

In memory of our dear friend Kristian Walianggen (1986–2017),  
Yali consultant for DocLing 2013.





## Introduction to the Special Issue on DocLing Workshop

NAKAYAMA, Toshihide

ILCAA, Tokyo University of Foreign Studies

JUKES, Anthony

University of the South Pacific, Fiji

This special issue celebrates the successful years of the Documentary Linguistics Workshop (DocLing). This workshop series was created and has been organized by the Linguistic Dynamics Science Project (LingDy) at ILCAA in collaboration with the Endangered Languages Academic Programme (ELAP) and the Endangered Languages Archive (ELAR) at SOAS (2008–2014) and later with the EL Training group (2015–2016).

Since its inception in 2008, DocLing has been offered every year for nine years. Under the guidance of Peter Austin and David Nathan and the continuous efforts of the LingDy staff members, DocLing has evolved into a world-class workshop on language documentation that now attracts participants from all over Japan and overseas.

Over these nine years, we have trained well over 150 researchers through DocLing, and have consequently gained a lot of experience and knowledge about language documentation training.

This publication reflects on the amazing nine years of DocLing. In addition to the core instructors, we invited the language consultants and participants to provide their different perspectives of DocLing.

We have come a long way, and the situation surrounding DocLing has changed significantly. We hope this publication is an opportunity to summarize what we have accomplished and to think about the way we can extend DocLing into the future.

Finally, I would like to thank all the people who have participated and supported DocLing in any capacity over the years. We could not have accomplished this without your support. Thank you!





## Documentary Linguistics Workshop: Its Beginning, Development, and Future

NAKAYAMA, Toshihide  
ILCAA, Tokyo University of Foreign Studies

This paper traces the history of the Documentary Linguistics Workshop to provide the background context and underlying motivations for the development of the workshop series.

**Keywords:** documentary linguistics, DocLing, training, endangered languages, language revitalization

1. Introduction
2. How it started: Context of the development of DocLing
3. Nine years of DocLing
4. Impact
5. The future

### 1. Introduction

The goal of this paper is to provide information about the academic and historical context in which the Documentary Linguistics Workshop (DocLing) was created and developed over the years. It is hoped that this paper will help readers better understand the needs and motivations that led to the initiation of the workshop series.

### 2. How it started: Context of the development of DocLing

First, I would like to describe the context in which the plan for the DocLing was developed in 2007. Documentary linguistics (or language documentation) as an academic research field has evolved out of traditional descriptive research on endangered languages. It is therefore useful to look at the changes progressing in endangered languages research at the time.

## 2.1. Trends in research on endangered languages

As the worldwide problem of language endangerment was rapidly intensifying, research on endangered languages was moving into a new era. The combination of the shrinking domain of use and the decline in the number of speakers drastically undercut the vitality of traditional languages in socially underprivileged minority communities. This development posed problems to two key areas of the study and revitalization of endangered languages.

One problem was affecting academic research for describing endangered languages. It had become rare to find opportunities for conducting field research in consultation with fluent speakers with rich firsthand knowledge of the traditional culture. Researchers were finding themselves relying more on previously published work and unpublished notes and records left by other researchers. Thus, archive research was becoming an important part of their research on endangered languages (see Nakayama 2007).

The other problem was affecting the local communities, particularly the efforts to preserve and revitalize the traditional languages. When it is spoken in everyday lives, a language is something that can be produced endlessly by its speakers. However, once the number of speakers falls below the critical point, it becomes difficult to obtain necessary information and resources about the language within the community. This makes it necessary for communities to look for information in outside sources, which are most commonly academic publications and unpublished materials collected by academics.

Thus, now that the opportunity for collecting firsthand information on traditional languages has become rare, the values of field research and its records are much higher and extend well beyond researchers' immediate goals and concerns. There are and there will be researchers with various interests in a language who want and need to rely on the records collected in someone else's fieldwork. Additionally, local community activists will look at such records for the information they need for their work to preserve and revitalize the language of their ancestors.

Linguistic records are now surrounded by a diverse range of needs, and at the same time, there is a growing expectation that academic researchers will contribute directly to, and make a tangible impact on, the situation in the endangered language community rather than simply advancing purely academic research and analyses (i.e., studying "about" the language and community). We could also say that the scarcity of data collection opportunities has created a strong need for authentic firsthand data that can be utilized for a variety of purposes. This attention to the needs of preservation and revitalization activities in the communities and also to the needs for multipurpose records was at the core of development of the documentary linguistic research paradigm.

The conceptual and methodological framework of documentary linguistics was

developed around the end of the 20th century, and the research was gaining momentum throughout the early 2000s as major sources of research funding started supporting documentary linguistic projects throughout the world. Such sources include the Documentation of Endangered Languages program in Germany, the Endangered Languages Documentation Programme in the United Kingdom, and the Documenting Endangered Languages grant program in the United States (Austin 2014). The field of documentary linguistics was reasonably well established and expanding its reach in the western world by the time the DocLing Workshop was developed in 2007.

## 2.2. Academic environment in Japan

Japan has a strong and long-standing academic tradition of fieldwork-based descriptive research. It was therefore no coincidence that when language endangerment became a major global issue in the late 1990s, Japan was among the first to respond with a launch of a large-scale project. The project, the Endangered Languages of the Pacific Rim (ELPR) Project, was one of the most notable and influential developments in the recent history of descriptive linguistic research in Japan. ELPR was a large-scale five-year project funded by the Japanese Ministry of Education, Culture, Sports, Science and Technology. It was launched in 1999 as a response to the growing global concern about language endangerment and was focused on field linguistic research on languages of the Pacific Rim region, with a particular emphasis on severely endangered languages. ELPR provided much-needed financial support to academic research on endangered languages and helped produce a wide range of materials, including text collections, grammatical descriptions, dictionaries and word lists, and collections of research papers.

With a large-scale descriptive research initiative like ELPR, it may appear that the development of active research programs in documentary linguistics was inevitable and well supported in Japan. However, this was not necessarily the case.

One of the issues was that the conceptual framework of documentary linguistics had not made its way into Japan. Endangered language research in Japan generally kept focusing on academic and descriptive research. There was an almost exclusive emphasis on analyses and grammatical descriptions rather than data themselves. There was interest in the collection of textual materials as a part of descriptive studies. However, textual materials were valued more as illustrations of grammatical rules and structure than as representative records of how language was being used. Thus, the data were considered valuable only in their relation to a grammatical description and were not necessarily valued themselves.

Another issue was the insufficient training opportunities in graduate programs. There were few graduate programs in Japan that could provide training in fieldwork-based linguistic research, let alone in documentary linguistics. It was almost impossible to obtain from graduate programs a systematic training in documentary linguistic theory or methodology. The issue was further complicated because of a lack of a community

of field linguists. There was not much of a network among field linguistics researchers beyond personal connections between individuals. Consequently, opportunities for information exchange and mutual support were limited. This situation posed a serious problem, particularly for students who were in small graduate programs. Such students did not have anyone to obtain advice from or consult with.

Thus, the foundation for the development of documentary linguistic research in Japan was not secure at all. It is true that the same problem exists elsewhere, but the situation was particularly serious in Japan. In other words, the prospect of the development of documentary linguistic research itself was endangered.

### 2.3. Linguistic Dynamics Science Project and DocLing as a response

The Linguistic Dynamics Science Project (LingDy), the project through which DocLing has been offered, was initiated to address issues that are hampering the development of documentary linguistics research in Japan. These are large and complex problems, and it is not possible to solve them in one go. Therefore, we have been focusing on the following four areas: (1) raising the profile of documentary linguistics research; (2) raising the awareness of the needs and values for a comprehensive record of the use of a language that can be utilized for a wide range of purposes (not only for linguistic analysis and description but also for language conservation and revitalization); (3) building capacity, especially that of junior researchers, in documentary linguistics research; (4) building a lasting academic infrastructure and collaboration network in documentary linguistics.

There are two reasons that we focused our attention on capacity building and infrastructure. First, the lack of training opportunities and support structure for documentary linguistic work was a critical problem in Japan that required immediate attention. Second, we needed to build self-sustaining growth mechanisms if we hoped to secure the long-lasting growth of documentary linguistic research. We needed to produce more researchers who could support themselves (i.e., who had relevant training and experience to secure funding for continuing documentary linguistic research). We also needed to build infrastructure and a research network that functioned as a mutual support mechanism for researchers. DocLing was developed as a centerpiece of the capacity building activity of LingDy.

## 3. Nine years of DocLing

The DocLing Workshop was first planned in fall 2007 as the LingDy project was undergoing preparation to start its operations in April 2008. The realization of the first workshop, as well as all the iterations that followed, was made possible by the full cooperation of Peter Austin and David Nathan at SOAS (formerly known as the School of Oriental and African Studies), University of London, who provided generous assistance in all aspects of the workshop from planning to instruction.

The first DocLing was held in February 2008 with 10 participants. The three-and-half day program covered a wide range of topics: the definition of documentary linguistics; creation of a corpus of linguistic data; audio recording; audio data transcription; data format and data management; software tools for data analysis; data archiving; mobilization of language data (creation of learning materials for language revitalization); grant proposal writing skills; and research ethics.

DocLing got a solid budgetary foundation when the LingDy project started in April 2008. The workshop series is now also a part of the training program scheme of the Research Institute for Languages and Cultures of Asia and Africa (ILCAA). This demonstrates a high level of commitment to training in documentary linguistics on the part of ILCAA.

We were able to expand DocLing gradually over the years as additional budgetary and administrative support became available. Starting with the 2010 workshop, we added Anthony Jukes to the regular teaching staff, expanding the workshop to four full days. This allowed us to add a component on video recording to the program. The teaching staff continued to grow: Sonja Riesberg, Hideo Sawada, and Toshihide Nakayama joined the regular staff, and we were also fortunate to have the participation of Nikolaus Himmelmann, John Bowden, and Honoré Watanabe.

The workshop program was substantially revised and expanded in 2011. It grew to a seven-day program and included work with language consultants who were speakers of minority languages. DocLing, since its second workshop, incorporated a group project component where participants engage in small-scale language documentation projects to actively utilize the knowledge and skills that they acquired in the lectures. However, the involvement of language consultants turned out to be one of the most significant improvements we could make to the workshop over the years: it had a substantial positive impact on the quality of the participants' learning experience by adding a high level of reality to the work that participants engaged in. The involvement of language consultants had an additional, equally important value: providing members of endangered language communities with an opportunity to participate in the workshop, allowing them to gain knowledge and skills useful for their language documentation and to participate in revitalization activities in their home communities. In that sense, this scheme served a double purpose, to improve the quality of learning experiences and to extend the reach of the workshop into language communities. Another recent successful addition to the program was individualized consultation sessions where participants could consult the instructors and support staff members for advice and assistance on their concerns and project ideas. This allowed us to provide participants with individualized assistance and learning experiences.

The profile of participants has also changed over the nine years. In the early years of DocLing, participants were mostly graduate students in descriptive linguistics from Tokyo and its surrounding areas. However, the range of participants the workshop attracts has broadened significantly since then. Now we regularly receive inquiries

and applications not only from all over Japan but also from other parts of the world. Moreover, the composition of participants is much more diverse, including dialect study researchers, sociolinguists, linguistic anthropologists, ethnographers, and language activists. We also have had “repeaters” who came back to participate in the workshop multiple times.

#### 4. Impact

Nine years ago, we set out to create DocLing with the goal of raising the profile of newly growing documentary linguistic research and to provide training opportunities and a place to get connected to other researchers with similar orientations. Although it is not easy to trace the exact amount of progress, it seems safe to say that we have made good progress toward these goals.

During these nine years, we trained about 150 participants. Many participants from the early years hold academic positions, and now their students have started to participate in the workshop. Participants are sometimes asked to offer a short workshop or lecture on documentary linguistic methodology. Thus, the benefits of the workshop are reaching beyond the participants themselves. Now the terms “documentary linguistics” or “language documentation” are commonly heard in conversation between researchers on fieldwork-based linguistics, especially those working on endangered languages and dialects. DocLing must have played a meaningful, if not exclusive, role in bringing about this situation.

The more tangible impact is felt in the changes in us. When we started organizing DocLing, we at ILCAA did not have the knowledge and abilities necessary for offering such workshop ourselves, and we had to rely on assistance and cooperation from Peter Austin and David Nathan on all aspects of the workshop. Through the nine years of DocLing, we have accumulated enough expertise to provide our own services and contributions: holding a number of language documentation workshops in Indonesia and Russia; organizing the Documentary Linguistics Seminar for the University of Hong Kong; organizing the Language Documentation Workshop for Japanese dialect researchers in collaboration with the National Institute for Japanese Language and Linguistics; and participating in training activities at international workshops such as the Institute on Field Linguistics and Language Documentation and the Institute on Collaborative Language Research.

#### 5. The future

The DocLing workshop series has been a great success, and we think that we have achieved good results, especially in regard to improving researchers’ capabilities. However, there are some limitations to the current form of the workshop. The most important among them is the limitation with regard to making a direct impact

on language documentation and revitalization activities in endangered language communities. This is an important issue, especially now that the focus of the host project, LingDy, has been shifted to capacity building in language communities.

We are now making a great deal of effort to collaborate with other institutions and with speaker communities to extend training opportunities to other parts of the world, especially to endangered language communities. The traditions of the DocLing workshop are being carried over into the new era.

## References

- Austin, Peter K. 2014. "Language documentation in the 21st century". *LIPP* 3. pp.57–71.
- Grenoble, Lenore. 2010. "Language documentation and field linguistics: The state of the field". In Lenore A. Grenoble and N. Louanna Furbie (eds.) *Language Documentation: Practice and values*. Amsterdam: John Benjamins Publishing Company. pp.289–309.
- Himmelman, Nikolaus P. 1998. "Documentary and descriptive linguistics". *Linguistics* 36. pp.161–195.
- . 2006. "Language documentation: What is it and what is it good for?". In Jost Gippert, Nikolaus P. Himmelman and Ulrike Mosel (eds.) *Essentials of Language Documentation*. Berlin: Mouton de Gruyter. pp.1–30.
- Nakayama, Toshihide. 2007. "Using written records to revitalize North American languages". In Osahito Miyaoka, Osamu Sakiyama and Michael Krauss (eds.) *The Vanishing Languages of the Pacific Rim*. Oxford: Oxford University Press. pp.91–106.





## Reflecting and Shaping the Evolution of Documentary Linguistics: Nine Years of DocLing Workshops

NATHAN, David

Groote Eylandt Language Centre

JUKES, Anthony

University of the South Pacific, Fiji

This paper discusses the evolution of the Documentary Linguistics Workshop (DocLing) organized by the Research Institute for Languages and Cultures of Asia and Africa (ILCAA) at Tokyo University of Foreign Studies. Held annually from 2008 to 2016, DocLing introduced nearly 200 students to the theory and methods of language documentation and made a unique contribution to the teaching of field linguistics in Japan, and to work on endangered and minority languages more generally. The changing nature of the workshops over time reflected changes in focus in the discourse surrounding language documentation.

**Keywords:** language documentation, training, ILCAA

1. Origins of the workshop
2. The early workshops
3. Social program
4. Staff
5. Group projects
6. Impacts and outcomes

### 1. Origins of the workshop

The DocLing workshops arose from an initiative of Dr. Toshihide Nakayama at ILCAA who recognised that there were many young linguistics scholars in Japan who were undertaking ambitious and widely ranging fieldwork on endangered and minority languages, but who were receiving little or no formal training explicitly aimed at equipping them for modern fieldwork in documenting such languages (see Nakayama, this volume). This paper gives an overview of nine annual cohorts of these linguists and their projects, and the evolution of an innovative training curriculum and methods that were significantly shaped by the responses and unique needs of these researchers, at

---

NATHAN, David and JUKES, Anthony. 2017. "Reflecting and shaping the evolution of documentary linguistics: Nine years of DocLing Workshops". *Asian and African Languages and Linguistics* 11. pp.11–22. [Permanent URL: <http://hdl.handle.net/10108/89204>]

the same time as the field of language documentation itself was developing at its most rapid pace.

Dr. Nakayama initially called on the expertise of Professor Peter K. Austin and David Nathan, both alumni of ILCAA from the 1990s, but more importantly who were key trainers in the Hans Rausing Endangered Languages Project (HRELP) at SOAS University of London. HRELP brought together a postgraduate academic program in Language Documentation and Description under the leadership of Peter Austin (ELAP), a digital language archive led by David Nathan (ELAR), and a granting program ELDP, together constituting the world's largest assemblage of activity and innovation on documenting endangered languages for the decade from 2003 onwards. At SOAS, Nathan and Austin realised that the potential of combining the intellectual foundations and teaching experience of the academic program, the technical and methodological strengths of the archive team, and the fostering of documenters through the granting program - all with significant international linkages and collaborations - would provide the ideal platform for training a "new breed" of language documenters. This also paralleled developments elsewhere such as the USA-based summer school InField which also had a significant focus on documenting endangered languages. Austin and Nathan, together with others in their teams and external guest experts, developed and ran the ELDP grantee training sessions in London from 2004 onwards, held once or twice a year for groups of 15–20, typically organised by the ELAR archive team under the auspices of ELDP.

This depth and breadth of experience seemed to meet the goals of Dr. Nakayama. While ILCAA has its main strength in field research across Asia, the SOAS team brought experience in teaching and training, as well as a certain international perspective (since it seems fair to say that some parts of Japanese academia are somewhat insular). And of course they also brought their already developed training curricula, materials, and methods. Additionally, both Austin and Nathan had some experience living in Japan and working in Japanese universities and with Japanese students.<sup>1</sup>

## 2. The early workshops

The first DocLing workshop took place over 4 days from 14–17 February 2008. It largely followed the model of the HRELP workshops, with this simple program (see below), where Day 1 introduced notions of language documentation, a practice largely unknown in Japan at the time, Day 2 focused on audio and related techniques and

---

<sup>1</sup> Austin studied Japanese at the Australian National University in the 1970s. Later, Austin held a post of Visiting Foreign Professor at ILCAA in 1996–1997. Nathan also held a post of Visiting Foreign Professor at ILCAA in 1997–1998. Nathan subsequently was Foreign Professor at the University of Tsukuba from 2002–2004. Anthony Jukes who joined the team later also had some experience in Japan and had learned Japanese at the University of Melbourne.

technologies, Day 3 on data management and linguistic annotation and processing, while Day 4 covered a variety of associated topics that were treated, at the time, as less core to the values and methodologies of documentary linguistics.

Thursday 14	Friday 15	Saturday 16	Sunday 17
Visiting lecturers meet the students	Audio principles	Data management & formats	Grant writing OR Mobilisation
Defining documentation; the documentation process	Audio practical	Data practical; metadata	Ethics & IP
Corpus creation	Transcriber	Toolbox	Archiving
Form work-groups	Audio evaluation & discussion	Dictionaries (advanced Toolbox & Lexique Pro)	Wrap up

As the courses evolved, planning took into account the opportunities as well as the constraints of holding them at ILCAA in Tokyo. Through the lifespan of the courses, we enjoyed the very capable support from the LingDy office for the logistics of advertising for and screening candidates, organising travel and accommodation, arranging welcome and other social events, and booking rooms.

There was initially very little training equipment, which, we learned, was due to a quite different approach to equipping researchers for fieldwork in Japan vs that with which we were familiar with in the UK and elsewhere. While the typical UK-based postgraduate student would generally rely on their department or funding source to provide field equipment, we learned that equivalent Japanese students had to find their own means and equipment, and we heard more than once of students who had worked in fast food restaurants to save up to buy their own audio recorders, laptops and other field equipment. In turn, this meant that their equipment was not always optimally chosen, nor was training and support available for its use - a situation amplified by the generally individualistic and solitary research practices typical of Japanese academia. This meant that for the earlier workshops, we needed to transport rather large cases of training equipment from London, although by mid-way through the series, ILCAA had acquired their own excellent range of equipment. Thus, and still focusing on the audio component, while for students at our UK and European workshops, sessions on audio tended to set high challenges for participants in terms of understanding microphone attributes and psychoacoustics, those at DocLing seemed to offer many participants

starker revelations about the possibility of selecting different microphones.

On the other hand, the individualistic nature of Japanese scholarship meant that students tended to be more self-reliant than comparable UK students, and as a result it very quickly became routine that participants would turn up to the courses with their own fieldwork kit, even if with limited training in its use.

One of ILCAA's goals in establishing DocLing was to give Japanese participants the opportunity of exposure to international trends and practices, and in the second course, in 2009, this was expanded though establishing an open public lecture as a flagship part of the training event. In 2009, David Nathan gave the first such lecture, based on his team's work in London, *Archiving endangered language materials*.

The 9 year span of DocLing was a hefty portion of the timespan over which documentary linguistics itself developed, and the evolution of DocLing itself provided a microcosm of that development.

### 3. Social program

The DocLing workshops had a number of social aspects. Although they brought together mainly Japanese participants, these came from various universities across Japan, and, in addition, there was a regular sprinkling of participants from other countries such as China, Finland, France, Indonesia, Italy, Korea, Mongolia, Myanmar, Russia (including Republic of Buryatia and Tuvan Republic), Taiwan, USA, Ukraine, and Vietnam. The LingDy office typically organised welcoming events and a workshop-final party. The increasing emphasis on group work (see Section 5) encouraged participants to get to know each other, share experiences, meet and work outside the formal workshop hours, and to keep in contact afterwards. Some participants went on to enrol in courses at Peter and David's university, several undertook some kinds of ongoing collaboration, and many are still contacts via social media such as Facebook. However, the workshops' greatest social accomplishment is that they introduced two people (one as participant, another as group consultant/leader) who subsequently married and have had (so far) two children.

### 4. Staff

Staffing of the workshops grew steadily. Initially the workshops were taught exclusively by Peter Austin and David Nathan. They were capably supported by several ILCAA staff, including Hideo Sawada, Toshihide Nakayama, and the LingDy office headed by secretary Sachiko Yoshida. In 2010, Anthony Jukes, then at La Trobe University, joined the "permanent" team bringing his various skills including fieldwork methodology, linguistic software, media production and a variety of Indonesian contacts. Then in 2013, especially in the context of growing collaboration with Indonesian colleagues, Sonja Riesberg (Cologne University) joined the team and added

strong experience in remote fieldwork, community-oriented research, and linguistic analysis and corpus creation.

The ILCAA Japanese staff also gradually joined the teaching effort over the project lifetime, with Hideo Sawada adding sessions on the use of photography in language documentation, Toshihide Nakayama on documenting conversations and Honoré Watanabe on fieldwork techniques. Finally, the workshop also occasionally included some ‘guest’ presenters, including Nikolaus Himmelmann, John Bowden, Iku Nagasaki, and Anna Berge.

## 5. Group projects

The most unique aspects of the workshop series were the increasing emphasis on group work, the facilitation of that group work by language specialists, and the evolution of the group activities from year to year. The idea of using groups as a locus of learning is not, of course new, and was already a core part of the training practice at HRELP and elsewhere. However, it is not merely a stereotype that in Japanese academia, people typically tend to work and study in solitary ways. While it is beyond the scope of this article to speculate how such solitary ways fit with a Japanese ethnography, a majority of our participants consistently reported that they had never before engaged in group activities, despite many of them being postgraduate students and academic staff. Group work started tentatively in the first couple of years, more or less as an experiment. David in particular was a driver for this work, drawing on his experience a few years earlier as an English professor at Tsukuba University, and many techniques he learnt from a high-impact teacher of Mandarin Chinese, Dr. Meili Fang. Several of these techniques involved providing the right balance between motivation, instruction, and theatrics in order to catalyse participation against the background of Japanese reticence to hold forth “*deru kui wa utareru*”.

In earlier workshops, group activities were rather like university linguistics Field Methods sessions. Group activities were preceded by formal classroom sessions on language documentation methods and tools (recording and software). Then, groups of 3 to 5 participants were each assigned a speaker of a language unknown to the participants (in the earlier years, these were mostly Mongolian speakers who were students at Tokyo University of Foreign Studies), and the groups attempted to “document” some feature(s) of their consultant’s language using the methods and tools they were exposed to in the more formal sessions.

Over the workshop series, we made several changes to this basic structure. Firstly, we made the group sessions more task- and outcome-based, so that the last afternoon was devoted to group presentations of their findings. We refined this to make the group activities more explicitly in the style of projects, so that groups were asked to prepare plans, negotiate roles, and present their plans to the class, thus providing a broader experience of working in teams. In fact, many of the participants reported in their

evaluations that this team-based work was not only their first experience of such activity but also the most stimulating, valuable and transforming aspect of the workshops (for examples of participant feedback, see the final section of this paper).

The next refinements, following trends in language documentation through the late 2000s, were to swing the major emphasis away from the more descriptive and formal linguistic side of documentation to more community-contextualised, humanistic and holistic approach to language documentation. We set this path through two mechanisms: firstly, by changing the roles and relationships with the groups' language consultants, and secondly by setting requirements on the group project goals. By 2012, the LingDy project was increasing its links with Indonesian linguists, and we were fortunate to have a cohort of four such linguists who could play the role of consultants to the groups. However, together with these linguists, we decided to adopt a kind of *realpolitik*, such that rather than "pretend" to be naive language consultants, these linguists would work with groups not only as language speakers but also as experts on their communities' language needs and contexts, and indeed as group *leaders* rather than servants of the groups. In addition, we prescribed that each group's project should include a community-oriented aspect, along a sociolinguistic, pedagogical, advocacy, or ethnolinguistic theme.

The teaching staff also provided assistance to the groups, as facilitators. In some cases, teachers were assigned to groups in order to provide particular skills to complement the groups' project goals. For example, those with an emphasis on video generally had access to Anthony Jukes who has experience with video; those who were developing websites had the assistance of David Nathan who is an experienced web developer.

Here is a description of the group work as presented to participants in the *DocLing Handbook 2016*:

*Practical group projects are an important part of the workshop. They provide a way to embed practical, specific, advanced techniques in a practical, holistic activity. Course participants work together in their group, with a consultant and staff, to practise and to further develop the skills covered in the formal sessions. Project work focuses on exploring language documentation through:*

- *roles and skills in group work*
- *planning in relation to documentation contexts and goals*
- *understanding workflow and documenting decisions*
- *collaboratively working towards concrete outcomes*

*Each group will give a public presentation describing their methods and presenting their documentation outcomes.*

*Groups will consist of approx 4 participants. Each group will be allocated a language consultant, and a staff member. The consultant will assist and advise the groups on language and community aspects, and the staff member will act as a mentor and help with skills. Each group will work to a specific theme, such as documenting a linguistic genre or feature, designing pedagogical materials, planning a project grant application, writing a sociolinguistic description etc. On the final day of the workshop, each group gives a public presentation about their planning, activities, decisions, outcomes, problems, and potential future work.*

The central role of group work can be seen in the program for DocLing 2016, which spanned 6 days (8–13 February). Group work related sessions fill 12 of the 24 session slots.

	10:00–11:20	11:40–13:00	14:00–15:20	15:40–17:00
Mon	Introduction to language documentation (PA)	Methods, materials & genres (SR, JB, TN)		Mobilisation: audience research & design (DN, AJ)
Tue	Ethics (PA)	Data management & archiving (lecture & practical) (DN)		Group projects startup: assign groups, themes, consultants, mentors
Wed	Audio (AJ, DN)	Software (PA)	Discussion forum - roles, languages & communities	Group work - plan & report
Thu	Still and moving image (HS, AJ) & discussion	Group clinic (topics as required)	Group work (facilitated)	
Fri	Discussion forum - documentation theory & methods	Group clinic (topics as required)	Group work (facilitated)	
Sat	Group work: finalise products and presentations		Group project presentations, Closing session	

As a result of this evolution, the workshops increasingly resulted in group projects completing resources that were of enduring and shareable value, and several of them were published on the web. Several examples can be seen at <http://www.el-training.org/outcomes/index.php#docling2015> and <http://www.el-training.org/outcomes/index.php#docling2016>.

A feature of the group work component was the afternoon of final presentations, held on the final afternoon of the workshop. The afternoon consisted of presentations from each group. Each group member was expected to contribute to the presentation, and the aims of the presentation were for the group to explain their plans, methods, difficulties faced, and to demonstrate the outcomes of their group sessions. More broadly, our motivations for requiring presentations were to (a) amplify the learning by sharing experiences and outcomes across all groups, (b) emphasise that language documentation and working with communities had broader practical applications and outcomes than narrow linguistic description and analysis, and (c) to round out the (for many) unique experience for Japanese scholars of working in collaboration and public speaking.

The presentation afternoons tended to take on a slightly festive atmosphere, fed by the adrenalin of working to a deadline and presenting in public and on occasion, to the culmination of groups' rather intense interactions over the week and impending departure, and some theatrics on the part of the teaching staff.

Another factor that made the workshops exciting was that, since the courses were rather autonomous, and not part of any accredited or mandated curriculum, we could experiment with content and activity structures. For example, in 2012, we ran an "experimental" group activity stream as a kind of distributed corpus preparation. Each group elicited and documented some language material, and in addition were tasked with negotiating and co-operating with other groups in order to understand the structural conventions and data management used by other groups so that they could interchange and combine the materials created by each individual group. Final presentations involved each group presenting the way that they had combined all the materials of the other groups with their own. The presentations were both impressive and diverse, with different but valid and creative approaches taken by each group.

The final DocLing saw a further development in the complementarity between the formal teaching sessions and the group work. In line with some trends in language documentation away from narrow emphasis on technical linguistic description and analysis, particularly a near-obsession with morphologically glossed written text, we decided to drop the formal sessions on some software tools (such as ELAN, Toolbox), and rather facilitate skills in these tools within the group work for those groups whose projects utilised them. In a small way, this change might be taken to represent the completion in a generational cycle of thinking about language documentation practice, evolving it from a bolting-on of some humanistic aspects (e.g. ethics, advocacy, pedagogy) to "classic" linguistic description and analysis, to a set of practices more



truly committed to responding to language endangerment.

## 6. Impacts and outcomes

Finally, after the conclusion of the workshop series, we are left to consider the impact and outcomes of the workshops. Perhaps the most revealing perspectives are those of the participants themselves, and we begin this section with selections from the feedback that we received from participants. The DocLing team emphasised continual improvement through seeking feedback from participants, both through discussion during the final wrap-up session, and follow-up surveys.<sup>2</sup>

Participants let us know about specific things they had valued learning, especially those which they had previously struggled with:

*“I learned differences between language archiving and descriptive linguistics, whereas both of them compliment each other”*

*“I always have a problem with data management. Attending the lectures help me to solve my problem”*

They frequently noted topics and concepts that were quite new to them. One frequently occurring feedback item was about the participants’ previous unfamiliarity with all aspects of audio recording and their surprise at how much they had gained from learning about it:

*“Before attending the work shop I only focused on recording necessary materials since I recognized audio data as merely an option to carry out descriptive language; however, I learned that it is very important to focus on a good sound quality as well”*

And other ‘discoveries’:

*“It was new to know that there is no such thing like an “idealistic documentation””*

*“An idea about finding supporters after making the locals understand what the researcher does in the field was very new to me”*

*“Although I was not interested in creating materials for language education, I found out it was more exciting than I thought and I discovered I can create educational materials that can be used by others”*

Some very useful feedback let us know how effective (or ineffective) our teaching was:

*“I think the ethical issues are really difficult to deal with because we have to learn legal issues to clarify the problems. However I could enjoy learning this*

---

<sup>2</sup> Note that we have lightly edited some of the feedback responses to improve readability.

*topic because it was more like a quiz program. I think it is an interesting and probably the best way to learn each case. It was a lot of fun"*

*"I was glad to learn little about XML. The demonstration about making xml files on excel was very easy to understand"*

The most enthusiastic comments were about group project work, in particular because it was often a novel or even first experience of collaboration for Japanese students and researchers:

*"Group discussion was a good training for us because we could learn so many things from each other, which was a refreshing experience"*

*"It was a great experience to listen to the presentation of other groups in terms of being able to learn the way to approach each issue from different aspects"*

*"I thought it would be so fantastic if I could work with my teammates when I actually work on archiving everyday!"*

*"This was the best part. I had to come to terms with my weakness, realized how group projects are productive, stimulating and challenging. More could have been done but it's not easy to be perfectly coordinated with people you just met a few days before. Overall, I learnt a lot from all the participants. This is a rewarding experience personally and professionally"*

*"The highlight of this workshop is the project work"*

*"It is the most enjoyable part of this workshop. We are very lucky to have great consultants, and also we are happy to be a team with other students"*

And of course for some, group work presented new challenges:

*"The difficulties in making communications internationally with people who do not have same mother tongue"*

Some comments noted that previous suggestions for improvement that had been addressed:

*"Since I participated in the previous DocLing 2011, I found the lecturers improved the direction of the project work. Last year the participants had to start from choosing topics and we had too many things to manage. However, this year the topic was already given and it was clear what to record, so we could concentrate on recording and data management"*

Participants frequently expressed wishes for more hands-on time:

*"Lectures were great but I would like to have a more time for practical sessions"*

Some expressed deeper wisdoms and transformations:

*"I knew that the technical methods of documentation are important so at first I*

*thought this workshop could be a good opportunity for me to learn how to record the language. However, it was a great discovery for me to know that it's much more important to understand what for we are recording the language"*

*"And the fact, or point of view, that language documentation and language description is not separate but connected was very interesting and enlightening for me. I think from my mind that it was very important step for me to have attended this seminar"*

Very encouraging resolutions were made:

*"In order to make a first step for managing data, I am going to organize my metadata"*

Some feedback was disarmingly frank, perhaps illustrating differences between Japanese and western cultural and expected teaching styles:

*"[following] my first experience of fieldwork, it was so instructive that I reflected my failures and learned important lessons. For example audio quality and detailed metadata are quite important. If I failed them, the irreplaceable materials would be lost. Especially when You Mr. Nathan told me face to face that the MP3 data I have was awfully bad and you looked a little angry about that recording, I felt my responsibility as a linguist by actual meaning. I will pay much more attention to record the sound materials with the best quality I can. I won't forget to write metadata"*

Overall, feedback indicated a very high level of satisfaction with content, delivery and their learning, and especially appreciative of teacher input, possibly because that is a less frequent phenomenon in Japan:

*"The advisers who had continuously supported our groupwork were great and kind. They are experts and continuously offered us the knowledge and skills to solve with the problems we faced. They helped us experiencing each activity regarding documentation and fieldwork. I would like to express my gratitude and appreciation for their support"*

Other measures of impact relate to the sheer reach of the workshops. Participants, although largely Japanese, also came from a range of other countries including Korea, China, Taiwan, Vietnam, Russia, Mongolia, Ukraine, Finland, France, Indonesia, Italy, Tuvan Republic, and the USA. And these participants were working on an even wider range of languages, including various Japanese dialects, Ainu, Ryukyuan, Uilta (Russia), Southern Min (China, Taiwan), Tibetan, Bende (Tanzania), Yakut (Russia), Mongolian, Korean, Persian, Basque (France), Kurdish, Turkmen, Coptic (Egypt), Swahili (Tanzania), Urdu (Pakistan), Dhivehi (Maldives), Roma (Czech), Breton (France) and several Indonesian languages.

The research and careers of many participants were aided by the workshops. For example, Robert Laub went on to complete his MA in Language Documentation and Description at SOAS University of London; Anna Bugaeva became a recipient of ELDP funding for documentation of Ainu language and collaborative project with the Endangered Languages Archive; and Michinori Shimoji went on to further document Ryukyuan languages. Those who participated as consultants also reported many positive outcomes, not just the chance to visit Tokyo (for some Indonesian visitors their first time to see snow) but also to learn more about their own languages, to learn the skills of language documentation, and to build links and friendships with the students and LingDy staff (see Yanti, this volume).

More broadly, the workshops cemented and initiated collaborations between ILCAA and other institutions, for example with SOAS University of London, and also with the University of Hong Kong, which has now seen a number of exchange events with the LingDy project. Many of the Indonesian consultants also played central roles in organizing the LingDy workshops across Indonesia which grew out of DocLing (see Jukes and Shiohara, this volume).

There are, of course, also some enduring teaching materials, most of which have been made public via the web. Several of these can be found at the following locations:

- <http://lingdy.aacore.jp/en/activity/docling.html>
- <http://www.el-training.org/courses/docling/2016/>
- <http://www.el-training.org/courses/docling/2015/>

And especially the participants' group work outcomes:

- <http://www.el-training.org/outcomes/index.php#docling2015>
- <http://www.el-training.org/outcomes/index.php#docling2016>

In the final years of the workshops, David and Anthony prepared handbooks for participants, elaborating on the program, content of sessions, links to further reading, and in particular with more detailed information about the group projects. These handbooks also help to leave a documented legacy of the workshops. But far more importantly, as a result of the DocLing workshops nearly 200 scholars have become wiser and better equipped for the important task of documenting endangered languages throughout the world.

## Language Documentation and Legacy Text Materials

AUSTIN, Peter K.

Linguistics Department, SOAS

For projects interested in documenting, describing or revitalizing languages, especially endangered languages, historically existing materials (whether digital or analog) like tape recordings made in earlier times or written materials collected years or even centuries ago may exist and may represent important sources of information, indeed, in some cases, the only information available. Making use of legacy text material raises many challenges that need to be confronted if we wish to include it in a corpus or to treat it along with other contemporary data. There are practical, technical, contextual, ethical, and political issues that legacy materials raise, and many questions which it may be difficult or even impossible to answer. There are also many opportunities to add value to legacy materials using documentary linguistics methods. We discuss and exemplify these throughout this paper. A conclusion that can be drawn from this work is that creating good metadata and meta-documentation for current written materials can potentially reduce legacy data problems for future researchers compared to the issues that we face today.

**Keywords:** language documentation, legacy text materials, metadata, meta-documentation, archiving

### Preamble

1. Introduction
2. Form, content and context issues
3. Stakeholder issues
4. Conclusions

### Preamble

This paper arises from teaching materials developed for sessions on data management for DocLing training courses at Tokyo University of Foreign Studies and is intended for beginning researchers to alert them to some of the challenges around working with legacy text sources and how documentary linguistics methods can be applied to confront these challenges and add value to such material. In some situations there may be no living speakers of the language recorded in such sources and legacy materials are all that exist; if processed and analysed well they can play an important role in language reclamation and revitalisation (see, for example, Amery 2000; Baldwin et al.

---

AUSTIN, Peter K. 2017. "Language documentation and legacy text materials". *Asian and African Languages and Linguistics* 11. pp.23–44. [Permanent URL: <http://hdl.handle.net/10108/89205>]

2016, 2017; Costa 2003, 2015; Giacon and Lowe 2016). The present paper expands on the training handouts and adds further examples and references to relevant literature.<sup>1</sup>

## 1. Introduction

Over the past 20 years a new area of research on human language called Language Documentation (or Documentary Linguistics) has developed, drawing on ideas from linguistics, anthropology, verbal arts, information science, media and recording arts, ethnoscience and other areas (Austin 2010, 2016; Himmelmann 1998, 2006; Lehmann 2001; Woodbury 2003, 2011). In his seminal article on the field, Himmelmann (1998: 161) presented its main goal as ‘to provide a comprehensive record of the linguistic practices characteristic of a given speech community’. Himmelmann (2006: v) restated this as concerning ‘the methods, tools, and theoretical underpinnings for compiling a representative and lasting multipurpose record of a natural language or one of its varieties’, while Woodbury (2011: 159) gives a similar definition: ‘language documentation is the creation, annotation, preservation and dissemination of transparent records of a language’. This approach emphasizes transparency and multifunctionality, as well as ethical engagement with a wide range of stakeholders, including speech community members. Himmelmann (1998: 161) also argued that language documentation ‘differs fundamentally from... language description [which] aims at the record of a language... as a system of abstract elements, constructions, and rules’. This view is critically discussed in Austin and Grenoble (2005) who argue that such a strong distinction is not useful or desirable.

Most of the theoretical and practical work on language documentation to date has tended to assume implicitly that it deals with audio-visual and textual data collected at the present time, and has not paid much attention to historically existing materials (whether digital or analog) like tape recordings made in earlier times or written materials collected years or even centuries ago (an exception is Himmelmann 2012 which discusses various data types). I will refer to such information sources as *legacy material* or *legacy data* in the discussion which follows, and concentrate on written textual materials in particular (what Himmelmann 2012 calls ‘original written document[s]’, which he considers to be *raw data*<sup>2</sup>). Pre-existing audio-visual recordings present many challenges to language documenters, including issues of digitisation, transcription, interpretation and contextualisation.<sup>3</sup> For example,

<sup>1</sup> Thanks to Lyle Campbell, Lisa Conathan, David J. Costa, Lise Dobrin, Andrew Garrett, Anthony Jukes, Susan Smythe Kung, David Nathan, Gabriela Perez-Baez, Anthony C. Woodbury and two anonymous reviewers for discussion and comments on earlier versions. I am also grateful to Lauren Gawne, Cristina Muru, Julia Sallabank and Candide Simard for feedback on a workshop presentation of some of this material at SOAS, 20 February 2017. None of these scholars can be held responsible for errors in the use of their feedback.

<sup>2</sup> The notion of ‘raw’ here is problematic as some, perhaps much, legacy written material is ‘processed’ in various ways, ranging from transcription and/or translation to full grammatical analysis (see examples in sections 2.1, 2.3 below).

<sup>3</sup> Many thanks to Lise Dobrin for discussion of these points.

digitisation problems may include dealing with mould on old tapes or the difficulty of finding machines which are able to play obsolescent formats like Hi8 video or minidisk audio. Specialist organisations such as *Paradisec*<sup>4</sup>, the *Phonogrammarchiv* of the Austrian Academy of Sciences<sup>5</sup> or the *British Library Sound Archive*<sup>6</sup> are set up to deal with tasks like cleaning the carriers, providing equipment, and digitising such material. Difficulties of transcription and translation can include being able to comprehend the speaker(s) and/or language in the recordings, differences between contemporary language use and that in the legacy recordings, mismatches between transcriptions and translations (if they exist) and what is in the recording (see Mosel 2014), and lack of sufficient background information to be able to make sense of the content and/or context. Fuller discussion of these issues in relation to audio or video recordings is beyond the scope of this paper, and I will concentrate on text materials only in the following.

Legacy text material raises many challenges that need to be confronted if we wish to include it in a documentary corpus or to treat it together with other contemporary data.<sup>7</sup> There are practical, technical, ethical, and political issues that legacy materials raise, and many questions which it may be difficult or even impossible to answer. There are also many opportunities to add value to legacy materials using documentary linguistics methods. We discuss and exemplify these below. When adding value to legacy materials, it is important to distinguish between adding structure (categories, entities, relationships), adding content, and adding format. Note that explicit and well-structured data (e.g. stored in a database, or marked up in extensible markup language (XML)) can have format added computationally, and also lends itself to repurposing for other uses and/or other users than the immediately intended audience (and so can be multifunctional, as language documentation proposes), see Gippert (2006: 358–361) and section 2.1 below for discussion and exemplification.

To make sense of legacy materials it is essential to explore the socio-cultural and historical context of the documents and their creation, including the biography of the author(s), especially what prior language knowledge and/or study and/or exposure they had, who their teachers/mentors/correspondents were, how long they worked on the language and at what point in their careers, how the work was funded and with what goals, whether there were previous studies of the language or the community that they could have had access to, and so on.<sup>8</sup> It is also important to explore aspects of the historical period during which the materials were created in terms of the kind and impact of contact between communities, including colonialists, and what descriptive categories and formats would have been known and might have influenced the author(s),

<sup>4</sup> See <http://www.paradisec.org.au/>, accessed on 2017-04-06.

<sup>5</sup> See <http://www.phonogrammarchiv.at/wwwnew/>, accessed on 2017-04-07.

<sup>6</sup> See <https://www.bl.uk/subjects/sound>, accessed on 2017-04-07 and Copeland (2008).

<sup>7</sup> See Goddard (1973) for discussion.

