Collaborative Project for Documenting Minority Languages in Indonesia and Malaysia

Jᴜᴋᴇs, Anthony

University of the South Pacific, Fiji

Sʜɪᴏʜᴀʀᴀ, Asako

ILCAA, Tokyo University of Foreign Studies

Yanti

Atma Jaya Catholic University of Indonesia

Indonesia and Malaysia are well-known for linguistic and cultural diversity. But many languages spoken in these countries remain under-described, and are also seeing drastic reductions in speaker numbers as communities shift to the national languages Bahasa Indonesia and Bahasa Malaysia, or to more vigorous regional languages. As a small-scale response to this situation, between 2013 and 2016 a series of nine 2–3 day training workshops on the theory and methods of language documentation were held in Indonesia and Malaysia, with the intention of increasing awareness and skills needed for documenting minority languages. Workshops were held in Denpasar (Bali), Samarinda (Kalimantan), Jambi (Sumatra), Kupang (Nusa Tenggara Timur), Manado (Sulawesi), and Kota Kinabalu (Sabah, Malaysia). These workshops were inspired by the DocLing training workshops, and like DocLing, they were financially supported by the Linguistics Dynamic Science (LingDy) Project. The trainers were Indonesian and foreign researchers active in the field of language documentation, while the trainees were a mix of students and established scholars. This paper reports on the workshops, discussing what was successful and how they changed over time.

Keywords: Indonesia, documentation, training, workshops

1. Background
2. The workshops
3. Summary and the team's plan in the future

1. Background

1.1. Regional diversity

In his review of the seminal language documentation handbook *Essentials of language documentation* (Gippert, Himmelmann and Mosel 2006), Evans points out

that most linguistic diversity is found in the developing world, and thus a 'central task facing the community of linguists is thus to recruit, train, and support scholars from developing countries in documenting this vast mosaic' (2008: 348). This 'vast mosaic' is readily apparent in Indonesia and Malaysia, which between them contain more than 800 languages - nearly 12% of the world's total (www.ethnologue.com). The majority of these are found in the eastern part of Indonesia, especially in the provinces of Papua and West Papua in the island of New Guinea, but western Indonesia and Malaysia also show considerable diversity. Many of the languages - even some with seemingly quite large numbers of speakers - have become endangered as communities shift, generally to Malay, whether in the form of the national languages Bahasa Indonesia / Bahasa Malaysia, or other regional varieties such as Kupang Malay or Manado Malay. At the same time, cultural and economic practices of indigenous peoples of the area are changing rapidly under the influences of globalization, technological change/progress, and migration, restricting and reducing the domains of traditional language use and hastening the loss of vocabulary relating to these practices (for example see Jukes 2011b which discusses changing agricultural practices in Minahasa, Indonesia).

Although there have been several projects over recent years which have aimed to document and describe some of these languages and their associated cultures,[1] most remain under-documented, or in many cases entirely undocumented. Florey and Himmelmann (2010: 123) estimate 'that fewer than 10% (and possibly as few as 5%) of the languages of Indonesia have been the subject of modern linguistic documentation'. It is clear that much is at stake. It is in this context that the Linguistic Dynamic Science (LingDy) Project supported a series of nine training workshops in language documentation in Indonesia and Malaysia between 2013 and 2017. For convenience we will refer to them as the LingDy workshops.[2]

1.2. Initial ideas

The idea for running the LingDy workshops grew out of conversations between Asako Shiohara and a Balinese linguist Ketut Artawa, while he was a visitor to ILCAA. Having attended DocLing 2011 as an observer, Artawa suggested that similar workshops could be run in Bali for staff and students of Udayana University.

The idea was further solidified after the participation in DocLing 2012 in Tokyo by four Indonesian visitors, who attended the workshop both as trainees and language consultants for group projects. DocLing 2011 had successfully included some students

---

[1] Several projects were funded by the Endangered Languages Documentation Programme (www.eldp.net), others by the Volkswagen Foundation's DoBeS programme (http://dobes.mpi.nl/) or by Max Planck Institute for Evolutionary Anthropology via the Jakarta Field Station (http://jakarta.shh.mpg.de/). There has also been a welcome move towards language documentation by the Indonesian Institute of Sciences (LIPI).

[2] Three LingDy workshops were also held in other areas: two in Russia (Buryat State University, November 2014; Kalmyk State University, May 2015) and one in Mongolia (Mongolian Academy of Sciences, September 2016). Though they were reportedly a great success, as these workshops were organized by other team members we cannot report on them in this paper.

from Mongolia and China as language consultants for some limited group work so that the students could get experience collecting language data in a language that they were not familiar with. This experiment was a success (see Nathan and Jukes, this volume), so the organisers decided to increase the amount of group work in subsequent years.

