was their 3\textsuperscript{rd} or 4\textsuperscript{th} language)
- I wrote my Honours dissertation on Diyari and then went on 1975-1978 to do my PhD on Diyari
Outcomes

- “A Grammar of Diyari, South Australia” 1981
- Several published articles on morphosyntax, and historical reconstruction
- Several published texts
- A study of literacy practices during the missionary period (1860 to 1915), Ben Murray biography
- Compiled materials for a dictionary (never published)
- Archived 50 hours of tape-recordings with AIATSIS
- Active research ended in late 1980s
Community developments 1990s-

- Formation of Dieri Aboriginal Corporation – 600 members in Maree, Lyndhurst, Broken Hill, Port Augusta, Whyalla
- DAC purchases properties, Port Augusta & Broek Hill
- Purchase of Maree Station and camp ground – handover at dawn 20th September 2008
Agreement with Santos 2011

Santos and Dieri sign milestone cultural heritage agreement

31 Jul 2011

Santos and the Dieri Aboriginal Corporation last week signed a cultural heritage management plan, recognising the significance of traditional owners in the management and protection of cultural heritage in the Dieri native title claim area of the Cooper Basin.

The signing ceremony – which took place at the Art Gallery of South Australia – represented the culmination of formal discussions that began in 2009.

Among the guests at the ceremony were Santos chief executive David Knox, Shane Kemp, Chairperson of the Dieri Aboriginal Corporation, Dieri Elders and the Hon Frances Bedford MP.

The plan brings Santos into best practice for cultural heritage management in its South Australian operations and...
Native title May 2012 (lodged 1997)
“Consent determination”

- Covers some 47,000 square kilometres of land, with part of its south-eastern boundary extending into the Strzelecki Regional Reserve and part of its western boundary extending into the Lake Eyre National Park
- 14% of Finland
- A second Consent Determination 26th February 2014 added to this land
Speakers of Aboriginal languages 2011
Revitalisation – the Dieri ILS project

- 2013 workshops: February Adelaide, March Port Augusta, April Adelaide, August Port Augusta
- Materials development with teacher-linguist Greg Wilson – songs, bilingual dictionary, Willsden Primary school language programme
- Blog dieriyawarra.wordpress.com
- 72 posts, 9,500 page views (as of 2015-06-01), still getting 20-30 views per day
- Community engagement process
Dieri mob February 2013
March 2013 workshop, 4 generations
kanungka
Variant: kadnungka. n. spectacled hare wallaby.
Lagorchestes conspicillatus.
Category: Mammals.

kanya n. ash of fire.
kanyangarri n. jail.
kanyangarriyanganka-rna ver. make go to jail.
kaparra n. boss, root.
Category: Plants.
See: mayatha.
kaparramaganha n. Kopperamanna.
Category: Placenames.
kaparraru int. come here!
Category: Interjections and particles.
kapawu int. look out!
Category: Interjections and particles.
kapi n. egg.
Category: Birds. miri kajiyamayi warlaya, kapi manilha “Climb up above to the nest to get the eggs!”
kapirri n. goanna. Varanus gouldii.
Category: Reptiles. ya kardiyali wama ya kapirri ya kadni nandrany. And the brother-in-law killed carpet snake, and goanna, and frill-necked lizard.
kapitha n. bandicoot.
Peramele morphala.
Category: Mammals.
kapurra n. ampit.
Category: Body parts and products.
kara part. perhaps, maybe.
Category: Interjections and particles.
yini kara wapalha nganayi nhingki yara ngalantha nhingki yara waparnanh. Maybe you go that way and we go this way.
kagakara n. close, nearby; nhayirna kagakara first thurruru nhuluhupularni, wardayari pula thuragarni parlikarnanhi, thupu nhayirna thangkuthangkuparna. At first (he) saw their fire close by where they were sleeping as they went along, watching the smoke in the morning. Pula wakarayi nhantuyali, thana mancarra pirkirnanhi pakarna kagakaralurama. They came on horseback, right close to where the girls were playing.
kara n. today, now, soon.
See: thangkuparna; walrawirli, walya karikamayi, kara n. wama thayilha “Wait a while to eat carpet snake (later) today.”
kararru n. moiety name. matrilineal moiety inherited from mother.
See: mathari.
Writing songs
ngapa-ngapa pirna ngariyi
ngarrimatha wakarayi
thalara pirna kurdayi
ngayanarni mithanhi
daku pirna thana
matya ngayana pankiyilha
ngapa pirna ngakayi
parru pirna pakarna

Lots of water is coming down
A flood is coming
Lots of rain is falling
In our country
There are big sandhills
So we are happy now
Lots of water is flowing
And big fish (are coming) too
Interim summary

- 40 years ago Dieri people were living in tin shacks on the margins of Marree, Port Augusta, Broken Hill
Today, 2 generations later, we have major changes:

- A clear corporate identity
- Ownership of land
- Recognition of traditional ownership and relationships with miners
- Strong political leadership, championing language issues
- Enthusiastic community participation (10% of DAC participating in each workshop)
For the language

- Desire and willingness to learn
- Good resource base – funds, recordings (Austin 50 hours, Hercus 12 hours, Wilson 2,000 files), grammar, dictionary, talented and well-trained community members (teachers, health professionals, singer etc.), highly experienced teacher-linguist, available linguist who worked with previous generations
Challenges

- DAC internal politics
- Fluent speakers all old and very shy, good semi-speakers shy and “expensive”
- Issues of planning, processes and flexibility
- Lack of staff with back office skills
- Monitoring and evaluation lacking
- School programme implementation
- Availability of teacher-linguist and linguist
So what did I learn about research?

- My work done 40 years ago turned out to be valuable and useful in ways that I never imagined at the time.
- The data and analysis could be repurposed.
- Metadata is an important key to usability.
- Tying to understand ideologies and beliefs and when they can get in the way is important.
- The theories promoted by linguists (e.g., pidginisation as a mechanism for revitalisation, intergenerational transmission as paramount) can be totally rejected by the community.
and …

- I had to leave my politics aside and let the participants in the process set agendas (eg. about songs)
- It was incredibly valuable that I had learnt how to speak Diyari (and could remember how to!)
- I took risks, even when I was not sure if things would work, and mostly they did
- I came to understand that sometimes you fail – people were incredibly forgiving and willing to laugh about it and move on
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 need to be explored and taken into account
Thank you!

To Dieri Aboriginal Corporation, Lise Dobrin, Anthony Jukes, David Nathan, Julia Sallabank, Candide Simard and other colleagues in discussions which informed the development of this paper.

But address all criticisms to us!

Emails: pa2@soas.ac.uk