<sup>8</sup> An example of this kind of detailed exploration is Silverstein (2015) on the historical development of Franz Boas' fieldwork methodology and products.

e.g. traditional grammar based on Latin or Greek models. For further discussion see section 2.4 below.

In the following sections we discuss and exemplify some of the issues to be confronted when working with legacy text materials in terms of formal, analytical, stakeholder and ethical matters.

## 2. Form, content and context issues<sup>9</sup>

In this section we discuss some of the challenges raised by the form, content, context and analysis of the written documents.

### 2.1. Issues with the form of the original

Reading and interpreting handwritten texts can be problematic as it can be difficult to deal with cursive writing and especially old writing styles that are no longer in use. The study and interpretation of old handwriting is called paleography, however most of the training material readily available on it<sup>10</sup> focuses on reading medieval or later manuscripts in European languages and there is nothing that I have been able to locate on reading other sources, especially non-Western materials.

An example of challenging linguistic materials is Figure 1 which is an extract from the Diyari grammar of Flierl (1880) where the grammatical explanation uses an archaic German script (the Diyari words are in regular Roman script).

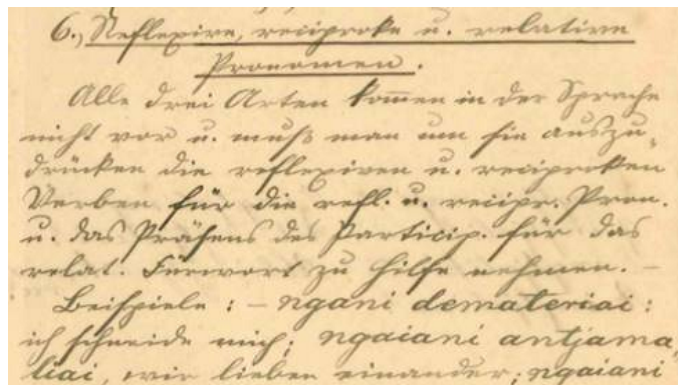


Fig. 1 Diyari Grammar extract from Flierl (1880), from Stockigt (2016: 80)

Sometimes, documents are written in non-Roman scripts such as ideographic or syllabic characters or phonetic scripts. Interpreting the symbols used in a document

<sup>9</sup> See also Bower (2003) for examples of some of the topics covered here. Andrew Garrett (p.c., 2010-02-05) notes that similar issues of interpretation can arise with contemporary 'born digital' data, and are not restricted to legacy materials in the sense intended here.

<sup>10</sup> For example the online course at <http://www.nationalarchives.gov.uk/palaeography/>, accessed on 2017-04-09.



much documentation was written in inaccessible transcriptions, often written phonetically (not phonemically) in the Finno-Ugric Transcription System, also known as the Uralic Phonetic Alphabet (UPA). It is important to note that the phonemic principle did not come into force until around 1930 or later. While the UPA is established among Uralic scholars, materials originally rendered in close phonetic transcription in this notation are formidable even for scholars accustomed to the UPA, and far worse still, are intimidating to the point of uselessness for speakers of the languages transcribed

nũ kũā' bõ: nu' ga' ? t' dšābaktõrāi' t' b' : . amnā' ? , dšābak-  
tõr t' b' bā : ? . amõ' rāit' t' pvd' qzõ' t' ta. amnā' ? , amõ' ra' ? .  
deñ i' p' k' i' p' c' t' a' p' o , k' a' j' a' k' nũā' , šamnā' k' i' g' . t' u' b'  
i' g' , munu' i' , u' j' a' , k' c' t' a' . amā' ? bar. mēn t' u' j' o' k' a-  
t' a' m' šumurā' n' o' , a' t' i' n' mā' n' a' t' amnō' . j' āt' o' m' a' . mēn m' b' i' t'  
b' u' b' i' t' n' a' n' t' e' t' i' t' . d' g' g' t' e' m' e' t' i' n' a' n' . t' c' a' d' i' n' b' o-  
b' i' : mēn t' u' r' a' n' b' i' . mēn m' b' i' a' m' : i' p' p' ? k' u' n' o' t' i' t' t' .  
b' i' t' t' e' . m' b' i' a' m' : amõ' ra' ? , pvd' amõ' rāit' t' qzõ' t' ta.  
d' e' m' p' i' : munu' i' amõ' rāit' t' n' . mēn t' u' r' a' n' a' m' b' i' a' m'  
d' i' m' . d' e' g' t' e' i' p' b' i' k' u' n' o' t' i' t' t' , m' a' n' i' p' p' i' b' i' a' m' .  
i' n' t' i' n' u' p' t' u' b' i' a' m' , šomurā' n' o' m' b' i' a' m' . šobā' m' , d' e' n' m' m-  
b' i' a' m' : u' p' t' a' ? , amõ' ra' ? , deñ namzõrā' s' i' e' t' . šut i' g' ,  
k' i' p' t' e' i' g' , munu' i' , i' p' k' . amõ' ra' ? , amnā' ? . b' i' b' i' a' m'  
d' e' m' p' i' , buzai' z' i' b' i' , d' e' g' t' e' amnō' b' i' a' m' : amõ' rāit' t' . d' i-  
g' i' t' i' k' a' t' a' d' i' n' b' i' . a' m' o' n' b' i' n' a' i' b' i' a' m' .

Another example of the transcription is a Saami paradigm for the word for ‘fish’ in Figure 3 from the website of the 3rd Sami Linguistics Symposium:<sup>11</sup>

<sup>11</sup> See <http://www.frias.uni-freiburg.de/de/veranstaltungen/frias-tagungen/saals-3-3rd-saami-linguistic-symposium>, accessed on 2017-05-17.

	<u>Sing.</u>	<u>Plur.</u>
nom.	gùo̅ll'ie̅	gùo̅ll'ie̅k <sup>x</sup>
gen.	gùo̅ll'ie̅	gùo̅ll'ie̅t
ack.	gùo̅ll'ie̅ <sup>up</sup>	gùo̅ll'ie̅t
ess.	gùo̅ll'ie̅m	gùo̅ll'ie̅m
iness.	gùo̅ll'ie̅m	gùo̅ll'ie̅m
lat.	gùo̅ll'ie̅st	gùo̅ll'ie̅st.
illet.	gùo̅ll'ie̅t	gùo̅ll'ie̅da - i̅t
kom.	gùo̅ll'ie̅(a)	gùo̅ll'ie̅t

Fig. 3 Pite Saami noun paradigm by Israel Ruong, archived at Institutet för språk och folkminnen, Uppsala.

Dobrin (p.c., 2017-05-11) gives a different example of problematic source materials for the Papuan language Arapesh:

The earliest Arapesh grammar was published in the journal *Anthropos* (Andreas Gerstner's *Grammatik der Alubänsprache* [Nordküste von Neuguinea]). It used the most bizarre notation for vowels, which I later learned was their standard house style. The only reason I could make sense of it was that it was modelled on the customary Massoretic pointillation of Hebrew.

The huge corpus of materials in the Meskwaki language spoken in Wisconsin, USA, is written in a unique syllabic script called papepipo which 'has a basic matrix of 48 syllables, comprising four vowels by themselves and compound characters for 11 consonant symbols combined with four vowel symbols' (Goddard 1996: 117).

These examples illustrate that sources need to be studied closely and that there may be different issues and problems to be faced depending on the particular situation. Interpretation of document orthography and spelling of a given language may require philological and linguistic training to be able to analyse the original and map it to a modern phonetic or phonological representation (see Broadbent 1957; Crowley and Austin 2005; Koch 2011). Austin (2008) discusses problems with the interpretation of William Ridley's 19th century publications on the Gamilaraay language of New South Wales, Australia. Ridley marks vowels in this language with or without a macron (e.g. ā versus a), however this is ambiguous and can represent vowel length

and/or stress. Issues that arise in reading and interpreting the corpus of Meskwaki have been discussed in Goddard (1996) and Bear and Thomason (2011) (see also Dahlstrom 2015). Bear and Thomason (2011) note the following:

Meskwaki papepipi is adapted to be wonderfully easy to write. It is much harder to read. Papepipi omits vowel length, the consonant ‘h’, and nearly all punctuation. This means that spoken Meskwaki has eight distinct vowels, but papepipi writes only four vowels; spoken Meskwaki has eleven distinct consonants, but papepipi writes only ten consonants; spoken Meskwaki has 29 distinct consonant clusters, but papepipi writes only 16 consonant clusters. As a result, there is a great deal of educated guesswork involved in transcribing papepipi into fully phonemicized words, phrases, and sentences. The edited and translated text of ‘Skunk and Grizzly Bear’ should be taken provisionally and read critically: there may be mistakes or infelicities in Y[oung] B[ear]’s original papepipi, in L[ucy] T[homason]’s reading of the papepipi, in L[ucy] T[homason]’s choices regarding phonemicization, word breaks, and sentence breaks, and in L[ucy] T[homason]’s choices regarding Meskwaki-to-English translation.

Goddard (p.c., 2017-05-26) notes that in his current work on an edition of a Harmony of the Gospels in Southern Unami (Oklahoma. Delaware; Lenape) produced by a Baptist missionary in 1837–1839:

The orthography is at the same time remarkably modern and seriously underdifferentiated (not distinguishing the long and short consonants), as well as inconsistent in the writing of the numerous vowels (not surprisingly). Reversing the odd choices of letters (‘v’ for /h/; ‘h’ for /č/; etc.) would still only produce an inconsistent mess of a transcription. But the translation is remarkably free and idiomatic, for example supplying enclitics that sounded right where the King James text has, of course, nothing. So here the edition has both the original and the phonemic transcription.

There may also be textual amendments (crossing out, additions), abbreviations, or other obscurities in the document, as in Figure 4 from Gerhard Laves’ fieldnotes on the Bardi language presented in Bown (2003):

Bown (2003) notes that Laves abbreviated words that occur frequently in his note (e.g. *g* for *ginyinggi* ‘this, he, she, it’, *g.on* for *ginyinggon* ‘and then’). Stephen Wurm’s 1955 fieldnotes of New South Wales Aboriginal languages contain glosses and translations in Hungarian shorthand.<sup>12</sup>

Careful retranscription of legacy documents into a modern form is necessary if they are to be made more useful and multifunctional. It may also be advisable to link

---

<sup>12</sup> AIATSIS archive item 002895A.



such transcription to images of the original documents so that readers may confirm the proposed analysis. A publication in print or online that aims to accurately reproduce all significant features in the original manuscript, including spelling and punctuation, abbreviations, deletions, insertions, and other alterations is called a diplomatic edition of a text. Online examples drawing on language documentation principles include colonial Zapotec documents curated by Lillehaugen et al. (2016), the Daisy Bates collection described in Thieberger (2016), and the website of the William Dawes manuscript of the Sydney language of Australia.<sup>13</sup> The underlying digital representation of the Dawes material uses XML extensible markup language and encodes crossing out, insertions, and amendments, as well as additions by the editors, such as spelling out abbreviations.<sup>14</sup> The transcription was created by the Dawes project team, with reference to Troy (1994), and then edited by tagging it in XML in order to:<sup>15</sup>

distinguish various text content structures (such as Sydney language vs English, and with the language content tagged for lemmas etc), person and place names, meanings, and commentaries. In addition, layout structures such as columns are represented.

From this XML source, XSLT transformations generate the content of each of the transcription pages (in two versions, edited and unedited). CSS is used to lay out the text visually on the web page.

An example of the resulting display which resembles a diplomatic edition (what the website somewhat misleadingly calls ‘unedited’) is shown in Figure 5.<sup>16</sup>

<sup>13</sup> See <http://www.williamdawes.org/>, accessed 2017-04-07.

<sup>14</sup> There is a long tradition in epigraphy of encoding this kind of information and value adding for creation of editions of ancient documents and stone inscriptions. For current approaches using XML see Bodard and Stoyanova (2016), the EpiDoc guidelines (<http://www.stoa.org/epidoc/gl/latest/>, accessed on 2017-05-29) and the cheat sheet available at <http://www.nesc.ac.uk/action/esi/contribution.cfm%3FTitle=964>, accessed on 2017-04-09.

<sup>15</sup> See <http://www.williamdawes.org/howtouse.html>, accessed on 2017-04-09.

<sup>16</sup> See <http://www.williamdawes.org/ms/msview.php?image-id=book-b-page-2>, accessed on 2017-04-09.

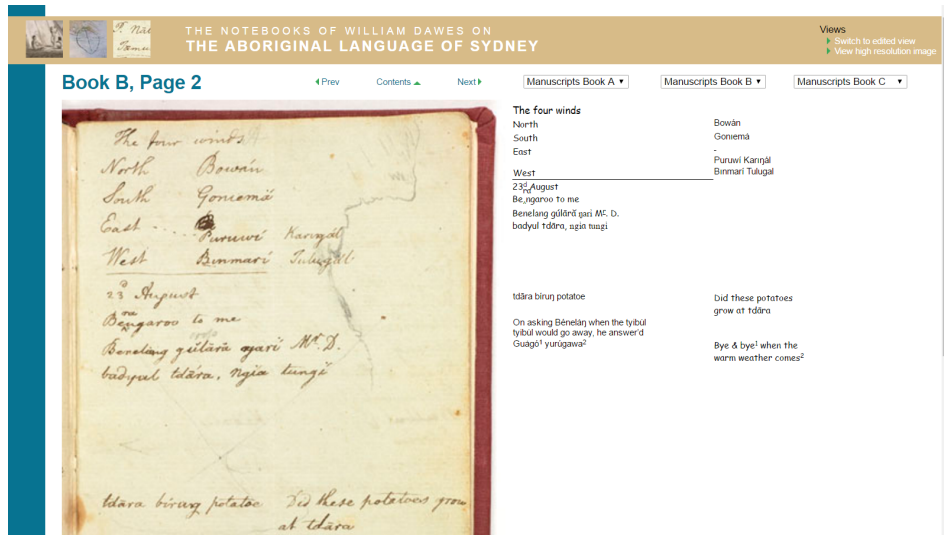


Fig. 5 Screenshot of William Dawes Notebook B page 2, unedited view.

A 'cleaner' (so-called 'edited') view of this page can be seen in Figure 6.<sup>17</sup>

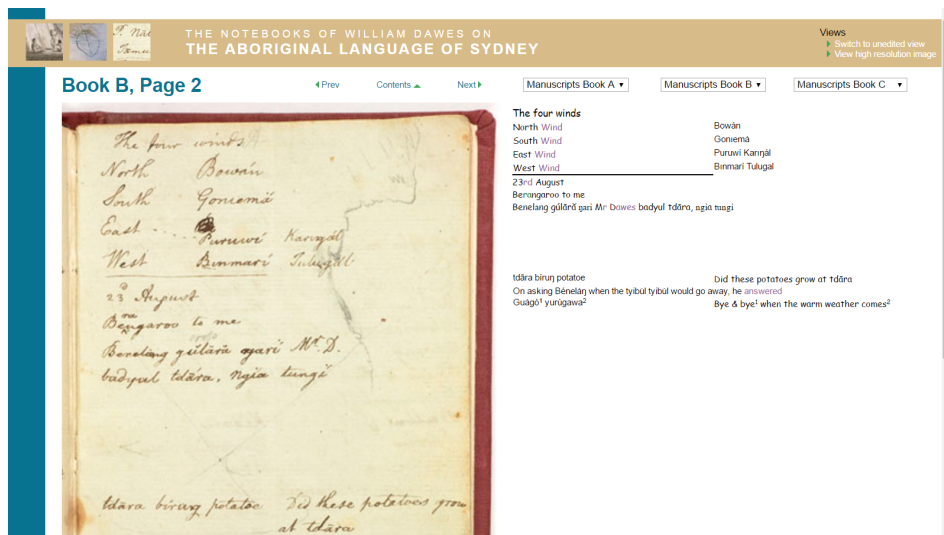


Fig. 6 Screenshot of William Dawes Notebook B page 2, edited view.

Note that Unicode fonts should be used for the representation in the corpus (see

<sup>17</sup> See <http://www.williamdawes.org/ms/msview.php?image-id=book-b-page-2&edited=true> (accessed on 2017-04-09).

Gippert 2006 on character encodings in language documentation).

Text material which is implicitly structured, such as by using typography or layout on the page to distinguish analytical categories or kinds of information, can be made more useful by encoding the structure separately from the form. For example, (1) presents a sample entry from the Scherer (1981) English translation of J. G. Reuther's four volume manuscript Diyari-German dictionary.<sup>18</sup>

(1) Entry for *banbana* in Scherer (1981).

5. *banbana* (v) = 'to stand still; to pause; to stop; to cease; to go no further', e.g.

- 1) *ninkida nau banbana warai* = 'here he stopped', i.e. he did not continue on.
- 2) With reference to the circulation of blood: *matja kumari banbai* = 'the [flow of] blood has already ceased'
- 3) With reference to water: *matja ngapa banbai* = 'the water is already stationary', i.e. has ceased to flow.
- 4) With reference to the star of the wise men from the East: *pungani miri ditji waka banbana wonti* = 'the star stood still above the house' (in Bethlehem).
- 5) With reference to a track: *ninkida paltu banbai* = 'here the track terminates, comes to an end'
- 6) With reference to a sandhill: *dako kajirani banbai* = 'the sandhill finishes up in the creek'

Nathan (2016) processed a digitized version of Scherer (1981) computationally and manually to create XML files where the different data types are tagged into a proper hierarchical structure, as shown in (2).

(2) XML representation for *banbana* entry in Scherer (1981).<sup>19</sup>

```
<?xml version="1.0"?>
<diyarilexicon version="20170409">
<entry label="5" num="5">
<lemma>
  <di>banbana</di>
  <pos type="lemma">v</pos>
</lemma>
  <gloss>to stand still; to pause; to stop; to cease; to go no further
</gloss>
```

<sup>18</sup> See <http://www.samuseum.sa.gov.au/collections/information-resources/archives/reuther-reverend-johann-georg-aa-266>, accessed on 2017-04-25.

<sup>19</sup> The presentation here is based on Nathan (2016) with amendments by the present author.

```

<eg num="1"> <di>ninkida nau banbana warai</di> <eg_gloss>here he
stopped, i.e. he did not continue on</eg_gloss>.
</eg>
<eg num="2"> With reference to the circulation of blood: <di>matja kumari
banbai</di> <eg_gloss>the [flow of] blood has already ceased</eg_gloss>
</eg>
<eg num="3"> With reference to water: <di>matja ngapa banbai</di>
<eg_gloss>the water is already stationary, i.e. has ceased to
flow</eg_gloss>.
</eg>
<eg num="4"> With reference to the star of the wise men from the East:
<di>pungani miri ditji waka banbana wonti</di> <eg_gloss>the star stood
still above the house (in Bethlehem)</eg_gloss>.
</eg>
<eg num="5"> With reference to a track: <di>ninkida paltu banbai</di>
<eg_gloss>here the track terminates, comes to an end</eg_gloss>
</eg>
<eg num="6"> With reference to a sandhill: <di>dako kajirani banbai</di>
<eg_gloss>the sandhill finishes up in the creek</eg_gloss>
</eg>
</entry>
</diyarilexicon>

```

Such an XML file can now be unambiguously searched for different types of data as well as presented in various formats, such as on a web page using XSLT and CSS to convert the data labels into HTML formatting. Nathan (p.c., 2015-10-02) noted the following issues arose when converting the original document files (which had been created by scanning and optical character recognition (OCR) of a printed copy of Scherer 1981):

- the documents had a massive over-use of quotation marks, for multiple purposes and often redundant or fatuous scare quotes. This made it hard to drive markup from them, even though in a huge number of cases they were the only clue to structure.
- scoping - the original documents had some mixed assumptions about scope, e.g. of entries, pages, footnotes etc. In particular footnotes and footnote references are relative to page units so in many cases they don't work properly (some are just in-line footnotes, which are OK).
- character data errors - mostly resulting from OCR errors.
- others, such as Scherer's spelling errors, inconsistent use of all kinds of punctuation and line breaks (e.g. some parsing could be driven by looking for strings like "i.e." but this appeared as i.e, i e., ie, i.e. Also, the part-of-speech



formulations are inconsistent and otherwise problematic.

Researchers need to be aware of possible complications such as these when working on adding value coding to documents.

A further problem with the form of legacy text materials may arise from cryptic glossing, or wrong glosses, because the author could not understand their language consultant's accent or pronunciation, or because the semantics of the source language terms were misunderstood. Crowley and Austin (2005: 60) give the examples in Table 1, drawn from various wordlists of Australian Aboriginal languages.

Table 1 Semantic problems in text materials, from Crowley and Austin (2005).

Wordlist meaning		Correct meaning
1.	Pronunciation problems	
	heart	hot
	wet	sweat
	moths	boss
	dung, shit	tongue
2.	Meaning problems	
a.	generic versus specific	
	grass	vegetation
	boy	uninitiated youth
	beard	hair
	day	now
	thumb	your hand
	girl	female
b.	related word	
	thighs	buttocks
	cloud	sky
	woman	wife
	hair	head
	frown	blind
	spider	to bite
	dig	drink

Another issue to pay attention to is that understandings and analyses by the text's author(s) of the language being recorded can be seen to change over time, and thus different parts of a collection of text material may show different spellings, translations etc. For example, Bower (2003) mentions that Gerhard Laves began to analyse the

Bardi material he was writing in his fieldnotes as he collected it and made a number of analytical mistakes as a result, that is, later parts of the document does not contain what he actually heard but rather what he thought he heard. Also, Steele (2005: 84) notes in relation to William Dawes' records of the Sydney language that 'in order to be in a position to make some assessment of the soundness of an interpretation of a word, expression or sentence provided by Dawes, it is useful to have an idea of at which stage of his language learning an entry was created'.

## 2.2. Issues with the content of the material

There can be issues of various types that arise due to the content of the original text. For example, some of the content may be inappropriate to discuss in front of particular individuals or groups within a community (for more on sensitivities see section 3.1). Bown (2003) notes that Gerhard Laves' Bardi fieldnotes contain names that should not be spoken aloud because they are the same as the name of a close relative of a community member who has passed away and thus subject to a death taboo. Innes (2010) mentions stories in legacy texts that are considered to be 'dangerous' by a given community and should not be reproduced or distributed. Bown (2003) also cites the problem that some of Laves' fieldnotes contain information about secret male rituals and that she, as a female, should not read them; in this case the challenge is that she, as a trained linguist, is the only person with knowledge to decipher and potentially pronounce particular forms in such texts.

There may also be content that is dated or inappropriate by contemporary standards. For example, Bown (2003) notes that manuscripts may also contain notes by the original author that were intended as private personal comments on the people who being worked with which other people, such as contemporary relatives, may find offensive. In addition, legacy materials may use ways of referring to indigenous people which were acceptable at the time they were written but which are offensive now. Authors may also have written down personal judgements about the nature of the material collected (e.g. 'superstitious rubbish') that do not match modern community judgements (e.g. interesting personal history). Andrew Garrett (p.c., 2010-02-05) notes that:

A lot of excellent examples of inappropriate content come up in early 20th century California fieldwork. People who never thought that any non-scholars, and certainly no Native people, would ever see these field notes, routinely wrote stuff like 'half-breed' to summarize genealogy, or commented on fluency levels in ways that people now would find insulting, or even wrote things like 'a syphilitic' next to people's names.

How and indeed if these kinds of content should be represented in the modern corpus will need to be decided (e.g. by censoring them in some way, or placing them in

password-protected files).<sup>20</sup>

### 2.3. Issues with analysis in the original

Difficulties with interpreting the content of legacy text materials can often arise if the original author either:

1. records what they think is a distinction (phonological, morphological, syntactic) that is not actually present in the language but may exist in the author's native language or one that they are familiar with; and/or
2. misses some crucial contrast because it is not made in the languages the author is familiar with (as a native language, languages studied).

Examples of the first type can be found in Reuther's Diyari dictionary where he distinguishes between voiced and voiceless consonants (e.g. <b> versus <p>, as in the entry for *banbana* in (1), however this contrast is only truly applicable for apico-domal (retroflex) stops. Similarly, Reuther writes vowels <e> and <o> although Diyari only has three contrastive vowels (*i*, *u*, and *a*). An example of the second type in Reuther's work is his failure to record the difference between stops and nasals at various points of articulation: apico-alveolar (*t*, *n* in the modern orthography), lamino-dental (*th*, *nh*), and apico-domal (*rt*, *rn*). Thus, we find a collapsing of distinctions in *nganha* 'me' (Reuther's *ngana*) and *nganarna* 'be.participle' (Reuther's *nganana*).

Examples of both types of misanalysis for grammar can also be found in missionary records of various Australian Aboriginal languages, including Diyari, as discussed in Stockigt (2016). For example, Reuther, and earlier missionaries, recorded a 'vocative' case for Diyari on the basis of Latin grammar models they knew, although the forms listed are actually phonological distortions used on shouted speech (see Austin 1981, 2015). On the other hand, they failed to notice that verb forms they identified as 'Modus Conditionalis' participate in a switch-reference system (encoding different subjects between two clauses) in paradigmatic contrast to verb forms they identified as infinitive (which encode same subject). Numerous examples of other such misanalyses can be found and the researcher needs to be careful when including legacy materials into their documentary corpus to label the original analysis and clearly distinguish it from their own, as well as cite the original source and where it can be found.

### 2.4. Issues arising from a lack of context

Some of the most difficult issues to deal with in legacy text materials relate to the lack of metadata (data about the data) and meta-documentation (information about the context of collection and analysis, see Austin 2013) which would help with understanding and analysing their form and content. It is often unclear, for example, where data comes from as speakers are not identified, and their geographical origins,

<sup>20</sup> A notorious example of this kind of material actually being published and causing negative reactions is Malinowski (1989).

social positions and relationships to other contributors are not specified. We are also not told whether they learnt the language as children or adults, and what other languages they might speak that could influence their knowledge of the language in the materials. As noted above, particularly useful can be information about the collector: their background, the languages they have some knowledge of, and their education history and prior language (and linguistic) study – sometimes this metadata is available in biographies or historical documents but often it is not. Darnell (1995) argues concerning anthropological fieldnotes:

A generation after the original research was carried out, the interpretative context in which the work was done – a context that was easily available to contemporaries – will require reconstruction. Such a task will only be possible if documents about the anthropologist, his/her research and professional milieu at the time have been preserved. Thus, records usually thought of as the history of anthropology are also crucial to the interpretation of anthropological records.

Also important is knowledge of the research training and methodology of the collector, including research methods and tools, what books and articles they were familiar with, who they studied and communicated with about the project (including mentors and colleagues), what the goals of the research were and where it fits within their own career trajectory, as well as their relationships with the language consultants and the community (see 3.1). For the latter, Good (2010) identifies ‘contact, consent, compensation, and culture’ as four important variables, that is how the community was contacted by the researcher, how consent for the work was given, what compensation was provided to participants and what were the cultural differences and expectations between the collector and the described community. For legacy materials these are often unknown or have to be reconstructed.

The issues we identify here are not unique to linguistic analysis of legacy text materials but, as Bishop (2006) points out, understanding context is a major challenge for secondary analysis of all qualitative data. Bishop (2006: 15–16) reminds us that the construction of knowledge is socially, culturally and historically embedded and needs to be (re-)studied as such:

Because secondary analysis is, in fact, re-contextualisation, as Moore (2005) suggests, this highlights how vital contextual information is to the process of reusing data. ... Both primary and secondary researchers have the responsibility to be reflexive in a manner suited to their specific projects. In the case of secondary analysis, reflexivity requires consideration of both the contemporary context and that of the original project (Fielding 2004). These challenges are significant, but not overwhelming; there may be lessons to learn from other disciplines. Historians do not lie awake nights agonising about not being able to do history because ‘they weren’t there’. They interrogate historical artefacts,

consider conditions that led to their production, while recognising that their own framework shapes what is seen (Scott 1990). Thinking deeply about context is a useful reminder that even the most knowing subject is never all-knowing. A perspective in which ‘data’ is reflexively constructed, contextualised, and re-contextualised helps us to acknowledge the inherently social character of knowledge.

### 3. Stakeholder issues

Documentation projects typically have many stakeholders who may have different kinds of interests in the materials collected and the analyses created. Issues of control, consultation, and decision-making are important when deciding what kind of documentary material to include in any corpus and how it can be used. This is also true of legacy sources, with the additional complication of possible mismatches between past situations and the present.

#### 3.1. Identifying stakeholders and relationships

As O’Meara and Good (2010) note, identifying who has a stake in a given document can be complicated by the fact that the current membership of a particular contemporary ‘community’ (however that is defined) may not coincide with past membership. Indeed, people who provided legacy materials may not even now be viewed as rightful members of a given group and therefore their information may be deprecated. It is also often unclear what agreements, if any, about such things as publication and distribution existed between the original collector and the community or particular individuals at the time (and whether these agreements were documented) as well as the relationship between any such agreements and arrangements that are currently being negotiated between the contemporary researcher and other stakeholders.

#### 3.2. Identifying rights in the materials

There may be a range of individuals and groups who could hold various types of rights with relation to legacy materials, and identifying them can be problematic. We can distinguish between:

1. *intellectual property rights*, which are legal rights that arise in relation to creations of the mind;
2. *copyright*, which relates to ownership and distribution of products with economic value. Copyright varies for different types of materials (text, sound, images, databases) and only applies to original works fixed in a tangible medium. It is a form of property law and relates to money and economic interest. As such, copyright can be inherited, given away or sold (for further details see Newman 2007);

3. *moral rights*, which concerns how representations affect reputations. As Article 6(1) of the Berne Convention identifies: ‘Independently of the author’s economic rights and even after the transfer of said rights, the author shall have the right to claim authorship of the work and to object to any distortion, mutilation or other modification of, or other derogatory action in relation to the said work, which would be prejudicial to his honour or reputation’.

Determining who holds these for legacy materials can be particularly problematic. Thus, O’Meara and Good (2010) raise the following questions:

1. who holds what rights? Are the rights documented? How do we establish rights retroactively? What if the researcher is not sure about speaker rights?
2. how do we determine rights when there are multiple contributors and data comes from multiple media?
3. who has inherited rights between the time of the original recording and now? (e.g. descendants of the original speakers, descendants of the original researcher)
4. what happens to ‘orphan works’ where the original stakeholders can no longer be identified? (e.g. materials passed from a researcher to a later researcher)

When analysing legacy data it is important to clearly document the various contributions to the work, including those of the original author, the linguist-editor, other researchers, and current community members. The data structures and metadata set up for a project that includes legacy materials should make this clear. An example is Bown (2003) who outlines her setup for the database (in Toolbox format) that she uses for the analysis of Laves’ Bardi data.

### 3.3. Deciding on access to materials

For language documentation projects it is essential that specification is made in relation to access to the corpus, that is spelling out permissions granted to individuals who do not have specific rights to given materials by setting out how they should be allowed to inspect or make copies of a particular resource or a representation of the content of that resource (e.g. a transcription or translation). Typically, researchers can choose between four levels of access to their data and analysis (e.g. when depositing materials in an archive, or distributing/publishing them):

1. *open access* anyone can view, use or copy the materials
2. *restricted access* access to the material is limited to a class of users, usually depending on the contributors, the type or content of materials, or the identity of the user
3. *requested access* the user must ask the researcher for permission
4. *closed access* only the researcher can view, use or copy the materials

Licence agreements, such as Creative Commons,<sup>21</sup> may be used to formalise how the materials may be used once access has been granted, e.g. copy but not change, change with attribution.

As O'Meara and Good (2010) point out, technological changes over time can introduce complications in deciding on access and use for legacy materials. So, for example, 'open access' in 1980 may well have meant users who were granted access would receive an analogue copy of a cassette tape or set of fieldnotes while in 2017 'open access' may mean anyone can download the materials from a website on the internet. It is unclear what 'open access' will mean in the future. In addition, for the categories of restricted or requested access, there is the general issue of how to verify the identity of users who want access (especially if the request comes from an internet address), and if access is restricted to a certain group (e.g. 'community member') how researchers (and archives) can identify such membership. To address these issues sensitive negotiations and extensive discussions with stakeholders, often over an extended time period, may be necessary.

#### 4. Conclusions

Working with legacy text materials from a language documentation perspective involves dealing with a range of often complex issues about the form, content, context and use of the original materials and analyses arising from them. However, there are many opportunities for researchers to add substantial value to legacy text materials, especially if they are able to work with other historical sources and/or contemporary knowledge holders to elucidate them and the context surrounding their creation, analysis and current status. Maximising such opportunities will require thinking about data entities, types and relationships and being explicit about them in the project design and application (e.g. in database design or XML tagging), with a very important role for metadata and meta-documentation. By creating good meta-documentation for written materials now we can hopefully reduce legacy data problems for future researchers, compared to the issues that we face today. Careful work with legacy text materials can also be very rewarding for researchers and communities, especially in the case of unique documents on languages/varieties or areas of knowledge that are no longer available, and that can serve as important sources for language support and revitalisation.

---

<sup>21</sup> See <https://creativecommons.org/>, accessed on 2017-04-09.

## References

- Amery, Rob. 2000. *Warrabarna Kurna! Reclaiming an Australian Language*. London: Taylor and Francis.
- Austin, Peter. 1981. *A Grammar of Diyari, South Australia*. Cambridge: Cambridge University Press.
- Austin, Peter K. 2008. "The Gamilaraay (Kamilaroi) language, Northern New South Wales — A brief history of research". In William McGregor (ed.) *Encountering Aboriginal Languages: Studies in the History of Australian Linguistics*. Canberra: Pacific Linguistics. pp.37–58.
- . 2010. "Current issues in language documentation". *Language Documentation and Description*. 7. London: SOAS. pp.12–33.
- . 2013. "Language documentation and meta-documentation". In Sarah Ogilvie and Mari Jones (eds.) *Keeping Languages Alive: Documentation, Pedagogy and Revitalization*. Cambridge: Cambridge University Press. pp.3–15.
- . 2015. *A Grammar of Diyari, South Australia. 2nd edition*. [Online at [http://www.academia.edu/2491078/A\\_Grammar\\_of\\_Diyari\\_South\\_Australia](http://www.academia.edu/2491078/A_Grammar_of_Diyari_South_Australia), accessed on 2017-04-09.]
- . 2016. "Language documentation 20 years on". In Martin Pütz and Luna Filipović (eds.) *Endangered Languages and Languages in Danger: Issues of Ecology, Policy and Human Rights*. Amsterdam: John Benjamins. pp.147–170.
- Baldwin, Daryl, David J. Costa and Douglas Troy. 2016. "Myaamiaataweenki eekincikoonihkiinki eeyoonki aapisaataweenki: A Miami language digital tool for language reclamation". *Language Documentation and Conservation* 10. pp.394–410.
- Baldwin, Daryl, Leanne Hinton and Gabriela Pérez Báez. 2017. "Chapter 18: The breath of life workshops and institutes". In Leanne Hinton, Leena Huss and Gerald Roche (eds.) *The Routledge Handbook of Language Revitalisation*. London: Routledge.
- Bear, Young and Lucy Thomason. 2011. "de ka ga . ka ka nwi ka de . Skunk and Grizzly Bear". Ms. Smithsonian Libraries. [Online at <https://repository.si.edu/handle/10088/17284>, accessed on 2017-05-29]
- Bishop, Libby. 2006. "A proposal for archiving context for secondary analysis". *Methodological Innovations Online* 1(2). pp.10–20.
- Bodard, Gabriel and Simona Stoyanova. 2016. "Epigraphers and encoders: Strategies for teaching and learning digital epigraphy". In Gabriel Bodard and Mateo Romanello (eds.) *Digital Classics Outside the Echo-Chamber: Teaching, Knowledge Exchange and Public Engagement*. London: Ubiquity Press. pp.51–68.
- Bowern, Claire. 2003. "Laves' Bardi texts". In Joe Blythe and R. McKenna Brown (eds.) *Maintaining the Links: Language, Identity and the Land*. Proceedings of FEL VII. Broome: Foundation for Endangered Languages. pp.137–143.
- Broadbent, Sylvia. 1957. "Rumsen I: Methods of reconstitution". *International Journal of American Linguistics* 23(4). pp.275–280.
- Campbell, Lyle and Bryn Hauk, with Panu Hallamaa. 2015. "Language endangerment and endangered Uralic languages". In Kaisa Leinonen, Sisko Brunn, Santeri Palviainen and Jari Sivonen (eds.) *Congressus Duodecimus Internationalis Fenno-Ugristarum, Oulu 2015 Plenary Papers*. Oulu: University of Oulu. pp.7–33.
- Copeland, Peter. 2008. *Manual of Sound Restoration Techniques*. London: British Library.
- Costa, David J. 2003. *The Miami-Illinois Language*. Lincoln: University of Nebraska Press.
- Costa, David J. (ed.) 2015. *New Voices for Old Words: Algonquian Oral Literatures*. Lincoln: University of Nebraska Press.
- Crowley, Terry and Peter Austin. 2005. "Interpreting old spelling". In Nicholas Thieberger (ed.) *Paper and Talk: A Manual for Reconstituting Materials in Australian Indigenous Languages from Historical Sources*. Canberra: AIATSIS. pp.54–102.
- Dahlstrom, Amy. 2015. "Highlighting rhetorical structure through syntactic analysis: An illustrated Meskwaki text by Alfred Kiyana". In David Costa (ed.) *New Voices for Old Words: Algonquian Oral Literatures*. Lincoln: University of Nebraska Press. pp.118–197.



- Darnell, Regna. 1995. "Documenting disciplinary history". In Sydel Silverman and Nancy J. Parezo (eds.) *Preserving the Anthropological Record. 2nd edition*. Wenner-Gren Foundation for Anthropological Research, Inc. [Online at <http://copar.org/par/index.htm>, accessed on 2017-05-27]
- Fielding, Nigel. 2004. "Getting the most from archived qualitative data: epistemological, practical and professional obstacles". *International Journal of Social Research Methodology* 7(1). pp.97–104.
- Flierl, Johann. 1880. "Dieri Grammatik [Comparative grammar of Diyari and Wangkangurru]". Unpublished ms. Lutheran Archives, Adelaide. Box 22 Immanuel Synod-Bethesda Mission. pp.306–510.
- Giacon, John and Kevin Lowe. 2016. "Key factors in the renewal of Aboriginal languages in New South Wales". In Peter K. Austin, Harold Koch and Jane Simpson (eds.) *Language, Land and Song: Studies in Honour of Luise Hercus*. London: EL Publishing. pp.523–538.
- Gippert, Jost. 2006. "Linguistic documentation and the encoding of textual materials". In Jost Gippert, Nikolaus Himmelmann and Ulrike Mosel (eds.) *Essentials of Language Documentation*. Berlin: Mouton de Gruyter. pp.337–361.
- Goddard, Ives. 1973. "Philological approaches to the study of Native North American languages: Documents and documentation". In Thomas A. Sebeok (ed.) *Linguistics in North America*. (Current Trends in Linguistics 10.) The Hague: Mouton. pp.727–745.
- . 1996. "Writing and reading Mesquakie". *The Papers of the Algonquian Conference / Actes du Congrès des Algonquistes* Volume 27. pp.117–134. [Online at <https://ojs.library.carleton.ca/index.php/ALGQP/article/view/528/430>, accessed on 2017-05-29]
- Good, Jeff. 2010. "Documenting consent, access and rights". Presentation at LSA Annual Meeting OLAC workshop on archiving, Baltimore. [Online at [http://www.ailla.utexas.org/site/lsa\\_archiving10.html](http://www.ailla.utexas.org/site/lsa_archiving10.html), accessed on 2017-04-09.]
- Himmelmann, Nikolaus P. 1998. "Documentary and descriptive linguistics". *Linguistics* 36. pp.161–195.
- . 2002. "Documentary and descriptive linguistics". In Osamu Sakiyama and Fubito Endo (eds.) *Lectures on Endangered Languages, Volume V*. Suita (Osaka): Osaka Gakuin University. pp.37–83.
- . 2006. "Language documentation: What is it and what is it good for?". In Jost Gippert, Nikolaus P. Himmelmann and Ulrike Mosel (eds.) *Essentials of Language Documentation*. (Trends in Linguistics. Studies and Monographs, 178.) Berlin: Mouton de Gruyter. pp.1–30.
- . 2012. "Linguistic data types and the interface between language documentation and description". *Language Documentation and Conservation* 6. pp.187–207.
- Innes, Pamela. 2010. "Ethical problems in archival research: Beyond accessibility". *Language and Communication* 30. pp.198–203.
- Koch, Harold. 2011. "George Augustus Robinson and the documentation of languages of South-Eastern New South Wales". *Language and History* 54(2). pp.140–163.
- Lehmann, Christian. 2001. "Language documentation: A program". In Walter Bisang (ed.) *Aspects of Typology and Universals*. Berlin: Akademie Verlag. pp.83–97.
- Lillehaugen, Brook, George Aaron Danielle, Michel R. Broadwell, Laurie Oudijk, May Plumb Allen and Mike Zarafonetis. 2016. Ticha: a digital text explorer for Colonial Zapotec, first edition. [Online at <http://ticha.haverford.edu/>, accessed on 2017-05-29.]
- Malinowski, Bronislaw. 1989. *A Diary in the Strict Sense of the Term. 2nd Edition with a New Introduction by Raymond Firth*. Stanford: Stanford University Press.
- Moore, N. 2005. "(Re)using qualitative data? Thoughts from the CRESC Qualitative Research Laboratory". Paper presented at Reusing Qualitative Data Workshop, CRESC, Manchester, 28 September, 2005.
- Mosel, Ulrike. 2014. "Putting oral narratives into writing — experiences from a language documentation project in Bougainville, Papua New Guinea". In Bernard Comrie and Lucia Golluscio (eds.) *Language Contact and Documentation / Contacto lingüístico y documentación*. Berlin: De Gruyter Mouton. pp.321–342.
- Nathan, David. 2016. The Diari dictionary of J. G. Reuther. XML tagged digital files.
- Newman, Paul. 2007. "Copyright essentials for linguists". *Language Documentation and Conservation* 1. pp.28–43.
- O'Meara, Carolyn and Jeff Good. 2010. "Ethical issues in legacy language resources". *Language and Communication* 30. pp.162–170. [Online at <https://www.acsu.buffalo.edu/~jcgood/OMearaGood-EthicalIssues.pdf>, accessed on 2017-04-09]

- Scherer, Philipp. 1981. *The Diari by J. G. Reuther*. Canberra: Australian Institute of Aboriginal Studies.
- Scott, John. 1990. *A Matter of Record: Documentary Sources in Social Research*. London: Polity.
- Silverstein, Michael. 2015. "From Baffin Island to Boasian Induction: How anthropology and linguistics got into their interlinear groove". In Regna Darnell, Michelle Hamilton, Robert L. A. Hancock and Joshua Smith (eds.) *Franz Boas as Public Intellectual — Theory, Ethnography, Activism*. Lincoln: University of Nebraska Press. pp.83–127.
- Steele, Jeremy. 2005. "The Aboriginal language of Sydney: A partial reconstruction of the indigenous language of Sydney based on the notebooks of William Dawes of 1790–91, informed by other records of the Sydney and surrounding languages to c.1905". Macquarie University MA thesis.
- Stockigt, Clara. 2016. "Pama-Nyungan morphosyntax: lineages of early description". PhD thesis, Adelaide University.
- Thieberger, Nick. 2016. "Daisy Bates in the digital world". In Peter K. Austin, Harold Koch and Jane Simpson (eds.) *Language, Land and Song: Studies in Honour of Luise Hercus*. London: EL Publishing. pp.102–114.
- Troy, Jakelin. 1994. *The Sydney Language*. Canberra: Australian Institute of Aboriginal and Torres Strait Islander Studies. [Online at [http://www.williamdawes.org/docs/troy\\_sydney\\_language\\_publication.pdf](http://www.williamdawes.org/docs/troy_sydney_language_publication.pdf), accessed on 2017-04-09.]
- Woodbury, Anthony C. 2003. "Defining documentary linguistics". *Language Documentation and Description* 1. London: SOAS. pp.35–51.
- . 2011. "Language documentation". In Peter K. Austin and Julia Sallabank (eds.) *The Cambridge Handbook of Endangered Languages*. Cambridge: Cambridge University Press. pp.159–186.