Thus, when Shiohara was given the task of planning DocLing 2012 she decided to invite some speakers of Indonesian regional languages as consultants. Ketut Artawa helped to select 3 consultants: Budi Sudarmanto from Balai Bahasa Palembang, I Wayan Budiarta from Mentari Foreign Language College, and Jermy Balukh from School of Foreign Languages, Cakrawala Nusantara Kupang. Another, Hendrik Paat from National University of Manado (UNIMA), was selected at the suggestion of Anthony Jukes. The languages they consulted on were Javanese, Balinese, Rote, and Tombulu, respectively.

The experiment with facilitated group work was so successful and enjoyable for both consultants and trainees (see Nathan and Jukes, this volume) that the practice of inviting consultants to assist with group work continued in the following years. From 2013 to 2016 another nine consultants came from Indonesia, together with six from other parts of the world.

It should be noted that although they were invited primarily as language consultants for group activities, they also learnt the theory and the method of language documentation in the same way other participants did. We could say that in some respects they learnt more than the regular participants, in that, through their experience, they came to realize the value of their languages and significance of documenting them. We could say that in this respect the DocLing workshop truly functioned as a place 'to recruit, train, and support scholars from developing countries' (Evans 2008: 348).

The Indonesian connections especially were instrumental in the decision to take language documentation workshops to Indonesia and Malaysia, and many were involved in the organization of the workshops. Jermy Balukh, Yanti, and Dominikus Tauk, who were invited to DocLing in 2012 and in 2016, respectively among others now play a core-role in LingDy workshops as local researchers.

1.3. Previous training workshops

It would be misleading to give the impression that the LingDy workshops we describe here were the first language documentation training workshops to be held in Indonesia. To our knowledge there have been at least three other series of workshops, and to various degrees these were also influential in the inception and planning of the LingDy courses. In addition there have been many workshops, summer schools, and longer courses in other parts of the world (for an overview see Jukes 2011a).

In 2006 and 2007 Nikolaus Himmelmann and Margaret Florey organized workshops on language documentation which were held in Ubud, Bali. These workshops, described in detail in Florey (2008) and Florey and Himmelmann (2010), were of an intensive, residential model, taking place over 10 (2006) and 7 (2007) days. There were

25 attendees and 11 instructors in 2006, and 11 attendees with 8 instructors in 2007.

The Center for Endangered Languages Documentation (CELD) at Universitas Negeri Papua (UNIPA) in Manokwari has organized occasional training sessions for local team members since 2009.

The Indonesian Institute of Sciences (LIPI) organized some in-house training from 2012–2013 for their documentation projects, for which they were assisted by John Bowden, then at the Jakarta Field Station of the Max Planck Institute for Evolutionary Anthropology.

In addition to these series of events, there was a one-off 2-week 'intensive community language documentation workshop' in Kupang in 2012, organized by Chuck Grimes, Mark Donohue, and Dan Kaufman from the Endangered Language Alliance (among others).[3]

There has been significant cross-over between the various workshops mentioned above, the workshops organized by LingDy, and also with DocLing.[4]

2. The workshops

Figure 1 shows the locations of the LingDy workshops: (from west to east) Jambi, Denpasar, Kota Kinabalu, Samarinda, Kupang, and Manado.

---

[3] http://austronesian.linguistics.anu.edu.au/timor/workshop/

[4] The following illustrates some of the connections.

Jermy Balukh was a student in Ubud 2006–2007, a consultant at DocLing 2012, and an instructor in the LingDy workshop at Kupang in 2016.

Deisyi Batunan was a consultant at DocLing 2015 and a local organizer for the LingDy workshop in Manado in 2015.

John Bowden helped with the LIPI training and also most of the LingDy workshops.

I Wayan Budiarta was a consultant at DocLing 2012 and part of the local organizing team in workshops held in Bali.

Nikolaus Himmelmann organized and taught at Ubud in 2006–2007 and was a trainer at DocLing 2013. He has also been involved with training at CELD.

Anthony Jukes was a trainer at Ubud in 2006, DocLing 2010–2016, and most of the LingDy workshops.

Yusuf Sawaki was a trainee in Ubud 2006–2007 and founded CELD in 2009.

Asako Shiohara was an organizer of DocLing and the LingDy workshops.

Antonia Soriente was an instructor in Ubud in 2006 and 2007, and also at most of the LingDy workshops.

Yanti was a consultant at DocLing 2016 and also taught at most of the LingDy workshops.