## Collaborative Project for Documenting Minority Languages in Indonesia and Malaysia

JUKES, Anthony

University of the South Pacific, Fiji

SHIOHARA, 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

---

JUKES, Anthony, SHIOHARA, Asako and Yanti. 2017. "Collaborative project for documenting minority languages in Indonesia and Malaysia". *Asian and African Languages and Linguistics* 11. pp.45–56. [Permanent URL: <http://hdl.handle.net/10108/89206>]

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](http://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,<sup>1</sup> 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.<sup>2</sup>

## 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

<sup>1</sup> Several projects were funded by the Endangered Languages Documentation Programme ([www.eldp.net](http://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).

<sup>2</sup> 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).<sup>3</sup>

There has been significant cross-over between the various workshops mentioned above, the workshops organized by LingDy, and also with DocLing.<sup>4</sup>

## 2. The workshops

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

<sup>3</sup> <http://austronesian.linguistics.anu.edu.au/timor/workshop/>

<sup>4</sup> 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.<sup>5</sup> Figure 3 shows the workflow employed at the workshop.

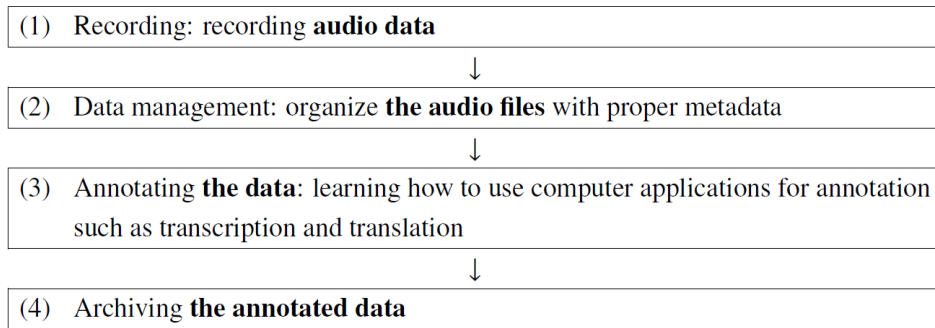


Fig. 3 The workflow employed in the workshop

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

<sup>5</sup> 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.<sup>6</sup>

---

<sup>6</sup> 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.<sup>7</sup>

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

<sup>7</sup> 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.

## The Importance of Documentary Linguistics Workshops: A Personal Account

BADAGAROV, Jargal  
Buryat State University

This paper offers an account of two Documentary Linguistics Workshops held in Tokyo based on the author's personal experience. The workshops have been held for nine consecutive years at the Research Institute for Languages and Cultures of Asia and Africa (ILCAA), Tokyo University of Foreign Studies (TUFS). The advantages and disadvantages of the courses are discussed in detail, and recommendations to students seeking similar programs are given.

**Keywords:** Documentary linguistics, training, fieldwork techniques, workflow, best practices

1. Introduction
2. Why documentary linguistics?
3. Documentary Linguistics Workshop
4. Conclusion

### 1. Introduction

I am a linguist working on my native tongue (Buryad) and related varieties (Mongolic) from both diachronic and synchronic perspectives. Unfortunately, Buryad is an endangered language, despite the fact that it is an official language of the Buryad Republic within the Russian Federation. The Buryad are Siberia's largest native nation and one of the only ones with a long literary tradition and complex social and religious practices prior to Russian colonization. Buryad has many dialects, most of which are severely endangered. For the most part, Buryad dialects lose out on the one hand to Standard Buryad and on the other to Russian, because over the decades, Standard Buryad proved to be inadequate to the contemporary needs of a modern language. Standard Buryad is the product of an early Soviet project that, being underfinanced, was ultimately never able to develop into a full-ranged standard linguistic variety supported by the media and education system. Russian has increased in usage due to abrupt industrialization and a drastic decrease in the proportion of the Buryad living on their native lands on the one hand, and the exclusion of Buryad language instruction

---

BADAGAROV, Jargal. 2017. "The importance of Documentary Linguistics Workshops: A personal account". *Asian and African Languages and Linguistics* 11. pp.57–69. [Permanent URL: <http://hdl.handle.net/10108/89207>]

from education curricula in the 1970s, on the other. The situation is not improving and activists are struggling for the survival of their language. This leads the field of Buryad linguistics to concentrate on issues aimed at documentation of the dialects and development of language-specific study materials. This eventually led me to language documentation, and I am very glad that I was able to experience the Documentary Linguistics Workshop in Tokyo. It allowed me a systematic education in the field of language documentation, which truly changed my life, both as a linguist and a member of an endangered language community.

## 2. Why documentary linguistics?

Documentary linguistics first formulated its principles and methods a couple of decades ago and has since been widely employed in the documentation of endangered languages (Austin 2003; Himmelmann 2006; Thieberger 2012; Austin and Sallabank 2015). It aims at a broad documentation of speech acts in an anthropological setting, the results of which can be used not only by the researcher but also by many others, both academic and layman, and especially by members of the speech community. A seasoned field linguist may ask, “What is the point of having a ‘documentary’ variant of what has long been known as field linguistics?” The key feature of documentary linguistics lies in its comprehensiveness: its theoretical and methodological underpinnings account for linguistic, ethnographic, ethical, technical, and many other aspects of the collection of speech data; ultimately, it is targeted at creating and preserving a full-scale cross-cut of a linguistic situation for current and future research. This becomes especially important in the context of the dramatically decreasing number of languages and dialects across the globe, a modern linguistic disaster.

## 3. Documentary Linguistics Workshop

As with others who have had some experience in linguistic fieldwork, I was not always sure if I was going about things in the right way. Thus, I wanted to learn proper methodology. Consequently, I read extensively on technical procedures and equipment. Additionally, I considered the types of questions a linguist should ask in the field, ways of communicating with informants, and effective methods of collecting and preserving data. However, despite the time and effort, I still lacked confidence. Therefore, I realized that to gain confidence I would have to learn the best practices of language documentation from experts. Fortunately, I found this in the Documentary Linguistics Workshop at the Research Institute for Languages and Cultures of Asia and Africa (ILCAA), Tokyo University of Foreign Studies (TUFS), which became a major turning point in my efforts.

The workshop had been running as a part of the LingDy project at ILCAA for



nine consecutive years. I appreciate this opportunity to discuss my impressions of the workshop and will be happy if this helps lead others into the field of language documentation and linguistics. First, I am going to give a general overview of the workshop, its aims, and organization, and then I will briefly concentrate on particular courses, drawing, where applicable, on my impressions and thoughts.

The goal of the workshop was to provide basic training in language documentation for field linguists, including methodological and technical training in various aspects of language documentation and archiving of endangered/minority languages. Topics covered included the following: introduction to language documentation, language archiving, hands-on training in audio and video recording and photography, data management and metadata, discussion of issues in field linguistic research, elicitation, and many other aspects of documentary linguistics. The teaching staff comprised experts who had worked in the field for many years, such as Peter Austin (School of Oriental and African Studies (SOAS), University of London), David Nathan (SOAS, University of London), Anthony Jukes (Centre for Research on Language Diversity (CRLD), La Trobe University), Sonja Riesberg (University of Cologne), and Hideo Sawada (ILCAA, Tokyo University of Foreign Studies). Project groups were led by native speaker consultants: Tshering Tashi (Royal Society for Protection of Nature, Board Director, Dzongkha native speaker (Bhutan)), Namgay Thinley (Dzongkha Development Commission, Senior Research Officer, Dzongkha native speaker (Bhutan)), Jargal Badagarov (Buryat State University, Lecturer, Buryat native speaker (Buryatia, Russian Federation)), Kristian Walianggen (Center of Endangered Language Documentation, State University of Papua, Language Consultant, Yali native speaker (Papua, Indonesia)).

The workshop achieved a unique balance between theory and practice through well-thought-out organization of theoretical classes with hands-on practice courses and group projects on the one hand, and through highly competitive selection of participants and language consultants on the other. This combination, which I believe was true for each of the nine workshops, led to the successful achievement of the workshop's mission and shared benefits for all participants, including teachers, language consultants, and organizers. The participants were mostly Japan-based MA and PhD students with a handful of international participants. The language of instruction was English. A certificate of attendance was given to students and language consultants.

In the following section, I will reflect on the benefits of the workshop as a whole and certain courses in particular.

### 3.1. Overall experience

A frequently asked question is "What skills am I going to gain from a workshop?" Well, firstly, participation in this workshop has helped fine-tune my knowledge and skills of audio and video recording and photography in language documentation, in

terms of both equipment and technique. I also learned about the ethical aspects of language documentation, regarding which systematic thinking is lacking in the Russian Federation and Mongolia. Data management and preservation was another aspect of this workshop, the importance of which cannot be underestimated. All of this material came together as a proven set of knowledge and skills during the implementation of a group project. I learned a lot about teamwork during the group project. The significance of teamwork is often underestimated in language documentation projects, because usually one person carries out both the fieldwork and the “postproduction.” Therefore, team projects were one of the cornerstones of this workshop, offering significant advantages over similar workshops I have since attended.

Theoretical and practical courses offered during the workshop were Introduction to Language Documentation; Planning Language Documentation and Group Projects; Ethics and Working with Communities; Audio & Video; Fieldwork Techniques & Elicitation; Software for Documentation—Survey & Practical Demonstration; Data Management; Multimedia & Mobilization; and Photography for Documentation. The importance of data management and metadata was very much stressed. Theory was discussed most intensively on the first day and again towards the second half of the workshop. Practical courses were demonstration oriented and had a carefully designed and effective structure.

About half the overall workshop time was devoted to group projects with language consultants, during which the following activities were conducted: preparing individual laptops/software, set-up of groups and projects, group project design and reports, and recording consultants.

There was also a public lecture called “Rethinking Language Documentation and Support for the 21st Century,” which enabled all the participants to take part in a discussion of the different aspects of language documentation.

### 3.2. Introductory class

The introductory class provided a stimulating start to the workshop thanks to Peter Austin’s clear-cut presentation style and well-structured slides discussing key concepts of language documentation theory. It walked the students through major aspects of language documentation, offering a basic understanding of the workflow and major challenges in language documentation projects, including corpus design, interlinear glossing, archiving, managing data and metadata, meta-documentation goals and methods, and sustainability issues in language documentation. The importance of metadata was constantly stressed throughout the class, sending a clear message to future experts in the field. The lecturer pointed out that “we need a theory of metadata,” that is, a theory that is applicable to the documentation of the process of language documentation. Sustainability, as covered in the class, was another under-theorized topic in language documentation. Through this, and the following lectures, the participants learned about the role of language consultant as compared to

that of an informant. A language consultant is an individual who actively participates in language documentation, often becoming the (co-)leader of the team. This role contrasts markedly with that of an informant, traditionally viewed as a mere conduit of information and someone assessed only on the “quality” of information provided, a standard set by the researcher alone. This major conceptual shift from a passive performer of communicative events to an active member of a language documentation project was very inspiring. In general, the presentation set out a very balanced understanding of the subject and its major problems and considered possible methods and directions to overcome them.

### 3.3. Research and group work planning

The Planning Language Documentation and Group Projects class by Anthony Jukes focused on the development of an actual language documentation project. During this process, he helped students understand topics such as what kind of projects granting agencies find most attractive and the major considerations behind the planning of a language documentation project, including project goals, timelines, audiences, outcomes and the researcher’s own skills and ways of combining them with the skills of other project members. The participants were advised to look for a balance between skills, time, and available resources, while at the same time understanding the fact that all documentation projects are different. The class also considered the possible goals of a language documentation project. A comparison of historical projects with more recent ones demonstrated that earlier projects were aimed at producing “a lasting, multipurpose record of a language” (Himmelman 2006: 1), whereas newer ones focused more on, for example, traditional ethnobotanical knowledge or traditional agricultural songs and stories. The students came to understand the importance of setting explicit timelines and that in any language documentation project, one will need to consider the audience and the possible outcomes of the project, such as grammars, linguistic/typological data, dictionaries, storybooks, schoolbooks, scholarly papers, websites, and accessible audio/video. Teamwork was another focus of the class: one should carefully reflect on one’s own skills and ways of combining these with others. In this class, I realized that participation in a language documentation project should not be restricted to linguists only. Language documentation can be all the more productive if members of the local community are involved not only for their language skills as informants or language consultants but also for their technical skills in assisting in the selection and use of equipment. Moreover, after the data collection and processing/annotation, assistance with technical equipment and archive curators will be necessary, and the most appropriate source of workforce is again from members of the community. This lecture was an essential part of the workshop and clarified very important aspects of language documentation.

### 3.4. Ethics and community interaction

Anna Berge (Alaska Native Language Center, University of Alaska Fairbanks), a guest presenter, continued the workshop with a presentation on ethics and community interaction in language documentation. Together with Peter Austin, who presented on Communities, Ethics and Rights, they covered ethical and legal concerns that might arise in the course of interaction between academia and a community during language documentation. The lecture successfully demonstrated the utmost relevance of ethical and legal considerations for all stages of language documentation. Language documentation was shown to be a partnership between different parties: communities, linguists, pedagogues, archivists/librarians, organizations, and funding bodies. A meaningful discussion of the community values and needs in the context of cultural differences and a community view of linguistic research was an essential part of the presentation and helped me gain a deeper understanding of community interests and the place of linguists in the process of language documentation. The concept of “giving back” on behalf of academia was new to me because of the differences in field research practices in the Russian Federation, rooted in Soviet tradition. At that time, the local community experienced linguists’ activities in the form of efforts to create orthographies and enacting positive changes in government policy in order to facilitate a better representation of the local language in society. This was especially true during the early Soviet period. Later on, there was only a shiny facade of “everlasting friendship” between the peoples of the Soviet Union, which more or less successfully concealed the gloomy reality of boarding schools, which prohibited children from speaking their mother tongue. This led to entire communities being deprived of their traditional ways of life and cultural practices. Naturally, this brought about the rapid degradation of languages and extinction of whole communities. In my view, the major weakness of the concept of “giving back” is that linguists, excluding a few, cannot devote their entire lives to the community they are studying. From the point of view of the community, a short-term effort is not an effort at all. Moreover, many community members would regard the whole enterprise to be a waste of time, money, and effort. However, this should not make us give up. We should remain steadfast in the belief that the preservation of a language and the linguistic and cultural practices of any endangered community is of the utmost benefit to humanity.

In his presentation, Peter Austin put forward a wider perspective of the concept of “giving back” as “establishing and maintaining relationships which (ideally) goes far beyond our departure from the field site.” Again, a specific realization of this concept may vary from project to project, enabling discussions and decisions to be made with the participation of the community. We can ask community members, especially those with whom we have worked extensively and/or can establish good relations, what the needs of their community are. Another important idea articulated in this lecture was the central role and conception of reciprocity between the researcher and community.

In other words, the researcher should negotiate the way a project is conducted, what the possible outcomes should be and in what way they will be interesting for the community. As Peter Austin stated in his presentation: “we need to question basic linguistic/academic/literacy assumptions,” with which I cannot agree more. Despite this having been recognized several years ago, it remains valid today. To conclude, I would like to quote the major principles that we should consider when working in the field (as formulated by Anna Berge): be honest, creative, flexible, and open; discuss issues; explain your work; give results back to the community; communicate, negotiate; and do not isolate yourself.

### 3.5. Audio recording

Next was Audio Theory and Practice for language documentation by David Nathan. He started by asking questions such as whether participants had ever recorded, published, or processed audio or if we thought digital or analogue was better for recording. This helped all participants to get off on the same foot, whether they were experienced or newbies. The lecturer formulated “big questions” to be kept in mind when recording (what, for whom, and why are we recording?) and explicitly formulated criteria for evaluating recordings, which proved to be instrumental in developing personal audio recording skills.

We learned a lot of technical information, which was presented in a clear and logical way with a less technically minded person in mind. Apart from the technical characteristics of devices and their settings, it is important to be able to adapt to a recording environment. For instance, trying to avoid hard, smooth, flat parallel surfaces, or facing away from noise sources, choosing a space away from doors, windows, traffic areas, and anything else noisy. The audio equipment selection depends on the goals of a specific documentation project. The lecturer gave a detailed description of microphone types and their advantages and disadvantages in different recording situations, although in the end there are only a few optimal choices. The best thing is to go directly to an expert. There was also detailed information provided about the correct way to use a microphone such as placement, distance, handling, and use in windy situations (e.g., using a deadcat).

The theoretical portion of the audio session concluded with an exhaustive description of audio workflow. The practical portion was truly impressive as the instructors demonstrated the use of some equipment such as a Superlux S502 full binaural microphone and a Rode NTG1 condenser shotgun microphone in the classroom.

### 3.6. Video and language documentation

Video and Language Documentation by Anthony Jukes concentrated on the advantages and disadvantages of video for language documentation. One of the major drawbacks of video footage as stated by the lecturer is that it “may seem ‘true’, but is actually less ‘authentic’ than audio” as it frames an environment rather than capturing

it.

When shooting video he recommends considering what is to be filmed, who is going to watch it, where/how they will watch it, how you will manage audio, how you are going to edit it, how much you are prepared to spend, and how much you are prepared to carry. Some basic technical information concerning lenses and sensors was explained in a succinct manner with examples of effects these two parameters produce on the final video. As is often the case, the key is to find a proper balance between price, specs, and picture when choosing a camera. Some specific models were discussed, from professional video cameras to camcorders; photo cameras from DSLR to cheap point-and-shoot cameras were also discussed, as they are becoming increasingly popular for shooting video. Aside from the best camera one can afford, one should always remember that “The best camera is the one you have with you.”

In addition to the technical side of video recording, the presentation addressed the basic principles required for quality video recordings, thoroughly discussing such principle components of an optimal video shot as framing, lighting, and camera movement. A quality camera and decent command of the basic principles of composition require appropriate stabilization equipment such as tripods, monopods, and camera rigs. When shooting a video, one should always take care to record the audio separately using suitable equipment because built-in audio recorders can be used only for syncing purposes and do not result in high-quality audio.

### 3.7. Fieldwork techniques and elicitation

Fieldwork Techniques and Elicitation by Sonja Riesberg started by providing some theoretical foundation to language documentation and reflection on its meaning. According to Himmelmann (2006: 7–10), the object of study is “linguistic practice and tradition” which are manifested in “1) the observable linguistic behavior and 2) the native speakers’ metalinguistic knowledge.”

While making recordings of linguistic practice and tradition seems to be quite natural and, technically speaking, straightforward, some issues are to be considered in order to make a comprehensive record of linguistic practice and tradition and provide a necessary quality of data. In reality, a record of natural communicative events is not quite possible, and therefore, in language documentation, a linguist has to manage observed or staged communicative events, or elicitation. Possible scenarios of all types of communicative events, both observed and staged, necessary to meet the requirements of a comprehensive record, were discussed in considerable detail with carefully selected examples and actual footage from real projects. Elicitation and use of stimuli to obtain explicit grammatical information, fill gaps in a paradigm, encourage narrative production, and elicit dialogues and speaker’s categorization of space and time all received a meticulous account. The material was presented in a systematic manner with the examples provided from the lecturer’s own experience. In general, the presentation managed to cover every important aspect of the topic, providing just the right amount

of detail.

### 3.8. Software tools

The lecture Software for Documentation (Peter Austin and David Nathan) and the following practical class focused on the processing of collected (recorded) information, the most time-consuming stage of documentation. This includes transcriptions, aligned annotations, interlinear glossing, vocabulary, and dictionaries. This is generally facilitated by specialized software tools covered in the lecture. Specific programs covered in this lecture include ELAN and Transcriber for preparing annotations, Toolbox and FLEEx for data management, parsing and interlinearization of the annotations from the previous stage, WeSay for collecting and organizing lexical data for dictionaries, and Arbil and SayMore for metadata management. The advantages and disadvantages of each program were discussed at length, and the comparison was summed up in a convenient table. Some possible ways to distribute the results of documentation, especially among community members, were also discussed.

The practical part of the class was devoted to working with Toolbox/FLEEx (Peter Austin) and ELAN (Anthony Jukes). In the beginning, participants were given an Ainu text and followed instructions to setup a new project in Toolbox/FLEEx with text and lexicon databases. We were told to pay attention to the naming of a project and its folders, numbering the entries of the lexicon database, adding extra fields to the text and lexicon databases to reflect the names of those who entered the data. Because of time limitations, it was not possible to get any profound knowledge or skill, but still it was very useful as a gentle introduction to the tool.

The latter part of the practical class described ELAN and its use for transcription and annotation. Students learned what this tool can and cannot do and gained knowledge of basic terminology behind the software and its workflow. Seven exercises were suggested to the students covering major stages of a language documentation project setup.

I believe that the knowledge and skills gained in the class were crucial for further mastering the software: I knew what courses and summer schools to look for in order to further my understanding of those tools, especially ELAN and metadata management tools. Thanks to the group projects, we were able to immediately apply the skills we had gained in this class.

### 3.9. Data management

We need to carefully consider strategies for storing the rapidly increasing amount of data we are able to store on our computers to facilitate quick identification and access. This was the major point of data management class by David Nathan. The presentation effectively demonstrated why file naming and folder structure strategies are vital, what a digital object is and its identity, how to create a collection of materials that can be used in the absence of the creator, and how to document decisions on file name and

folder structure conventions creating an additional layer of metadata.

Three folder structure models were presented along with a discussion of their pros and cons: 1) a tree of descriptive folder and file names; 2) one folder with descriptive filenames; and 3) one folder with numerical filenames. These recommendations were very useful since I had had experience creating what turned out to be badly managed data sets.

In the final part of the presentation, the lecturer discussed the issue of encoding, presenting some possible problems and ways to avoid them. After the class, the researcher should be able to design a well-organized system of folders. These skills were immediately put into use during the group projects.

### 3.10. Photography

Photography in language documentation by Hideo Sawada concentrated on the roles of photography in language documentation and its technical aspects. The technical aspect focused on camera types and lenses. Participants learned about possible problems that make images unusable, including lighting, blurring, defocusing, and obscurity. There were also some tips on how to obtain the best picture possible under different circumstances.

During the lecture, the importance of tripods was highlighted again (see the paragraph on Audio and video class). The lecturer gave very helpful recommendations on choosing photography equipment, including tripods, and provided amusing demonstrations clearly illustrating possible problems in the field and ways to overcome them. Camera types, such as DSLR, mirrorless, high-end and point-and-shoot compact cameras, mobile and smartphone built-in cameras, and PC built-in cameras, received a detailed account along with a discussion of their advantages and disadvantages in a fieldwork context. Students learned to take into account lighting conditions to adapt equipment to different working situations. The role of close-ups in language documentation was examined, and useful tips for good quality close-up photos were given.

The students learned how to manage photos in a language documentation, about naming of files and software for bulk renaming, such as Flexible Renamer for MS Windows or Rename for Macs. Personally, I prefer using Total Commander on Windows and Terminal on Mac. Photo management software to enable keyword searches was discussed by the lecturer, including Picasa, StudioLine Photo Basic, and Adobe Lightroom. It should be added that newer versions of Windows and OS X enable users to organize files with tags. There is a study by Civan et al. (2008) showing that combining folders and tags (labels) is better than using only one of these methods for organizing files.



### 3.11. Digital archiving

The Endangered Language Archive (ELAR) and Digital Archiving for Documentation of Endangered Languages by David Nathan gave a very important perspective to the whole workshop, clarifying some intricate “secrets” of language documentation for the inexperienced. It began with a discussion of the definition of a digital archive, its functions, and the peculiar features of language archives rooted in the diverse and complex nature of language itself. Among the advantages of language archives mentioned were security, preservation, discovery, protocols, sharing, acknowledgement, mobilization, quality, and standards. Their drawbacks partly originate from the disadvantages of digital data: it is fragile and ephemeral.

Listeners were introduced to the architecture of language archives using the examples of the Open Archival Information System (OAIS) and ELAR models, and the need for redefining the digital ELAR was postulated. After the introductory section of the presentation, the lecturer discussed at length various aspects of endangered language archiving using ELAR and its inventory as an example. The lecture concluded with practical information on formats and standards. All of this requires a variety of computer skills. Therefore, one may benefit from acquiring such skills or collaborating with someone who has them.

### 3.12. Group projects

The group projects proved to be a critical element of the Documentary Linguistics Workshop. They helped the participants to consolidate their knowledge and strengthen their skills by working together on a project of their choice. The group in which I worked consisted of three highly motivated individuals I would like to mention by name—Sami Honkasalo, Robert Laub, and Kunio Kinjo—who concentrated on something that could be part of an actual language documentation project. We recorded a Buryad fairy tale and transcribed it in ELAN, collaborated on a translation in Google Docs, and then published it on YouTube with subtitles in Buryad and English.

Similar to most of the teams, we had at first experienced difficulties deciding on a topic. We had tried several other options, for example, discussing tense forms in Buryad, before finally deciding on the fairy tale. This proved to be an appropriate decision as the allocated time of four days was insufficient for anything too ambitious. We found Google Docs to be very useful in the course of our project, although it is not always usable in the field. We learned a lot from interacting within the group, as well as with other participants and teaching staff.

The culminating point of the group projects and the workshop itself was the presentation and discussion of the group projects. The groups presented successful projects and received a lot of feedback from all the participants. I believe that the opinions and advice we received at this point from our experienced instructors was extremely important.

### 3.13. Workshop atmosphere

The whole workshop was about free communication of ideas, knowledge, and skills. This concluded in an individual clinic and consultations, during which a student could ask for the advice of any lecturer or any language consultant. Some used this opportunity to set up a project in Toolbox or FLE<sub>x</sub>, while others chose to discuss their MA or PhD projects, or project proposals, or receive firsthand advice on equipment and its use. The range of possible topics was limitless, and I believe everyone was able to effectively use this phase of the workshop.

## 4. Conclusion

In conclusion, I would like to emphasize the role such workshops play. They have a significant influence on future careers since they provide excellent opportunities to interact with experts and senior specialists in the field and share in their firsthand experience. The Documentary Linguistics Workshop at ILCAA, TUFS (Tokyo) succeeded in this and granted its participants self-confidence and inspiration in their language documentation efforts.

There was a good balance between theory and practice, group projects, individual consultations, expert teaching staff, carefully selected language consultants, and students. Similar to any other event of this type, organizers tried hard to counterbalance contents and activities with very limited time. However, it is difficult to meet the expectations of every participant. Among minor imperfections was the limited coverage of documentation software, though I do realize that within the six-day timeframe, this was not possible. Therefore, students who seek further training in software tools for language documentation should look for more specialized workshops. However, it was possible to consult individually with instructors about the software.

In sum, I should say that this workshop proved to be a very solid introduction to language documentation. It provided a strong foundation to further training in the field and for starting individual projects. I believe that this type of workshop is very important for researchers and endangered language communities around the globe. Fortunately, the new iteration of the LignDy project implements a series of outreach activities as a continuation of the Documentary Linguistics Workshop in different parts of the world, including Indonesia, Mongolia, Russia, and China. This facilitates the dissemination of knowledge and skills and sets up long-term collaboration amongst academics and communities all over the world.

For anyone who participates in similar workshops, I would recommend that while enjoying the welcoming and relaxed atmosphere of such a workshop, participants take notes and read the recommended literature, or at least some of the papers one notes as important. I also recommend that participants download the workshop materials and the webpage to be able to use it offline.

Lastly, I would like to express my sincerest gratitude to the LingDy team headed by Toshihide Nakayama. My participation in the workshop would not have been possible without the continued effort and careful arrangements of Yoshida-san (Sachiko Yoshida) and without the ongoing support and friendly advice of Professor Tokusu Kurebito, to whom I would also like to express my appreciation.

## References

- Austin, Peter K. 2003. "Introduction (LDD 1)". In Peter K. Austin (ed.) *Language Documentation and Description* 1. London: SOAS. pp.6–14.
- Austin, Peter K. and Julia Sallabank, (eds.) 2011. *The Cambridge Handbook of Endangered Languages*. Cambridge: Cambridge University Press.
- Civan, Andrea, William Jones, Predrag Klasnja and Harry Bruce. 2008. "Better to organize personal information by folders or by tags?: The devil is in the details". *Proceedings of the American Society for Information Science and Technology* 45 (1). pp.1–13.
- Himmelmann, Nikolaus. 2006. "Language documentation: What is it and what is it good for?". In Jost Gippert, Nikolaus P. Himmelmann and Ulrike Mosel (eds.) *Essentials of Language Documentation*. Berlin: Mouton de Gruyter. pp.1–30.
- Thieberger, Nicholas. 2012. *The Oxford Handbook of Linguistic Fieldwork*. Oxford: Oxford University Press.



## Documentary Linguistics Workshop (DocLing) 2016: A Reflection

Yanti

Atma Jaya Catholic University of Indonesia

The Research Institute for Languages and Cultures of Asia and Africa (ILCAA) at Tokyo University of Foreign Studies (TUFS), in collaboration with Endangered Language (EL) Training, conducted a Documentary Linguistics (DocLing) Workshop from 2008 through 2016. I had the opportunity to attend DocLing 2016, which was hosted at the institute from February 8 through 13, 2016. This paper shares my experience from DocLing 2016 and presents my reflection on the program.

**Keywords:** language documentation, collaboration, reflection, workshop

1. Introduction
2. Overview of DocLing 2016 workshop
3. Zooming in sessions in DocLing 2016
4. Personal reflection
5. Epilogue

### 1. Introduction

Language documentation is a subfield of linguistics which has grown very rapidly in this decade. Language documentation emerged as endangerment of the world's languages became a serious concern. In recent years, language documentation has become not only an interest of those who are concerned about language endangerment, but also of those who are working in all areas of linguistics. Linguists in general started to be aware that linguistic theories cannot be generalized based on a few well-known languages, such as English, German, French, etc. or on limited data based entirely on elicitation. Primary data collected through carefully planned and well-conducted documentation of various languages would also be useful to serve as the empirical foundation for the development of linguistic theories. For instance, this is the position taken by Himmelmann (2006: 1), who defines language documentation as “a field of linguistic inquiry and practice in its own right which is primarily concerned with the compilation and preservation of linguistic primary data and interfaces between primary

---

Yanti. 2017. “Documentary Linguistics Workshop (DocLing) 2016: A reflection”. *Asian and African Languages and Linguistics* 11. pp.71–78. [Permanent URL: <http://hdl.handle.net/10108/89208>]

data and various of types of analyses based on these data.”

In 2013, a documentation workshop was hosted collaboratively by the Research Institute for Languages and Cultures of Asia and Africa (ILCAA) and Udayana University in Bali. I participated in the workshop to share with others the practical experience I had gotten from my own work in language documentation (Yanti 2010; Yanti, Tadmor, Cole and Hermon 2015). I came to realize that providing language documentation training to members of language communities was an effective way to empower them in documenting their own languages and to ensure the sustainability of the activity. Since then, I have been actively collaborating with the ILCAA group to give documentary workshops hosted in various places in Indonesia, such as Bali, Jambi, Samarinda, and Kupang (see Jukes, Shiohara and Yanti, this volume).

After lecturing in several workshops, I realized that what I shared in those events was mainly based on my field experience in language documentation, and I had never been to a structured and professionally organized language documentation workshop. Therefore, in one of my email exchanges with Asako Shiohara from ILCAA, I expressed my interest in participating in the documentary workshop that ILCAA was hosting in February 2016. Asako Shiohara then invited me to participate in the workshop.

This paper aims to share my experience during DocLing 2016 and to present my reflection on the program. Section 2 briefly presents an overview of DocLing 2016. Section 3 zooms in on the sessions in the workshop. Section 4 shares my personal reflection on the program and Section 5 is the epilogue of the paper.

## 2. Overview of DocLing 2016 workshop

DocLing 2016 was hosted on February 8–13, 2016 at ILCAA, Tokyo University of Foreign Studies. The workshop aimed “to provide methodological and technical training in various aspects of language documentation research, including audio and video recording, data analysis, metadata, data management, data mobilization, archiving and research ethics” (<http://lingdy.aacore.jp/en/activity/docling.html>).

As a series organized collaboratively by ILCAA and Endangered Languages (EL) Training group, DocLing 2016 presented four of the main speakers from EL Training: Peter K. Austin, David Nathan, Anthony Jukes, and Sonja Riesberg; as well as three ILCAA scholars who have ample experience in documenting minority languages: John Bowden, Toshihide Nakayama, and Hideo Sawada.

The program covered a wide range of topics which are essential for those who are interested in working on language documentation projects. These topics were presented in the forms of lectures, discussion forums, and group projects. The early sessions of the workshop mainly consisted of lectures, and covered topics like “an introduction to language documentation”, “methods, materials and genres”, “mobilization: audience research and design”, “ethics”, “data management and archiving”, “audio”, “software”,

as well as “still and moving images”. Discussion forums allowed the lecturers and participants to exchange their thoughts and experience about “roles, languages, and communities”, as well as “documentary theory and methods”. Some last sessions were devoted to group projects.

To provide the participants with a real sense of language documentation fieldwork, three of the participants whose native language was not known by the other participants were invited to serve as language consultants. The consultants speak different languages spoken in Indonesia. The first language consultant was Dominikus Tauk, whose native language is Helong, a language spoken in Kupang, East Nusa Tenggara. The second one was Hesti Widyawati Wieringa, who speaks Javanese, a language which is spoken by more than 80 million people (Simons and Fennig 2017). Finally, I was the third language consultant (Yanti, a native speaker of Jambi Malay spoken in Jambi City, Sumatra). Each of these speakers worked with several other participants in a group on projects based on the speakers’ respective languages.

### 3. Zooming in sessions in DocLing 2016

The six-day DocLing 2016 was a resource for those who were interested in documenting languages. The workshop presented both theoretical and practical aspects of language documentation, from planning to dissemination.

Some sessions were devoted to teaching the participants basic knowledge of language documentation. These sessions mainly consisted of lectures and covered topics such as what language documentation is, what it is concerned with, and why language documentation is important. In addition, topics like methods of collecting data, what materials to collect, what ethical issues one may encounter when doing language documentation, and what should be considered before starting a language documentation project were also discussed.

Some other sessions were designed to teach more practical and technical issues in language documentation, such as equipment for making audio and video recordings, how to make good recordings, software for data annotation, data management, and archiving. In addition to lecturing, the instructors showed the participants various types of equipment, such as audio and video recorders, microphones, tripods, etc. The use of software for data annotation and for making documentation products (e.g. dictionaries), however, was not practiced much unless the participants needed help for their group project.

Finally, some last sessions were allocated for group projects and involved participants’ concrete and active participation in conducting a so-called language documentation project. The participants were divided into three groups and each group worked with one language consultant.

In what follows, I will highlight the nature of the group project which the participants needed to work on. As the language consultant in a group working on Jambi Malay,

my native language, I worked with four other participants. They were Akiko Tokunaga, Linjing Li, Yuko Morokuma, and Norifumi Kurosima. Our group decided to make a video containing a story told in Jambi Malay, my first language. The video would have subtitles in three languages: Jambi Malay, English, and Japanese.

To work on the project, we could use any resources available at the ILCAA, including the audio recorder, microphone, and the well-equipped recording studio. The use of the studio was really a privilege for us. The lecturers were available to help if any group needed assistance or to discuss problems. Our group members worked together and applied what we learned from the workshop in conducting our project. What was coincidentally unique about our group was that each member seemed to have his/her own pre-assigned roles. I told the story in Jambi Malay, edited the audio recording using Audacity software, and provided the Jambi Malay and English texts for subtitles.<sup>1</sup> Norifumi Kurosima was responsible for recording me telling the story in the studio and translated the English text into Japanese. Yuko Morokuma is very talented in drawing, and, thus, she was the one who prepared the illustrations in the video using Firealpaca software.<sup>2</sup> Linjing Li was in charge of putting together and adjusting all the illustrations and the audio recording to produce the video, as well as taking pictures for documentation. Last but not least, Akiko Tokunaga inserted the subtitles in three languages on the video. She was also the one who uploaded the video to YouTube.<sup>3</sup>

It was not only our group which produced an outcome from the training. The group working with Dominikus Tauk, produced a website containing information about Helong.<sup>4</sup> In addition, the group working with Hesti Widyawati Wieringa, made a power point presentation about *Bebek Ayu* (Lit. 'pretty duck').<sup>5</sup> We were all very pleased with what we produced in such limited time.

#### 4. Personal reflection

The DocLing 2016 workshop provided the participants with a plentitude of materials and information about language documentation. The workshop combined lectures, discussions, and group work, and these activities made the workshop interesting and lively. Although I had some prior experience in doing language documentation, I found that my experience as a participant and a language consultant at the same time was still invaluable.

My personal reflection on my participation in the workshop resulted in my reaching two main realizations. The first realization is about language documentation in general and the second one is about the documentary workshop, especially in the context of

<sup>1</sup> The Audacity software can be downloaded from <http://www.audacityteam.org>

<sup>2</sup> The Firealpaca software can be downloaded from <http://firealpaca.com/en>

<sup>3</sup> Online at <https://www.youtube.com/watch?v=ByOjvNI-LmY>, accessed on 2016-12-19.

<sup>4</sup> Online at <http://www.el-training.org/outcomes/docling2016/helong/>, accessed on 2016-12-19.

<sup>5</sup> Online at [http://www.el-training.org/docs/docling2016/2016-02-13\\_dogling2016.Presentation%20Slides\\_Group\\_3.pdf](http://www.el-training.org/docs/docling2016/2016-02-13_dogling2016.Presentation%20Slides_Group_3.pdf), accessed on 2016-12-19.



Indonesia.

With respect to language documentation in general, I would like to make four main points.<sup>6</sup>

a. Planning is important

Planning is the earliest and crucial stage in language documentation. The planning stage in language documentation includes, among other things, setting the goals of the project, deciding on who will collaborate on the project, what funding resources and other resources are available, setting the time frame for the project, and deciding what kind of data are going to be collected, etc.

Before we decided to make a video as the goal of our project, we considered the resources and the time we had and what kind of project might seem useful and realistic to carry out. Then, we decided that we would make a video and started to discuss issues, such as what the contents of the video should be, who would be the target audience, how would the video be accessed by the audience later, what contributions each member of group could provide, the process that our group would go through in making the video, etc.

The planning we carried out was really important as it helped us focus on our goal and it served as guidance about what to do to reach the goal, and how to find a solution to any problems we might encounter. When we started to make the audio recording, for example, we were not sure which microphone to use and how far it should be located from the speaker. We tried two microphones and decided to use the stereo one. We also had to adjust the distance of the microphones several times until we got the best sound quality in the recording. We realized that we needed a good recording because we were making a video for the public.

b. Language community involvement

Speakers of a language know best about their language and people are always excited talking about their language because they own it. Including the active participation of members of a language community in a documentation project can be valuable for documenting the language itself. Native speakers of a language usually know of interesting materials to collect and can contribute in collecting data as well as transcribing the data.

In addition, inviting members of a language community to participate in a documentation project can also raise the awareness of the community about language endangerment. Dalby (2002) points out that “a language dies every two weeks.” Researchers are not in a position to force members of a language community to use and pass on their own language to the next generation.

---

<sup>6</sup> These are not novel points with respect to language documentation. A number of publications about language documentation have pointed these out, for example, Himmelman, Gippert and Mosel 2006, Austin 2010.

Members of the community should decide what they think they should do with their language. If members of a language community are aware of how important their language is and are part of a documentation project, they are likely to be agents for protecting their language from endangerment.

c. Archiving

Archiving raw data and unpublished fieldwork notes has been a common practice for field linguists. In the Annual Business Meeting of the Linguistic Society of America in 2010, members of LSA concluded the following resolution

“the Linguistic Society of America supports the recognition of these materials as scholarly contributions to be given weight in the awarding of advanced degrees and in decisions on hiring, tenure, and promotion of faculty. It supports the development of appropriate means of review of such works so that their functionality, import, and scope can be assessed relative to other language resources and to more traditional publications”<sup>7</sup>

By archiving recordings and texts, as permitted by the language community, a field researcher has shown respect to the language community, especially those who participate in the work. As pointed out by Dwyer (2006: 40), “...; disseminating or at least properly archiving collected data is far more respectful to a speaker community than piling it in the back of a closed.”

Archiving raw data and unpublished notes may also allow the materials to be further used by the future researchers as well as researchers from other fields of study.

d. Dissemination

The final stage in a language documentation project is dissemination. Disseminating the products of a documentation project is very important. Otherwise, the entire documentation project will not be useful. Products from language documentation can be in a form of dictionaries, grammar books, story books, videos, etc and they need to be published.

In the context of Indonesia, it is important that products from language documentation of the local languages, such as storybooks, picture dictionaries, videos and other potential materials for school local contents be made available to the members of the language community. The native speakers of these languages are excited when they find their language in such products. When I went back to the community with the storybook we published from

---

<sup>7</sup> <http://www.linguisticsociety.org/resource/resolution-recognizing-scholarly-merit-language-documentation>

documentation of Jambi Malay (Yanti, Tadmor, Cole and Hermon 2015), for example, I could see how excited and proud the people were when they saw their language appeared in print. We hope that this kind of efforts can help motivate the language community to maintain their language and pass their language to the next generation.

Regarding the documentary workshop, I suggest that workshops similar to DocLing be hosted in more places, especially in places where minority languages are spoken. Hosting documentary workshops in these places is a concrete effort to support the idea of involving the members of language community in documenting their own language and culture. Furthermore, collaborating with members of community can be beneficial to build language corpora for various types of research.

Learning from DocLing 2016, planning and holding future documentary workshops, especially in the context of Indonesia should pay attention to the following. First, a documentary workshop should allow enough time to deliver basic lectures about language documentation and how to do language documentation as well as to give the opportunity to the participants to familiarize themselves with various equipments needed in documenting languages (e.g. voice recorders, video recorders, and microphones), and to plan and work on a small project based on their native language. The experience to do a documentation project may inspire them to continue working on language documentation after the workshop. In previous documentation workshops hosted in Indonesia, we usually allocated two days and I think the workshops went well.<sup>8</sup> Most of the participants claimed that the workshop was useful and inspiring.<sup>9</sup> They also practiced recorded themselves/each other speaking or singing in their native languages and then transcribed part of the recordings. Nonetheless, they did not have a chance to design their own project and make a product which they could see and share at the end of the workshop. Therefore, adding two to three days to future workshops should be considered as it will give the participants the opportunity to plan and work on their own project during the workshop.

Second, documentary workshops hosted in different places should not be a hit-and-miss project. We hope that after the participants get some experience in doing a documentation project during the workshop, they will continue doing language documentation. Nevertheless, the participants may need help or guidance when they work on their projects. Therefore, it would be ideal if there is a follow-up program after the workshop. This follow-up program aims to ensure that the participants who decide to do language documentation could get supervision or guidance if they encounter problems in their projects, to nurture those who have passion in language documentation, as well as to seek the opportunity to collaborate with local communities

---

<sup>8</sup> A half day seminar about language documentation for bigger audience usually preceded the actual documentation workshop.

<sup>9</sup> Based on the answers provided by the participants on questionnaires we distributed at the end of the workshop we hosted in Kupang in 2016 as well as personal communication with some participants in previous workshops.

for future projects.

## 5. Epilogue

I benefited a lot from DocLing 2016. The experience during the workshop has enriched my knowledge and skills about language documentation. The knowledge and skills, together with the reflection points I made, will be useful for upcoming documentation projects and documentary training I will be involved with at my university in Jakarta or elsewhere.

## Acknowledgement

I would like to thank the Linguistic Dynamic Science Project researchers at Research Institute for Languages and Cultures of Asia and Africa (ILCAA), especially Asako Shiohara for inviting me to DocLing 2016. I would also like to thank Toshihide Nakayama and Anthony Jukes, the editors of the featured topic “DocLing” and Yasuhiro Yamakoshi, the editor of this volume, as well two anonymous reviewers for their insightful comments and suggestions.