Fig. 1   Locations of workshops

2.1. The early workshops

The first of the LingDy workshops was the 'Workshop on Language Documentation' held at Udayana University in Denpasar, Bali on 5–6 August 2013. It was initially planned as a one-off event, arranged at the suggestion of Ketut Artawa. The 24 attendees were mostly staff and graduate students of Udayana, though two exchange students from L'Orientale in Naples, Italy also attended at the suggestion of Antonia Soriente. The workshop was successful in that it aroused the participants' interest in language documentation, which had not been a very common research area in Indonesia, and strengthened the intention to establish a research network between researchers and students in and out of Indonesia.

Fig. 2   The LingDy workshop at Udayana, August 2013. Instructors (seated L to R): Anthony Jukes, Antonia Soriente, Ketut Artawa (standing), John Bowden, Asako Shiohara, Atsuko Utsumi.

After the first workshop in Bali, the team decided to continue the attempts and launched an ILCAA joint research project titled "Constructing a research network for documenting minority languages in and around Indonesia" (April 2014–March 2017), which was funded by LingDy. The project was organized by researchers in Japan and Indonesia and also included members from other parts of world, such as Italy, Australia, and the US. In the first two years of the project term, the project held seven workshops:

- Universitas Malay Sabah, Kota Kinabalu, Malaysia, 11–12 August 2014
- Udayana University, Denpasar 15–16 August 2014
- Mulawarman University, Samarinda, Indonesia, 19–21 August 2014
- Seloko Institute, Politeknik Jambi, Indonesia, 15–16 June 2015
- Manado State University, Indonesia, 6–7 August 2015
- Balai Bahasa Denpasar, 15–16 August 2015
- The University of Nusa Cendana, Kupang, Indonesia 10–12 August 2015

All the workshops basically constituted of two parts: introductory lectures on theory and methods of language documentation, followed by practical training of recording techniques and annotation software such as ELAN.

Table 1 shows the program of the workshop held on 15–16 August 2014 at Udayana University, Denpasar Bali. Here, more time was assigned to lectures than practical

training to fill the needs of participants, many of whom were established scholars.

Table 1 The program of the workshop held on 15–16 August 2014 in Udayana University

15 Aug	
9:00–9:20	Registration, Opening talk
9:20–9:50	Lecture 1: What is language documentation and why do we do it?
9:50–10:50	Lecture 2: What kind of data do you collect?
10:50–12:30	Lecture 3: How do you make recordings?
12:30–14:00	Lunch
14:00–14:50	Lecture 4: Data management
15:00–16:30	Practice 1: Group activity : Recording and producing (Making recordings of monologue and do free translation, and produce metadata)
16:30–17:30	Evaluation of the recordings
16 Aug	
9:30–10:00	Lecture 4: Ethics in language documentation
10:00–11:45	Lecture 5: Software and tools (including short practice of ELAN)
11:45–12:30	Session 3: Orthography
12:30–14:00	Lunch
14:00–16:00	Session 3: Group work: Practice of ELAN
16:00–17:00	Wrap up and Evaluation

The earlier workshops were intended to introduce the idea of language documentation and give very basic training, with the primary aim of building networks with local institutes and researchers. Later workshops included the aim of conducting substantial work of actual language documentation during the workshops.

The turning point was the seventh workshop held at the University of Nusa Cendana, Kupang in August 2015. The participants there were distinct from the majority of the attendees of previous workshops, in that most of them were graduate or undergraduate students from a language community of endangered or minority languages spoken in Nusa Tenggara Timur Province, a particular hotspot of Indonesian diversity. This workshop saw the 'vast mosaic' (Evans 2008: 348) notably present in the workshop venue, with every student showing personal interest and motivation for gaining knowledge and skills for documenting their own language. They also had strong support from the local lecturers.

According to their needs, we modified the program we had employed in the preceding workshops. We established a workflow in which the participants could have concrete outcomes at the end of the workshop. For that purpose, the audio data that the participants recorded in the earlier part of the workshop played a core-part. The details

of the new structure were as follows.

- Compared to the previous workshops, more time was taken for practical training than lectures. It was a three-day event and we separated the seminar on the first day from the practical workshop on the second and third days. General lectures for a wider audience were given in the seminar, and then a restricted number of participants - who were nominated by their lecturers or otherwise showed that they had serious interest in documenting their languages - attended the workshop days, where they could focus on intensive documenting work.
- The session on data management, which was given as a lecture in the previous workshops, was changed to include some hands-on training in which they organized their own recorded data with appropriate metadata, which would be a core of the workshop outcome.
- In the practical sessions of annotation software ELAN, the participants were encouraged to take more time to annotate their own recordings, after they had learned the basics of the software with a very short sample exercise.