## References

- Austin, Peter K. 2010. “Current issues in language documentation”. In Peter K. Austin (ed.) *Language Documentation and Description* 7. London: SOAS. pp.12–33.
- Dalby, Andrew. 2002. *Language in Danger: How Language Loss Threatens Our Future*. London: The Penguin Press.
- Dwyer, Arienne M. 2006. “Ethics and practicalities of cooperative fieldwork and analysis”. In Himmelmann, Nikolaus P., Josh Gippert and Ulrike Mosel (eds.) *Essentials of language documentation*. Berlin: Mouton de Gruyter. pp.31–66.
- Himmelmann, Nikolaus P. 2006. “Language documentation: What is it and what is it good for?”. In Jost Gippert, Nikolaus P. Himmelmann and Ulrike Mosel (eds.) *Essentials of Language Documentation*. Berlin: Mouton de Gruyter. pp.1–30.
- Himmelmann, Nikolaus P. Jost Gippert and Ulrike Mosel. 2006. *Essentials of Language Documentation*. Berlin: Mouton de Gruyter.
- Jukes, Anthony, Asako Shiohara and Yanti 2016. “Training for documenting minority languages in Indonesia and Malaysia”. *Asian and African Languages and Linguistics* 11. pp.45–55.
- Simons, Gary F. and Charles D. Fennig (eds.) 2017. *Ethnologue: Languages of the World, Twentieth Edition*. Dallas, Texas: SIL International. [Online at <http://www.ethnologue.com>, accessed on 2016-12-19.]
- Yanti. 2010. “A reference grammar of Jambi Malay”. Doctoral dissertation, University of Delaware, Newark, DE.
- Yanti, U. Tadmor, P. Cole and G. Hermon (eds.) 2015. *Critò Kitò: A collection of Jambi Stories in the Sabrang Dialect*. Jakarta: The Linguistic Society of Indonesia.

## Documenting Language Use: Remarks on some theoretical and technical issues for language documenters

KIMOTO, Yukinori

Nagoya University, Japan

This paper reviews the documentary linguistics workshop DocLing2013 from the viewpoint of a participant and discusses future prospects on documentary linguistics.<sup>1</sup> The workshop was valuable in that it provided the participants with several key ideas and a rich amount of information crucial for documenting endangered or lesser-known languages. It had a great influence on my own project for the documentation and description of the Arta language; in particular what I learned from the workshop was of great help when I applied for a grant on language documentation. At least two challenges to be addressed are also noted, one of which is relevant to the use of ELAN. I attempt to solve the problem by outlining a systematic use of ELAN in the documentary linguistic context from tier-structuring to the exporting transcripts to other software.

Keywords: DocLing 2013, grant application, interdisciplinary approach, ELAN

1. Introduction
2. DocLing 2013 from a participant's viewpoint
3. Personal benefits from the documentary linguistics workshop
4. Documentary linguistics as an interdisciplinary enterprise
5. Towards a standardization of ELAN
6. Concluding remarks

### 1. Introduction

I have been working on Arta, an endangered language, spoken by ten fluent speakers and 30–50 second language speakers in the northern part of Luzon in the Philippines. My on-going project (since 2012) concerns the documentation of the

---

KIMOTO, Yukinori. 2017. "Documenting language use: Remarks on some theoretical and technical issues for language documenters". *Asian and African Languages and Linguistics* 11. pp.79–94. [Permanent URL: <http://hdl.handle.net/10108/89209>]

<sup>1</sup> I would like to thank DocLing lecturers and staff including Peter Austin, David Nathan, and Anthony Duke, from whom I was able to learn a lot about documentary linguistics during DocLing2013. I am indebted to Daisuke Yokomori, Hiromichi Hosoma, Kohei Kikuchi, Mandana Seyfeddinipur, and Sophie Salfner for sharing their knowledge and skills of ELAN. I also thank two anonymous reviewers for comments on this paper. Remaining errors are of course my responsibility. This paper is partially supported by Endangered Languages Documentation Programme (SG0380 "A Documentation and Description of the Arta language").

language (including the recording and analysis of various kinds of discourse), and a description of the phonological system and morphosyntax, which will be submitted as a Ph.D dissertation. When I began this project, I had no systematic knowledge of linguistic documentation, and I struggled to find better ways to record, analyze, and archive the data by myself. It was at that time that fortunately I was informed that a workshop on language documentation (DocLing 2013) would be held at the ILCAA, the Research Institute for Languages and Cultures of Asia and Africa, Tokyo University of Foreign Studies. In fact, the ideas I learned during the workshop had a great influence on my project: it led me to reconsider the ways of documenting the language, from the choice of microphone to future plans for archiving the data.

In this paper, I will review DocLing 2013 from the viewpoint of a participant, and will discuss prospects for language documentation research. In §2, the Documentary Linguistics Workshop 2013 will be reviewed focusing on the lectures and some features characteristic of DocLing; in §3, I will remark on the benefits I have had from the workshop, in particular, I focus on how it has had an impact on my research, in particular, when I applied for an international grant on endangered language documentation. In §4 and §5, future prospects for documentary linguistics, concerning both theoretical and practical aspects, will be discussed.

## 2. DocLing 2013 from a participant's viewpoint

DocLing 2013 was held for six days in 11th–16th February, 2013, at the Research Institute for Languages and Cultures of Asia and Africa (ILCAA), Tokyo University of Foreign Studies. Lectures were given by Peter Austin (SOAS, University of London), David Nathan (SOAS, University of London<sup>2</sup>), Anthony Jukes (CRLD, La Trobe University), Sonja Riesberg (University of Cologne), and Hideo Sawada (ILCAA, Tokyo University of Foreign Studies). Participants devoted almost half of the time to group projects on language documentation, collaborating with the following native speakers whose languages were not familiar to the participants: Tshering Tashi,<sup>3</sup> Namgay Thinley,<sup>4</sup> Zhargal Badagarov,<sup>5</sup> and Kristian Walianggen.<sup>6</sup> ILCAA staff also assisted the participants during the workshop. The workshop also encouraged the participants to interact with each other and with staff; participants were given sufficient break times to communicate with each other informally.

Let me briefly review the lectures that I found quite impressive; the program of

<sup>2</sup> Currently Co-ordinator of the Centre for Australian Languages and Linguistics, at the Batchelor Institute of Indigenous Tertiary Education.

<sup>3</sup> Royal Society for Protection of Nature, Board Director, Dzongkha native speaker (Bhutan)

<sup>4</sup> Dzongkha Development Commission, Senior Research Officer, Dzongkha native speaker (Bhutan)

<sup>5</sup> Buryat State University, Lecturer, Buryat native speaker (Buryatia, Russian Federation)

<sup>6</sup> Center of Endangered Language Documentation, State University of Papua, Language Consultant, Yali native speaker (Papua, Indonesia)

DocLing 2013 is given in Figure 1<sup>7</sup>:

	10:00-11:20	11:40-13:00	14:00-15:20	15:40-17:00
<b>Day 1</b>	Introduction to language documentation (PA)	Planning language documentation and group projects (AJ, DN)	Public lecture: "Rethinking language documentation and support for the 21st century" (PA, DN, AJ)	Ethics and working with communities (PA, guest presenter Anna Berge)
<b>Day 2</b>	Audio & video (DN, AJ)	Audio & video (AJ, DN)	Set up groups & projects (DN)	Group projects design & report (group & staff)
<b>Day 3</b>	Fieldwork techniques & elicitation (SR)	Recording with consultants (groups & staff)	Software for documentation: survey (PA, AJ)	Software for documentation: practical demonstration (AJ)
<b>Day 4</b>	Data management (DN)	Data management (DN)	Multimedia & mobilisation (AJ, DN)	Photography for documentation (HS)
<b>Day 5</b>	Archiving (DN)	Group project development (group & staff)	Group project development (group & staff)	Individual clinic / Group project (group & staff)
<b>Day 6</b>	Individual clinic / Group project (group & staff)	Group project finalisation (group & staff)	Project presentation (groups)	Project presentations (groups), Wrap-up (all)

Fig. 1 Program of DocLing 2013

**Overview of Documentary linguistics (Day 1, 1st and 3rd classes, Austin 2010a "Current Issues in Language Documentation"):** The lectures and the reading material distributed before the workshop enhanced our knowledge of language documentation or documentary linguistics, including the definition of documentary linguistics, differences between documentary and descriptive linguistics, the current situations of documentary projects in the world and the availability of research fundings, such as ELDP (Endangered Languages Documentation Programme) at SOAS, University of London, the Volkswagen Foundation DoBes project (Dokumentation bedrohter Sprachen), and DEL (Documenting Endangered Languages), an interagency program of the National Science Foundation and the National Endowment for the Humanities. Since this academic field has few introductory textbooks, we found his lectures and the paper to be excellent introductions to the field.

**Ethics in linguistic documentation (Day 1, 1st and 3rd classes, Austin 2010b):** Although linguistic fieldwork presupposes a relationship between researchers and members of the community, the field of linguistics had not begun to address ethical issues seriously until fairly recently. This lecture and the reading material addressed

<sup>7</sup> From <http://lingdy.aacore.jp/en/activity/docling/2013.html> (accessed at Sep 9th, 2016); PA: Peter Austin, DN: David Nathan, AJ: Anthony Jukes, SR: Sonja Riesberg, HS: Hideo Sawada.

such ethical issues as legal aspects (copyright, intellectual rights, and so forth), principles of informed consents, and how to establish a good relationship with a community. In Japan, grant applications often require an explicit statement about ethical matters, but we do not have opportunities to focus on such matters in detail in academic institutions. It was thus fortunate that we were able to discuss the matter in this class and with other participants.

**Audio and video recording, photography (Day 2, 1st and 2nd classes, Day 4, 4th class, Nathan 2010):** Probably one of the most significant academic impact of the field of documentary linguistics in general is that it has made it possible to discuss the technical and practical issues on audio and video recording as an academic pursuit, and has made it possible to accumulate and share the knowledge. In this lecture and the reading material, we learned about epistemological principles of recording (especially the importance of human ears in evaluating recordings), the basics of audio and video recordings, and the typology of microphones, we also saw practical demonstrations of using microphones and learned about various ways of taming noises. All of the participants realized how important it is to choose appropriate microphones based on the conditions of the field, and to adjust the settings carefully for better recordings.

**Software for documentation (Day 3, 3rd and 4th classes):** Two lectures were given concerning software used for language documentation. The first lecture provided a brief overview of various pieces of software, including the purposes of using each one, and their advantages and disadvantages. The second lecture, the participants practiced using ELAN (annotation software for media files) and Toolbox (software for conducting grammatical analysis and building a lexicon). It was of great help for those who wanted to use software in language documentation. However, the lecture was only the first step for using the software; thus slightly more systematic teaching would also be necessary.

**Data management (Day 4, 1st and 2nd classes):** On Day 4, two lectures were given on data management, in which participants were systematically taught how to manage various kinds of data files (such as .eaf (ELAN), .mp4, .wav, and so on). One of the important ideas relates to the labeling of file names. File names are identifiers that should uniquely distinguish one file from another. It is thus advantageous to give a simple file name with ASCII characters, which are structured systematically across files (e.g. arta0031). Only a limited amount of semantic information should be included in the file names; meaningful information should mainly be stored in a metadata file, which is linked to the files using an identifier. This idea has influenced my project in terms of my management of various kinds of files. I still remember “Always backup the data!” and “A filename is an identifier, not a data container”.

In sum, the workshop provided us with a rich amount of important ideas for language documentation, most of which were not taught in universities, probably for the simple



reason that they have not been established as a discipline.

### 3. Personal benefits from the documentary linguistics workshop

I received many benefits from the documentary linguistics workshop, the first of which was the knowledge of how to produce strong grant applications. I applied for an Endangered Languages Documentation Programme (ELDP) grant funded by SOAS, University of London, and was awarded a one-year grant titled “A Documentation and Description of the Arta Language” (SG0380; Yukinori Kimoto, 2015–2016). The grants require the applicants to write a long description of the project, and the ELDP expects them to be familiar with documentation methods. Figure 2 shows the actual questionnaires in the 2015 version of the ELDP grant. For this application form, specific descriptions are expected for several questions, including the summary of the project (Q10, max. 100 words), the description of specific outcomes (Q11, max. 750 words), the detailed description of the project including language context, documentation methods, community context, and research outcomes (Q12, max. 2000 words, thus around 400 words per subcategory), ethical aspects of the project (Q13, max. 750 words), the project work plan (Q14, max. 750 words), and financial details of the support requested (Q15). One can easily imagine that applicants are required to have a systematic knowledge of language documentation. Thanks to the documentary linguistics workshop, I understood what the applicants were expected to describe. I reviewed the slides and papers that had been shared during the documentary linguistics workshop, and was able to compose my proposal in a clear and suitable manner.

Q1 Applicant details (name, address, his/her current position, etc.)
Q2 Host institution which will administer the award (including a letter of acceptance by head/dean of the institution)
Q3 Title of the project
Q4 Duration of the project
Q5 Proposed start date
Q6 Previous ELDP grants? (whether s/he have held a grant from ELDP)
Q7 Is this re-submission?
Q8 Other funding applications
Q9 Commitment to the project (the intended average number of hours per week, etc.)
<b>Q10 Summary of the proposed project (Maximum 100 words)</b>
<b>Q11 Specific outcomes (Maximum 750 words)</b>
<b>Q12 Detailed description of the project (Maximum of 2,000 words across the 4 headings)</b> (i) Language context, (ii) Documentation methods, (iii) Community context, (iv) Research outcomes
<b>Q13 Ethical aspects of the project (Maximum 750 words)</b>
<b>Q14 Project work plan (Maximum 750 words)</b>
<b>Q15 Financial details of the support requested</b> (Specify the details with calculations, and state the justifications of the items.) (a) Replacement teaching costs/research salaries, (b) Technical staff, (c) Language consultants, (d) Equipment, (e) Travel and subsistence, (f) Consumables, (g) Other costs
Q16 Referee (Referee details)
Q17 Curriculum Vitae

Fig. 2 ELDP grant application form (c.a. 24 pages)

Second, and perhaps more importantly, the workshop played a significant role in connecting field researchers in Japan and across countries. The one-week workshop includes a group session as an integral part of it, in which the participants were encouraged to conduct collaborative work. As mentioned above, during the longer breaks between lectures we communicated with each other informally. Our social relationships established in DocLing 2013 is still maintained.

Furthermore, since the workshop was strongly connected to SOAS, it was easy for me to extend my social network to people in SOAS. I attended two workshops held at SOAS: “Plants, Animals, Words”: a workshop on ethnobiological, linguistic, and anthropological studies (September, 2013), and an ELDP training workshop (September, 2015), and I shared fruitful experiences with the other participants, particularly with young field researchers.

In the next two sections, the scope of discussion is broadened into documentary linguistics as a whole. Focusing on the documentation of language use, I will explore the theoretical significance and a practical challenge of documenting language use. In §4, I reconsider the interdisciplinary nature of documentary linguistics, and will suggest that any documentary project without theoretical considerations would be useless to researchers in related fields; we should learn related fields wherein the documentation of language use is of great significance. In §5, a practical issue of annotating discourse data with ELAN, annotation software, is addressed. Based on the technical workshop I gave in Kyoto University in Japan, some basic method and advanced use of ELAN will be explicated.

#### 4. Documentary linguistics as an interdisciplinary enterprise

Although documentary linguistics is supposed to “work in interdisciplinary terms” (Austin 2010a: 13), it seems that there are so far few interdisciplinary projects involving or applying documentary linguistics, with the primary focus of most other research in the field remaining on the interactions between language documentation and language description (or documentary vs. descriptive linguistics).<sup>7</sup> If language documentation is aimed at the preservation of primary data for long-term storage in interdisciplinary terms, trainings on language documentation should show participants how our data can contribute to research in other disciplines. Whether or not the documenters have such knowledge may seriously affect the quality of data collected.

Consider a hypothetical situation in which a linguist used an IC recorder to record a procedural text of traditional cooking, and archived the data with the transcription: “*we put that here, and do it like this, or like this, and after that, we put it there*”. It would be no problem if he just wants to write a grammar, but language documenters should note

<sup>7</sup> In fact, in DocLing 2013, all the lecturers and staff were linguists, and almost all the participants in DocLing 2013 were linguists. We participants were thus quite unclear as to how we can develop our work “in interdisciplinary terms” possibly with researchers of other fields.

that non-verbal information might be a crucial key to understand their cultural practices in cooking when none of the community member lost such cultural knowledge.

It is necessary for language documenters to at least be familiar with other (broad) areas of studies in languages because ignorance of other academic fields results in data collections that are totally useless to researchers involved in linguistic or come in other fields. There is no theory-neutral documentation (**theory-ladenness of observation**). After all, we are the one who “constructs” and “frames” the data from a particular viewpoint.

Actually, the primary data of language use are not useful only for descriptive linguistics, but are also of great use for other fields. Some typical fields to which the primary data of language use can contribute are shown below:

- usage-based approach in cognitive linguistics (Langacker 1987, Bybee 2010)
- gesture studies in psychology (Kendon 1988, McNeill 1992, Clark 1996, Kita 2003)
- research on language socialization in anthropology (Ochs et al. 2001, Takada 2012)
- conversation analytic studies in sociology (Sacks, et al. 1974, Schegloff 2007)
- interactional linguistics—an interdisciplinary enterprise between conversation analysis and functional linguistics (Couper-Kuhlen and Selting 2001, Hayashi 2003)

The attempt to apply documentary linguistic studies to these fields will reveal illuminating facts on language, cognition, social interaction and culture.<sup>8</sup> If practitioners of language documentation are informed of such orientations, their data are potentially of great use to these fields.

## 5. Towards a standardization of ELAN

The second challenge of documenting language use is relevant to a technical issue, concerning the standardization of ELAN.<sup>9</sup> It seems that the practical skills in ELAN that are necessary for language documentation are not always taught in the context of language documentation trainings. Since DocLing ended, I learned more about how to use the software through trial and error with the help of the manual and other users of ELAN; this has sometimes required me to radically change the format of some of my annotated files to make them consistent.

---

<sup>8</sup> Mark Dingemanse, Nick Enfield, and their colleagues, for example, surveyed typological variations of so-called “other-initiated repair” (Dingemanse and Enfield 2015), revealing remarkable commonalities across languages/cultures.

<sup>9</sup> ELAN was developed by the Max Planck Institute for Psycholinguistics; it was designed for a number of different users such as linguists, language documenters, gesture and sign language researchers, and for their collaborations (Brugman and Russel 2004). It is powerful and enables various methods of annotation, so that it has become an essential tool for language documentation at the moment.

In this section, I attempt to provide one model for how to teach ELAN to language documenters and the students of linguistics in general. The model tutorial presented below was gradually modified through the process of writing an ELAN tutorial for language documentation (Kimoto (to appear)), and technical workshops on ELAN, which was held three times in the summer of 2016 designed for linguistics students at Kyoto University, Japan.

For ELAN users in the context of language documentation or linguistic analysis, the following skills and types of knowledge are especially relevant:

1. A basic method of annotating media files.
2. The conceptual difference between linguistic types and tiers, and their practical usage.
3. Advanced usage: automatic numbering, word-breaking, controlled vocabulary

Two-hour workshop was designed to include the above contents: the outline of the workshop is given in Figure 3. Prior to the workshop, participants had finished installing ELAN, and had been provided with a movie file (.mp4), and a sound file (.wav) extracted from the movie file. The movie file had been recorded in my field session on the Arta language.

In the technical workshop, I demonstrated the method to use ELAN in my field situation, where a researcher is annotating speech in Arta, spoken by Speaker A, and translating it into Ilokano (a lingua franca widely used in the northern Luzon area of the Philippines) and English. Participants practiced ELAN within the exemplar situation.

1. Basics of ELAN:
  - 1.1. A simple method of annotation
  - 1.2. Using ELAN transcriptions as a corpus: searching across annotation files
2. Constructing tier structure
  - 2.1. Conceptual difference between linguistic types and tiers
  - 2.2. Setting up linguistic types
  - 2.3. Adding tiers
  - 2.4. Saving as a template
  - 2.5. Adding participants from a single participant template
3. Advanced uses of ELAN
  - 3.1. Synchronizing video and audio data
  - 3.2. Notes on video data
  - 3.3. Autonumbering of annotations
  - 3.4. Controlled vocabulary
4. Exporting
  - 4.1. Exporting as an interlinearized text file
  - 4.2. Exporting to FLEx (FieldWorks Language Explorer)

Fig. 3 Technical workshop on ELAN for linguistic analysis

In what follows, I provide one model for teaching ELAN to language documenters, which was modified based on the workshop I gave to linguistics students.

### Session 1: Basics of ELAN

**A simple method of annotation** As a first step, participants are directed to notice how simply they can annotate media files. After the movie and sound files are imported into ELAN, the participants are directed to annotate files in the simplest way (without making any change in linguistic type and tier). They learn the following points:

- how to use shortcut keys. In particular, annotations cannot be moved or expanded/contracted without using the ALT key (Table 1).
- media player options; e.g. arranging the playback speed
- representation of the waveform (it is often the case that if the sound is recorded in a low volume, the waveform is not visible clearly): right-click on the waveform viewer > vertical zoom > select 300% (Note that waveforms are not shown if the media type is video!)

Table 1 Shortcut keys in ELAN

Start/Pause the playback:	CONTROL + SPACE
Play the selected interval:	SHIFT + SPACE
Enter an annotation into a selection:	ALT + N (active tier only)
Edit a selected annotation:	ALT + M (active tier only)
Save the current annotation:	ENTER
Move an annotation:	ALT + drag the middle of the annotation
Expand/contract the length of an annotation:	ALT + drag the border of the annotation

**Using ELAN transcriptions as a corpus: searching across annotation files** The next session introduces methods of searching in ELAN, particularly because the ability to search across files is one of the main advantages of using ELAN; this makes it possible to use our data as a corpus. Here just a short introduction may be provided, including a method of defining the corpus (“define domain”), and the availability of regular expressions (with reference to Mosel 2015).

### Session 2: Constructing tier structure

When using ELAN, it is essential to understand the conceptual difference between linguistic types and tiers. But this is not always easy for learners to understand quickly. In the workshop I gave, I used illustrations to make it easier to grasp the ideas (Figure 4).

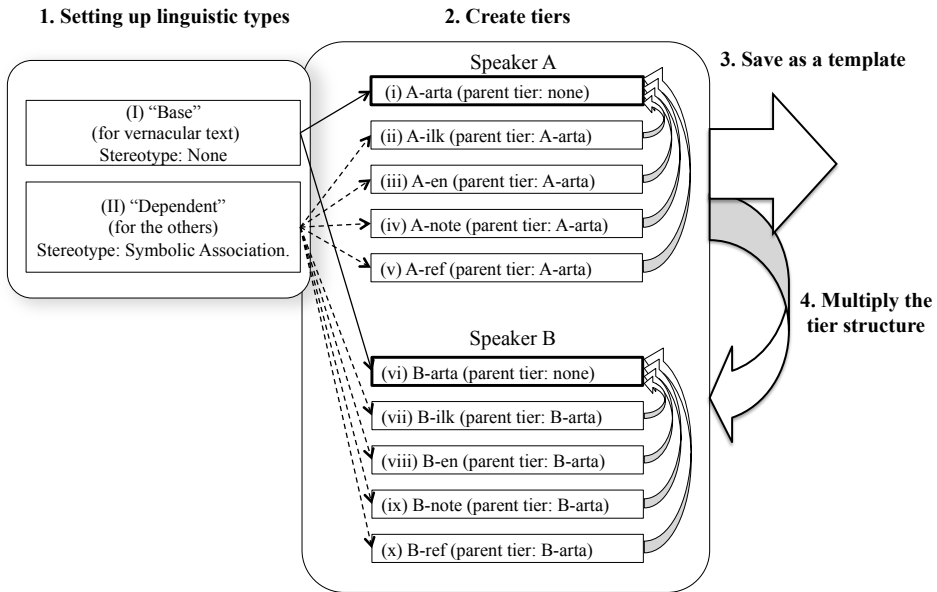


Fig. 4 Tier-structuring process

In the workshop, vernacular texts (in Artá) were assigned to the “base” linguistic type, and the others (Ilokano translation, English translation, grammatical/lexical notes, reference number) the “dependent” linguistic type. Participants were directed to create a tier structure for one speaker, and to save it as a template for further use of the same tier structure. They also learned to multiply the tier structure in case more than one speaker is relevant to the transcription. This process seems more advantageous than manual additions particularly because users can create the same tier structure in quite a simple way.

**Conceptual difference between linguistic type and tier** I illustrated the conceptual difference between linguistic type and tier by explaining the following to students. Suppose you need five kinds of information for each annotation, that is, (i) a transcription of the object language (here “Artá”), (ii) a translation in the lingua franca (Ilokano), (iii) an English translation, (iv) grammatical/lexical notes, and (v) the utterance ID. In this case, these five layers should be prepared in the setting. But on a more abstract level, they can be classified into two general types: (I) one in which you can freely change the intervals and positions within the timeline (that is, (i) Artá), and (II) another in which the intervals and positions are dependent on the first tier type (that is, (ii)–(v)). These two general tier types are called **linguistic types**, and the five actual layers are called **tiers** in ELAN. If the media file to annotate includes three speakers, 15 “tiers” should be prepared, whereas two “linguistic types” are sufficient for the purpose.

**Setting up linguistic types and adding tiers.** After reaching a conceptual understanding of the two different levels of tier structure, participants learn other key concepts such as “stereotype”, “parent tier”, “symbolic association”, and inheritances between linguistic types and tiers, guided by the illustration in Figure 4.

**Saving as a template.** After a tier structure for one speaker is constructed, the tier structure is saved as a template (File > Save as Template). Doing so eliminates the need to re-create the same tier structure for each annotation file, and allows for consistent tier structures across files.<sup>10</sup>

**Adding a new participant.** If the media file being annotated has two or more speakers, the tier structure for one speaker can be multiplied in a relatively simple manner (rather than manually adding tiers). Click on Tier menu > Add New Participant... (then a dialog window will appear). In the dialog window, (i) check if the tier structure to copy is selected; (ii) enter the name of the speaker to add in “Specify the new participant”; (iii) select the relative position of the name value specified in the label in “Specify whether the suffix or prefix of the selected tier structure(s) to be changed” (for a new tier structure like “B-arta, B-ilk, ...” from “A-arta, A-ilk, ...”, select “prefix”, or for “arta-B, ilk-B, ...”, select “suffix”); (iv) enter the name value of the speaker in the original tier structure (e.g. A) in “Enter the value to be replaced”; (v) enter the name value of the new speaker (e.g. B) in “Enter the new value for replacement”; (vi) select “OK”, and the result will be shown in the timeline viewer.

### Session 3: Advanced uses of ELAN

During the workshop I organized, participants worked with some of the more advanced uses of ELAN, such as annotation autonumbering, controlled vocabulary, and word/morpheme-breaking; they also practiced synchronizing multiple media files. These advanced functions are not necessarily required when using ELAN, but are useful in some cases. Here I just make some short remarks on the synchronization of media files.

**Synchronizing video and audio data.** Although ELAN provides ways to synchronize more than one media file (e.g. video-video, video-audio), researchers should first consider whether this is really necessary. If a single session has been recorded with two videocameras, it would make sense to synchronize those two video files. However, if one hopes to synchronize video and audio files (e.g. suppose a situation in which session was both video-recorded with a camcorder, and audio-recorded with an IC-recorder), one should reconsider whether that is the only option available. Since synchronization in ELAN can be a troublesome task, I always

<sup>10</sup> Even if tier structures are inconsistent across files, ELAN does work. The problem arises when you search across files, however. One of the advantages of cross-file searching is that a particular tier or speaker can be specified. If tier structure are inconsistent across files, you cannot search across files according to these attributes.

record sessions with a videocamera and an external microphone; this seems sufficient to get a clear sound with less noise.

Various pieces of software can be used to extract .wav files from .mp4 or other video files (e.g. Avidemux (Windows), QuickTime Player 7 (Mac), Audacity (Windows, Mac)); using .wav files enables ELAN to show waveforms along the timeline.<sup>11</sup>

#### Session 4: Exporting interlinearized texts into FLE<sub>x</sub> (FieldWorks Language Explorer)

FLE<sub>x</sub> or FieldWorks Language Explorer is software used for building a lexicon, analyzing texts, and studying morphology and syntax, distributed by SIL International. Although it is also powerful software, and ELAN and FLE<sub>x</sub> have an interface for exporting/importing each other's files, it is quite a complex procedure. The manual "Working with ELAN and FLE<sub>x</sub> together" is available under "Third party resources" on the ELAN website. It begins with instructions to set up a tier structure and to annotate a media file in a special way, which do not really answer what we want to know; if you already have several annotated ELAN files to export into FLE<sub>x</sub>, this method does not work at all.

This is the background motivation to explicate what is probably the most realistic way to export annotated ELAN files into FLE<sub>x</sub> below in detail.

**Step 1. Add the linguistic type of the title.** FLE<sub>x</sub> cannot import the data properly if the labels of linguistic types do not follow a designated format and if the data do not have a title tier. Here you need to add a linguistic type for the title of the annotated file. Go to "Type" > "Add Linguistic Type" and create a new linguistic type with the name "Title", and the stereotype "None". Next, go to "Tier" > "Add tier" and create an additional tier with the name "Interlinear-title-[ISO639]" (see also Table 2).

**Step 2. Assign participant information to the annotations.** Interlinear texts in FLE<sub>x</sub> are not adopted to multi-participant discourses so far. If the session annotated is a conversation by more than one speaker, such information on the speakers should be put in the "notes" tier. Create a new tier with the dependent linguistic type, and add new annotations of speaker information below each utterance (e.g. "A", "B", etc. for each annotation). It is easy to make empty annotations automatically by clicking "Tier" > "Create Annotations on Dependent Tiers..." and entering the speaker information, "A" or "B", over the same tier by clicking "Tier" > "Label And Number Annotations..."

**Step 3. Change the tier names.** FLE<sub>x</sub> requires tier names to follow a particular format to be able to read them properly. Change the names of relevant tiers (Table 2). In particular, ISO-639 codes for an object language or vernacular language (e.g. *atz* for Arta), and a translation language (e.g. *en* for English) are required.<sup>12</sup> You can consult

<sup>11</sup> Note that the waveform does not appear if the annotation file is linked only to video file(s).

<sup>12</sup> It seems that this information should correspond to the ISO codes you used when a FLE<sub>x</sub> project was built.



*Ethnologue*<sup>13</sup> or even *Wikipedia* to get the ISO codes.

Table 2 Tier names for FLEx exporting

	Format	Example
Title	<b>Interlinear-title-[ISO639-of-object-lang.]</b>	<b>Interlinear-title-atz</b>
Transcription	[speaker]- <b>txt</b> -[ISO639-of-object-lang.]	<b>A-txt-atz</b>
Translation	[speaker]- <b>gls</b> -[ISO639-of-translation-lang.]	<b>A-gls-en</b>
Speaker	[speaker]- <b>note</b> -[ISO639-of-translation-lang.]	<b>A-note-en</b>

The whole tier structure modified here is shown in Figure 5, where the difference between the one constructed here and the one illustrated in Figure 4 is shown in bold.

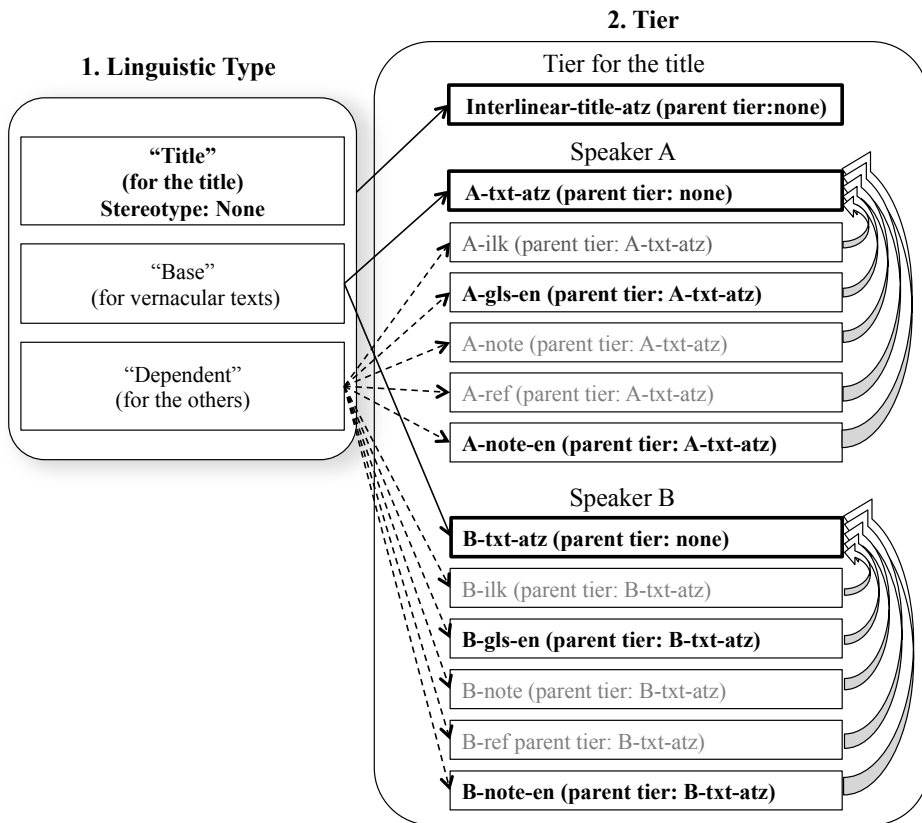


Fig. 5 Tier structure for FLEx exporting

<sup>13</sup> Online version: <https://www.ethnologue.com/>.

Note that if the transcription tier is not a parent tier, but if another tier (e.g. a reference number tier) is a parent tier within the tier structure, then the parent-child relation must be reversed. ELAN does not seem to have the function of changing the tier structure directly, but the “copy tier” function can be used to do this. First, copy the transcription tier (i.e. a dependent tier), selecting “Transcription (no parent)” to make it an independent tier, then copy the original parent tier (e.g. the reference number tier), placing it under the new parent tier. After editing the tier structure in this way, delete the original tier structure.

**Step 4. Annotate the title.** Create one annotation of the `Interlinear-title-xxx` tier in an arbitrary location, and enter the title of the file, which may correspond to the filename (e.g. “arta0032”), or which may be a more informative title (e.g. “building a traditional house”). This annotation becomes the title of the interlinear text in FLEEx (but sometimes this does not work, and you might need to input the title in FLEEx manually!).

**Step 5 Export** Choose File > Export as > FLEEx file .... and an “Export as FLEEx File” window will appear:

**Screen 1 Setting up the Title/Base linguistic type and tier(s)**

- In the upper box “Element mapping”, select “Export interlinear text tier” (but not “Export paragraph tier”), and then select “Interlinear-title-xxx” in the drop-down menu. Confirm that “phrase” in the “FLEEx Element Name” column corresponds to “Base” in “Corresponding Tier Type” column. If not, select “Base” from among the candidates.
- In the lower box “Select tiers to be exported”, select all of the transcription tiers (e.g. “A-txt-atz” and “B-txt-atz” in Figure 5) in the “phrase” column, and unselect the other tiers.

**Screen 2 Setting up the Dependent linguistic type(s) and tier(s)**

- In the upper box, check if “phrase\_item” in the “Item Mapping” includes the “Dependent” type.
- In the lower box (“Select tiers to be exported”), select all of the dependent tiers necessary (e.g. translation tiers: “A-gls-en”, “B-gls-en”; and participant-value tiers: “A-note-en”, “B-note-en”), and unselect the others.

**Screen 3** Nothing to arrange. Click “Next”.

**Screen 4** Specify the location for saving the file where FLEEx can access it. Click “Finish”.

Now the file has been exported. To import the file in FLEEx, open the project where the data will be imported. Select “View” > “Texts & Words” > “Interlinear Text” (or select “Texts & Words” on the lower left side of the window). Go to “File” > “Import” > “FLEEx Text Interlinear”. Specify the location of the file and open it.<sup>14</sup> Now you can

<sup>14</sup> At present (Sep.11th, 2016), there is a bug such that the text cannot be recognized as a vernacular language. In

view data from ELAN in FLEx.

## 6. Concluding remarks

In this paper, I reviewed the documentary linguistics workshop DocLing 2013, where I looked back at the classes that were particularly beneficial from a participant's perspective. I also described how the workshop had an influence on my research; the lectures on documentary linguistics that I attended as part of the workshop were of great help to me when I applied for the Endangered Languages Documentation Programme (ELDP) grant. Finally, I mentioned some current challenges in documentary linguistics, focusing on the documentation of language use, i.e. the lack of interdisciplinary approach, and the absence of standardization of ELAN. For the latter problem, I attempted to explicate a better way to use ELAN with a special focus on tier-structuring and the interface between ELAN and FLEx. I hope the discussions provided here will provide some perspectives and insights for the further development of documentary linguistics in general and projects around the world.

## References

- Austin, Peter K. 2010a. "Current issues in language documentation". In Peter K. Austin (ed.) *Language Documentation and Description* 7. London: SOAS. pp.12–33.
- Austin, Peter K. 2010b. "Communities, ethics and rights in language documentation". In Peter K. Austin (ed.) *Language Documentation and Description* 7. London: SOAS. pp.34–54.
- Brugman, Hennie, and Albert Russel. 2004. "Annotating multi-media / multi-modal resources with ELAN". In: *Proceedings of LREC 2004, Fourth International Conference on Language Resources and Evaluation*.
- Bybee, Joan. 2010. *Language, Usage and Cognition*. Cambridge: Cambridge University Press.
- Clark, Herbert. 1996. *Using Language*. Cambridge: Cambridge University Press.
- Couper-Kuhlen, Elizabeth and Margret Selting. 2001. "Introducing interactional linguistics". In Margret Selting and Elizabeth Couper-Kuhlen (ed.) *Studies in Interactional Linguistics*. London: SOAS. pp.1–22.
- Nathan, David. 2010. "Sound and unsound practices in documentary linguistics: Towards an epistemology for audio". In Peter K. Austin (ed.) *Language Documentation and Description* 7. London: SOAS. pp.262–284.
- Dingemanse, Mark and Nick J. Enfield. 2015. "Other-initiated repair across languages: Towards a typology of conversational structures". *Open Linguistics* 1(1). pp.96–118.
- Hayashi, Makoto. 2003. *Joint Utterance Construction in Japanese Conversation*. Philadelphia: John Benjamins Publishing.
- Kendon, Adam. 1988. "How gestures can become like words". *Cross-Cultural Perspectives in Nonverbal Communication* 1. pp.131–141.
- Kimoto, Yukinori. to appear. "Gengo dokyumenteesyon notameno ELAN siyoo (ELAN for language documentation)". In Hosoma, Hiromichi and Kohei Kikuchi (eds.) *ELAN Nyuumon (An Introduction to ELAN)*. Tokyo: Hituzi Syobo Publishing.
- Kita, Sotaro (ed.) 2003. *Pointing: Where Language, Culture, and Cognition Meet*. New Jersey: Lawrence Erlbaum Associates.

---

such a case, select all of the text and choose the language value in the drop-down menu of the languages appearing in the second-from-the-top row.

- Langacker, Ronald W. 1987. *Foundations of Cognitive Grammar: Vol.1 Theoretical Prerequisites*. Stanford: Stanford University Press.
- McNeill, David. 1992. *Hand and Mind: What Gestures Reveal about Thought*. Chicago: University of Chicago Press.
- Mosel, Ulrike. 2015. "Searches with regular expressions in ELAN corpora". [Online at <https://tla.mpi.nl/tools/tla-tools/elan/thirdparty/> under the description of "Regular expressions for ELAN users", accessed on 2016-09-12].
- Ochs, Elinor and Bambi Schieffelin 2001. "Language acquisition and socialization: Three developmental stories and their implications". In Alessandro Duranti, Elinor Ochs and Bambi B. Schieffelin (eds.) *Linguistic Anthropology: A reader*. Chichester: Blackwell. pp.263–301.
- Sacks, Harvey, Emanuel A. Schegloff and Gail Jefferson 1974. "A simplest systematics for the organization of turn-taking for conversation". *Language* 50.4. pp.696–735.
- Schegloff, Emanuel A. 2007. *Sequence Organization in Interaction: Vol.1: A Primer in Conversation Analysis*. Cambridge: Cambridge University Press.
- Takada, Akira. 2012. "Preverbal infant–caregiver interaction". In Alessandro Duranti, Elinor Ochs and Bambi B. Schieffelin (eds.) *The Handbook of Language Socialization*. Chichester: Blackwell. pp.56–80.

## DocLing from the Participant's Perspective

SHIMOJI, Michinori

Kyushu University

This short essay examines how DocLing has positively impacted the research career of the present author. It also gives a brief introduction to several documentation and revitalization projects in which the present author or former participants of DocLing are working in collaboration.

**Keywords:** DocLing, Language documentation, Language revitalization, Ryukyuan

1. Introduction
2. DocLing and Ryukyuan linguistics
3. Ryukyuan documentation/revitalization projects
4. Concluding remarks

### 1. Introduction

As a DocLing project participant, I have much to say about how DocLing has positively impacted my own research career. After reviewing DocLing and its impact on my research within the context of endangered language studies in Japan, I go on to mention several documentation and revitalization projects in which the present author or former participants of DocLing are working in collaboration. I have been working on the description of Ryukyuan for about ten years now. Ryukyuan is a group of endangered Japonic languages, spoken on a chain of islands called the Ryukyu archipelago. According to UNESCO's report on language endangerment (<http://www.unesco.org/languages-atlas/>), there are at least eight endangered languages in Japan: Ainu, Hachijō, and six Ryukyuan languages (Amami, Kunigami, Okinawa, Miyako, Yaeyama, and Yonaguni).

In 2008, I wrote a descriptive grammar of Irabu, a Southern Ryukyuan language, as a PhD thesis submitted to the Australian National University (Shimoji 2008, 2017). It was just after submitting the thesis that I first joined DocLing. At that time, I had some elementary knowledge of the emerging field of research called “documentary linguistics,” having read the foundational work by Himmelmann (1998), but that

was almost all the information I had about documentary linguistics, and it was only theoretical knowledge. As was usual for many field linguists in those days, I identified myself as a “descriptive linguist,” with my primary interest being in compiling comprehensive grammar, text material, and a dictionary of the subject language, and I did not have any actual experience, or even interest, in conducting fieldwork that was informed by the theory of documentary linguistics. I had the misconception that documentary linguistics is a mere subdivision of descriptive linguistics, or “over-engineered” descriptive linguistics.

The experience of studying a range of concepts pertaining to the theory of documentary linguistics in a week-long intensive workshop had a big impact on me. As a participant of DocLing, I learned how description and documentation differ in their goals, and how the knowledge of one field of research is essential for the practice of another. In Section 3, I will briefly mention my own documentation project, which is clearly based on what I learned in the DocLing workshop.

## 2. DocLing and Ryukyuan linguistics

Around the time of the first DocLing (2008), there was an emerging “movement” of Ryukyuan studies, where younger scholars started to work on individual Ryukyuan dialects in the context of language endangerment, aiming to describe each individual Ryukyuan dialect in its own right. This new approach was distinct from the traditional Ryukyuan study, which tended to view Ryukyuan merely as useful dialectal data for the description of synchronic variation, or for the historical reconstruction of the Japonic family.

The departure from the traditional dialectological approach was quite independent of DocLing, but it was DocLing that introduced the important concept of documentation, which should not be confused with description, into the new approach of Ryukyuan studies. As a researcher involved in the emergence of Ryukyuan studies, I was aware that our new approach was more description-oriented. That is, our concern was primarily to describe each individual Ryukyuan dialect in the form of descriptive grammar (Shimoji 2008, Pellard 2009, Hayashi 2013, Niinaga 2014, etc.). Further, we did not have a clear picture of how the languages we were working on should be documented, and how the data should be managed in such a way that the data collected can be accessible for the wider community of researchers and for the local people. We also wanted to make our data usable for language revitalization if the local community wants such data.