The attempts above worked well and the participants succeeded in making recordings of nine minority languages with transcription and translation. The results were deposited at the language archive PARADISEC.[5] Figure 3 shows the workflow employed at the workshop.

(1)	Recording: recording **audio data**

↓

(2)	Data management: organize **the audio files** with proper metadata

↓

(3)	Annotating **the data**: learning how to use computer applications for annotation such as transcription and translation

↓

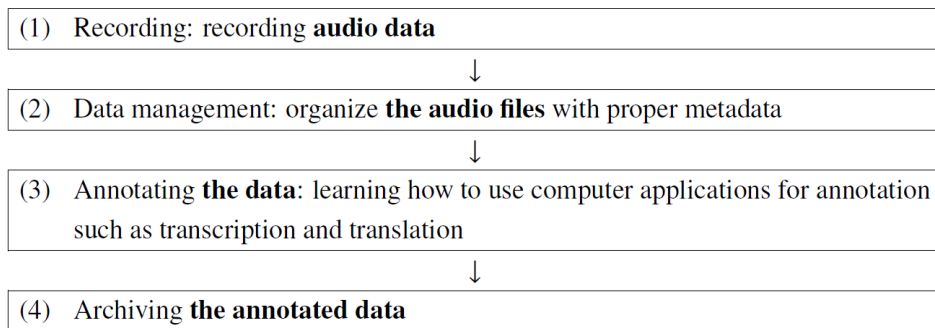| (4)	Archiving **the annotated data**

Fig. 3   The workflow employed in the workshop

Before closing this section, some technical aspect of the early workshops should be noted. For the first few training courses, it was considered useful to carry LingDy's high-quality recording equipment (digital recorders, video cameras, microphones) and many sets of closed headphones from Tokyo, so that students were able to hear the results of different recording equipment and techniques (based on the similar practice

---

[5] The outcome can be seen on the website below;
http://catalog.paradisec.org.au/collections/NTT2015

used during DocLing in Tokyo). However, this soon turned out to be overly expensive and onerous, and further workshops were run using a smaller number of microphones, small Zoom H1 recorders, and earbuds (supplied, or the students' own). An ancillary benefit was that the Zoom H1 recorders, purchased in Japan at the authors' expense, were able to be sold at cost to trainees for their future use, as similar recorders are not readily available in Indonesia.

A final note regarding language: in the first few workshops, English was the medium of instruction, as most of the lecturers initially had teaching materials such as slides and handouts only in English. Furthermore, some trainees had requested English instruction to give them more exposure to the international language. As we could easily expect, however, it soon became apparent that teaching in Indonesian (or Malaysian in Malaysia) was much more effective and the language of the workshops shifted from English to Indonesian.

2.2. Workshops for substantial collaborations in documenting languages

The experience in Kupang in 2015 made us re-evaluate the aim of the project and we decided to focus more on collaboration for language documentation with the workshop participants, more precisely, collecting a significant amount of linguistic data and creating metadata and annotations during the workshops. With the revised aims, three workshops were held in 2016 and 2017.

(1) "Workshop on languages spoken in Sabah state, Malaysia," Skypod Hostel at Kota Kinabalu, Malaysia 9–13 August 2016

(2) "International seminar on documenting languages *What, why, and how?*" and "International workshop on documenting minority languages: *Theory and practice*", Artha Wacana Christian University, Kupang, 28–30 November 2016

(3) "International workshop for documenting endangered languages in Nusa Tenggara Timur Province, Indonesia", ILCAA, TUFS, Tokyo, 24–30 March 2017

Event (1) was a new attempt in that it was something like a language documentation camp. The participants were several linguists and non-academic language consultants, some of whom are local language activists, who were involved in activities of language conservation and description. They worked together for 5 days in collecting the data of indigenous languages in Sabah state, such as Iranun, Sūg (Tausug), Liwan dialect of Dusun, Brunei Malay and Sabah Malay. The language consultants learned how to use digital recorders and the annotation software ELAN, and made substantial contributions to documenting their languages resulting in the following outcomes; (i) basic vocabulary and two annotated narratives in Dusun, (ii) narrative with transcription and translation in Iranun, (iii) narrative with transcription and translation in Sūg.[6]

---

[6] The outcomes can be seen on the website below;

In event (2), the team returned to Kupang to continue the collaboration we had started in 2015. On the first day there were general lectures about language documentation attended by about 60 students and teachers from language departments at Artha Wacana Christian University and The University of Nusa Cendana. Following this there were two days of practical workshops attended by 25 of the participants from the earlier session. As with the earlier workshop at Kupang, most of the students were speakers of regional languages of the province of East Nusa Tenggara as well as Kupang Malay. To illustrate the diversity, in the group recording practice sessions recordings were made of Roti, Kupang Malay, Lamaholot, and Helong.