Today, we are looking at a well-balanced research trend in which both description and documentation are emphasized. As will be shown in the following section, there are a number of research projects that focus on the documentation and revitalization of Ryukyuan languages.

### 3. Ryukyuan documentation/revitalization projects

The last ten years have seen the emergence of a number of documentation and revitalization projects by Ryukyuan researchers. They are not necessarily led by former participants of DocLing, but most projects similar to the ones I introduce in this section are in collaboration with the former participants of DocLing. In what follows, I briefly give an overview of three such projects, two of which aim at revitalization, and one of which aims at documentation.

#### 3.1. Devising an orthography

The project called “A unified writing system for the Ryukyuan languages” is a collaborative project granted by the Toyota Foundation. The project is led by Dr. Shinji Ogawa (Kumamoto Prefectural University), with twelve collaborative researchers. The project aims to establish an orthographic system for Ryukyuan languages. The outcome has been published as a book (Ogawa (ed.) 2015), and the syllabary (i.e., the full list of symbols used for the orthographic system), with the sound(s) represented by each symbol, is listed on the project's website (<http://www.ryukyuan-writing.net/>).

When we think about language revitalization in Japan, we must take into account the fact that the local people are all bilingual with Japanese as one of the languages, and that Japanese has its own writing system, the *kana* system, which gives one letter for each syllable/mora (e.g., あ for /a/ and か for /ka/). All Ryukyuan researchers agree that the local people find it easier to use the *kana* system, than learning to use an IPA-based alphabetic system uniquely devised for the language they speak. The problem is that the syllable structure and phoneme inventory of a given Ryukyuan language are often far more complex than those of Japanese, and the existing *kana* system devised for Japanese cannot fully represent the sound system of Ryukyuan. For example, Irabu Ryukyuan has a number of syllabic consonants that begin in /b/, like /bz/ [bʒ] and /br/ [br̥], in addition to regular syllables that begin in /b/, like /ba/, /bi/, and /bu/. The *kana* system provides characters for the latter only, ば /ba/, び /bi/, and ぶ /bu/.

Thus, if we use the Japanese *kana* system for Ryukyuan languages, we need to devise a strategy to cover the syllables that cannot be represented by the existing *kana* system. In addition, it is useful if such a modified *kana* system is applicable to *all* local varieties of Ryukyuan, so that any local community of Ryukyuan may use the system when they want to create a textbook or some other written material for their revitalization efforts. The problem here is that the dialectal difference in Ryukyuan is striking, making it difficult to provide a unified, cross-dialectal *kana* system applicable for all Ryukyuan languages and their sub-dialects. The project solves this problem by providing a large but consistent set of auxiliary symbols in addition to the regular *kana* syllabary for Japanese, whereby we can write in *kana*, with additional symbols where necessary. In the Irabu orthography, for example, the syllables /bz/ and /br/ are written with the

regular characters 𐄇 plus the small symbols 𐄈 and 𐄉 respectively, as in 𐄇𐄈 for /bz/ and 𐄇𐄉 for /br/, where the small symbols represent syllabic consonants. The use of small symbols to represent syllabic consonants is consistent for all other Ryukyuan dialects.

The project is a key to achieve active involvement of the local community, which is essential in language revitalization. The orthography enables education in a minority language, which would otherwise be impossible. It also encourages local writers and poets to publish their works in their own language. In this way, the language revitalization process gains sustainability, another important feature of a successful revitalization project.

### 3.2. Port language revitalization

The orthography project noted above has significant consequences for the revitalization of Ryukyuan languages, since it is now possible for subsequent revitalization projects to concentrate on content (i.e., products and outcomes such as textbooks, etc.) by utilizing the orthographic system for their target languages.

Dr. Masahiro Yamada (National Institute for Japanese Language and Linguistics) provides an excellent example of a language revitalization project using the orthographic system. In pursuit of revitalizing Yoron, Okinoerabu, and Yonaguni, all of which are in an imminent danger of extinction, Yamada is working in collaboration with linguistic researchers, designers, and the local community. His project is called *Port Language Revitalization* (<http://plrminato.wixsite.com/webminato>). The concept here is to provide a platform for content-based language revitalization. That is, his project serves as a platform (or “port” in his metaphor) where various kinds of content (e.g., fun books, storybooks, etc.) are created for the local community to utilize for language revitalization activities. Five Ryukyuan researchers, one specialist in the science of design, four designers, one photographer, and one illustrator are involved in this project, and more than ten content items have been created and made available for use.

### 3.3. The interdisciplinary dictionary project of Irabu

I am currently working on a documentation project that aims to document the lexicon of Irabu Ryukyuan with the help of specialists of different research disciplines<sup>1</sup>. The interdisciplinary project team consists of a linguist (i.e., myself), a geographer, and a biologist. In this project, the local flora-fauna terms and place names are collected and documented in the form of a dictionary. The lexical knowledge of these semantic fields is much more effectively documented by collaborating with specialists of geography and of biology, than by a single linguist like me. In addition, the dictionary so compiled will be usable for the wider research community of biologists, geographers, as well as

<sup>1</sup> The project is supported by QR Program of Kyushu University (2016 to current)



linguists and anthropologists. Even though the project has a fixed term of two years, I will continue the project by getting another grant after the current term has terminated, so that different semantic fields can be added to the existing lexicon (i.e., flora-fauna terms and place names) with the help of a new documentation team consisting of different specialists.

#### 4. Concluding remarks

DocLing has provided a firm basis on which Ryukyuan researchers can have a basic knowledge of documentary linguistics and its theory, methodology, and technology in an effective way. Thanks to DocLing, it is now common for a field linguist who works on Ryukyuan to pay fair attention to language documentation and revitalization. Being a free workshop that takes place in Japan, DocLing has allowed a number of young scholars and students based in Japanese institutions, especially those who work on Ryukyuan, to gain the basic knowledge of language documentation at a surprisingly low cost. When I joined DocLing, the majority of participants were postgraduate students and post-docs, and they all agreed that they would not have been able to attend such a workshop, if it had taken place overseas and/or with a high registration fee. The positive result, or “investment effect,” of the free workshop is obvious. A number of documentation and revitalization projects are in progress as we noted in Section 3.

#### References

- Hayashi, Yuka. 2013. “Minami Ryukyu Miyakogo Ikema hogen no bunpo [A grammar of the Ikema dialect of Miyako, of Southern Ryukyuan]”. Unpublished PhD thesis, Kyoto University.
- Himmelmann, Nikolaus P. 1998. “Documentary and descriptive linguistics”. *Linguistics* 36. pp.161–195.
- Niinaga, Yuto. 2014. “A grammar of Yuwan, a Northern Ryukyuan language”. Unpublished PhD thesis, University of Tokyo.
- Ogawa, Shinji. ed. 2015. *Ryukyu no Kotoba no Kakikata [How to Write Ryukyuan Languages]*. Tokyo: Kurosio Shuppan.
- Pellard, Thomas. 2009. “Ogami: Éléments de description d’un parler du sud des Ryukyus”. PhD thesis, École des hautes études en sciences sociales, Paris, France.
- Shimoji, Michinori. 2008. “A grammar of Irabu, a Southern Ryukyuan language”. Unpublished PhD thesis, Australian National University.
- . 2017. *A Grammar of Irabu, a Southern Ryukyuan Language*. Fukuoka: Kyushu University Press.



## Vocabulary of Shingnyag Tibetan: A Dialect of Amdo Tibetan Spoken in Lhagang, Khams Minyag

SUZUKI, Hiroyuki

IKOS, Universitetet i Oslo / National Museum of Ethnology

Sonam Wangmo

IKOS, Universitetet i Oslo

Lhagang Town, located in Kangding Municipality, Ganzi Prefecture, Sichuan Province, China, is inhabited by many Tibetan pastoralists speaking varieties which are similar to Amdo Tibetan even though it is located at the Minyag Rabgang region of Khams, based on the Tibetan traditional geography. Among the multiple varieties spoken by inhabitants living in Lhagang Town, the Shingnyag dialect is spoken in the south-western part of the town. It is somewhat different from other Amdo varieties spoken in Lhagang Town in the phonetic and phonological aspects. This article provides a word list with ca. 1500 words of Shingnyag Tibetan.

**Keywords:** Amdo Tibetan, Minyag Rabgang, dialectology, migration pattern

1. Introduction
2. Phonological overview of Shingnyag Tibetan
3. Principal phonological features of Shingnyag Tibetan

### 1. Introduction

This article aims to provide a word list (including ca. 1500 entries) with a phonological sketch of Shingnyag Tibetan, spoken in Xiya [Shing-nyag]<sup>1</sup> Hamlet, located in the south-western part of Tagong [lHa-sgang] Town (henceforth Lhagang Town), Kangding [Dar-mdo] Municipality, Ganzi [dKar-mdzes] Tibetan Autonomous Prefecture, Sichuan Province, China (see Figure 1). Lhagang Town is in the easternmost part of Khams based on the traditional Tibetan geography, however, it is inhabited by many Tibetans whose mother tongue is Amdo Tibetan.<sup>2</sup> Referring to Qu and Jin (1981), we can see that it is already known that Amdo-speaking Tibetans live in

---

SUZUKI, Hiroyuki and Sonam Wangmo. 2016. "Vocabulary of Shingnyag Tibetan: A dialect of Amdo Tibetan spoken in Lhagang, Khams Minyag". *Asian and African Languages and Linguistics* 11. pp.101–127. [Permanent URL: <http://hdl.handle.net/10108/89211>]

<sup>1</sup> The original Tibetan orthography in a Wylie transliteration is attached in square brackets for each proper name with a transcription of Chinese.

<sup>2</sup> For this reason, it is often misunderstood that the Amdo-speakers are the original inhabitants in Lhagang Town; that is why Suzuki and Sonam Wangmo (2015a, b) emphasise the existence of the local vernacular of Khams spoken there.

some parts of Khams, however, varieties of Amdo Tibetan spoken in Khams have not been well investigated so far, especially those spoken in the easternmost Khams region. Wang (2012) includes a phonological description of a Amdo variety spoken in Luhuo [Brag-'go], which is, indeed, a variety spoken in a different area from Kangding. The gSerkha dialect of Amdo Tibetan (Suzuki 2015) is a variety considered as a member of Amdo Tibetan in Kangding.



Fig. 1 Geographical position of Lhagang Town. (=marked)<sup>3</sup>

As shown in Suzuki and Sonam Wangmo (2015b), Lhagang Town is a multilingual area, and they claim that there are at least four Tibetic varieties spoken in the town: two Khams varieties and two Amdo varieties. In addition, Suzuki and Sonam Wangmo (2016b) report an existence of a newly recognised non-Tibetic language, Lhagang Choyu, spoken in a hamlet within Lhagang Town. All of them are spoken in the centre of Lhagang Town, called Lhagang Village, because it is inhabited by Tibetans who are originally sedentary settlers in the village as well as immigrants from the surrounding rural, pastoral area due to the rapid urbanisation and the governmental resettlement policy.<sup>4</sup>

Figure 2 displays a distribution of regiolects within Lhagang Town recognised and

<sup>3</sup> Figure 1 is designed with Googlemaps (<https://www.google.co.jp/maps/>; accessed on 25th June 2015). Also employed in Suzuki and Sonam Wangmo (2015b, 2016a).

<sup>4</sup> See Sonam Wangmo (2013) and Suzuki and Sonam Wangmo (2017) for details.

classified by the present authors. There are three *languages*<sup>5</sup> in Lhagang Town: Khams (Tibetic; Lhagang Village and the westernmost area), Amdo (Tibetic; central area of Lhagang Town except for Lhagang Village), and Lhagang Choyu (Qiangic; the westernmost area). This figure displays that Amdo is distributed the most widely. The geographical point of Shingnyag Hamlet is given in a specific symbol. The location of Shingnyag Hamlet is far from Lhagang Village, and relatively close to another Khams Tibetan-spoken hamlet called Tage [Thabs-mkhas], where Lhagang Choyu is also distributed.<sup>6</sup> Shingnyag Tibetan, dealt with in this article, is, in fact, excluded from the two varieties of Amdo reported by Suzuki and Sonam Wangmo (2015b), and it should be regarded as a new variety of Amdo Tibetan recognised in Lhagang Town.



Fig. 2 Distribution of regiolects in Lhagang Town.<sup>7</sup>

According to the folk tradition, the element *nyag* in Shingnyag is related to that in Nyagchu (Yalongjiang [Nyag-chu] River) or Nyagrong (Xinlong [Nyag-rong] County), however, it awaits confirmation. This means that ancestors of Tibetans living in Shingnyag Hamlet might have immigrated from a certain place along the Yalongjiang

<sup>5</sup> As for the 'Tibetan language', we follow the definition of 'Tibetic languages' coined by Tournadre (2014).

<sup>6</sup> However, Lhagang Choyu has hardly functioned as a communication language any more. See Suzuki and Sonam Wangmo (2016b).

<sup>7</sup> Figure 2 is designed with the online geocoding system provided at the following site: <http://ktgis.net/gcode/lonlatmapping.html> (accessed on 29th March 2016).

River, located at the west to Lhagang Town (present Yajiang [Nyag-chu-kha] and Xinlong counties). Amdo-speakers certainly live in the mountainous area along Yalongjiang, especially around the border area between Litang [Li-thang] and Xinlong counties. According to the oral history narrated by pastoralists in Litang, their ancestors belong to the *Wa shul*<sup>8</sup> tribe, whose originally inhabiting place was around the present Seda [gSer-rta] County. If the relation between Shingnyag Tibetans and Litang pastoralists is confirmed, the language spoken by Shingnyag Tibetans will be related to that of the *Wa shul* tribe, which is different from the varieties spoken in Gongrima and gYukhyim hamlets within Lhagang Town.

## 2. Phonological overview of Shingnyag Tibetan

### 2.1. Phonemes and suprasegmentals

The phonological inventory, following the *pandialectal phonetic transcription* defined in Tournadre and Suzuki (forthcoming), adapted to the description of this article is as follows:

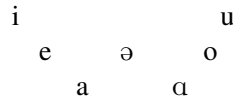
#### Consonantism

		lab.	alveo.	retr.	palatal		vel.	uvul.	glot.
					pre-	post-			
plosive	asp.	p <sup>h</sup>	t <sup>h</sup>	t̪ <sup>h</sup>		c <sup>h</sup>	k <sup>h</sup>	q <sup>h</sup>	
	vl.	p	t	t̪		c	k	q	ʔ
	vd.	b	d	d̪		ɟ	g	ɢ	
affricate	asp.		ts <sup>h</sup>		tɕ <sup>h</sup>				
	vl.		ts		tɕ				
	vd.		dz		dʒ				
fricative	asp.		s <sup>h</sup>	ʂ <sup>h</sup>	ɕ <sup>h</sup>		x <sup>h</sup>		
	vl.	ɸ	s	ʂ	ɕ		x	χ	h
	vd.		z		ʒ		ɣ	ʁ	ɦ
nasal	vd.	m	n		ɳ		ŋ		
	vl.	m̥	n̥		ɳ̥		ŋ̥		
liquid	vd.		l	r					
	vl.		l̥						
semivowel	vd.	w, ʋ				j			

(asp.: voiceless aspirated; vl.: voiceless non-aspirated; vd.: voiced; lab.: labial; alveo.: alveolar; retr.: retroflex; vel.: velar; uvul.: uvular; glot.: glottal)

<sup>8</sup> Generally pronounced as /ɣa x<sup>h</sup>u/ in Lithang Amdo.

## Vocalism



No contrasts between ‘short’ and ‘long’, and ‘plain’ and ‘nasalised’ are attested.

## Suprasegmentals

No suprasegmental distinctions are attested as in the majority of Amdo Tibetan varieties.<sup>9</sup> Stress, realised as a higher pitch and a clearer articulatory gesture of the initial, on the first syllable of a polysyllabic word is often prominent as a prosodic feature.

### 2.2. Phonotactics

The phonotactics of Shingnyag Tibetan is a little complicated. A maximum sound structure of a syllable is described by using the manner of Suzuki (2005) as follows:

$${}^{CC}C_iGVC$$

However, it is rare that a double preinitial ( ${}^{CC}$ ) appears; hence an ordinary sound structure should be  ${}^CC_iGVC$ . A minimum structure is  $C_iV$ .

All the consonant phonemes are able to appear at the position of main initial ( $C_i$ ) except for /ɦ/. A preinitial ( ${}^C$ ) may be prenasalised elements (i.e., homorganic nasal), bilabial nasals ( ${}^m, {}^ɱ$ ), bilabial stops ( ${}^p, {}^b$ ), bilabial continuants ( ${}^ɸ, {}^w$ ), alveolar liquids ( ${}^l, {}^ɭ$ ), retroflex continuants ( ${}^ʂ, {}^ɻ$ ), velar fricative ( ${}^x$ ), or preaspirations ( ${}^h, {}^{ɦ}$ ). A double preinitial ( ${}^{CC}$ ) may be a combination of a prenasalisation/preaspiration and others within the elements listed above. A final ( $C$ ) can be occupied by one of /p, t, k, q, m, n, ŋ, l, r/.

## 3. Principal phonological features of Shingnyag Tibetan

We will list up principal phonological features characterising Shingnyag Tibetan below from a typological viewpoint.<sup>10</sup>

- Uvular sounds

<sup>9</sup> Wang (2012) reports that there is a variety called rMa-stod which possesses a suprasegmental distinction (pitch-tone).

<sup>10</sup> A dialectological analysis of Tibetic languages requires a diachronic view by comparing the synchronic data with Written Tibetan (WrT) forms. However, the discussion provided here concentrates on a synchronic description. An exhaustive description of a comparison with WrT will be separately provided in another article.

Shingnyag Tibetan has three uvular plosives /q<sup>h</sup>, q, ɢ/ and two uvular fricatives /χ, ʁ/. The existence of uvular plosives in a Tibetic language is noted by many scholars, such as Huang (2012) and Wang (2012).

Examples of uvular plosives are following:

/q<sup>h</sup>a ta/ ‘crow’, /q<sup>h</sup>əp/ ‘needle’, /qwar k<sup>h</sup>oŋ/ ‘window’, /qap/ ‘hide (oneself)’, /<sup>N</sup>go/ ‘ride’, etc.

Examples of uvular fricatives are following:

/p<sup>h</sup>ə ʁa/ ‘shoulder’, /ʁa/ ‘fox’, /rə ʁoŋ/ ‘rabbit’, etc.

- Labiodental semivowel

A labiodental semivowel /ʋ/ is a typologically rare sound within the varieties of Amdo Tibetan. Examples are following:

/fi<sup>h</sup>vin ŋgo/ ‘knee’, /m<sup>h</sup>tɕə<sup>h</sup> ʋəq/ ‘lip’, etc.

- Labial/labiodental sounds

A labial fricative /ɸ/ as a main initial rarely appears. However, as a preinitial, it appears more, and its voicing counterpart, labial semivowel /w/, also appears frequently. /ɸ/ and /p/ as a preinitial are distinctive in a few cases. Examples are following:

/ɸu ts<sup>h</sup>əp/ ‘chili/Sichuan pepper’, /ɸa ro/ ‘over there’;

/ɸ<sup>h</sup>tən/ ‘cloud’ (cf. /p<sup>h</sup>təŋ/ ‘breast’), /ɸsə/ ‘body hair’;

/w<sup>h</sup>la/ ‘thigh’, /w<sup>h</sup>ri/ ‘snake’, etc.

- Vowel contrast between /a/ and /ɑ/

Shingnyag Tibetan has a contrast between two vowels /a/ and /ɑ/, which can form a minimal pair. Examples are following:

/k<sup>h</sup>a/ ‘mouth’ - /k<sup>h</sup>ɑ/ ‘snow’;

/ɸ<sup>h</sup>kwa/ ‘order’ - /ɸ<sup>h</sup>kwa/ ‘portion’, etc.

- Contrast between /<sup>l</sup>C<sub>i</sub>/ and /<sup>r</sup>C<sub>i</sub>/

This feature certainly characterises Shingnyag Tibetan. The actual difference of phonetic realisations between these two combinations is not regarding the articulation of the preinitial ([l, l̥] or [r, r̥]), but regarding its articulatory manner: the preinitial of /<sup>l</sup>C<sub>i</sub>/ is pronounced with a contact of the tongue tip to the alveolar, while that of /<sup>r</sup>C<sub>i</sub>/, without a contact of the tongue tip to the alveolar.<sup>11</sup> Examples are following:

/l̥ŋa/ ‘five’ - /r̥ŋa/ ‘drum’;

/l̥tɕəq/ ‘iron’ - /<sup>ʁ</sup>na t̥tɕə/ ‘nose ring’ - /<sup>ʁ</sup>tɕa/ ‘hair’;

/l̥ta/ ‘look’ - /m̥tɕ<sup>h</sup>o t̥ten/ ‘stūpa’ - /<sup>ʁ</sup>ta/ ‘horse’, etc.

- Double preinitials (<sup>CC</sup>)

A double preinitial (<sup>CC</sup>) rarely appears. The preinitial part is a combination of either preaspiration + a labial plosive or prenasalisation + a labial plosive.

<sup>11</sup> The preinitial /<sup>l̥</sup>/ can be distinguished from /<sup>ʁ</sup>/, however, an accoustic analysis will be required for this phenomenon.



Examples are following:

/<sup>fi</sup>b dɔŋ/ ‘sugar’, /<sup>mp</sup>t<sup>h</sup>ə ŋa/ ‘beads’, etc.

### Word list of Shingnyag Tibetan

The entries of vocabulary are ordered following Hua ed. (2002), in which selected items are described here, as: Nouns (classified into several semantic categories, such as Astrology and Geology, Body, Person, Animals, Plants, Food, Clothings, Housing, Instrumentals, Cultural objects, Space and Time), Numbers, Pronouns, Adjectives, and Verbs.<sup>12</sup> A small number of words excluded in Hua ed. (2002) are also added. Evident loanwords are indicated by † following a word with a footnote.

#### Astrology and Geology

sky	<sup>fi</sup> nam	dust	t <sup>h</sup> ɑ
sun	ŋə ma	mud	<sup>n</sup> da m <sup>b</sup> əq
light	wet / wot	water	tɕ <sup>h</sup> ə
moon	<sup>fi</sup> dza / <sup>fi</sup> dza wa	forest	nəq / nəq ts <sup>h</sup> əl
star	<sup>h</sup> kar ma	glassland	<sup>s</sup> tsa t <sup>h</sup> əŋ
cloud	ɕtən	gold	<sup>x</sup> ser
thunder	t <sup>h</sup> oq / <sup>m</sup> dək	silver	<sup>s</sup> ŋi
wind	<sup>ʋ</sup> loŋ / <sup>ʋ</sup> loŋ k <sup>h</sup> a	copper	rəq
rain	<sup>fi</sup> nam / tɕ <sup>h</sup> ar / tɕ <sup>h</sup> a rwa	iron	<sup>l</sup> tɕəq
rainbow	<sup>ɳ</sup> dza	rust	<sup>h</sup> tsa
snow	k <sup>h</sup> ɑ	coal	s <sup>h</sup> u ki
hail	s <sup>h</sup> e ra	ash	t <sup>h</sup> ɑ
ice	tɕ <sup>h</sup> ɑ rəm	place	s <sup>h</sup> a tɕ <sup>h</sup> a
fire	<sup>m</sup> ŋe	Lhasa	la s <sup>h</sup> a
smoke	to fia	Khams	k <sup>h</sup> am ba s <sup>h</sup> a tɕ <sup>h</sup> a
earth	s <sup>h</sup> a	Amdo	ʔa m <sup>do</sup>
mountain	rə	Lhagang Monastery	la r <sup>g</sup> əŋ <sup>fi</sup> gon ba
river	tɕ <sup>h</sup> ə	town	r <sup>dz</sup> oŋ
lake	<sup>m</sup> ts <sup>h</sup> o	street	toŋ nəŋ
road	lam	village	<sup>fi</sup> de wa
soil	s <sup>h</sup> a	bridge	zam ba
stone	r <sup>do</sup>	homeland	ha jəl
sand	ɕɕe ma		

<sup>12</sup> Several verbs might have a stem alternation, which, however, does not regularly appear any more. Another article will be needed for a detailed description regarding this issue.

## Body

body	<sup>fi</sup> zu xu	index finger	koŋ <sup>m</sup> dzək
head	<sup>m</sup> go	middle finger	<sup>h</sup> ci <sup>m</sup> dzək
hair	<sup>s</sup> tɕa	small finger	<sup>m</sup> dzək tɕ <sup>h</sup> oŋ ma
front	t <sup>h</sup> o pa	nail	s <sup>h</sup> en mo
eyebrow	<sup>fi</sup> dzə ma	fist	k <sup>h</sup> ə ts <sup>h</sup> ər
eyelash	ɣ <sup>ni</sup> <sup>ɸ</sup> sə	skin	ʔo x <sup>h</sup> a / ləq <sup>fi</sup> jep
eye	ɣ <sup>ni</sup> k	wrinkle	<sup>fi</sup> ge <sup>fi</sup> ɲer
nose	na	mole	nə xo
ear	<sup>r</sup> nɔ	sore	<sup>r</sup> ma
face	k <sup>h</sup> a ŋo	wound	<sup>r</sup> ma
mouth	k <sup>h</sup> a	blood	c <sup>h</sup> əq
lip	<sup>m</sup> tɕ <sup>h</sup> ə vəq	pulse	<sup>s</sup> tsa
moustache	k <sup>h</sup> a <sup>ɸ</sup> sə	brain	<sup>fi</sup> la pa
beard	<sup>fi</sup> ja <sup>ɸ</sup> sə re	bone	rə pa
jaw	ma ne	joint	ts <sup>h</sup> əq t <sup>h</sup> o
neck	<sup>s</sup> ke	tooth	s <sup>h</sup> o
shoulder	p <sup>t</sup> <sup>h</sup> əq qa / p <sup>t</sup> <sup>h</sup> ə ɛa	tongue	<sup>l</sup> tɕe
back	na <sup>n</sup> təq	throat	<sup>m</sup> ni p <sup>h</sup> a
armpit	ts <sup>h</sup> ək k <sup>h</sup> əŋ	lung	<sup>fi</sup> lwa
breast	p <sup>t</sup> əŋ	heart	ŋəŋ
milk	ɣo ma	liver	<sup>m</sup> tɕ <sup>h</sup> ən ba
belly	hwa	kidney	<sup>m</sup> k <sup>h</sup> wa ma
navel	<sup>h</sup> te ja	gall bladder	<sup>m</sup> c <sup>h</sup> ə <sup>ɸ</sup> sa
waist	<sup>r</sup> jəp	stomach	p <sup>t</sup> əŋ <sup>13</sup>
buttocks	tam <sup>fi</sup> go	intestine	<sup>fi</sup> jə ma
thigh	<sup>w</sup> la	excrement	<sup>h</sup> cə xa
knee	<sup>fi</sup> vin ŋgo	urine	<sup>s</sup> tɕən
leg	na ji	fart	<sup>fi</sup> jə
foot	<sup>s</sup> ka ŋa	sweat	<sup>fi</sup> ŋi tɕ <sup>h</sup> ə
elbow	tɕə <sup>m</sup> dzək	sputum	tɕ <sup>h</sup> ə ma
hand	ləq pa / la xa	nasal mucus	<sup>s</sup> ŋəp
arm	ts <sup>h</sup> ək t <sup>h</sup> o	tear	ɣ <sup>ni</sup> k tɕ <sup>h</sup> ə
finger	<sup>m</sup> dzə ɣə	voice	<sup>s</sup> kat
palm	ləq <sup>m</sup> t <sup>hi</sup> <sup>h</sup> ci <sup>h</sup> kə	life	<sup>s</sup> hoq
thumb	<sup>m</sup> dzək tɕ <sup>h</sup> en ma		

<sup>13</sup> This word form corresponds to WrT *brang* 'breast'.

## Person

person   <sup>m</sup>ṇə  
 Tibetan   wo pa  
 Han Chinese   <sup>r</sup>ja  
 child   za ʒi / zə lə  
 old man   <sup>fi</sup>ga <sup>r</sup>ge  
 man   zi lə  
 woman   zi mo  
 doctor   <sup>o</sup>man ba  
 military   <sup>fi</sup>məq  
 pastoralist   <sup>m</sup>dzo ʒa  
 beggar   həq <sup>fi</sup>jo  
 friend   ro ʒwa  
 teacher   <sup>r</sup>ge <sup>r</sup>gen  
 lunatic   ṇon ba  
 grandfather   ʔa <sup>m</sup>ṇe  
 grandmother   ʔa ji  
 father   ʔa p<sup>h</sup>a  
 mother   ʔa ma  
 parents   ha ma  
 son   zə lə  
 bride   <sup>fi</sup>na ma

daughter   zi mo  
 bridegroom   wo t<sup>h</sup>əŋ  
 grandson   ts<sup>h</sup>a wo  
 granddaughter   ts<sup>h</sup>a mo  
 elder brother   ʔa <sup>r</sup>ja  
 elder sister   ʔa tce  
 younger brother   c<sup>h</sup>a ṇi  
 younger sister   s<sup>h</sup>əŋ mo  
 paternal elder uncle   ʔa k<sup>h</sup>ə  
 paternal elder uncle's wife   ʔa ne  
 paternal younger uncle   ʔa k<sup>h</sup>ə  
 nephew   ts<sup>h</sup>a wo  
 brother   wo <sup>ɸ</sup>sən  
 sister   wo mo <sup>h</sup>pən <sup>u</sup>dji  
 maternal uncle   ʔa zəŋ  
 maternal uncle's wife   ʔa ne  
 family   c<sup>h</sup>əm <sup>fi</sup>jət  
 relatives   ne fia  
 husband   <sup>fi</sup>ga ho  
 wife   <sup>fi</sup>gan mo  
 twins   <sup>m</sup>ts<sup>h</sup>e lə

## Animals

cattle   zoq  
 bull   <sup>fi</sup>ləŋ <sup>m</sup>bi  
 yak   <sup>fi</sup>jəq  
 female yak   <sup>m</sup>də  
 mdzo<sup>14</sup>   <sup>m</sup>dzo  
 female mdzo   <sup>m</sup>dzo mo  
 calf   wi li / wi  
 ox   ʔa <sup>r</sup>qo  
 cow   mo zoq  
 milk cow   <sup>m</sup>də

cattle's wet dung   <sup>l</sup>təwa  
 horn   ra tce  
 skin   kwa / wə ʒa  
 body hair   <sup>ɸ</sup>sə  
 tail   <sup>r</sup>ja ma  
 horse   s<sup>h</sup>ta  
 pony   <sup>h</sup>ti <sup>h</sup>tsi  
 stallion   <sup>h</sup>ta ho  
 mare   <sup>fi</sup>gon ma  
 sheep   lək

<sup>14</sup> A hybrid of ox and female yak.

ewe       lək ma  
 goat       ra ma  
 baby goat   ri<sup>h</sup>tsi  
 lamb       lə ɣə  
 wool       lək wɑ / wɑ  
 sheep's dung   ri ma  
 mule       p<sup>t</sup>wi  
 donkey     kə rə  
 pig        həq  
 sow        həq mo  
 boar       ho həq  
 piglet     həq p<sup>t</sup>hək  
 dog        c<sup>h</sup>ə  
 bitch      c<sup>h</sup>ə mo  
 cat        mo<sup>n</sup>dzə  
 rabbit     rə ʋəŋ  
 chicken    ɸɕa  
 cock       ɸɕa p<sup>h</sup>o  
 hen        ɸɕa mo  
 wing       <sup>h</sup>ɕo ɣa  
 down      ɸɕa<sup>fi</sup>ɖo / <sup>fi</sup>ɖo  
 duck      tɕ<sup>h</sup>ə ɸɕa  
 pigeon    mə<sup>h</sup>kə  
 tiger      <sup>h</sup>təq  
 lion       s<sup>h</sup>ɛŋ gi  
 dragon    m<sup>d</sup>ɖək  
 claw      <sup>fi</sup>rwa wa

monkey    ʔa<sup>r</sup>ge  
 bear       tom  
 deer       ɕ<sup>h</sup>wa  
 mouse     tsə ɣə  
 wolf       ɸswəŋ k<sup>h</sup>ə  
 fox        ʋa  
 bird       ɸɕa  
 eagle      <sup>fi</sup>ləq  
 vulture    ɸɕa<sup>r</sup>got  
 swallow   k<sup>h</sup>əŋ ɸɕi  
 sparrow    ɸɕi<sup>h</sup>tse  
 crow       q<sup>h</sup>a ta  
 peacock   <sup>fi</sup>ma ɸɕa  
 snake      <sup>w</sup>ri  
 frog       <sup>w</sup>rwa wa / <sup>w</sup>rwa<sup>n</sup>ʔə  
 fish       ɳa  
 insect     m<sup>b</sup>bə  
 flea       <sup>l</sup>dɖwa  
 louse      ɕ<sup>h</sup>ik  
 fly        ɕ<sup>h</sup>a m<sup>b</sup>bə  
 mosquito   <sup>h</sup>tsa m<sup>b</sup>bə  
 spider     m<sup>b</sup>bə qa ra / tɕoq qa ra  
 centipede   <sup>h</sup>kə məŋ m<sup>b</sup>bə  
 ant        coq ma  
 bee        <sup>fi</sup>wəŋ ma  
 butterfly   ɸɕ<sup>h</sup>e ma la<sup>p</sup>tse

### Plants

tree       <sup>l</sup>tɕəŋ ma<sup>15</sup>  
 root       <sup>h</sup>tswa  
 leaf       lo ma  
 flower     me toq  
 conifer    ɕ<sup>h</sup>ə ɣa  
 pine       <sup>fi</sup>ɖon ma  
 bamboo    ɳik ma  
 throne     ts<sup>h</sup>er ma

fruit       ɕ<sup>h</sup>əŋ t<sup>h</sup>oq  
 peach      k<sup>h</sup>am bə  
 lotus flower   pan ma me toq  
 saffron     kəŋ gəm me toq  
 walnut     <sup>h</sup>tar ga  
 crop       lo t<sup>h</sup>oq  
 foodstuff   m<sup>d</sup>ɖə rik  
 rice       m<sup>d</sup>ɖe

<sup>15</sup> This word form corresponds to WrT *lcang ma* 'willow'.

seed	s <sup>h</sup> a wən	edible seed	kwa tsə / kə <sup>m</sup> də
wheat	co	pea	s <sup>h</sup> an ma
highland barley	ne	grass	s <sup>h</sup> tsa
chili	φu ts <sup>h</sup> əp	mushroom	ç <sup>h</sup> a mo
garlic	f <sup>i</sup> go χwa	pteridophyte root	co ma
ginger	s <sup>h</sup> təa <sup>r</sup> ga		

## Food

food	<sup>m</sup> də / za ma / <sup>n</sup> du <sup>†16</sup>	cheese	tç <sup>h</sup> ə ra
porridge	<sup>m</sup> də ts <sup>h</sup> a	first milk after the birth	zo çə
wheat flower	tçəq f <sup>i</sup> dze	tsampa <sup>17</sup>	s <sup>h</sup> tsam ba
steamed bun	f <sup>i</sup> dzor na	beef	zoq ç <sup>h</sup> a
noodle	pə <sup>h</sup> təq	salt	ts <sup>h</sup> a
steamed stuffed bun	poq tsə	sugar	f <sup>i</sup> b <sup>i</sup> dəŋ
breakfast	nəŋ tç <sup>h</sup> ə	Sichuan pepper	φu ts <sup>h</sup> əp
lunch	tçə tçə	egg	r <sup>h</sup> goŋ wa
dinner	nəp tç <sup>h</sup> ə	soup	k <sup>h</sup> wa
milk tea	γo tçə	alcohol	tç <sup>h</sup> əŋ
meat	ç <sup>h</sup> a	hot water	tç <sup>h</sup> ə ts <sup>h</sup> a
lean meat	ç <sup>h</sup> a r <sup>h</sup> əq	tea	tçə
oil	s <sup>h</sup> əmə	tobacco	tə wa
salad oil	mar nəq	medicine	mən
fat oil	ts <sup>h</sup> i	snuff	ŋə tə
butter	mar		
yoghurt	zo		

## Clothings

thread	<sup>h</sup> kə pa	button	f <sup>i</sup> jo xo
cloth	re	trousers	tor ma
monk's cloths	çan t <sup>h</sup> əp	hat	za
cloths	ko ze	belt	tor t <sup>h</sup> əq / <sup>h</sup> ke rəq
Tibetan woolen cloths	çə sə	shoe	xaj <sup>†18</sup>
collar	nəŋ koŋ	comb	<sup>h</sup> tə p <sup>h</sup> ç <sup>h</sup> ət
sleeve	φç <sup>h</sup> ə koŋ	jewel	nor wə

<sup>16</sup> Loan from Lhagang Choyu /<sup>n</sup>du/ 'food'. See Suzuki and Sonam Wangmo (2016b).

<sup>17</sup> It just denotes 'fried highland barley flour' in Shingnyag.

<sup>18</sup> Loan from Sichuanese (Southwest Mandarin) xai 鞋 'shoe'.

coral     $\phi\epsilon$  rə  
 turquoise     $^{\text{fi}}\text{jə}$   
 pearl    mə t<sup>h</sup>ək  
 ivory    pa s<sup>h</sup>o  
 amber     $^{\text{h}}\text{pu}$  ɕ<sup>h</sup>i

earring     $^{\text{fi}}\text{nə}$  loŋ  
 necklace     $^{\text{h}}\text{kə}$  r<sup>h</sup>jet  
 ring     $^{\text{m}}\text{dzəp}$   $^{\text{h}}\text{ci}$   
 bracelet    ləq  $^{\text{h}}\text{ci}$

## Housing

blanket     $^{\text{h}}\text{tan}$   
 pillow     $^{\text{m}}\text{go}$  r<sup>h</sup>əe  
 cushion     $^{\text{h}}\text{tan}$   
 house    k<sup>h</sup>ə  $^{\text{m}}\text{ba}$   
 roof    q<sup>h</sup>əŋ t<sup>h</sup>oq  
 inn     $^{\text{m}}\text{dɔn}$  q<sup>h</sup>əŋ  
 kitchen    tɕa q<sup>h</sup>əŋ  
 storied house    ts<sup>h</sup>ək t<sup>h</sup>oq  
 upstairs    q<sup>h</sup>əŋ t<sup>h</sup>oq  
 downstairs    q<sup>h</sup>əŋ joq  
 store house     $^{\text{m}}\text{dzot}$  q<sup>h</sup>əŋ  
 cowshed    nor q<sup>h</sup>əŋ / zoq q<sup>h</sup>əŋ  
 pigsty    həq q<sup>h</sup>əŋ  
 horse fence    s<sup>h</sup>ta q<sup>h</sup>əŋ  
 sheepfold    lək ra  
 chicken coop     $\phi\epsilon$ a ts<sup>h</sup>əŋ

wall     $^{\text{h}}\text{tsə}$  ɣa  
 log    ɕ<sup>h</sup>aŋ  
 plank    ɕ<sup>h</sup>aŋ ləp  
 pillar    kwa  
 gate     $^{\text{fi}}\text{go}$   
 threshold     $^{\text{fi}}\text{go}$  t<sup>h</sup>i  
 entrance door     $^{\text{fi}}\text{go}$  tɕ<sup>h</sup>in  
 window    qwar k<sup>h</sup>oŋ  
 stair     $^{\text{h}}\text{ki}$   
 beam     $^{\text{fi}}\text{doŋ}$  ma  
 step    r<sup>h</sup>do s<sup>h</sup>ki  
 tent    kər  
 yak-woolen tent    <sup>w</sup>ra / <sup>w</sup>rwa  
 garden    rwa / ra wa  
 toilets    tɕ<sup>h</sup>ap q<sup>h</sup>əŋ

## Instrumentals

thing    tɕa k<sup>h</sup>a  
 table    s<sup>h</sup>tɕoq tse  
 chair    s<sup>h</sup>kəp t<sup>h</sup>əq  
 bed    ɲa t<sup>h</sup>ə  
 box    r<sup>h</sup>gam  
 cabinet    tɕ<sup>h</sup>wa  
 glass    ɕ<sup>h</sup>el  
 mirror    ɕ<sup>h</sup>e r<sup>h</sup>go  
 bloom     $^{\text{m}}\text{ɕ<sup>h</sup>əq}$  ma  
 light     $^{\text{fi}}\text{luq}$   
 candle    p<sup>h</sup>twa  $^{\text{fi}}\text{dɔn}$

firewood     $^{\text{m}}\text{bə}$  tɕi  
 coal    s<sup>h</sup>u ki  
 flint     $^{\text{m}}\text{ɲe}$  r<sup>h</sup>do  
 match    jaŋ xwo<sup>†19</sup>  
 torch     $^{\text{m}}\text{ɲe}$  ɕ<sup>h</sup>u  
 incense     $^{\text{h}}\text{pu}$   
 garbage     $\phi\epsilon$ <sup>h</sup>ə r<sup>h</sup>ge  
 cooking stove    t<sup>h</sup>ap k<sup>h</sup>a  
 iron pan    t<sup>h</sup>o  
 frying pan    t<sup>h</sup>o  
 steamer     $^{\text{fi}}\text{lo}$  ɲi

<sup>19</sup> Loan from Sichuanese (Southwest Mandarin) yanghuo 洋火 ‘match’.

lid	k <sup>h</sup> a ləp	rein	ʃ <sup>h</sup> am m <sup>h</sup> da
knife	cə	whip	l <sup>h</sup> təq <sup>h</sup> tse
ladle	ʃ <sup>h</sup> coq	nose ring	ʃ <sup>h</sup> na t <sup>h</sup> təq
spoon	k <sup>h</sup> ək <sup>h</sup> f <sup>h</sup> de	glasses	ɕ <sup>h</sup> e m <sup>h</sup> ɲik
gourd ladle	ʃ <sup>h</sup> coq	oxhide string	m <sup>h</sup> dəŋ na
wooden bowl	ɕ <sup>h</sup> aŋ tɕe	ship	tɕə zan
bowl	tɕa ɲe	airplane	f <sup>h</sup> nam ɕ <sup>h</sup> ca
dish	f <sup>h</sup> der ma	bicycle	l <sup>h</sup> təq <sup>h</sup> t <sup>h</sup> ta
chopsticks	za t <sup>h</sup> ər	instrument	tɕ <sup>h</sup> ər tɕ <sup>h</sup> a
bottle	tam bi	axe	ʃ <sup>h</sup> ta re
pot	r <sup>h</sup> dza ma	hammer	t <sup>h</sup> ɑ
jar	r <sup>h</sup> dza ma	chisel	m <sup>h</sup> bək
kettle	tem / ts <sup>h</sup> a tem	saw	ʃ <sup>h</sup> oq le
bucket	tɕ <sup>h</sup> ə ɣ <sup>h</sup> zi	plough	ʃ <sup>h</sup> ko ma
wooden tray	f <sup>h</sup> zəŋ ma	leather bag	f <sup>h</sup> ɲwa
basket	pi t <sup>h</sup> u	carrying pole	p <sup>h</sup> t <sup>h</sup> əq ɕ <sup>h</sup> aŋ
tripet	ʃ <sup>h</sup> kəŋ <sup>h</sup> səm	handle	jə fia
leather bellows	k <sup>h</sup> u mo	grip	r <sup>h</sup> dza loŋ
suspender	h <sup>h</sup> kər t <sup>h</sup> əq	rope	t <sup>h</sup> ə Ɂa
steelyard	r <sup>h</sup> ja ma	fertiliser	lət
money	ta jin <sup>†20</sup>	sickle	zo ra
ruler	t <sup>h</sup> ə tsə	sieve	ts <sup>h</sup> əq
needle	q <sup>h</sup> əp	millstone	roŋ n <sup>h</sup> dəq
nale	l <sup>h</sup> təq <sup>h</sup> n <sup>h</sup> dzer	loom	zo t <sup>h</sup> ə
scissors	tsən təq	conch	h <sup>h</sup> pe
ladder	h <sup>h</sup> ki	lance	m <sup>h</sup> dəŋ
umbrella	ɕ <sup>h</sup> oq f <sup>h</sup> dək	sheath	tɕə ɕ <sup>h</sup> əp
lock	f <sup>h</sup> go l <sup>h</sup> təq <sup>h</sup>	gun	wu
key	l <sup>h</sup> de m <sup>h</sup> ɲik	bullet	m <sup>h</sup> de ɣə
wheel	ʃ <sup>h</sup> k <sup>h</sup> or lo	arrow	m <sup>h</sup> da
rod	f <sup>h</sup> ji ɣa	poison	tək
saddle	ʃ <sup>h</sup> ta r <sup>h</sup> ga	net	tɑ
halter	m <sup>h</sup> t <sup>h</sup> ər f <sup>h</sup> go	present	xə m <sup>h</sup> bə
belly band	f <sup>h</sup> lo	notebook	tep
bit	ʃ <sup>h</sup> ap	quilt	p <sup>h</sup> u f <sup>h</sup> gi <sup>†21</sup>
stirrup	jəp tɕ <sup>h</sup> en	matress	t <sup>h</sup> an dzə
horseshoe	r <sup>h</sup> mik l <sup>h</sup> təq <sup>h</sup>		
manger	k <sup>h</sup> a ts <sup>h</sup> əq		

<sup>20</sup> Perhaps loan from Sichuanese (Southwest Mandarin) dayin 大銀 or dayang 大洋 'money, cash'.

<sup>21</sup> Loan from Sichuanese (Southwest Mandarin) beigai 被盖 'quilt'.

## Cultural objects

script	ji ɣe	Buddha	s <sup>h</sup> əŋ f <sup>i</sup> dzi
letter	ji ɣe	soul	r <sup>i</sup> nam ɕ <sup>h</sup> i
alphabet	<sup>h</sup> sa βzɪt	spirit	<sup>w</sup> la
picture	rə mo	incarnation	<sup>n</sup> di
book	ji ɣe	next life	ts <sup>h</sup> e ɕ <sup>h</sup> ɕ <sup>h</sup> e ma
paper	ɕ <sup>h</sup> o ɸo	reincarnation	ɸ <sup>i</sup> k <sup>h</sup> or wa
pen	ɸə ɣə	fortune	le lam
ink	s <sup>n</sup> əq ts <sup>h</sup> a	destiny	le
school	ɕwo t <sup>h</sup> ɑŋ	charity	f <sup>i</sup> ge wa
knowledge	jon den	evil omen	<sup>l</sup> ti ŋen
talk	s <sup>n</sup> ket ɕ <sup>h</sup> a	lama	<sup>w</sup> la ma
Spoken Tibetan	wot f <sup>i</sup> ket	reincarnated lama	<sup>h</sup> təl <sup>h</sup> kə
Written Tibetan	wo jik	abbot	<sup>m</sup> k <sup>h</sup> wan bo
name	<sup>m</sup> ɸaŋ	monk	t <sup>h</sup> wa / t <sup>h</sup> wa wa
family name	ri <sup>m</sup> ɸaŋ	nun	tɕo mo
sign	s <sup>n</sup> təq	housekeeper	f <sup>i</sup> ɸer wa
newspaper	ts <sup>h</sup> əq <sup>h</sup> pər	monk resuming secular life	<sup>h</sup> ka loq
story	f <sup>i</sup> na rək	donor	<sup>w</sup> zen b <sup>d</sup> əq
proverb	<sup>h</sup> tam ɣwe	donation of tea	məŋ tɕa
joke	k <sup>h</sup> a <sup>m</sup> ts <sup>h</sup> ər	sorcerer	ʔa <sup>m</sup> ɕ <sup>h</sup> ot
riddle	ʔa ke mə ket	fortune teller	mo wa
voice	f <sup>i</sup> də	hell	<sup>m</sup> ɸal wa
song	f <sup>i</sup> lə	monastery	f <sup>i</sup> gon ba
dance	p <sup>t</sup> o	congregation	lə q <sup>h</sup> əŋ
drum	r <sup>i</sup> ɸa	religious institution	t <sup>h</sup> a ts <sup>h</sup> əŋ
flute	f <sup>i</sup> dzwa f <sup>i</sup> lon	meditation cell	<sup>m</sup> ts <sup>h</sup> am q <sup>h</sup> əŋ
bell	tə lwa	maṇi	ma ne jik tək
trumpet	toŋ / ja loŋ	burning incense	ɕsəŋ
thangka, Buddhist drawing	t <sup>h</sup> əŋ k <sup>h</sup> a	stūpa	<sup>m</sup> ɕ <sup>h</sup> o f <sup>i</sup> ten
mask	<sup>m</sup> bəq / <sup>m</sup> go	Buddha statue	lə f <sup>i</sup> kə
religion	tɕ <sup>h</sup> u lək	butter lamp	<sup>m</sup> ɕ <sup>h</sup> o mar
belief	ta pa	ceremonial scarf	k <sup>h</sup> a tar
deity	lə	amulet box	ɕsəwəŋ <sup>h</sup> kor
female deity	lə mo	release of animals	ts <sup>h</sup> e t <sup>h</sup> ar
ghost	<sup>n</sup> də	mantra	mə ŋen
female ghost	<sup>n</sup> də mo	sūtra	tɕ <sup>h</sup> u
evil	b <sup>d</sup> ət	beads	<sup>m</sup> p <sup>t</sup> <sup>h</sup> ə ŋa
Nāga	f <sup>i</sup> lə	vajra	r <sup>i</sup> do r <sup>i</sup> dze



maṇi wheel    ma ne ṽk<sup>h</sup>or lo  
 donation    w<sup>z</sup>ən ba  
 position    ko s<sup>h</sup>a  
 power    f<sup>i</sup>ɛəŋ  
 life    n<sup>ɪ</sup>ts<sup>h</sup>o wa  
 salary    hoq  
 portion    ɸkwa  
 market    ts<sup>h</sup>oŋ ra  
 priority    han t<sup>h</sup>oq  
 cause    r<sup>ɟ</sup>ə m<sup>ɪ</sup>ts<sup>h</sup>en  
 answer    lan  
 famine    mə ge  
 suffering    f<sup>i</sup>duk r<sup>ɟ</sup>jal  
 mistake    no ṽt<sup>h</sup>əl  
 danger    n<sup>ɪ</sup>en k<sup>h</sup>a  
 distinction    c<sup>h</sup>et tɛ<sup>h</sup>u  
 space    h<sup>ɪ</sup>səŋ  
 lucky    p<sup>t</sup>a ɛ<sup>h</sup>i  
 thanks    ɸkwa d<sup>ɪ</sup>n tɛ<sup>h</sup>e  
 crack    ke se  
 trace    ɛ<sup>h</sup>i  
 shadow    cəp s<sup>h</sup>o

colour    m<sup>ɪ</sup>doq qa  
 dream    m<sup>ɪ</sup>ɲi lam  
 spirit    jit ṽk<sup>h</sup>am  
 idea    ɸsam ts<sup>h</sup>əl  
 appearance    w<sup>z</sup>o l<sup>ɪ</sup>ta  
 affair    ton dəq  
 means    w<sup>ɪ</sup>lo t<sup>h</sup>əp  
 strength    ɛ<sup>h</sup>e ɛ<sup>h</sup>ik  
 order    ɸkwa  
 prison    ɸtswən k<sup>h</sup>əŋ  
 rumour    h<sup>ɪ</sup>tam c<sup>h</sup>er  
 sin    ɲi pa  
 bare foot    f<sup>ɪ</sup>kəŋ r<sup>ɟ</sup>dzin  
 drawer    n<sup>ɪ</sup>t<sup>h</sup>em f<sup>i</sup>gam  
 weapon    m<sup>ɪ</sup>ts<sup>h</sup>wən tɛ<sup>h</sup>a  
 victory    r<sup>ɟ</sup>al k<sup>h</sup>a  
 country    r<sup>ɟ</sup>an ṽk<sup>h</sup>am  
 experience    ɲam m<sup>ɪ</sup>ɲoŋ  
 conference    ts<sup>h</sup>oq n<sup>ɪ</sup>də  
 distance    war t<sup>h</sup>əq  
 walking    f<sup>ɪ</sup>kəŋ n<sup>ɪ</sup>ə

### Space and Time

time    ti ts<sup>h</sup>ot  
 today    te raŋ  
 yesterday    k<sup>h</sup>a f<sup>ɪ</sup>tsəŋ  
 day before yesterday    k<sup>h</sup>e ɲin bə  
 three days before    k<sup>h</sup>e z<sup>i</sup> ɲin bə  
 tomorrow    s<sup>h</sup>əŋ ɲin  
 day after tomorrow    f<sup>i</sup>ne ɲin  
 three days later    f<sup>i</sup>z<sup>i</sup> ɲin  
 this evening    to k<sup>h</sup>o  
 tomorrow evening    s<sup>h</sup>a r<sup>ɟ</sup>oŋ  
 last night    m<sup>ɪ</sup>dəŋ r<sup>ɟ</sup>oŋ  
 daytime    ɲin h<sup>ɪ</sup>kər  
 morning    ɲa mo  
 noon    ɲin f<sup>i</sup>goŋ  
 sunset time    h<sup>ɪ</sup>sar ɛ<sup>h</sup>ot

evening    f<sup>i</sup>goŋ mo  
 night    m<sup>ɪ</sup>ts<sup>h</sup>an ki  
 midnight    m<sup>ɪ</sup>ts<sup>h</sup>an ɸɛ<sup>h</sup>it  
 animal of year    lo s<sup>ɪ</sup>ta  
 mouse year    tsə ɣə  
 ox year    zoq  
 tiger year    s<sup>ɪ</sup>təq  
 rabbit year    jor  
 dragon year    m<sup>ɪ</sup>dək  
 snake year    b<sup>ɪ</sup>dəl  
 horse year    s<sup>ɪ</sup>ta  
 sheep year    lək  
 monkey year    p<sup>t</sup>wə ji  
 cock year    ɸɛa  
 dog year    c<sup>h</sup>ə

pig year      p<sup>h</sup>əq  
 day      ts<sup>h</sup>e w<sup>w</sup>za  
 first day      ts<sup>h</sup>e h<sup>h</sup>tɕik  
 second day      ts<sup>h</sup>e ɱi  
 month      ɱi dza  
 in the morning      ɱa mo  
 in the afternoon      ɕɕ<sup>h</sup>ə ʈo  
 January      ɱi da wa təŋ bo  
 February      ɱi da wa ɱi pa  
 December      ɱi da wa p<sup>t</sup>tɕə ɱi pa  
 beginning of a month      ɱi da t<sup>t</sup>ot  
 middle of a month      ɱi dza wa t<sup>t</sup>cit  
 end of a month      ɱi dza wa m<sup>m</sup>dzək  
 birthday      s<sup>s</sup>ci h<sup>h</sup>kər  
 year      lo  
 recently      ɱe rəŋ  
 this year      ta lo  
 last year      na nəŋ  
 last two years      ɱi zi nəŋ lo  
 next year      s<sup>h</sup>əŋ lo  
 next two year      ɱi nəŋ lo

before      ɱi ma ɱi na ɱi na / ɱi na ɱa mo  
 long time ago      ɱi na ti  
 now      ta t<sup>h</sup>əŋ / ta t<sup>l</sup>ta  
 future      ma ɔŋ pa  
 beginning      ɱgo n<sup>n</sup>dzək  
 Monday      ɱi za ɱi da wa  
 Tuesday      ɱi za ɱi mək mar  
 Wednesday      ɱi za pa s<sup>h</sup>əŋ  
 Thursday      ɱi za h<sup>h</sup>pen ba  
 Friday      ɱi za p<sup>h</sup>ər wə  
 Saturday      ɱi za ləq pa  
 Sunday      ɱi za ɱi ma  
 spring      s<sup>s</sup>tɕət ka  
 summer      ɱi jar k<sup>h</sup>a  
 autumn      s<sup>s</sup>ton k<sup>h</sup>a  
 winter      r<sup>r</sup>gən k<sup>h</sup>a  
 new year      lo h<sup>h</sup>sər  
 solar eclipse      ɱə n<sup>n</sup>dzən  
 lunar eclipse      ɱi da n<sup>n</sup>dzən  
 festival      ti tɕ<sup>h</sup>in

## Numbers

1	h <sup>h</sup> tɕik	17	p <sup>t</sup> tɕi b <sup>b</sup> dən
2	ɱi ɱi	18	p <sup>t</sup> tɕu b <sup>b</sup> jet
3	h <sup>h</sup> səm	19	p <sup>t</sup> tɕə r <sup>r</sup> gə
4	w <sup>w</sup> zə	20	ɱi xə
5	l <sup>l</sup> ɱa	21	s <sup>s</sup> tɕa h <sup>h</sup> tɕik
6	ʈək	28	ɱi ɕ <sup>h</sup> ə s <sup>s</sup> tɕa b <sup>b</sup> jet
7	b <sup>b</sup> dən	30	s <sup>h</sup> əm tɕə
8	b <sup>b</sup> jet	32	s <sup>h</sup> əm tɕə s <sup>h</sup> o ɱi
9	ɱi gə	38	s <sup>h</sup> əm tɕə s <sup>h</sup> o b <sup>b</sup> jet
10	p <sup>t</sup> tɕə	40	w <sup>w</sup> zə p <sup>t</sup> tɕə
11	p <sup>t</sup> tɕə h <sup>h</sup> tɕik	43	w <sup>w</sup> zə p <sup>t</sup> tɕə zə h <sup>h</sup> sum
12	p <sup>t</sup> tɕə ɱi	50	l <sup>l</sup> ɱa p <sup>t</sup> tɕə
13	p <sup>t</sup> tɕə h <sup>h</sup> səm	54	l <sup>l</sup> ɱa p <sup>t</sup> tɕə ɱa w <sup>w</sup> zə
14	p <sup>t</sup> tɕi w <sup>w</sup> zə	60	ʈək tɕə
15	p <sup>t</sup> tɕwa ɱa	65	ʈək tɕə re l <sup>l</sup> ɱa
16	p <sup>t</sup> tɕə h <sup>h</sup> ʈək	70	b <sup>b</sup> dən tɕə

76	<sup>b</sup> dən tɕə ton tək
80	<sup>b</sup> ja tɕə
87	<sup>b</sup> ja tɕə ca <sup>b</sup> dən
90	<sup>fi</sup> gə <sup>p</sup> tɕə
98	<sup>fi</sup> gə <sup>p</sup> tɕə ko <sup>b</sup> jet
99	<sup>fi</sup> gə <sup>p</sup> tɕə ko <sup>r</sup> gə
100	<sup>b</sup> ja
101	<sup>b</sup> ja də <sup>h</sup> tɕik
108	<sup>b</sup> ja də <sup>b</sup> jet
880	<sup>b</sup> jet <sup>b</sup> ja <sup>b</sup> ja <sup>p</sup> tɕə

1000	<sup>s</sup> toŋ <sup>h</sup> tɕik
10000	<sup>t</sup> hə <sup>h</sup> tɕik
100000	<sup>m</sup> bəm
1000000	<sup>ɕ</sup> ce wa
10000000	<sup>s</sup> ha ja
100000000	toŋ <sup>ɕ</sup> ce <sup>h</sup> ər
half	<sup>ɕ</sup> ce <sup>h</sup> et
first	təŋ bo
second	<sup>fi</sup> ni pa

### Pronouns

I	ŋa
we two	ŋe <sup>fi</sup> ni
we (exclusive)	ŋa ts <sup>h</sup> o
you (singular)	<sup>c</sup> ho
you (honorific)	<sup>c</sup> het
yow two	<sup>c</sup> he <sup>fi</sup> ni
you (plural)	<sup>c</sup> ho ts <sup>h</sup> o
he / she / it	<sup>k</sup> ho
they two	<sup>k</sup> ho <sup>fi</sup> ni
they	<sup>k</sup> ho ts <sup>h</sup> o
we (inclusive)	ŋa rəŋ ts <sup>h</sup> o
everyone	ta <sup>m</sup> tɕe <sup>h</sup> et ts <sup>h</sup> əŋ ma
self	rəŋ nət
other person	mə <sup>w</sup> zan ba
this	<sup>n</sup> də
these	<sup>n</sup> də ts <sup>h</sup> o
here	<sup>n</sup> də na
here around	<sup>n</sup> də <sup>ɕ</sup> ce <sup>h</sup> oq
these two	<sup>n</sup> də <sup>fi</sup> ni ka
like this	<sup>n</sup> də <sup>n</sup> ɕa

it	te
that	te
those	te ts <sup>h</sup> o
there	ter
over there	ɕa ro / te <sup>ɕ</sup> ce <sup>h</sup> oq
like that	te <sup>h</sup> tar
who	<sup>s</sup> hə
who (plural)	<sup>s</sup> hə <sup>s</sup> hə
what	tɕe <sup>h</sup> ə ze <sup>n</sup> ɕe
where	kəŋ na
when	nam
how	tɕe <sup>h</sup> ə zə ma / kəŋ <sup>n</sup> ɕa
how much	ka ts <sup>h</sup> ot
other	<sup>fi</sup> zen ba
each	rəŋ rəŋ <sup>s</sup> ho <sup>s</sup> ho
whole	<sup>m</sup> t <sup>h</sup> a tək
all	ts <sup>h</sup> əŋ ma
this time	ta t <sup>h</sup> əŋ
someday	<sup>h</sup> kam ŋga re

### Adjectives

big	tɕe
small	tɕe <sup>h</sup> oŋ
wide	<sup>fi</sup> bom
thin	<sup>p</sup> t <sup>h</sup> a
high	<sup>m</sup> t <sup>h</sup> o
low	<sup>fi</sup> ma

protruding    <sup>m</sup>bər <sup>m</sup>bər  
 sunken    <sup>fi</sup>gɔŋ <sup>fi</sup>gɔŋ  
 convex    <sup>m</sup>ba rə <sup>m</sup>bə re  
 long    raŋ wo  
 short    t<sup>h</sup>oŋ  
 far    t<sup>h</sup>əq raŋ  
 near    t<sup>h</sup>əq ɲe  
 middle    <sup>m</sup>dwaŋ ɲa  
 wide    r<sup>ɟ</sup>a tɕ<sup>h</sup>i bo  
 narrow    r<sup>ɟ</sup>a tɕ<sup>h</sup>oŋ tɕ<sup>h</sup>oŋ  
 thick    <sup>n</sup>t<sup>h</sup>ək  
 thin    ʂ<sup>h</sup>ap  
 deep    zap  
 shallow    mə r<sup>ɟ</sup>dəŋ  
 full    kəŋ  
 vacant    <sup>s</sup>toŋ ɲa  
 many    maŋ wo  
 little    ɲoŋ  
 square    t̚ə <sup>w</sup>zə ma  
 round    ɓo ɓo ɕ<sup>h</sup>ə  
 flat    lep lep  
 pointed    <sup>h</sup>tse <sup>m</sup>buk  
 bald    <sup>m</sup>go r<sup>ɟ</sup>do  
 level    p<sup>t</sup><sup>h</sup>a <sup>s</sup>ɲəŋ  
 front    t̚əŋ mo  
 reverse    <sup>l</sup>doq <sup>ɕ</sup>ɕ<sup>h</sup>oq  
 slanting    jor  
 horizontal    <sup>mp</sup>t<sup>h</sup>ət  
 vertical    ru ɲo  
 straight    t̚waŋ bu  
 curved    kək ro  
 much curved    ja re jo re  
 black    na ɓo  
 pitch dark    nəq <sup>fi</sup>dəŋ <sup>fi</sup>dəŋ  
 white    <sup>h</sup>ka ro  
 red    <sup>fi</sup>ma ro  
 brilliant red    <sup>fi</sup>mar ləp ləp  
 yellow    <sup>sh</sup>er po  
 glistering yellow    <sup>sh</sup>er p<sup>t</sup>i p<sup>t</sup>i  
 green    <sup>fi</sup>dzəŋ k<sup>h</sup>ə  
 blue    <sup>s</sup>ɲəŋ bu

skyblue    <sup>s</sup>ɲo r<sup>ɟ</sup>a r<sup>ɟ</sup>a  
 grey    <sup>s</sup>tɕa  
 bright    <sup>h</sup>sa ro  
 shining    wot lam nam  
 dark    mən nəq  
 heavy    <sup>l</sup>dzə mo  
 light (weight)    jəŋ mo  
 fast    <sup>m</sup>jo qa  
 slow    ka le  
 early    <sup>s</sup>ɲa mo  
 late    <sup>ɕ</sup>ɕ<sup>h</sup>ə  
 sharp    <sup>fi</sup>no  
 dull    mə r<sup>ɟ</sup>no  
 clear    t̚əŋ mo  
 muddy    ɲo ɓo  
 fat    ts<sup>h</sup>on bo  
 thin    ɕ<sup>h</sup>a t̚kəm  
 dry    <sup>h</sup>kam bo  
 wet    <sup>fi</sup>lən ba  
 tight    ts<sup>h</sup>əq tam bo  
 thin    ʂ<sup>h</sup>ap ro  
 hard    ʂ<sup>h</sup>a mo  
 soft    r<sup>ɟ</sup>ə mo  
 smooth    <sup>n</sup>dzəŋ bo  
 rough    <sup>h</sup>tsəp po  
 slippery    <sup>w</sup>zəq  
 tight    tam bo  
 loose    lo bo  
 solid    ʂ<sup>h</sup>a mo  
 disorder    ɲoq <sup>fi</sup>zəŋ  
 chaotic    tɕ<sup>h</sup>a rə tɕ<sup>h</sup>ə re  
 correct    <sup>n</sup>dək  
 incorrect    mə ɲen kə  
 true    ɲo ma  
 false    <sup>fi</sup>dzəŋ ma  
 raw    <sup>fi</sup>lən ba  
 new    <sup>h</sup>sa ra  
 old    <sup>fi</sup>ɲa ɲa  
 good    ji ɓa  
 bad    ɲan ba  
 weak    <sup>fi</sup>dək

expensive <sup>h</sup>kon tɕ<sup>h</sup>e γə  
 cheap <sup>h</sup>kon <sup>b</sup>de γə  
 old (year) <sup>fi</sup>gwa rge  
 young lo tɕ<sup>h</sup>oŋ  
 beautiful ji ɤa  
 ugly <sup>h</sup>tso ɣa  
 hot (temperature) tɔ γə  
 cold <sup>fi</sup>ɕ<sup>h</sup>əq  
 lukewarm <sup>n</sup>dʒam ja nə  
 warm tɔn bu  
 cool <sup>ɕ</sup>səl wo  
 hard (work) <sup>s</sup>ka mo  
 easy <sup>s</sup>tsa mo  
 fragrant tɔ ʒəm  
 smelly tɔ ŋan  
 tasty ʒəm bo  
 sour <sup>r</sup>cu ru  
 sweet <sup>fi</sup>ŋar mo  
 bitter q<sup>h</sup>a mo  
 hot (taste) k<sup>h</sup>a ts<sup>h</sup>a γə  
 salty ts<sup>h</sup>a q<sup>h</sup>a γə  
 tasteless ts<sup>h</sup>a mə tɕ<sup>h</sup>oq  
 astringent <sup>r</sup>cu ru  
 oily <sup>s</sup>ŋəm <sup>s</sup>coŋ <sup>h</sup>tɕik ʒəq <sup>h</sup>tɕik  
 not busy k<sup>h</sup>om ba  
 busy <sup>p</sup>twi wa  
 rich <sup>ɕ</sup>ɕ<sup>h</sup>i γo  
 poor <sup>fi</sup>wol wo  
 clean <sup>h</sup>tsəŋ ma  
 dirty <sup>p</sup>tso ɣa  
 living <sup>h</sup>son bo  
 fresh <sup>h</sup>sar wa  
 dead ɕ<sup>h</sup>ə wo

clear təŋ mo  
 good-looking <sup>h</sup>ta s<sup>h</sup>a jəq qə  
 noisy <sup>s</sup>kat cəq qə  
 bored s<sup>h</sup>em mə <sup>r</sup>cit kə  
 hastily <sup>m</sup>jo ɤa  
 colourful t<sup>h</sup>a t<sup>h</sup>a  
 wise t<sup>h</sup>a mo  
 foolish <sup>w</sup>len ba  
 honest təŋ mo  
 sly <sup>fi</sup>jo <sup>r</sup>jə <sup>fi</sup>dʒen  
 careful s<sup>h</sup>em ʒəm mo  
 gentle k<sup>h</sup>a <sup>n</sup>dʒam bo  
 arrogant ŋa <sup>r</sup>je tɕ<sup>h</sup>e  
 suitable <sup>n</sup>dək pa  
 severe ŋan ba  
 courteous <sup>n</sup>dʒam <sup>n</sup>dʒam  
 industrious <sup>fi</sup>go pu  
 lazy jor co  
 clumsy <sup>w</sup>len ba  
 well-behaved k<sup>h</sup>wa ŋen mo  
 hardworking <sup>p</sup>tsən <sup>n</sup>dʒi  
 pitiful <sup>s</sup>ŋaŋ <sup>r</sup>dʒe  
 glad <sup>fi</sup>ga xə  
 happy <sup>b</sup>de <sup>r</sup>cit  
 peaceful <sup>b</sup>de <sup>n</sup>dʒəq  
 sad <sup>s</sup>co <sup>r</sup>dək  
 proficient <sup>m</sup>k<sup>h</sup>e <sup>ɕ</sup>sa  
 kind <sup>m</sup>dza <sup>r</sup>tse  
 disagreeable ŋa na mo  
 single <sup>n</sup>tɕə ŋa / k<sup>h</sup>e <sup>r</sup>cəŋ  
 cliffy <sup>fi</sup>zar wo  
 so-so s<sup>h</sup>a rə s<sup>h</sup>ə γə  
 rare ja <sup>m</sup>ts<sup>h</sup>an

## Verbs

love <sup>fi</sup>ga  
 decoct <sup>h</sup>ku  
 pull up wa  
 arrange <sup>p</sup>twa mo <sup>fi</sup>dək

swing <sup>ŋ</sup>gi  
 worship <sup>ɕ</sup>ɕ<sup>h</sup>əŋ <sup>n</sup>ts<sup>h</sup>a  
 move (house) <sup>h</sup>po  
 move (a thing) k<sup>h</sup>ər

help roq ram <sup>ɸ</sup>çet  
 tie up <sup>fi</sup>dom  
 wrap tɕup <sup>ɸ</sup>çet  
 keep secret <sup>h</sup>səŋ  
 protect <sup>h</sup>tɔŋ cup  
 be full hwa <sup>b</sup>ʒəp  
 hug wu lən <sup>n</sup>dep  
 peel off <sup>ɸ</sup>çə  
 recite <sup>w</sup>la <sup>n</sup>don  
 carry on the back <sup>k</sup>hər  
 compare <sup>r</sup>dər  
 close <sup>r</sup>tsəm  
 weave la  
 change (something) <sup>n</sup>ʒər <sup>fi</sup>doq  
 (something) change <sup>fi</sup>ʒər  
 be sick na tsə  
 repair lan po <sup>b</sup>ʒəp  
 donate <sup>w</sup>ʒən ba <sup>h</sup>tɔŋ  
 wipe <sup>ɸ</sup>çəhə  
 wipe up ha rə <sup>ç</sup>i  
 guess ʔa k<sup>h</sup>e mə k<sup>h</sup>et  
 step on <sup>m</sup>nəq  
 participate ʒək  
 hide <sup>h</sup>koŋ  
 scratch <sup>mp</sup>t<sup>h</sup>ək  
 unseam ha rə <sup>ɸ</sup>çək  
 destroy <sup>ɸ</sup>çək  
 collapse <sup>fi</sup>ʒik  
 mix <sup>n</sup>dɕe  
 taste <sup>p</sup>tɔ wa <sup>n</sup>on  
 sing lon  
 quarrel wor <sup>b</sup>ʒəp  
 fry <sup>h</sup>sə  
 fight <sup>n</sup>dzən ri <sup>ɸ</sup>çet  
 sink t<sup>h</sup>əm  
 measure <sup>p</sup>çəp  
 prize <sup>p</sup>tot ʔa  
 open (an umbrella) <sup>ç</sup>hə <sup>ʔ</sup>dək hək  
 complete <sup>n</sup>dɕup ts<sup>h</sup>a  
 fill <sup>w</sup>lək  
 admit k<sup>h</sup>e lon

purify təŋ ŋik <sup>ɸ</sup>çet  
 eat za  
 sprint tɕ<sup>h</sup>oŋ  
 flush <sup>ɸ</sup>çət  
 take out <sup>n</sup>den  
 smoke <sup>n</sup>t<sup>h</sup>en  
 whip <sup>fi</sup>doŋ  
 go out <sup>fi</sup>gwa s<sup>h</sup>oŋ  
 (sun) rise <sup>ç</sup>har  
 come out <sup>ɸ</sup>çəhə <sup>fi</sup>a <sup>ç</sup>oq  
 hoe up <sup>p</sup>ko  
 wear kon  
 thread (a needle) <sup>b</sup>ʒəŋ  
 transmit <sup>b</sup>ʒən <sup>n</sup>dzən  
 infect <sup>n</sup>gu  
 blow (something) hor  
 hit <sup>fi</sup>doŋ  
 hurt by stinging ts<sup>h</sup>a  
 urge tet  
 get wrong nor  
 answer k<sup>h</sup>e lan  
 make gesture ləq <sup>b</sup>da <sup>w</sup>ze  
 hunt <sup>r</sup>ŋon <sup>b</sup>da  
 shoot (with a gun) me <sup>m</sup>da <sup>b</sup>ʒəp  
 hit (on the target) hoq  
 scatter <sup>h</sup>tor  
 separate <sup>n</sup>t<sup>h</sup>or  
 fetch lon  
 nod <sup>fi</sup>ŋi joŋ  
 yawn <sup>w</sup>lap rəŋ <sup>ɸ</sup>çet  
 hiccup ri ŋo  
 open k<sup>h</sup>a <sup>ɸ</sup>çəhə  
 thunder bolt breaks t<sup>h</sup>oq <sup>b</sup>ʒəp  
 thunder <sup>m</sup>dək <sup>b</sup>ʒəp  
 take with oneself <sup>w</sup>zoŋ  
 go to war <sup>fi</sup>də <sup>n</sup>dzəŋ  
 knot <sup>m</sup>dət pa <sup>b</sup>ʒəp  
 snore hop rəŋ ne  
 sneeze hap tɕ<sup>h</sup>ə <sup>b</sup>ʒəp  
 look after ŋək  
 lead t<sup>h</sup>ət

waer (a turban) <sup>h</sup>ti  
 wear (a bracelet) <sup>h</sup>təq  
 be born <sup>h</sup>t<sup>h</sup>oŋ  
 delay <sup>ŋ</sup>gor  
 keep off <sup>k</sup>hoq  
 fall down loq  
 topple <sup>h</sup>dzoq  
 pound to pieces t<sup>h</sup>əq  
 reverse <sup>ŋ</sup>go <sup>m</sup>dzək loq  
 arrive t<sup>h</sup>on / ləp  
 obtain <sup>ŋ</sup>t<sup>h</sup>op  
 wait <sup>h</sup>gək  
 (earth) quake s<sup>h</sup>a <sup>ŋ</sup>gi  
 lower the head <sup>m</sup>go <sup>h</sup>gər  
 shake head <sup>m</sup>go <sup>h</sup>jək / <sup>m</sup>go <sup>p</sup>tək  
 burn <sup>m</sup>bar  
 turn light on <sup>h</sup>loq <sup>h</sup>jon  
 fall down loŋ  
 hook <sup>n</sup>doq  
 fish <sup>n</sup>dzən  
 tumble loq  
 flod up <sup>h</sup>doq  
 bite s<sup>h</sup>o <sup>h</sup>tap  
 hammer in <sup>n</sup>dep  
 lose (something) wor s<sup>h</sup>oŋ  
 understand ɕ<sup>h</sup>i / ko  
 freeze <sup>h</sup>k<sup>h</sup>əŋ  
 stir <sup>ŋ</sup>gi  
 read <sup>n</sup>don  
 block up <sup>ŋ</sup>ts<sup>h</sup>əŋ / <sup>n</sup>dam  
 cross (a river) <sup>r</sup>gal  
 (thread) snap t<sup>h</sup>et  
 (stick) snap t<sup>h</sup>əq  
 snap (a thread) <sup>h</sup>tət  
 snap (a stick) <sup>h</sup>təq  
 hide (oneself) qap  
 chop <sup>h</sup>tsap  
 be hungry <sup>h</sup>toq  
 press <sup>h</sup>nən  
 occur <sup>m</sup>dzoŋ  
 develop <sup>m</sup>p<sup>h</sup>əl <sup>b</sup>git

be angry ts<sup>h</sup>ə ɣa za  
 swear <sup>h</sup>na <sup>h</sup>ce  
 shiver <sup>h</sup>t<sup>h</sup>ək <sup>n</sup>dər zik  
 ferment <sup>s</sup>ɲal  
 have a fever ts<sup>h</sup>a pa <sup>b</sup>jəq  
 worry <sup>r</sup>dək ri <sup>ɕ</sup>et  
 sprout <sup>m</sup>ɲə ɣə <sup>m</sup>bi  
 punish t<sup>h</sup>ə pa <sup>p</sup>tət  
 turn inside out ts<sup>h</sup>ə la <sup>h</sup>jər  
 turn over <sup>h</sup>zu ɣu <sup>h</sup>jər  
 oppose ŋo <sup>r</sup>gul <sup>ɕ</sup>et  
 translate <sup>h</sup>jər  
 leave (something) <sup>n</sup>dzoq  
 put in <sup>n</sup>dep  
 put out to pasture <sup>n</sup>ts<sup>h</sup>o  
 fly <sup>m</sup>p<sup>h</sup>ər  
 share <sup>h</sup>go  
 separate k<sup>h</sup>a k<sup>h</sup>a <sup>ɕ</sup>et  
 be mad <sup>h</sup>on / <sup>h</sup>o  
 sew <sup>m</sup>ts<sup>h</sup>em / <sup>w</sup>zo  
 apply <sup>ɕ</sup>ə  
 put one's hand <sup>s</sup>cor  
 satisfy <sup>m</sup>t<sup>h</sup>ən  
 rot re s<sup>h</sup>oŋ  
 cover <sup>ŋ</sup>gep  
 be dry <sup>h</sup>kam bo ret ts<sup>h</sup>ə  
 work le <sup>h</sup>ka le  
 go to market t<sup>h</sup>oŋ naŋ ton <sup>n</sup>jo  
 catch a cold t<sup>h</sup>am sə na  
 dare hot  
 say zer  
 cut up <sup>h</sup>tət  
 cut off t<sup>h</sup>et s<sup>h</sup>oŋ  
 mow <sup>p</sup>tək  
 give <sup>w</sup>zən / <sup>h</sup>ter  
 follow <sup>h</sup>dzi <sup>h</sup>tet  
 plough <sup>n</sup>dep  
 be enough <sup>n</sup>dəŋ  
 assess ha lam  
 hire <sup>h</sup>la  
 shave <sup>h</sup>toq

(wind) blow <sup>fi</sup>loŋ <sup>fi</sup>jək  
 hang jar <sup>p</sup>kwa  
 turn light off <sup>h</sup>tɑ  
 care about <sup>s</sup>həm <sup>k</sup>hər <sup>φ</sup>çet  
 close <sup>r</sup>ʒəq  
 enclose <sup>t</sup>çət  
 manage to tam <sup>φ</sup>çet  
 irrigate <sup>fi</sup>ləq  
 kneel <sup>fi</sup>vi <sup>h</sup>tsək <sup>t</sup>çet  
 rotate <sup>n</sup>ɖe loq <sup>φ</sup>çet / <sup>b</sup>ʒe  
 celebrate New Year lo <sup>h</sup>sər <sup>h</sup>toŋ  
 (time) pass <sup>ŋ</sup>gor <sup>s</sup>həŋ  
 be ashamed ŋo <sup>t</sup>ʃa  
 be afraid <sup>ʃ</sup>cəq  
 shout <sup>m</sup>bot  
 drink <sup>n</sup>tʰəŋ  
 be appropriate <sup>mp</sup>tʰək <sup>mp</sup>tʰək  
 hate ŋan lɑ  
 regret <sup>ɲ</sup>ɔt pa <sup>φ</sup>çet  
 paddle <sup>h</sup>toŋ  
 draw (a picture) <sup>p</sup>twa  
 be pregnant <sup>fi</sup>ŋwar <sup>t</sup>çə <sup>φ</sup>wa  
 doubt to ʒa zə  
 repay ho rən <sup>h</sup>sət  
 return (a thing) <sup>h</sup>ter  
 change <sup>b</sup>dze  
 wave <sup>fi</sup>jək  
 return (home) loq  
 recall çə <sup>n</sup>dən  
 reply lan <sup>n</sup>de  
 destroy <sup>h</sup>tsa met to <sup>h</sup>toŋ  
 know (to do) çʰi  
 muddle <sup>ŋ</sup>əq  
 be alive <sup>s</sup>hə <sup>m</sup>bo ret <sup>t</sup>hɑ  
 raise <sup>h</sup>so  
 gain <sup>n</sup>tʰop  
 gather <sup>n</sup>tsʰoq  
 assemble <sup>h</sup>pɑŋ <sup>n</sup>ɖək  
 press <sup>p</sup>tsər  
 squeeze <sup>w</sup>zo  
 remember ʃan

deposit <sup>p</sup>tçu  
 send <sup>h</sup>kər  
 envy <sup>p</sup>tʰəq toq <sup>φ</sup>çet  
 fasten <sup>p</sup>tçəŋ  
 pick up <sup>fi</sup>nən  
 sollect <sup>fi</sup>zəm  
 reduce mar çʰət  
 cut with scissors <sup>h</sup>tɑ  
 tell <sup>φ</sup>çet  
 descend mar loŋ  
 exchange <sup>fi</sup>zə <sup>fi</sup>zi <sup>φ</sup>çet  
 make friends with <sup>fi</sup>ɖək  
 irrigate <sup>fi</sup>lək  
 be burnt <sup>n</sup>tsʰə <sup>fi</sup>zəŋ  
 chew <sup>l</sup>dat  
 teach loq  
 (bird) crow <sup>φ</sup>ça çəq <sup>fi</sup>di  
 (cat) meow mo <sup>n</sup>dzə <sup>h</sup>ket  
 (donkey) bray woŋ wə ŋar  
 (horse) neigh <sup>ʃ</sup>ta <sup>n</sup>tsʰər  
 (cow) moo zoq ŋar  
 (dog) bark çʰə zək  
 (pig) grunt həq tsʰər <sup>h</sup>ket <sup>h</sup>toŋ  
 (sheep) baa lək <sup>m</sup>ba  
 (tiger) growl <sup>ʃ</sup>təq ŋar  
 (wolf) howl <sup>φ</sup>səŋ <sup>k</sup>hə ŋar  
 be named <sup>m</sup>ŋəŋ <sup>n</sup>doq  
 lift up <sup>φ</sup>çʰe  
 bear (fruit) <sup>m</sup>ɖe wu <sup>n</sup>tʰoq  
 freeze <sup>t</sup>çə <sup>r</sup>əm <sup>ʃ</sup>çʰoq  
 marry <sup>ʃ</sup>ton mo ji  
 untie <sup>fi</sup>dot pa <sup>φ</sup>çet  
 borrow <sup>fi</sup>jar  
 soak <sup>b</sup>dzəŋ  
 forbid <sup>h</sup>kor <sup>ŋ</sup>go <sup>φ</sup>çet  
 submerge <sup>t</sup>həm  
 enter naŋ <sup>fi</sup>ə ze  
 pass <sup>b</sup>ʒet  
 shock <sup>h</sup>təq  
 be shocked <sup>n</sup>ɖoq  
 rescue <sup>h</sup>çop <sup>t</sup>çet



live <sup>fi</sup>dot / <sup>fi</sup>dək  
 raise (a hand) <sup>p</sup>tcek  
 saw <sup>sh</sup>oq le <sup>p</sup>tceɔ  
 complete <sup>n</sup>dzom  
 roll up <sup>fi</sup>d̥i t̥ei  
 curl up <sup>s</sup>kəm <sup>fi</sup>gəm <sup>ɕ</sup>et  
 block <sup>ɳ</sup>gəq  
 open <sup>ɕ</sup>ce<sup>h</sup>e  
 be boiling <sup>ɳ</sup>k<sup>h</sup>u  
 blossom ce  
 drive <sup>h</sup>toŋ  
 start <sup>ɳ</sup>go <sup>n</sup>dzəq  
 hew <sup>h</sup>t̥et / <sup>p</sup>tsəp  
 look <sup>l</sup>ta / <sup>l</sup>tu  
 show <sup>h</sup>ton  
 see rək  
 see a doctor na <sup>h</sup>ton  
 shoulder k<sup>h</sup>ər / <sup>p</sup>ceq  
 warm oneself <sup>s</sup>ho  
 rely <sup>s</sup>ten  
 cough <sup>fi</sup>lo lə  
 be thirsty k<sup>h</sup>a <sup>h</sup>kom  
 carve <sup>s</sup>qo  
 agree <sup>n</sup>t<sup>h</sup>əq t̥e  
 gnaw <sup>ɕ</sup>ha <sup>s</sup>qe  
 dig out with a finger <sup>m</sup>t̥e<sup>h</sup>ət  
 button up <sup>fi</sup>jo  
 be idle k<sup>h</sup>om ba jot  
 weep ɳə  
 be sleepy <sup>fi</sup>ɳi jəŋ k<sup>h</sup>ə  
 pull <sup>n</sup>t<sup>h</sup>en  
 defecate <sup>s</sup>ce ɣa <sup>h</sup>təŋ  
 be hot (taste) k<sup>h</sup>a ts<sup>h</sup>a wa  
 be missing loq / li  
 come joŋ  
 drag <sup>ɕ</sup>ce<sup>h</sup>ə rə nəŋ  
 be old <sup>fi</sup>ge  
 be tired <sup>h</sup>ka / t̥e<sup>h</sup>at  
 connect <sup>m</sup>t<sup>h</sup>ət  
 dry <sup>h</sup>kem / <sup>h</sup>kam  
 chat k<sup>h</sup>a <sup>b</sup>da <sup>ɕ</sup>et

split open ke  
 drench <sup>fi</sup>lən ba ret t<sup>h</sup>ə  
 flow <sup>fi</sup>zər  
 reserve <sup>n</sup>dzəq  
 be deaf <sup>fi</sup>na won  
 embrace wə rə <sup>h</sup>tap  
 leak zəq  
 mess up <sup>h</sup>t̥ək  
 be messy <sup>ɳ</sup>t̥<sup>h</sup>ək  
 pile up <sup>p</sup>tsək  
 (sun) set ɳi ma rə wə <sup>fi</sup>ge / nəp  
 be numb <sup>fi</sup>d̥ət  
 scold <sup>fi</sup>dzəŋ  
 bury s<sup>h</sup>a lep <sup>b</sup>jəq  
 buy ɳo  
 sell <sup>h</sup>tsəŋ  
 be full <sup>h</sup>kəŋ  
 not to have me  
 cover up <sup>p</sup>kap  
 die out <sup>h</sup>ta  
 (bird) chirp t̥eəq  
 close <sup>m</sup>t̥e<sup>h</sup>ə <sup>p</sup>tsək <sup>p</sup>tsəm <sup>ɕ</sup>et  
 understand ha ko  
 stroke rək  
 whet <sup>r</sup>dar  
 grind <sup>n</sup>t<sup>h</sup>əq  
 hold len  
 scratch <sup>fi</sup>ja <sup>p</sup>t̥ək  
 can t<sup>h</sup>əp  
 congeal <sup>h</sup>qəq <sup>h</sup>qəq  
 twist <sup>h</sup>t̥eə <sup>r</sup>ʒər / <sup>p</sup>t̥eər  
 vomit <sup>s</sup>t̥ək  
 crawl ja ra <sup>ɳ</sup>go  
 climb <sup>fi</sup>gwər  
 clap <sup>fi</sup>d̥ep  
 line up <sup>fi</sup>d̥ək  
 dispatch <sup>fi</sup>ŋəq  
 circle <sup>ɳ</sup>k<sup>h</sup>or  
 run <sup>fi</sup>ʒək  
 make (tea) <sup>fi</sup>dzəŋ  
 compensate <sup>ɕ</sup>ceə ɣa