Fig. 4   Kupang, November 2016. Instructors (seated L to R): Jermy Balukh, Antonia Soriente, Anthony Jukes, Yanti, Asako Shiohara. Not pictured: June Jacob.

During the two workshops in Kupang the team became convinced that it is one of the most ideal locations for this type of workshop, because it is in the centre of such linguistic diversity, and many potential collaborators (native speaker researchers and students) are present at the local universities.[7]

Iranun story "A story of Kandalayang, a king of the sea" told by Mabulmaddin Shaiddin
https://www.youtube.com/watch?v=JLIo3Tim7A4
Sūg story "A story of cow stone"
https://www.youtube.com/watch?v=qvv7nV8TR2o

[7] We were not the first to have this realization, e.g. the two-week workshop held in 2012 by Grimes, etc. (as mentioned in 1.3).

Having seen the presence of vast numbers of lesser documented or undocumented languages in Kupang, the team (especially Shiohara and Yanti) decided that in order to go to the next step of the documentation they should select one or two languages to focus on and take substantial time for them. They selected Rote and Helong, as they had native speakers of the each language in the team, namely Jermy Balukh and Dominikus Tauk, both of whom were alumni of DocLing held in Tokyo.

Event (3) was held at Tokyo for that purpose, in 'documentation camp' style. Jermy Balukh and Dominikus Tauk each selected a speaker from their language community who was relatively less-exposed to Indonesian and Kupang Malay, and accompanied them to Tokyo. Antoinette Schapper, a specialist of the Alor-Pantar-Timor languages spoken in NTT province, also participated in the workshop with two language consultants from Alor and Pantar, for documenting Teiwa (Alor) and Nedabang (Pantar) respectively. The documentation camp succeeded in recording several genres of each language, such as traditional stories and conversations.

3. Summary and the team's plan in the future

In summary, the activities conducted by the team can be summarized as (1)–(3) below.

(1) DocLing workshop in Tokyo; native speaker linguists participated in the workshops. They played a role as language consultants and shared the knowledge of their language as well as learnt the skills and knowledge necessary for documenting languages. They also realized the value of their languages and significance of documenting them (section 1.2)

(2) Seminars and Workshops to talk about the value of indigenous minority languages, the significance of documenting them, and establish a network of researchers interested in language documentation (section 2.1)

(3) Workshops for substantial work of actual documentation collaborating with native speakers (section 2.2)

Each step was related to previous steps in that the local researchers recruited in step (1) had contributed in organizing the events of step (2), as leaders of the local academic and linguistic community, and then played a central role in collaboration of step (3) as native speaker linguists.

The ultimate goal would be to document as many minority languages as is feasible, and collect and archive data with appropriate metadata and annotation. This process takes a great deal of time, but concerted collaborative workshops of type (3) have proven to be very productive. Events such as (1) and (2), however, are also useful and resources should be assigned to them; recruiting local leaders through events such as (1) and establishing academic networks through events such as (2) helps to support the documentary work of both local and overseas researchers.

Although both DocLing and the 'traveling' training workshops of type (2) have finished as of March 2017, the members will continue activities of type (3) based in Kupang and Tokyo for at least 5 years, supported by LingDy.

References

Evans, Nicholas. 2008. "Review of Gippert, Himmelmann and Mosel (2006) *Essentials of language documentation*". *Language Documentation and Conservation* 2/2. pp.340–350.

Florey, Margaret. 2008. "Language activism and the 'new linguistics': expanding opportunities for documenting endangered languages in Indonesia". In Peter K. Austin (ed.) *Language Documentation and Description* 5. London: SOAS. pp.120–135.

Florey, Margaret and Nikolaus P. Himmelmann. 2010. "New directions in field linguistics: Training strategies for language documentation in Indonesia". In Margaret Florey (ed.) *Endangered languages of Austronesia*. Oxford: Oxford University Press. pp.121–140.

Gippert, Jost, Nikolaus P. Himmelmann and Ulrike Mosel. 2006. *Essentials of Language Documentation*. Berlin; New York: Mouton de Gruyter.

Jukes, Anthony. 2011a. "Researcher training and capacity development". In Peter K. Austin and Julia Sallabank (eds.) *The Cambridge Handbook of Endangered Languages*. Cambridge: CUP.

———. 2011b. "Culture documentation and linguistic stimulus". In Nicholas Thieberger et al. (eds.) *Sustainable Data from Digital Research*. Melbourne: University of Melbourne.