bump    t<sup>h</sup>ək  
 float    l<sup>1</sup>dəŋ  
 splash    m<sup>1</sup>tɕ<sup>h</sup>ot  
 be torn    t<sup>h</sup>et  
 split up    tɕ<sup>h</sup>əq  
 be damaged    h<sup>1</sup>tɕoq  
 be broken    tɕ<sup>h</sup>əq  
 break    f<sup>1</sup>doq s<sup>h</sup>əp  
 bully    p<sup>1</sup>tsot h<sup>1</sup>təp ɕ<sup>1</sup>et  
 deceive    m<sup>1</sup>go h<sup>1</sup>kor h<sup>1</sup>toŋ  
 ride    n<sup>1</sup>go  
 stand up    jar ləŋ  
 lead along    c<sup>h</sup>et  
 owe    f<sup>1</sup>go  
 rob    m<sup>1</sup>p<sup>1</sup>t<sup>h</sup>oq  
 knock    f<sup>1</sup>doŋ  
 raise (the tail)    f<sup>1</sup>jik  
 cut up    h<sup>1</sup>təp  
 kiss    po ji / po ɕ<sup>1</sup>et  
 look down    tɕ<sup>h</sup>oŋ h<sup>1</sup>təq ɕ<sup>1</sup>et  
 request    zə wa ɕ<sup>1</sup>et  
 drive out    b<sup>1</sup>da  
 fetch    len  
 marry (a woman)    lon  
 go    n<sup>1</sup>jo  
 heal    təq  
 be complete    f<sup>1</sup>dəŋ  
 dye    p<sup>1</sup>tɕa t<sup>1</sup>tsə ɕ<sup>1</sup>ək / p<sup>1</sup>tsu ku  
 give way    lam f<sup>1</sup>zər  
 heat up    ts<sup>h</sup>a h<sup>1</sup>ko  
 recognise    ŋo ɕ<sup>h</sup>i  
 throw    m<sup>1</sup>p<sup>h</sup>ən / h<sup>1</sup>hen  
 dissolve    zə  
 knead    f<sup>1</sup>dzə  
 endure    s<sup>h</sup>on  
 tan    m<sup>1</sup>net  
 spray    n<sup>1</sup>dep  
 scatter seeds    s<sup>h</sup>a won n<sup>1</sup>dep  
 end    ci  
 come loose    f<sup>1</sup>zik  
 sweep    ɕ<sup>h</sup>əq

kill    ɕ<sup>1</sup>sot  
 seave    h<sup>1</sup>cap  
 dry in the sun    n<sup>1</sup>i ma h<sup>1</sup>tə  
 sunbathe    n<sup>1</sup>i ma n<sup>1</sup>de  
 lightning    f<sup>1</sup>loq n<sup>1</sup>ts<sup>h</sup>ək  
 injure    f<sup>1</sup>mi  
 consult    t<sup>1</sup>u ɕ<sup>1</sup>et  
 go upstairs    ja rə n<sup>1</sup>jo  
 shoot    m<sup>1</sup>p<sup>h</sup>en  
 have shot    h<sup>1</sup>hoq  
 stretch out    s<sup>1</sup>cəŋ  
 extend    s<sup>h</sup>aŋ  
 grow    s<sup>1</sup>ce  
 rust    h<sup>1</sup>tsa γo  
 grow a boil    r<sup>1</sup>ma wət  
 give birth    s<sup>1</sup>ce  
 remain    ləq  
 rise    ɕ<sup>h</sup>ar  
 lose (something)    wor  
 release    h<sup>1</sup>təŋ  
 try    ts<sup>h</sup>ot l<sup>1</sup>ta  
 be (equative)    ret  
 harvest    p<sup>1</sup>təq  
 receive    m<sup>1</sup>p<sup>1</sup>t<sup>h</sup>ot  
 close    p<sup>h</sup>ap  
 put in order    f<sup>1</sup>di h<sup>1</sup>soq r<sup>1</sup>ja  
 defend    s<sup>h</sup>oŋ məq  
 comb    ɕ<sup>1</sup>et  
 lose (a game)    ham  
 be cooked    n<sup>1</sup>ts<sup>h</sup>u  
 be ripe    p<sup>1</sup>tse  
 get thinner    ɕ<sup>h</sup>a tɕ<sup>h</sup>et  
 count    p<sup>1</sup>t<sup>1</sup>wəŋ h<sup>1</sup>ka / p<sup>1</sup>t<sup>1</sup>wəŋ t<sup>1</sup>tsə  
 rinse    ɕ<sup>1</sup>ɕa  
 fall    ləŋ  
 toss    m<sup>1</sup>p<sup>h</sup>en  
 tie    h<sup>1</sup>təq  
 sleep    n<sup>1</sup>a  
 fall asleep    f<sup>1</sup>nət  
 speak    ɕ<sup>1</sup>et  
 tear up    p<sup>1</sup>tet

die ɕʰə  
 count ʰtsi ʳjəp  
 break into pieces zik  
 spoil ʰtɕək  
 collapse zik  
 trample ʰdɔp  
 lift up jar ʰcəq  
 shed ʰɲik tɕʰə ɕʰər  
 lie down loq  
 burn (the hand) ʃʰəq  
 run away ʰtɔ  
 beg (for food) ʃcə  
 have a headache ʰmgo na  
 kick ʃkəŋ tʰoq ʳjap  
 shave ʰwzar  
 be cloudy ʰnam nəq  
 be fine day ʰnam tɔ  
 dawn ɲi ma ɕʰar  
 get dark mən nəq  
 lick ʳdəq  
 carry with a pole ʰtə ʁa ʰkʰər  
 choose ʰdam  
 dance ʰtɔ ʰdzam  
 jump ʰtɕʰoŋ  
 pulse ʰpʰəq  
 paste ʰjar  
 listen ɲon  
 hear ko  
 stop ʰtsʰam ʰdzəq  
 notify ʰda ʰdzər  
 steal ʰkə  
 hurl ʰjik  
 spit ʰdep  
 push ʰhit  
 offer as an excuse ʰkʰəq ʰzəq  
 move back ʰɕʰə nət  
 swallow ʰnət  
 drag ʰtət  
 dislocate tsʰək hət  
 carry on the back ʰkwa  
 dig ʰko

become bent kək  
 bend kək kə ʰdzək  
 finish tsʰar  
 play ʰtse mo ʰtse  
 forget ʰdzet  
 violate ʰgəl  
 feed ʰlək  
 smell ɲəm  
 ask ʰtə  
 grasp ʰmzə  
 cover səm  
 inhale ʰden  
 be accustomed kom ɕʰor  
 wash ʰtə  
 be blind ʰɲik za ra  
 go downstairs ʰdze ʰde mar  
 lay (an egg) ʰtəŋ  
 rain tɕʰa rwa ʰbap  
 scare ʃcəq ʃʰa ʰɣat  
 envy jil ʃmon tɕʰor  
 trust ji tɕʰi  
 think ʰsam / ʰtan  
 recall ʰtan kʰə  
 want ʰsam  
 resemble ʰɲa mo  
 digest zə  
 disappear jal / met pa ʰhoŋ  
 subside ʃcəŋ re  
 peel ʰzəq  
 be careful ka le / ʰzap ʰzap  
 laugh ʰgot  
 write ʰtə  
 have diarrhea ʰɕə  
 wake up ʰnət ʰsʰet  
 rest mɔ ʰso  
 embroider ʰtsəq  
 learn ʰdzəŋ  
 smoke tə wa ʰdɔp  
 look for ʰtswa  
 push down ʰnən  
 itch za ʰpʰtʰək / za

shake <sup>fi</sup>jik  
 bite <sup>sho</sup> <sup>h</sup>tap  
 scoop up <sup>p</sup>tɕə  
 need <sup>fi</sup>go  
 lead <sup>t</sup>hət  
 rely <sup>h</sup>ten <sup>sh</sup>a  
 overflow <sup>ɕ</sup>ɕʰə <sup>h</sup>tɕʰer  
 conceal <sup>h</sup>səŋ  
 win <sup>k</sup>he  
 greet <sup>fi</sup>ga <sup>ɕ</sup>sə <sup>ɕ</sup>ɕet  
 embrace <sup>fi</sup>wu <sup>fi</sup>lən <sup>n</sup>dep  
 swim <sup>t</sup>ɕʰə <sup>ŋa</sup> <sup>r</sup>ʃap  
 have <sup>j</sup>ot  
 be (existential) <sup>j</sup>ot  
 meet <sup>t</sup>hək  
 agree <sup>t</sup>u <sup>ɕ</sup>ɕet  
 cross <sup>m</sup>tɕʰoŋ <sup>ra</sup> <sup>m</sup>bət  
 be dizzy <sup>m</sup>go <sup>jər</sup> <sup>h</sup>kʰor  
 permit <sup>t</sup>ɕʰoq  
 plant <sup>n</sup>dzək  
 increase <sup>jar</sup> <sup>ŋən</sup>  
 stick into <sup>p</sup>tsək  
 blink <sup>fi</sup>ŋik <sup>fi</sup>də  
 extract <sup>p</sup>tsək  
 pluck <sup>n</sup>tʰo  
 stick down <sup>b</sup>ʃar  
 stand up <sup>jar</sup> <sup>lən</sup>  
 open <sup>k</sup>hə <sup>fi</sup>dəŋ  
 grow up <sup>t</sup>ɕʰi <sup>bo</sup> <sup>ret</sup>  
 feel bloated <sup>hwə</sup> <sup>w</sup>ru  
 catch fire <sup>m</sup>ŋe <sup>ɕ</sup>ʰor  
 catch cold <sup>h</sup>ɕʰəq

convene <sup>m</sup>tsʰwə <sup>b</sup>ʃə / <sup>n</sup>tsʰoq  
 have found <sup>fi</sup>ŋet / <sup>rək</sup>  
 sting <sup>b</sup>ʃap  
 shake <sup>ŋ</sup>gil  
 fight for <sup>m</sup>pʰtʰoq <sup>rə</sup> <sup>ɕ</sup>ɕet  
 steam <sup>fi</sup>lən <sup>p</sup>tso <sup>ji</sup>  
 know <sup>k</sup>o  
 weave <sup>n</sup>tʰəq  
 point at <sup>s</sup>ton  
 be swelling <sup>s</sup>cəŋ  
 cook <sup>p</sup>tso  
 use a walking stick <sup>fi</sup>loq <sup>t</sup>ten <sup>k</sup>ʰər /  
<sup>h</sup>kʰen  
 wish <sup>s</sup>mon <sup>ləm</sup> <sup>n</sup>dep  
 grab <sup>fi</sup>zoŋ  
 turn around <sup>k</sup>hə <sup>h</sup>kʰor  
 turn a corner <sup>k</sup>hə <sup>s</sup>kor  
 turn <sup>h</sup>kʰor  
 pack <sup>naŋ</sup> <sup>fi</sup>a <sup>t</sup>ɕʰək  
 chase after <sup>n</sup>det / <sup>b</sup>də  
 prepare <sup>t</sup>a <sup>b</sup>dək  
 peck at <sup>n</sup>tʰə  
 walk <sup>n</sup>ʃo  
 curse <sup>fi</sup>ŋo <sup>ŋan</sup>  
 go through <sup>n</sup>dzəl / <sup>n</sup>dzi  
 be drunk <sup>fi</sup>ʃəq  
 sit <sup>fi</sup>dot  
 do <sup>le</sup>  
 dream <sup>fi</sup>ŋi <sup>lu</sup> <sup>fi</sup>ŋi / <sup>m</sup>ŋi <sup>lam</sup> <sup>fi</sup>ŋi  
 do business <sup>n</sup>tsʰoŋ <sup>fi</sup>ʃəq

## References

- Hua, Kan (ed.). 2002. *Zangyu Anduo Fangyan Cihui [Vocabulary of Tibetan Amdo dialects]*. Lanzhou: Gansu Minzu Chubanshe.
- Huang, Bufan. 2012. “Zangmianyu de xiaosheyin [Uvulars in Tibeto-Burman]”. *Yuyanxue Luncong* 45. pp.157–174.

- Qu, Aitang and Xiaojing Jin. 1981. "Zangyu fangyan de yanjiu fangfa [Methodology of the research on Tibetan dialects]". *Xinan Minzu Xueyuan Xuebao* 3. pp.76–84.
- Sonam Wangmo. 2013. "Lhagang Monastery in Myth, History and Contemporary Society". Oslo: Universitetet i Oslo, M.Phil. Thesis.
- Suzuki, Hiroyuki. 2005. "Tibettogo onsetsu koozoo no kenkyuu [Study on the Tibetan syllable structure]". *Journal of Asian and African Studies* 69. pp.1–23.
- . 2015. "Kamu tiiki de hanasareru Amudotibetto-go Seka [gSerkha] hoogen no onsei kizyutu [Phonetic description of the gSerkha dialect of Amdo Tibetan spoken in the Khams region]". *Kyoto University Linguistic Research* 34. pp.89–107.
- Suzuki, Hiroyuki and Sonam Wangmo. 2015a. "Quelques remarques linguistiques sur le tibétain de Lhagang. «l'endroit préféré par le Bodhisattva»". *Revue d'études tibétaines* 32. pp.153–175.
- . 2015b. "Lhagang Tibetan of Minyag Rabgang Khams: Vocabulary of two sociolinguistic varieties". *Asian and African Languages and Linguistics (AALL)* 10. pp.245–286.
- . 2016a. "Kamu-Tibettogo Tagong [Lhagang] hoogen no bunpoo suketti [Grammatical sketch of the Lhagang dialect of Khams Tibetan]". *Journal of Kijutsuken* 8. pp.21–90.
- . 2016b. "Lhagang Choyu: A first look at its sociolinguistic status". *Studies in Asian Geolinguistics* 2. pp.60–69.
- . 2017. "Language evolution and vitality of Lhagang Tibetan, a Tibetic language as a minority in Minyag Rabgang". *International Journal of Sociology of Language* 245. [Originally presented at the workshop on the linguistic minorities of the Chinese Tibetosphere (Uppsala, 2014)]. pp.63–90.
- Tournadre, Nicolas. 2014. "The Tibetic languages and their classification". In Thomas Owen-Smith and Nathan W. Hill (eds.) *Trans-Himalayan Linguistics: Historical and Descriptive Linguistics of the Himalayan Area*. Berlin: Walter de Gruyter. pp.105–129.
- Wang, Shuangcheng. 2012. *Zangyu Anduo Fangyan Yuyin Yanjiu [Study on the phonetics of Tibetan Amdo dialects]*. Shanghai: Zhongxi Shuju.



## A Classified Lexicon of Shan Loanwords in Jinghpaw\*

KURABE, Keita

Japan Society for the Promotion of Science / ILCAA, Tokyo University of Foreign Studies

Jinghpaw is a Tibeto-Burman language primarily distributed in northern Burma, while Shan is a Tai-Kadai language whose distribution partially overlaps with that of Jinghpaw. The aim of this paper is to provide a classified lexicon of Shan loanwords in Jinghpaw, which are borrowed into Jinghpaw due to close cultural and linguistic contact. This paper also provides a brief overview of linguistic situation in the Jinghpaw-speaking area, followed by descriptions of linguistic properties of Shan loanwords in terms of phonology, morphology, syntax and semantics.

**Keywords:** Jinghpaw, Shan, language contact, loanwords, lexical borrowing

1. Introduction
2. Linguistic situation in northern Burma
3. Linguistic properties of Shan loanwords
4. Classified lexicon of Shan loanwords

### 1. Introduction

Jinghpaw is a Tibeto-Burman (TB) language primarily distributed in northern Burma (Myanmar), but whose distribution is broad, stretching from the upper Brahmaputra valley of northeastern India across northern Burma, and beyond the Sino-Burmese border into far western Yunnan. The Jinghpaw people have had a long-term symbiotic relationship with the Tai-speaking Shan people whose distribution partially overlaps with that of the Jinghpaw. Although Jinghpaw and Shan are genetically unrelated, Jinghpaw has absorbed a large number of lexical items from Shan, with which it has been in close cultural and linguistic contact for the past centuries. The aim of this paper is to provide a classified lexicon of Shan loanwords adopted by Jinghpaw, mainly collected by the author as a part of historical-comparative and contact linguistic

---

KURABE, Keita. 2017. "A classified lexicon of Shan loanwords in Jinghpaw". *Asian and African Languages and Linguistics* 11. pp.129–166. [Permanent URL: <http://hdl.handle.net/10108/89212>]

\* A part of this paper is based on work that was presented at the 48th International Conference on Sino-Tibetan Languages and Linguistics held at University of California, Santa Barbara, August 21–23, 2015. I would like to express my gratitude to the participants of the conference for their helpful discussion and comments. I am also grateful to the two anonymous reviewers for their careful reading and constructive comments on an earlier draft of this paper. My fieldwork was supported in part by a Grant-in-Aid for JSPS Fellows (Nos. 24-2938 and 26-2254) from the Japan Society for the Promotion of Science (JSPS).

research on Jinghpaw and its dialects. I will also offer a brief introduction to linguistic situation in the Jinghpaw-speaking area and descriptions of linguistic properties of Shan loanwords in Jinghpaw in terms of phonology, morphology, syntax and semantics.

The organization of this paper is as follows. Section 2 provides an introduction to linguistic situation in the Jinghpaw-speaking region. This is followed by sections providing brief linguistic sketches of Shan loanwords in Jinghpaw, beginning with phonology (Section 3.1), and followed by morphology (Section 3.2), syntax (Section 3.3) and semantics (Section 3.4). Section 4 is devoted to providing Shan loanwords in Jinghpaw classified in terms of semantic fields, with some corresponding forms from relevant languages that have close cultural and/or linguistic relationships with Jinghpaw and/or Shan. In the remainder of this section, I provide a brief review of literature, sources of the linguistic data and a brief description of Tai varieties in northern Burma.

### 1.1. Previous studies

Major previous studies that provide and/or discuss Shan loanwords in Jinghpaw include: Hanson (1906), Maran (1964) and Dai and Xu (1995). Hanson (1906), a dictionary of Jinghpaw, is a significant contribution to lexical borrowing between Shan and Jinghpaw in that it identifies a number of Jinghpaw lexical items of Shan origin. Hanson (1906), however, does not offer corresponding Shan forms. Maran (1964) investigates bilingualism in acculturation in Jinghpaw, surveying the varying degrees of bilingualism and cultural borrowing from Shan. His findings show that the structural resistance to the Shan loanwords in Jinghpaw can be seen at phonic, lexical and grammatical levels, at the last level of which the Shan influence is quite minor, as Shan loanwords follow the rules of Jinghpaw grammar. Dai and Xu (1995: 259–265) identify over one hundred Jinghpaw words of Tai origin, listing half of them with corresponding forms of Dehong Tai, a Tai dialect closely related to Shan. They point out several linguistic properties of these loanwords that will be noted in relevant sections below.

### 1.2. Data

The secondary data of the relevant languages are, unless otherwise noted, taken from the following sources. The Jinghpaw data based on Maran (1978) are given with slight modification according to the phonemic transcription employed in Kurabe (2016). The Shan data based on Sao Tern Moeng (1995), an updated version of Cushing (1881 [1914]), are transcribed according to the system used in SEALang Library Shan Dictionary.<sup>1</sup> The data for Colloquial Burmese are transcribed according to the system illustrated by Kato (2008) and the transcription of Written Burmese is based on the system outlined by Duroiselle (1916).

- Burmese (Colloquial): Myanmar Language Commission ed. (2009)
- Burmese (Written): Myanmar Language Commission ed. (2009)

---

<sup>1</sup> <http://www.sealang.net/shan/dictionary.htm> (accessed on 2016-08-31)



- Dehong: Luo (1999); supplemented by Dai and Xu (1995) marked by (DX) and Meng (2007) marked by (M)
- Ganan: Huziwara (2012a); supplemented with the help of Dr. Huziwara Keisuke marked by (H)
- Hpun: Henderson (1986)
- Jinghpaw: Maran (1978)
- Kadu: Sangdong (2012); supplemented with the help of Dr. Huziwara Keisuke marked by (H)
- Khamti: Harris (1976); supplemented by Pulu (1997) marked by (P) and Weidert (1977) marked by (W)<sup>2</sup>
- Lacid: my field notes
- Langsu: Dai (2005)
- Leqi: Dai and Li (2007)
- Lhaovo: Sawada (2003, 2004); supplemented by my field notes marked by (K)
- Lisu: Fraser (1922)
- Ngochang: Nasaw Sampu et al. (2005)
- Numhpuk: Morey (2007b)
- Pali: Rhys Davids and Stede, (eds.) (1921–1925)
- Proto-Tibeto-Burman: Matisoff (2003)
- Rawang: LaPolla and Sangdong (2015)
- Sanskrit: Macdonell (1929)
- Shan: Sao Tern Moeng (1995)
- Siamese: Haas (1964)
- Turung: Morey (2007a)
- Zaiwa: Lustig (2010); supplemented by Duoshi et al. (1992) marked by (D)

### 1.3. Tai varieties in northern Burma

Northern Burma is inhabited by Tai peoples whose languages and dialects are closely related, such as Shan (Tai Long or Tai Yai), Tai Mao, Tai Laing and Tai Khamti. This paper, as noted earlier, is based on Tai data from Shan because of the availability of its copious data (Cushing 1881 [1914], Sao Tern Moeng 1995). While it would also be possible that the direct source of Tai items in Jinghpaw has been of other Tai varieties closely related to Shan, and some mismatches between Jinghpaw and Shan forms might be attributed to this fact, this question is not easily solved due to the lack of sufficient data for other Tai varieties and dialects spoken in northern Burma. The attempt of this paper should thus be viewed as a preliminary approximation toward studies in Tai-Jinghpaw contact linguistics. Despite this situation, it is still true that Tai varieties in northern Burma have close relationship with each other, and that Jinghpaw words of Tai origin well correspond to those of Shan in many respects (see Section 3).

<sup>2</sup> The tone marks employed in Weidert (1977) are modified in accordance with the system used in Harris (1976).

Before we move on, a brief discussion of some of the other possible donor languages is in order. Dai an Xu (1995) treat Dehong Tai (Chinese Shan) spoken in the southwestern part of Yunnan province as the direct source of Tai words in Jinghpaw. This does not seem to hold, however, as can be seen from the following comparison where Dehong reflects initials *\*n-* and *\*hn-* of Proto-Tai with /l/ in contrast to other varieties that reflect them with /n/, which Jinghpaw has borrowed.<sup>3</sup>

Table 1 Reflexes of Proto-Tai initials *\*n-* and *\*hn-*

	Shan	Dehong	Khamti	Siamese	Jinghpaw	Proto-initials
‘rice field’	naa <sup>4</sup>	laa <sup>2</sup>	naa <sup>3</sup>	naa	nà	PT <i>*n-</i>
‘lady’	naaŋ <sup>4</sup>	laaŋ <sup>2</sup>		naaŋ	nàŋ	PT <i>*n-</i>
‘face’	naa <sup>3</sup>	laa <sup>4</sup>	naa <sup>5</sup>	nâa	na	PT <i>*hn-</i>
‘heavy’	nak <sup>4</sup>	lak <sup>1</sup>	nak <sup>4</sup>	nàk	nák	PT <i>*hn-</i>
‘water’	nam <sup>5</sup>	lam <sup>5</sup>	nam <sup>2</sup>	nám	nàm	PT <i>*nl/r-</i>

Tai Khamti (Khamti Shan) is another possible candidate for the direct donor language which has introduced Tai lexical items into Jinghpaw given the partial overlap between the Khamti and Jinghpaw peoples in the northern part of what is present-day Kachin State, as well as the early migration history of Jinghpaw which suggests a north-to-south migration. The following comparison, however, suggests that Khamti, which reflects the proto-initial *\*ʔd-* with /n/ unlike many other Tai varieties which reflect it with /l/ (Li 1977: 107–108, Edmondson 2008: 199–200), is not the direct source language because Tai loanwords in Jinghpaw have /l/ for these lexical items. Compare:

Table 2 Reflexes of proto-initial *\*ʔd-*

	Shan	Dehong	Khamti	Siamese	Jinghpaw	Proto-initials
‘silk’	laaj <sup>3</sup>	laai <sup>4</sup>	naay <sup>5</sup>	dâay	lày	PT <i>*ʔd-</i>
‘mountain’	lɔj <sup>1</sup>	lɔi <sup>6</sup>	nɔy <sup>1</sup>	dɔɔy	loy	PT <i>*ʔdl/r-</i>
‘red’	lɛŋ <sup>1</sup>	lɛŋ <sup>6</sup>	nɛŋ <sup>1</sup>	dɛɛŋ	leŋ	PT <i>*ʔdl/r-</i>
‘month’	lɯn <sup>1</sup>	lən <sup>6</sup>	nɯn <sup>1</sup>	dɯan	lun	PSWT <i>*ʔd-</i>

<sup>3</sup> Proto-Tai (PT) initials in Tables 1 and 2 are based on Li (1977: 108, 111, 114, 129, 131) and Proto-Southwestern Tai (PSWT) initials in Table 2 on Edmondson (2008: 200).

## 2. Linguistic situation in northern Burma

The Jinghpaw-speaking region in northern Burma is a site of intensive contact in which cultural and language contact among intra- and extra-TB speakers has been a long-standing phenomenon. The intra-TB contact is represented by the Kachin people who consist of several TB linguistic groups, of which the Jinghpaw is a primary member (2.1). The Kachin people, including the Jinghpaw, have also had a long symbiotic relationship with the Tai-speaking Shan people, from whom they have borrowed a number of lexical items (2.2). Burmese and Chinese are two dominant languages in the Jinghpaw-speaking region today. Their influence on Jinghpaw, however, is diachronically quite limited (2.3). Within the Sino-Tibetan language family, Jinghpaw is closely related to the Luish (Asakian) languages, some of which are distributed in northern Burma. While they are not in direct contact relationship with Jinghpaw, they are in contact with Shan (2.4).

### 2.1. Jinghpaw and Kachin

The Kachin people are recognized as one of the major ethnic groups in Burma. Linguistically, the Kachin are not a monolith group and exhibit internal diversity, consisting of speakers of languages belonging to several TB branches. In spite of internal linguistic diversity, the Kachin people form more or less a coherent socio-cultural complex of shared cultural traits such as a marriage-alliance system. In Burma, this Kachin grouping consists of speakers of languages such as Jinghpaw, Zaiwa, Lhaovo, Lacid, Ngochang and Rawang, and includes some Lisu speakers as well. In the Kachin region, especially in the southeastern part where non-Jinghpaw Kachin population is great, as noted by Bradley (1996), it is not difficult to find Kachin villages (*kăhtawng*), village clusters (*măre*) and communities inhabited by several linguistic groups. Leach (1954: 63–100) describes a Kachin community of some 500 people in the Kachin Hills of Burma situated close to the Burma-China border, which consists of diverse linguistic groups speaking Jinghpaw, Gauri, Zaiwa, Lhaovo, Lisu and Chinese.

Aside from common inheritance, members of the Kachin grouping share linguistic as well as cultural traits that have arisen as a result of intensive contact. Of particular importance is the fact that Jinghpaw serves as a lingua franca among the Kachin people, being spoken not only by the Jinghpaw people but also by other groups whose native tongues belong to distinct branches of TB. Jinghpaw, as a lingua franca, has provided many words to non-Jinghpaw Kachin languages, which form the areal lexicon in the Kachin cultural area. Table 3 provides a few examples of lexical items of Jinghpaw origin borrowed into other Kachin languages. Jinghpaw loanwords in other Kachin languages are also identified by Yabu (1982), Sawada (2003, 2004), Lustig (2010) and Matisoff (2013).

Table 3 Selected words of Jinghpaw origin in several Kachin languages

	Jinghpaw	Zaiwa	Lhaovo	Ngochang	Rawang
‘mistake’	ɕút	syut <sup>5</sup>	šat <sup>H</sup>	shuot	shut
‘different’	ɕàɣ	syai <sup>31</sup>	šay <sup>H</sup>	shaih	sháy
‘correct’	jò	zyo <sup>11</sup>	co <sup>F</sup>		jò
‘song’	məkhón	me <sup>1</sup> -k/hon <sup>31</sup>			mvkún
‘bless’	ɕəmán		šăman <sup>H</sup>	shvmanh	shvmán
‘public’	ɕəwà		šăva <sup>F</sup>	shvwas	shvwà

## 2.2. Jinghpaw and Shan

The Kachin people, including the Jinghpaw, have also had a long-term symbiotic relationship with the Tai-speaking Shan people. Although the situation has been changing, in general, Kachins are highlanders occupying hills and mountains where they practice slash and burn agriculture, while Shans are lowlanders occupying river valleys where they practice rice cultivation in irrigated fields (Leach 1954: 1). In spite of the contrastive ecological settings they occupy, “Kachins and Shans are almost everywhere close neighbours and in the ordinary affairs of life they are much mixed up together” (ibid.: 2). The ethnological fluidity in the region can be seen in the fact that it is not uncommon to observe that a Kachin “becomes a Shan” (ibid.: 30). Leach (1954) shows that Kachin communities in the first half of the twentieth century were “oscillating” between an egalitarian system and a Shan feudal system.

The Shan influence on the Kachin people is also reflected in a large number of Shan loanwords in their languages. The borrowing relationship is seemingly hierarchical, as witnessed by the fact that, in contrast to the situation in which Kachin languages have adopted a number of Shan words, Shan seems to have borrowed a very few lexical items from them. Some words of Shan origin, as shown in Table 4, are shared across several Kachin languages. Section 4 provides Shan loanwords in Jinghpaw, together with relevant data from other Kachin languages.

Table 4 Selected Shan loanwords in several Kachin languages

	Shan	Jinghpaw	Zaiwa	Lhaovo	Rawang
‘bag’	t <sup>h</sup> oŋ <sup>1</sup>	thuŋ	tung <sup>31</sup>	thauŋ <sup>F</sup>	d̀v̀nggúng
‘rabbit’	paaj <sup>1</sup> taaj <sup>4</sup>	praŋtáy	bang <sup>11</sup> dvai <sup>31</sup>	pyaj <sup>F</sup> tay <sup>L</sup>	bàngday
‘country’	mvŋ <sup>4</sup>	múŋ	meng <sup>11</sup>	muk <sup>L</sup>	móng
‘pond’	nɔŋ <sup>1</sup>	nónŋ	nvong <sup>31</sup>	nauŋ <sup>F</sup>	núng
‘hook’	met <sup>4</sup>	myít	mit <sup>5</sup>		mit

Although it is not always easy to tell whether non-Jinghpaw Kachin languages

borrowed lexical items directly from Shan or through Jinghpaw, there are at least some examples that indicate the latter scenario. These examples come from Shan items with non-etymological Jinghpaw elements as in (1), and shared semantic shifts as in (2).

- (1) Shan *kaa*<sup>5</sup> ‘to trade’ > Jinghpaw *phəga* ‘trade (n.)’ > Zaiwa *pe*<sup>5</sup>*ga*<sup>35</sup> ‘trade (n.)’, Lhaovo *phăka*<sup>L</sup> ‘trade (n.)’, Lacid *pha:ga* ‘trade (n.)’, Rawang *pvga* ‘goods’
- (2) Shan *maj*<sup>5</sup>*s<sup>h</sup>aw*<sup>3</sup> ‘slender piece of wood’ > Jinghpaw *màysàw* ~ *màwsàw* ‘paper’ > Zaiwa *mau*<sup>11</sup>*sau*<sup>11</sup> ‘paper’, Lhaovo *muk*<sup>F</sup>*suk*<sup>H</sup> ‘paper’, Lacid *moug*: *soug* ‘paper’, Rawang *mèsò* ‘paper’

As noted above, Jinghpaw, as a lingua franca among the Kachin people, performs the function of transferring lexical items of languages of high prestige in the region, such as Shan, into non-Jinghpaw Kachin languages. Matisoff (2013: 24) provides the borrowing chains across several language families given in (3), remarking that the Jinghpaw lexicon has a large number of Shan elements, and “[s]ome of these Shan items were themselves from Burmese, and in turn some of these were originally from Indo-Aryan”. Jinghpaw, as a lingua franca among the Kachin people, introduced some of these items into non-Jinghpaw Kachin languages such as Rawang.

- (3) Borrowing chains  
Pali (IA) → Burmese (TB) → Shan (Tai) → Jinghpaw (TB) → Rawang (TB)

Table 5 shows some of the lexical items of Pali origin shared across several TB and Tai languages discussed above, including Written Burmese (WB), Shan and Jinghpaw.<sup>4</sup>

Table 5 Selected Pali loanwords in several TB and Tai languages

	Pali	WB	Shan	Jinghpaw	Lhaovo	Rawang
‘caution’	<i>sati</i>	<i>sati</i>	<i>s<sup>h</sup>a<sup>1</sup>ti<sup>5</sup></i>	<i>sədi?</i>	<i>săti?</i> <sup>F</sup>	<i>svdiq</i>
‘luck’	<i>kamma</i>	<i>karin</i>	<i>kaam</i> <sup>2</sup>	<i>gàm</i>	<i>kam</i> <sup>F</sup>	<i>gàm</i>
‘machine’	<i>sakka</i>	<i>cak</i>	<i>tsaak</i> <sup>3</sup>	<i>jàk</i>	<i>cak</i> <sup>F</sup>	<i>jvk</i>
‘intellect’	<i>ñāṇa</i>	<i>ñāṇ</i>	<i>jaan</i> <sup>2</sup>	<i>nyàn</i>		<i>nyan</i>
‘camel’	<i>kula-oṭṭha</i>	<i>kulā</i> <sup>3</sup> <i>ut</i>	<i>ka<sup>1</sup>laa<sup>4</sup>ʔuk<sup>5</sup></i>	<i>golaʔúk</i>	<i>ko<sup>L</sup>la<sup>L</sup>ʔuk<sup>F</sup></i>	<i>golawu</i>
‘rich’	<i>setṭhi</i>	<i>sūthe</i> <sup>3</sup>	<i>s<sup>h</sup>a<sup>5</sup>t<sup>h</sup>e<sup>4</sup></i>	<i>səthí</i>	<i>săthe</i> <sup>H</sup>	<i>svté</i>

<sup>4</sup> It is often the case that the original meaning has been obscured in recipient languages as a result of semantic change. Table 5 provides glosses in the recipient languages. The original Pali meanings are respectively as follows: ‘memory, recognition, consciousness’, ‘the doing, deed, work’, ‘able, possible’, ‘knowledge, intelligence, insight’, ‘clan-camel’, ‘foreman of a guild, treasurer’. Note that Written Burmese *kulā<sup>3</sup>ut* (lit. Indian-camel), whose roots are not combined in the donor language, seems to be a novel compound coined by Burmese. The first syllable of Pali *setṭhi* which does not convey any meaning is replaced by *sū* ‘person’ in Burmese due to folk etymology.

### 2.3. Jinghpaw and other dominant languages

The Jinghpaw-speaking region is also inhabited by Burmese and Chinese speakers. The linguistic influence from Burmese and Chinese on Jinghpaw is, diachronically speaking, less significant than that from Shan, as reflected in the relative lack of Burmese and Chinese loanwords in Jinghpaw as compared to those of Shan. Many of the Burmese loanwords seem to have been introduced into Jinghpaw through Shan, given that many of them are also found in Shan, and that intensive direct contact between Jinghpaw and Burmese seems to be dated back no further than the early 1800s before the Anglo-Burmese wars. The fact that some Jinghpaw words of Burmese origin have additional non-etymological Shan elements also suggests that these words are introduced into Jinghpaw through Shan. Written Burmese *un*<sup>3</sup> ‘coconut’, for example, is borrowed into Jinghpaw as *məʔûn* (cf. Shan *maak*<sup>2</sup>*ʔun*<sup>1</sup>) with a Shan class term *maak*<sup>2</sup> ‘fruit’, which frequently occurs in Shan fruit names (see Section 3.4).

Table 6 shows some selected Burmese loanwords in Jinghpaw with corresponding Shan forms. Observe in the table that Burmese lexical items borrowed into Jinghpaw retain phonological properties of Written Burmese (WB), which have undergone significant sound changes in Colloquial Burmese (CB). The Written Burmese liquid *r* which has merged with *y* in Colloquial Burmese, for example, appears as *r* in Shan and Jinghpaw. As such, data of Burmese loanwords in Shan and Jinghpaw, together with evidence from the writing system and conservative Burmese dialects, offer clues for reconstruction of phonetic values of Old Burmese phonemes.

Table 6 Burmese loanwords in Shan and Jinghpaw

	WB	CB	Shan	Jinghpaw
‘building/box’	tuik	taiʔ	tyk <sup>3</sup>	dèk
‘capital’	arañ <sup>3</sup>	ʔəyín	ʔa <sup>1</sup> raaŋ <sup>4</sup>	ʔáráŋ
‘certificate’	lakmhat	lɛʔhmaʔ	laak <sup>3</sup> maat <sup>3</sup>	làkmàt
‘bell’	khoñ <sup>3</sup> loñ <sup>3</sup>	kháunláun	k <sup>h</sup> ɔŋ <sup>4</sup> loŋ <sup>4</sup>	khonloŋ
‘respite’	khyam <sup>3</sup> sā	chánḍà	k <sup>h</sup> jaam <sup>4</sup> s <sup>h</sup> aa <sup>2</sup>	khyámsà
‘murraya plant’	sanapkhā <sup>3</sup>	ṭānākhá	s <sup>h</sup> a <sup>1</sup> naap <sup>3</sup> k <sup>h</sup> aa <sup>4</sup>	sənapkhá

Sino-Jinghpaw words are much more restricted, as Hanson (1913: 29) puts it: “[v]ery few Chinese terms have been incorporated, although the Kachins [Jinghpaws] for centuries have been in close contact with their powerful and intelligent neighbours.” Some words of Chinese origin are shared between several Kachin languages. Table 7 shows some selected Chinese loanwords in pinyin found in Jinghpaw and some other Kachin languages.<sup>5</sup> Note that some Chinese items are also found in Shan, leaving the

<sup>5</sup> The Lhaovo data in orthographic forms are from my field notes.

possibility that they were introduced into Jinghpaw through Shan, e.g. *s<sup>h</sup>ɔn<sup>2</sup>* ‘calculate’, *lɔ<sup>3</sup>* ‘mule’.

Table 7 Chinese loanwords in some Kachin languages

	Chinese	Jinghpaw	Zaiwa	Lhaovo	Rawang
‘cut with scissors’	jiǎn	zèn	zuen <sup>11</sup>		zìn
‘calculate’	suàn	son	son <sup>55</sup>		sun
‘chopsticks’	kuàizi	khoydzè	koi <sup>55</sup> zvue <sup>11</sup>		
‘mule’	luózi	lòdzè	lo <sup>11</sup>	lo ze	lòzè
‘eel’	huángshàn	khàŋcan	hang <sup>11</sup> syān <sup>55</sup>	khang: shan:	
‘boss / rich man’	lǎobǎn	làwbàn		lau ban	lòbòn

While the Burmese and Chinese influence on Jinghpaw is historically less significant, the situations are rapidly changing, with increased demands for Burmese in Burma and Chinese in China as the medium of education and communication. The majority of the Kachin people in Burma today, including Jinghpaw, are fully bilingual in Burmese from childhood.

#### 2.4. Jinghpaw and Luish

Jinghpaw, as noted earlier, belongs to the TB branch of the Sino-Tibetan language family, as evidenced by well-established sound correspondences between Jinghpaw and the proto-language as well as by a large number of Jinghpaw lexical items inherited from the proto-language. Within TB, Jinghpaw is closely related to the Luish (Asakian) languages such as Cak (Sak), Kadu, Ganan, Andro and Sengmai, which are distributed in small discontinuous pockets situated across three countries: northwestern Burma, southeastern Bangladesh and northeastern India. Together these constitute the Jinghpaw-Luish (Jinghpaw-Asakian) branch of TB, which covers a widespread but discontinuous area in the northwestern part of Greater Mainland Southeast Asia (Huziwara 2012b, 2014, Matisoff 2013).

Jinghpaw and Luish languages, being geographically separated, have not been in contact relationship for the past centuries. It should be noted, however, that both Jinghpaw and some Luish languages in northern Burma such as Kadu and Ganan have been in contact with Shan independently. Shan influence on Kadu, for example, is reflected in its lexicon where basic lexical items such as some numerals and kinship terms are of Shan origin (Sangdong 2012: 139–141, 236–238). Some Shan loanwords in Kadu and Ganan, due to this situation, are shared with some languages of the Kachin including Jinghpaw, some examples of which are presented in Table 8. Section 4, when relevant, provides Shan loanwords in Kadu and Ganan shared by Jinghpaw.

Table 8 Selected Shan loanwords in Kadu and several Kachin languages

	Shan	Kadu	Jinghpaw	Zaiwa	Rawang
‘bag’	t <sup>h</sup> oŋ <sup>1</sup>	shíthaúng	thuŋ	tung <sup>31</sup>	dǝnggúng
‘rabbit’	paan <sup>1</sup> taaj <sup>4</sup>	pángtaí	praŋtáy	bang <sup>11</sup> dvai <sup>31</sup>	bàngday
‘shoes/slippers’	k <sup>h</sup> ep <sup>4</sup> tin <sup>1</sup>	heúttín	khyèpdin		chøpdòn
‘duck’	pet <sup>4</sup>	aūmpeút	khaypyék	bvyet <sup>5</sup>	kabit
‘castrate’	ton <sup>1</sup>	taún	don		dun
‘mosquito net’	s <sup>h</sup> ut <sup>4</sup>	sūt	sút	sut <sup>5</sup>	

### 3. Linguistic properties of Shan loanwords

This section deals with a brief overview of linguistic characteristics of Shan loanwords in Jinghpaw in terms of phonology (3.1), morphology (3.2), syntax (3.3) and semantics (3.4).

#### 3.1. Phonology

The mapping of vowels and initial consonants from Shan to Jinghpaw can be summarized as follows.

<i>Vowels</i>	<i>Stops</i>	<i>Sonorants</i>	<i>Fricatives</i>
i > i	p > p, b	m > m	s <sup>h</sup> > s
e > i	t > t, d	n > n	h > kh, (h)
ɛ > e	ts > c, j	ɲ > ny	
a > a	k > k, g	ŋ > ŋ	
aa > a	p <sup>h</sup> > ph	l > l	
ɔ > o	t <sup>h</sup> > th	r > r	
ɣ > i, u, uy	k <sup>h</sup> > kh	w > w	
o > u	? > ?	j > y	
uu > i, u			
u > u			
au > aw			

Some remarks on the correspondence summarized above are in order. Jinghpaw has six monophthongs, including a marginal /ə/, while Shan has many more vowel contrasts. Many of the Shan vowels, thus, are replaced in Jinghpaw by close equivalents. Shan has contrastive mid vowels /e/, /ɛ/, /o/ and /ɔ/, while Jinghpaw has only /e/ and /o/. Shan open-mid vowels /ɛ/ and /ɔ/ are mapped to mid vowels /e/ and /o/ in Jinghpaw, because Jinghpaw mid vowels are phonetically open-mid vowels. Shan close-mid vowels /e/ and /o/, on the other hand, are adapted as high



vowels /i/ and /u/ in Jinghpaw. Shan vowels /ɿ/ and /ʊ/, which are lacking in the Jinghpaw vowel inventory, are mapped to /i/ or /u/, the choice of which seems to be unpredictable, e.g. *nàmlíj* ‘dropsy’ (< Shan *nam<sup>5</sup>lɿŋ<sup>1</sup>*), *lun* ‘lunar month’ (< Shan *lɿn<sup>1</sup>*), *khík* ‘a kind of basket’ (< Shan *k<sup>h</sup>ɯk<sup>2</sup>*), *lùksùk* ‘soldier’ (< Shan *luk<sup>3</sup>s<sup>h</sup>ɯk<sup>4</sup>*). Vowel length is not phonemic in Jinghpaw while Shan has a length contrast for /a/. Both of Shan /a/ and /aa/ are mapped to /a/ in Jinghpaw without any trace. The Shan diphthong /au/ is replaced by /aw/ in Jinghpaw. Shan vowels /a/, /i/ and /u/ are faithfully retained in Jinghpaw. Table 9 shows some examples of vowel correspondences between Shan and Jinghpaw.

Table 9 Some vowel correspondences

	Shan		Jinghpaw
‘hook’	met <sup>4</sup>	>	myít
‘jewel’	s <sup>h</sup> ɛŋ <sup>1</sup>	>	seŋ
‘bean’	t <sup>h</sup> o <sup>2</sup>	>	thù
‘rivulet/ditch’	hoŋ <sup>3</sup>	>	khòŋ
‘water’	nam <sup>5</sup>	>	nàm
‘barren’	haam <sup>4</sup>	>	khám

Jinghpaw faithfully retains Shan voiceless aspirated stops, glottal stop and sonorants, which are mapped straightforwardly to corresponding Jinghpaw consonants without any substitution. The two Shan fricatives /s<sup>h</sup>/ and /h/ are mapped to Jinghpaw /s/ [s<sup>h</sup>] and /kh/ or /h/. The mapping /h/ > /kh/ can be accounted for in terms of phonological nativization as Jinghpaw /h/ is marginal in the native phonology, being mostly restricted to interjections and onomatopoeic words. The Shan voiceless unaspirated stops and a voiceless affricate /ts/ (represented by /c/ in Lengtai 2009 and by /s/ in Hudak 2000) are mapped to either voiceless or voiced counterparts in Jinghpaw, the conditioning factor of which is unknown, e.g. *té* ‘ridge’ (< Shan *tɛ<sup>4</sup>*), *dépkhá* ‘thatch comb’ (< Shan *tɛp<sup>4</sup>k<sup>h</sup>aa<sup>4</sup>*). There are some cases where Jinghpaw shows hesitation between voiceless and voiced stops, e.g. *kày* ~ *gày* ‘fowl’ (< Shan *kaj<sup>2</sup>*), *maypàw* ~ *maybàw* ‘a kind of tree’ (< Shan *maj<sup>5</sup>paw<sup>4</sup>*). Table 10 shows some examples of initial stops, affricates and sonorants between Shan and Jinghpaw.

Table 10 Some correspondences between initial consonants

	Shan		Jinghpaw
‘palace’	hɔ <sup>1</sup>	>	kho
‘yoke’	ʔɛk <sup>2</sup>	>	ʔék
‘epilepsy’	maa <sup>3</sup> mu <sup>1</sup>	>	màmu
‘difficult’	jaak <sup>3</sup>	>	yàk
‘garden’	s <sup>h</sup> on <sup>1</sup>	>	sún
‘shellfish’	hɔj <sup>1</sup>	>	khoy

The rhyme systems of Jinghpaw and Shan are similar in many respects, for example, in that they have basically the same set of final consonants, and that the vowel-coda combination is basically free in both languages, except when the coda is a glide. This fact accounts for the fact that final consonants of Shan are faithfully retained in Jinghpaw without modification. Note additionally that Jinghpaw has regularly lost the final \*k of Proto-Tibeto-Burman (PTB), reducing it to a glottal stop /ʔ/ in inherited words (Benedict 1972: 14), e.g. wàʔ ‘pig’ (< PTB \*p<sup>w</sup>ak). The Shan final /k/ is imported as /k/ in Jinghpaw, filling the structural gap in the Jinghpaw native phonology once created by the sound change. These facts indicate that Jinghpaw borrowed these items from Shan after the PTB \*k was reduced to /ʔ/ in Jinghpaw. As such, Jinghpaw words with final /k/ are good candidates for loanwords (Matisoff 1974: 157, 2013: 24). Table 11 shows some correspondences of final consonants between Shan and Jinghpaw.

Table 11 Some correspondences between final consonants

	Shan		Jinghpaw
‘cave’	t <sup>h</sup> am <sup>3</sup>	>	thàm
‘model’	laaj <sup>4</sup> len <sup>4</sup>	>	làylen
‘poison’	kɔŋ <sup>5</sup>	>	guŋ
‘uncooked fish’	paa <sup>1</sup> lip <sup>4</sup>	>	balíp
‘animal’	to <sup>1</sup> s <sup>h</sup> at <sup>4</sup>	>	dùsàt
‘compartment’	lɔk <sup>4</sup>	>	lòk
‘gourd’	taw <sup>3</sup>	>	tàw
‘small scale’	jɔj <sup>4</sup>	>	yòy

Jinghpaw has a native phonotactic restriction against combining an initial /w/ or /ŋ/ with front vowels. An illicit combination of /w/ plus front vowels has arisen as a result of a sound change \*w > /y/ before front vowels (Kurabe 2014), e.g. yi ~ yì ‘female’ (< PTB \*pwi(y)-n ‘female’). Shan loanwords having such illicit combinations in the Jinghpaw native phonology are imported into Jinghpaw, e.g. wíŋ ‘enclosure’ <

Shan *wɛɹ̃*<sup>4</sup> ‘town’ (possibly from Burmese), *jàŋŋín* ‘silversmith’ < Shan *tsaəñ³ŋuñ*<sup>4</sup>, the former example of which suggests the borrowing took place after the sound change \*w > /y/.

Shan has more contrastive tones than Jinghpaw. These Shan tones are adapted in Jinghpaw according to the Jinghpaw tone system. The tonal correspondences between Shan and Jinghpaw are summarized in Table 12.<sup>6</sup>

Table 12 Tonal correspondences in checked and unchecked syllables

Unchecked syllables			Checked syllables		
(ma <sup>2</sup> )	low	>	low, (mid)	(mak <sup>2</sup> )	low > high, (low)
(ma <sup>3</sup> )	mid	>	low, (mid, high)	(mak <sup>3</sup> )	mid > low, high
(ma <sup>4</sup> )	high	>	high, (low, mid)	(mak <sup>4</sup> )	high > high, (low)
(ma <sup>5</sup> )	falling	>	mid, low	(mak <sup>5</sup> )	falling > low, (high)
(ma <sup>1</sup> )	rising	>	mid, (low, high)		
(ma <sup>6</sup> )	emphatic	>	no example		

As can be seen, the correspondence is not straightforward. In some cases the tone assignment seems to be based on faithfulness to the input (e.g. the Shan high tone in unchecked syllables), while in other cases it seems to be based on a default assignment mechanism by which the default low and mid tones are assigned to loanwords in contrast to the less frequent Jinghpaw high tone (Matisoff 1974: 159). It would be also possible that the non-straightforwardness of tonal correspondences might be attributed to dialectal differences within Shan.

There are also some puzzling cases in which Shan forms are repaired in Jinghpaw even though there is no necessity. Examples include: Shan /k-/ > Jg. /kh-/ (e.g. ‘duck’, below); Shan /m-/ > Jg. /l-/ (e.g. ‘mango’); Shan ∅ > Jg. /-ʔ/ (e.g. ‘mix’); Shan /-ŋ/ > Jg. /-k/ (e.g. ‘lima bean’); Shan /-t/ > Jg. /-k/ (e.g. ‘duck’, ‘red pepper’, ‘hook’, ‘black pepper’); the insertion of a non-etymological medial /r/ (e.g. ‘rabbit’, ‘red pepper’).

<sup>6</sup> Table 12 is based only on tones on monosyllabic words because Jinghpaw sometimes shows tonal alternations involving polysyllabic words whose rules are not thoroughly understood.

Table 13 Unnecessary repairs

	Shan		Jinghpaw
‘duck’	kaj <sup>2</sup> ‘fowl’, pet <sup>4</sup> ‘duck’	>	khaypyék
‘mango’	maak <sup>2</sup> moŋ <sup>3</sup>	>	məmûŋ ~ ləmuŋ
‘lima bean’	t <sup>h</sup> o <sup>2</sup> poŋ <sup>1</sup>	>	thùbúk
‘hook’	met <sup>4</sup>	>	myít ~ ?myék
‘rabbit’	paan <sup>1</sup> taaj <sup>4</sup>	>	praŋtáy
‘mix’	l <sup>4</sup> le <sup>4</sup>	>	lò?lè?
‘red pepper’	maak <sup>2</sup> p <sup>h</sup> it <sup>5</sup>	>	ŋprík
‘black pepper’	maak <sup>2</sup> p <sup>h</sup> it <sup>5</sup> pəm <sup>3</sup>	>	mákphyíkbòm

### 3.2. Morphology

The unnecessary repair noted in Section 3.1 can also be observed at the morphological level. As noted by Dai and Xu (1995: 248–249), some Shan words are borrowed into Jinghpaw with non-etymological meaningless presyllables, the function of which are unclear.

Table 14 Addition of non-etymological meaningless presyllables

Shan			Jinghpaw
paan <sup>1</sup>	‘gathering place’	>	dəbaŋ ‘camp’
s <sup>h</sup> ɛ <sup>3</sup>	‘bolt’	>	sùmsè ‘bolt’
pi <sup>2</sup>	‘pipe, flute’	>	sumpyi ‘flute’
paan <sup>2</sup>	‘level as a tract of land’	>	dùmbàŋ ‘substantial tract of land’

Jinghpaw has verb-deriving reduplication whereby verbs are derived from nouns by means of partial reduplication. This process, as pointed out by Diehl (1988) and Dai and Xu (1995: 249–250), is also applicable to loanwords. Thus, Shan nouns such as *kaat<sup>2</sup>* ‘market’, *k<sup>h</sup>ɛp<sup>4</sup>tin<sup>1</sup>* ‘sandals’ (cf. *tin<sup>1</sup>* ‘foot’), *tsaw<sup>3</sup>kɔŋ<sup>3</sup>* ‘gunner’ (lit. master-gun) and *s<sup>h</sup>ɔŋ<sup>1</sup>hon<sup>1</sup>* ‘cholera’ are involved in the noun-verb reduplicative construction in Jinghpaw, the morphological process of which is not attested in Shan. Examples:<sup>7</sup>

#### (4) Reduplicative noun-verb constructions

<i>gát</i> ‘market’	→	<i>gàt gàt</i> ‘open, as a market’
<i>khyépdin</i> ‘shoes’	→	<i>khyèpdin din</i> ‘put on shoes’
<i>jàwgòŋ</i> ‘hunter’	→	<i>jàwgòŋ gòŋ</i> ‘hunt (v.)’
<i>soŋkhun</i> ‘cholera’	→	<i>soŋkhun khun</i> ‘be infected with cholera’

<sup>7</sup> Reduplication may trigger irregular tone alternations.

Shan is a head-initial language where the head of noun-noun compounds precedes its modifier while Jinghpaw is a head-final in noun-noun compounding, the default head position of which is the right position. Many Shan noun-noun loanwords in Jinghpaw follow the rules of Shan, e.g. *bà-lúk* ‘catfish’ (lit. fish-pit), *màk-pháy* ‘fire fruit’ (lit. fruit-fire), *moŋ-din* ‘socks’ (lit. bag-foot), *nam-sum* ‘vinegar’ (lit. water-sour), suggesting that they were borrowed into Jinghpaw as a whole. There are also a few examples which show the reverse modifier-head order, suggesting that they were borrowed into Jinghpaw part by part and then compounded in Jinghpaw in accordance with its morphological rule, e.g. *khaw-nà* (lit. rice-paddy) ‘rice field’ (cf. Shan *naa<sup>1</sup>-k<sup>h</sup>aw<sup>3</sup>*), *khàw-gát* (lit. rice-market) ‘bazaar’ (cf. Shan *kaat<sup>2</sup>-k<sup>h</sup>aw<sup>3</sup>*), *nà-lòk* (lit. paddy-compartment) ‘compartment of a rice field’ (cf. Shan *lok<sup>4</sup>-naa<sup>4</sup>*).

### 3.3. Syntax

Borrowed numerals sometimes function as classifiers in Jinghpaw. As pointed out by Dai and Xu (1995: 245), Jinghpaw round numbers over ‘thousand’ are loanwords from neighboring languages. Observe this in the following list of Jinghpaw numerals where some numerals exhibit semantic shifts.<sup>8</sup>

Table 16 Round numbers in Jinghpaw

çi	‘ten’	PTB	*ts(y)i(y) ɤ*tsyay	‘ten’
tsa	‘hundred’	PTB	*b-r-gya	‘hundred’
khyiŋ	‘thousand’	Shan	heŋ	‘thousand’
mùn	‘ten thousand’	Shan	mum <sup>2</sup>	‘ten thousand’
sèn	‘hundred thousand’	Shan	s <sup>h</sup> en <sup>1</sup>	‘hundred thousand’
wàn	‘million’	Chinese	wàn	‘ten thousand’
rì	‘ten million’	Chinese	yì	‘hundred million’
gədì	‘hundred million’	Burmese	gǎdè	‘ten million’

The borrowed round numbers are distinguished from inherited round numbers in terms of the relative position they occur. Consider the contrast in Table 17. These borrowed round numbers can be interpreted as classifiers based on their position, e.g. *məçà məray məli* (person-CLF-four) ‘four persons’. This is supported, in part, by the fact that many of the Jinghpaw classifiers are of foreign origin (Xu 1987).

<sup>8</sup> Shan *heŋ<sup>1</sup>* and *s<sup>h</sup>en<sup>1</sup>* may be originally from Chinese *qiān* ‘thousand’ and from Written Burmese *sin<sup>3</sup>* ‘hundred thousand’, respectively. Burmese *gǎdè* is of Indo-Aryan origin, i.e. Pali *koṭṭi* ‘the end’ (Dr. Huziwara Keisuke, p.c., 2016).

Table 17 Asymmetrical distribution of inherited and borrowed round numbers in Jinghpaw

‘forty’	məli ɕi	*ɕi məli
‘four hundred’	məli tsa	*tsa məli
‘four thousand’	*məli khyiŋ	khyiŋ məli
‘forty thousand’	*məli m̀n	m̀n məli
‘four hundred thousand’	*məli sèn	sèn məli
‘four million’	*məli wàn	wàn məli
‘forty million’	*məli rì	rì məli
‘four hundred million’	*məli gədi	gədi məli

### 3.4. Semantics

Shan loanwords in Jinghpaw recur in several specific semantic fields. Dai and Xu (1995: 259–265) point out that many Jinghpaw words associated with wet-rice cultivation, trading, fauna and flora (esp. fruit, vegetable and fish), and cultural items (esp. food and daily utensils) are of Tai origin, since the Jinghpaw people did not have these objects before they came into contact with the Tai people. Hanson (1896: 91) points out the existence of some Shan morphemes which occur recurrently in the Jinghpaw lexicon such as *hɔ*<sup>1</sup> ‘palace’, *tsaw*<sup>3</sup> ‘master’ and *nam*<sup>5</sup> ‘water’. Some of these recurrent items appear with Shan class terms such as *maak*<sup>2</sup> ‘fruit’, *paa*<sup>1</sup> ‘fish’, *nam*<sup>5</sup> ‘water’, *maj*<sup>5</sup> ‘wood’ and *k<sup>h</sup>aw*<sup>3</sup> ‘rice’. For example, Jinghpaw, as shown in Table 18, has a number of lexical items associated with fruits involving a morpheme *màk* ~ *mák* ~ *mə*, which has its diachronic source in the recurrent Shan class term *maak*<sup>2</sup> ‘fruit’.

Table 18 Plant names with *maak*<sup>2</sup> ‘fruit’

	Shan	Jinghpaw
‘jackfruit’	<i>maak</i> <sup>2</sup> <i>laaŋ</i> <sup>4</sup>	<i>màk</i> <i>laŋ</i>
‘fire fruit’	<i>maak</i> <sup>2</sup> <i>p<sup>h</sup>aj</i> <sup>4</sup>	<i>màk</i> <i>pháŋ</i>
‘sweet orange’	<i>maak</i> <sup>2</sup> <i>waan</i> <sup>1</sup>	<i>màk</i> <i>wan</i>
‘bitter gourd’	<i>maak</i> <sup>2</sup> <i>ʔaaŋ</i> <sup>1</sup> <i>k<sup>h</sup>aaŋ</i> <sup>1</sup>	<i>màk</i> <i>ʔaŋkhaŋ</i>
‘pomegranate’	<i>maak</i> <sup>2</sup> <i>man</i> <sup>5</sup>	<i>màk</i> <i>mân</i>
‘quince’	<i>maak</i> <sup>2</sup> <i>waaw</i> <sup>4</sup>	<i>màk</i> <i>wáw</i>
‘orange’	<i>maak</i> <sup>2</sup> <i>tsək</i> <sup>4</sup>	<i>màk</i> <i>còk</i>
‘black pepper’	<i>maak</i> <sup>2</sup> <i>p<sup>h</sup>it</i> <sup>5</sup> <i>pəm</i> <sup>3</sup>	<i>mák</i> <i>phyíkbòm</i>
‘hog’s plum’	<i>maak</i> <sup>2</sup> <i>kək</i> <sup>2</sup>	<i>mək</i> <i>ók</i>
‘mango’	<i>maak</i> <sup>2</sup> <i>mɔŋ</i> <sup>3</sup>	<i>məm</i> <i>ûŋ</i>
‘pear’	<i>maak</i> <sup>2</sup> <i>kɔ</i> <sup>3</sup>	<i>mə</i> <i>go</i>
‘plum’	<i>maak</i> <sup>2</sup> <i>k<sup>h</sup>ɔ</i> <sup>1</sup>	<i>mək</i> <i>hó</i>

Some of the loanwords have undergone semantic changes in Jinghpaw. As an illustration, consider a Shan word *ho<sup>1</sup>k<sup>h</sup>am<sup>4</sup>* ‘royal palace’ which has undergone metonymic semantic change, meaning ‘king’ in Jinghpaw (i.e. *khokhám*). Other examples include:

Table 19 Semantic shifts of loanwords

Jg. màysàw	‘paper’	<	S. maj <sup>5</sup> s <sup>h</sup> aw <sup>3</sup>	‘slender piece of wood’
Jg. phày	‘flint’	<	S. p <sup>h</sup> aj <sup>4</sup>	‘fire’
Jg. khám	‘gold leaf’	<	S. k <sup>h</sup> am <sup>4</sup>	‘gold’
Jg. sùŋ	‘deep’	<	S. s <sup>h</sup> uŋ <sup>1</sup>	‘high, tall’
Jg. ʔəroŋ	‘honor’	<	S. ʔa <sup>5</sup> rəŋ <sup>2</sup>	‘brightness’ (< B. ‘color’)
Jg. sasəná	‘mission’	<	S. s <sup>h</sup> a <sup>2</sup> s <sup>h</sup> a <sup>1</sup> naa <sup>2</sup>	‘religious system’ (< B. < Pali)

#### 4. Classified lexicon of Shan loanwords

This section provides a lexicon of Shan loanwords in Jinghpaw classified in terms of semantic fields based on Aung Kyaw et al. (2001): nature (4.1); animals (4.2); plants (4.3); food and drink (4.4); clothing and adornments (4.5); dwelling (4.6); tools (4.7); community, occupation and production (4.8); commerce and trade (4.9); communication and transportation (4.10); culture and entertainment (4.11); cults, customs and socializing (4.12); human body (4.13); life, sickness and death (4.14); types of people (4.15); activity and mental activity (4.16); state and quality (4.17); time (4.18); number (4.19).

The lexicon, including Numhpuk and Turung, two Jinghpaw dialects spoken in northeastern India, also provides corresponding lexical items of several relevant languages from Tai-Kadai (TK) and Tibeto-Burman (TB) which are outlined in Section 2. They include: Dehong [TK] and Khamti [TK] (genetically having a close relationship to Shan); Hpun [TB] (culturally having a close relationship to Shan); Lacid [TB], Langsu [TB], Leqi [TB], Lhovo [TB], Lisu [TB], Ngochang [TB], Rawang [TB] and Zaiwa [TB] (culturally having a close relationship to Jinghpaw and Shan); Kadu [TB] and Ganan [TB] (genetically having a close relationship to Jinghpaw and culturally to Shan). Among these, Lacid and Leqi, and Langsu and Lhaovo can be viewed as closely related dialects of a single language, respectively.

Abbreviations for the languages in the lexicon are as follows.

- |                           |              |               |
|---------------------------|--------------|---------------|
| • B. Burmese (Colloquial) | • Hp. Hpun   | • Lan. Langsu |
| • C. Chinese              | • Ka. Kadu   | • Le. Leqi    |
| • D. Dehong               | • Kh. Khamti | • Lh. Lhaovo  |
| • G. Ganan                | • Lac. Lacid | • Li. Lisu    |

- Ng. Ngochang
- Nu. Numhpuk
- P. Pali
- R. Rawang
- S. Shan
- Sk. Sanskrit
- T. Turung
- Z. Zaiwa

Transcription systems depend on data sources: phonological (Colloquial Burmese, Dehong, Ganan, Hpun, Kadu, Khamti, Langsu, Leqi, Lhaovo by Sawada 2003 and 2004, Lisu, Numhpuk, Shan, Turung, Zaiwa); orthographic (Chinese, Lacid, Lhaovo by the author, Ngochang, Rawang); literal (Pali, Sanskrit and Written Burmese). For the phonological systems and/or correspondences between phonological and orthographic transcriptions of Lacid, Lhaovo, Ngochang and Rawang, the readers are referred to Wannemacher (2011), Sawada (2006), Nasaw Sampu et al. (2005), and LaPolla and Sangdong (2015), respectively.

Notes on the lexicon are as follows. The first line provides English translations for corresponding Jinghpaw words based on Hanson (1906) and Maran (1978). The abbreviation *bf.* stands for a bound form, *clf.* for a classifier and *v.* for a verb. The second line offers Jinghpaw words in the orthographic forms followed by phonemic transcriptions enclosed in square brackets. The third line provides corresponding Shan forms followed, if any, by corresponding forms from other relevant languages. No glosses are provided for these words when all the given items denote the same meanings.

#### 4.1. Nature

cave	<i>htam</i> /thəm/	S. <i>t<sup>h</sup>am<sup>3</sup></i> , D. <i>tham<sup>4</sup></i>
earth oil	<i>nam myin</i> /nəmmyin/	S. <i>nam<sup>5</sup>men<sup>1</sup></i> ; R. <i>nəmmin</i>
enamel	<i>nam ya</i> /nàmyà/	S. <i>nam<sup>5</sup>jaa<sup>3</sup></i> ; R. <i>nəmmya</i> ‘blue’
ditch	<i>hkawng</i> /khòŋ/	S. <i>hɔŋ<sup>3</sup></i> ‘rivulet’, Kh. <i>hɔŋ<sup>2</sup></i> ‘channel’ (W), D. <i>hɔŋ<sup>6</sup></i> ; Z. <i>xoŋ<sup>21</sup></i> (D), R. <i>tikùŋ</i> (water-ditch)
ditch	<i>nam hkawng</i> /nàmkhòŋ/	S. <i>nam<sup>5</sup>hɔŋ<sup>3</sup></i> ‘stream of water’
flood	<i>nam htum</i> /nàmthùm/	S. <i>nam<sup>5</sup>t<sup>h</sup>om<sup>3</sup>mɔŋ<sup>4</sup></i>
gold leaf	<i>hkam</i> /khám/	S. <i>k<sup>h</sup>am<sup>4</sup></i> ‘gold’, Kh. <i>k<sup>h</sup>am<sup>3</sup></i> ‘gold’, D. <i>xam<sup>2</sup></i> ‘gold’
lake <sup>9</sup>	<i>nawng</i> /nón/	S. <i>nɔŋ<sup>1</sup></i> , Kh. <i>nɔŋ<sup>4</sup></i> (W), D. <i>lɔŋ<sup>1</sup></i> ; Z. <i>nvong<sup>31</sup></i> , Lh. <i>yit<sup>F</sup>nauw<sup>H</sup></i> , Lac. <i>gyid</i> <i>noung</i> ”, R. <i>tínúŋ</i> , G. <i>nàuw</i> (H)
mountain <sup>10</sup>	<i>loi</i> /loy/	S. <i>lɔj<sup>1</sup></i> , Kh. <i>nɔy<sup>1</sup></i> , D. <i>loi<sup>6</sup></i> ; Z. <i>loi<sup>11</sup>lung<sup>35</sup></i> ‘Loilung village’

<sup>9</sup> This word is often preceded by *khà?* ‘water’ in Jinghpaw. The same holds for Lhaovo, Lacid and Rawang where *yit<sup>F</sup>*, *gyid* and *tí* mean ‘water’, respectively.

<sup>10</sup> “a large number of Kachin hills and villages have names derived from the Shan; thus *Loije*, *Loijau*, *Loihkang* and *Loilung*; ... *Loi sam sip*, the traditional thirty Hills in the *Sinli* district” (Hanson 1906:357)



ocean	<i>nam muk dāra</i> /nàmmùkdāra/ <sup>11</sup>	S. <i>nam</i> <sup>5</sup> ‘water’, Sk. <i>samudra</i> , P. <i>samudda</i> ; B. <i>θāmou?dāya</i> (WB <i>samuddarā</i> ), Z. <i>nam</i> <sup>31</sup> <i>moq</i> <sup>1</sup> <i>de</i> <sup>1</sup> <i>ra</i> <sup>35</sup> , Lh. <i>nam mug dara</i> : (K), Lac. <i>nam mug dara</i> ., Ng. <i>nvmuk-dra</i> , R. <i>nvmukdvra</i>
Salween river solder	<i>sap hkung</i> /sàpkhuŋ/ <i>nam san</i> /nàmsan/	S. <i>k<sup>h</sup>oŋ</i> <sup>4</sup> , D. <i>xoŋ</i> <sup>2</sup> D. <i>lām</i> <sup>53</sup> <i>san</i> <sup>35</sup> (M); Z. <i>nam</i> <sup>51</sup> <i>san</i> <sup>55</sup> (D)
sky	Nu. <i>kanghkau</i> /kaŋ <sup>4</sup> khau <sup>4</sup> /	S. <i>p<sup>h</sup>aa<sup>5</sup>kaaŋ<sup>1</sup>haaw<sup>1</sup></i>
tract of land	<i>dumbang</i> /dùmbàŋ/	S. <i>paat</i> <sup>2</sup> ‘be level as a tract of land’
water	<i>nam</i> /nàm/ <sup>12</sup>	S. <i>nam</i> <sup>5</sup> , Kh. <i>nam</i> <sup>2</sup> , D. <i>lam</i> <sup>5</sup>
water (oozing) <sup>13</sup>	<i>nam chyim</i> /nàmcím/	S. <i>nam</i> <sup>5</sup> <i>tsum</i> <sup>4</sup> , D. <i>lam</i> <sup>5</sup> <i>tsim</i> <sup>2</sup>
water rust	<i>nam hkan</i> /nàmkhan/	S. <i>nam</i> <sup>5</sup> ‘water’, <i>k<sup>h</sup>aan</i> <sup>1</sup> ‘formation of lime, as a stalactite’

#### 4.2. Animals

animal	<i>du sat</i> /dùsàt/	S. <i>to</i> <sup>1</sup> <i>s<sup>h</sup>at</i> <sup>4</sup> , Kh. <i>to</i> <sup>1</sup> ‘classifier for animals’, D. <i>sat</i> <sup>1</sup> ; Z. <i>du</i> <sup>11</sup> <i>sat</i> <sup>1</sup> ~ <i>du</i> <sup>11</sup> <i>sat</i> <sup>5</sup>
bird, <i>bf.</i>	<i>kai</i> /kày/ ~ <i>gai</i> /gày/ ~ <i>hkai</i> /khày/	S. <i>kaj</i> <sup>2</sup> ‘fowl’, Kh. <i>kay</i> <sup>4</sup> ‘fowl’, D. <i>kai</i> <sup>3</sup> ‘chicken’
capon	<i>kai dawn</i> /kàydon/	S. <i>kaj</i> <sup>2</sup> ‘fowl’, <i>tən</i> <sup>1</sup> ‘to geld’, D. <i>kai</i> <sup>3</sup> <i>phu</i> <sup>4</sup> <i>tən</i> <sup>6</sup>
catfish	<i>bāluk</i> /bəlúk/	S. <i>paa</i> <sup>1</sup> <i>luk</i> <sup>4</sup> , Kh. <i>nuk</i> <sup>4</sup> ‘catfish’; Lh. <i>balug</i> (K), Lac. <i>ba lug</i>
dipper	<i>kai nam</i> /káynam/ ~ <i>gai nam</i> /gáynam/	S. <i>kaj</i> <sup>2</sup> ‘fowl’, <i>nam</i> <sup>5</sup> ‘water’
duck	<i>pyek</i> /pyék/ <sup>14</sup>	S. <i>pet</i> <sup>4</sup> ‘duck’, Kh. <i>pet</i> <sup>4</sup> , D. <i>pet</i> <sup>3</sup> ; Z. <i>bvyet</i> <sup>5</sup> , Lac. <i>gai pyed</i> , R. <i>kabit</i> , Ka. <i>aūmpeút</i> , G. <i>ʔáuppét</i> (H)
duck	<i>hkai pyek</i> /khaypyék/	S. <i>kaj</i> <sup>2</sup> ‘fowl’, <i>pet</i> <sup>4</sup> ‘duck’

<sup>11</sup> This hybrid word is created in Jinghpaw by combining etymologically Tai and Indo-Aryan parts: Shan *nam*<sup>5</sup> ‘water’ and Written Burmese *samuddarā* (< Sk. *samudra*).

<sup>12</sup> This morpheme is found in many river and place names in Kachin and Shan State, as can be seen in Nam Ti Sang, Nam Li Hka, Nam Yin Hka, Nam Si, Nam Mi Lawng, Nam Mun, Nam Ma, Nam Hkam, Nam Hpat Kar (see Sawada 2011).

<sup>13</sup> ‘water oozing through the ground’

<sup>14</sup> This word is recorded with final *t* as *u pyet* in Scott (1900:664) and *kaipet* in Grierson (1928:521), suggesting a sporadic sound change of *t* to *k*. Also note Numhpuk *khaí<sup>4</sup>pet<sup>4</sup>* and Turung *kai<sup>3</sup>pet<sup>4</sup>*.

elephant (tuskless male)	<i>jang dwi</i> /jǎŋduy/	S. <i>tsaaj<sup>5</sup>tuj<sup>4</sup></i> , Kh. <i>caaj<sup>2</sup></i> ‘elephant’, D. <i>tsaaj<sup>5</sup></i> ‘elephant’
fish, <i>bf.</i>	<i>ba</i> /ba/	S. <i>paa<sup>1</sup></i> , Kh. <i>paa<sup>1</sup></i> , D. <i>paa<sup>6</sup></i>
geld	<i>dawn</i> /don/	S. <i>tɔn<sup>1</sup></i> , Kh. <i>tɔn<sup>1</sup></i> , D. <i>tɔn<sup>6</sup></i> ; Z. <i>ton<sup>21</sup></i> (D), R. <i>dun</i> , Ka. <i>taún</i>
horse	<i>ma</i> /mà/ <sup>15</sup>	S. <i>maa<sup>5</sup></i> , Kh. <i>maa<sup>2</sup></i> , D. <i>maa<sup>5</sup></i>
horse (black)	<i>ma gam</i> /màgàm/	S. <i>maa<sup>5</sup></i> ‘horse’, <i>kam<sup>2</sup></i> ‘be black’
leech (water)	<i>bying</i> /byìŋ/	S. <i>piŋ<sup>1</sup></i> , Kh. <i>piŋ<sup>1</sup></i> , D. <i>piŋ<sup>6</sup></i> ; Z. <i>toŋ<sup>21</sup>pjin<sup>21</sup></i> (D), Lh. <i>na bying</i> (paddy-leech) (K), Lac. <i>na byen</i> (paddy-leech), R. <i>nabìng</i> ~ <i>naqbìng</i> (paddy-leech), Hp. <i>pəiŋ</i> ~ <i>peìŋ</i>
parrot	<i>kai ke</i> /kayke/ ~ <i>hkai ke</i> /khayke/	S. <i>kaj<sup>2</sup></i> ‘fowl’, <i>nok<sup>5</sup>k<sup>h</sup>ew<sup>1</sup></i> ‘parrot’, D. <i>lok<sup>5</sup>xeu<sup>1</sup></i> ; Z. <i>gvai<sup>55</sup>gvyeq<sup>1</sup></i> ‘kind of parrot’, Lac. <i>khai: ke:</i> , Ka. <i>mákheú</i> , G. <i>mak<sup>h</sup>é</i> (H)
rabbit	<i>prang tai</i> /praŋtáy/	S. <i>paaj<sup>1</sup>taaj<sup>4</sup></i> , Kh. <i>paŋ<sup>1</sup>taay<sup>3</sup></i> , D. <i>paaj<sup>6</sup>taai<sup>2</sup></i> ; Z. <i>bang<sup>11</sup>dvai<sup>31</sup></i> , Lh. <i>pyaŋ<sup>F</sup>tay<sup>L</sup></i> , Lan. <i>paŋ<sup>35</sup>tai<sup>55</sup></i> , R. <i>bàngday</i> ~ <i>brvngday</i> , Hp. <i>pàŋ thwaí</i> , Ka. <i>pángtai</i> , G. <i>páŋté</i>
red catfish	<i>ba leng</i> /balen/	S. <i>paa<sup>1</sup></i> ‘fish’, <i>lɛŋ<sup>1</sup></i> ‘be red’
ride (horse)	<i>hkyi</i> /khyì/ <sup>16</sup>	S. <i>k<sup>h</sup>i<sup>2</sup></i> , Kh. <i>khí<sup>4</sup></i> , D. <i>xí<sup>3</sup></i>
shellfish	<i>hkoi</i> /khoy/	S. <i>hɔj<sup>1</sup></i> , Kh. <i>hɔI<sup>4</sup></i> , D. <i>hɔi<sup>1</sup></i> ; Z. <i>hoi<sup>55</sup></i> , Lh. <i>lǎxø<sup>L</sup></i> , Lan. <i>lǎ<sup>31</sup>xɔi<sup>35</sup></i> , Lac. <i>la</i> <i>foi:</i> , Le. <i>lǎ<sup>31</sup>xɔi<sup>33</sup></i>
stork	<i>gawn</i> /gon/	S. <i>kɔn<sup>2</sup></i> , D. <i>kɔn<sup>3</sup></i> ; Z. <i>gon<sup>35</sup></i>
turtle	<i>tau</i> /taw/	S. <i>taw<sup>2</sup></i> , Kh. <i>taw<sup>4</sup></i> , D. <i>tau<sup>3</sup></i>
turtle	<i>tau kaw</i> /tawkók/	S. <i>kək<sup>4</sup>taw<sup>2</sup>?</i> ‘upper shell of a turtle, oil dipper’; Z. <i>dvau<sup>55</sup>-gvop<sup>1</sup></i> , Lh. <i>tau: kog</i> (K), Lac. <i>tau: kug</i> , Ng. <i>thomh tau-kuk</i>

#### 4.3. Plants

banana (kind of)	<i>gwi hkawm</i> /gùykhom/	S. <i>koj<sup>3</sup>hɔm<sup>1</sup></i> , Kh. <i>koy<sup>5</sup></i> ‘banana’, D. <i>koi<sup>4</sup></i> ‘banana’
banana (kind of)	<i>gwi lung</i> /gùyluŋ/	S. <i>koj<sup>3</sup></i> ‘banana’, <i>loŋ<sup>1</sup></i> ‘forest’

<sup>15</sup> Possibly from Chinese *mǎ* ‘horse’ through Shan.

<sup>16</sup> Possibly from Chinese *qí* ‘to ride’ through Shan.

banana (kind of)	<i>gwi sum</i> /gùysùm/	S. <i>koj<sup>3</sup>s<sup>h</sup>om<sup>3</sup></i>
bean	<i>htu</i> /thù/	S. <i>t<sup>h</sup>o<sup>2</sup></i> , Kh. <i>tho<sup>4</sup></i> , D. <i>tho<sup>3</sup></i>
betel, <i>bf.</i>	<i>doi hkawm</i> /doykhòm/ <sup>17</sup>	S. <i>tɔj<sup>2</sup></i> ‘creeper the leaf of which is used for chewing instead of the betel leaf’, <i>k<sup>h</sup>om<sup>1</sup></i> ‘be bitter’
betel	<i>ya man</i> /yàmàn/	S. <i>jaa<sup>3</sup></i> ‘grass’, D. <i>yaa<sup>4</sup></i> ‘grass’
bitter gourd	<i>mak ang hkang</i> /màkʔaŋkhaŋ/	S. <i>maak<sup>2</sup>ʔaaŋ<sup>1</sup>k<sup>h</sup>aaŋ<sup>1</sup></i>
black pepper	<i>mak hpyik bawm</i> /mákphyík bòm/	S. <i>maak<sup>2</sup>p<sup>h</sup>it<sup>5</sup>pòm<sup>3</sup></i> , D. <i>maak<sup>3</sup>phit<sup>5</sup></i> ; Z. <i>mɿ<sup>55</sup>phjik<sup>55</sup>pòm<sup>21</sup></i> (D)
Bodhi tree leaf	T. <i>nyong</i> /nyoŋ <sup>3</sup> /	S. <i>maj<sup>5</sup>ɲoŋ<sup>2</sup></i>
coriander	<i>hpak kyi</i> /phákkyí/ ~ <i>hpăkyi</i> /phăkyí/	S. <i>p<sup>h</sup>ak<sup>4</sup>ki<sup>4</sup></i> ; Z. <i>phɿ<sup>55</sup>kjɿ<sup>51</sup></i> (D), Lh. <i>pha jhi</i> : (K), Lac. <i>pha: jhi</i> ”, R. <i>banzi</i> , Ka. <i>cīhaúm</i> ‘flower’, G. <i>cīhɔm</i> (H)
eggplant	<i>ba kyi</i> /bàkyí/	S. <i>maak<sup>2</sup>k<sup>h</sup>ɿ<sup>1</sup></i> , Kh. <i>ma<sup>4</sup>khy<sup>4</sup></i> , D. <i>maak<sup>3</sup>xə<sup>1</sup></i> ; Lh. <i>pa<sup>F</sup>khyi<sup>L</sup>šɿ<sup>L</sup></i> ‘tomato’, R. <i>baki</i>
fig	<i>mai hpan</i> /màypħaŋ/	S. <i>maj<sup>5</sup>p<sup>h</sup>aj<sup>5</sup>?</i> ‘touchwood, decayed wood’; Ng. <i>maiphang</i>
fire fruit	<i>mak hpai</i> /màkpháy/	S. <i>maak<sup>2</sup>p<sup>h</sup>aj<sup>4</sup></i>
fruit, <i>bf.</i>	<i>mak</i> /màk/	S. <i>maak<sup>2</sup></i> , Kh. <i>maak<sup>4</sup></i> , D. <i>maak<sup>3</sup></i>
garlic	<i>hpăraw</i> /phəro/	S. <i>p<sup>h</sup>ak<sup>4</sup></i> ‘pod’, D. <i>phak<sup>1</sup>lo<sup>1</sup></i> ; Z. <i>pe<sup>5</sup>ro<sup>35</sup></i> , R. <i>gvro</i>
ginger	<i>hkying hka</i> /khyíŋkha/ <sup>18</sup>	S. <i>k<sup>h</sup>ij<sup>1</sup></i> , Kh. <i>khiŋ<sup>4</sup></i> , D. <i>xij<sup>1</sup></i>
gourd	<i>tau</i> /taw/ ~ <i>tau ba</i> /tawbà/	S. <i>taw<sup>3</sup></i> , D. <i>tau<sup>4</sup></i> ; Z. <i>dvau<sup>55</sup></i> ‘bottle’, R. <i>dobá</i>
hog’s plum	<i>măkaw</i> /măkók/	S. <i>maak<sup>2</sup>kək<sup>2</sup></i> ; Lh. <i>mho kaug</i> (K), Lac. <i>ma koug</i>
jackfruit	<i>mak lang</i> /màklaŋ/ ~ <i>mălang</i> /məlaŋ/	S. <i>maak<sup>2</sup>laaŋ<sup>4</sup></i> ; R. <i>mvlángshí</i> , Hp. <i>màŋtàn</i>
lotus	<i>tawng mu</i> /tóŋmu/	S. <i>mo<sup>1</sup></i> , Kh. <i>mo<sup>1</sup></i> (W), D. <i>mo<sup>6</sup></i>
lima bean	<i>htu buk</i> /thùbúk/	S. <i>t<sup>h</sup>o<sup>2</sup>poŋ<sup>1</sup></i>
mango	<i>lămung</i> /ləmuŋ/ ~ <i>mămung</i> /məmũŋ/	S. <i>maak<sup>2</sup>moŋ<sup>3</sup></i> , Kh. <i>ma<sup>4</sup>moŋ<sup>1</sup></i> , D. <i>maak<sup>3</sup>moŋ<sup>6</sup></i> ; Lh. <i>lamung</i> (K), Lac. <i>lamung</i> , Ng. <i>lvung</i> , R. <i>lvúng</i>

<sup>17</sup> This word occurs only as the couplet of *pínlāŋ* ‘betel’ (< Chinese *pīnglang* ‘betel’).<sup>18</sup> Possibly related to Burmese *jín* (WB *khyai<sup>3</sup>*) ‘ginger’.

maize	<i>hkau hpa</i> /khàwphà/	S. <i>k<sup>h</sup>aw<sup>3</sup>paaj<sup>3</sup></i> , Kh. <i>khaupa</i> (P), D. <i>xau<sup>4</sup>faa<sup>5</sup></i>
myrobalan fruit	<i>măna</i> /mənâ/	S. <i>maak<sup>2</sup>naa<sup>5</sup></i> ; Ka. <i>maʔnâci</i> (H), G. <i>mənà<sup>h</sup>i</i> (H)
orange	<i>mak chyawk</i> /màkcòk/	S. <i>maak<sup>2</sup>tsək<sup>4</sup></i> , D. <i>maak<sup>3</sup>tsək<sup>5</sup></i> ; Z. <i>mak<sup>1</sup>zyok<sup>5</sup></i> , Lan. <i>mak<sup>31</sup>tʃək<sup>31</sup></i> , Le. <i>mak<sup>31</sup>tʃək<sup>55</sup></i>
pear	<i>măgaw</i> /məgo/	S. <i>maak<sup>2</sup>kə<sup>3</sup></i> ; Z. <i>mē<sup>5</sup>ko<sup>55</sup></i> , R. <i>mvgo</i>
plum	<i>măhkaw</i> /məkhó/	S. <i>maak<sup>2</sup>k<sup>h</sup>ɔ<sup>1</sup></i> ; Lh. <i>mhokho</i> ” (K), Lac. <i>mo:kho</i> ” ~ <i>moo:kho</i> ”, Ng. <i>mvkhoh</i>
plum	<i>mak man</i> /màkmân/	S. <i>maak<sup>2</sup>man<sup>5</sup></i> , D. <i>maak<sup>3</sup>man<sup>5</sup></i> ; Ka. <i>máʔmánçi</i> (H)
pomegranate	<i>măchyang</i> /məcaŋ/	S. <i>maak<sup>2</sup>tsaŋ<sup>3</sup></i>
quince	<i>mak wau</i> /màkwáw/ ~ <i>măwau</i> /məwáw/	S. <i>maak<sup>2</sup>waaw<sup>4</sup></i>
red lentil	<i>htu leng</i> /thùlɛŋ/	S. <i>t<sup>h</sup>o<sup>2</sup></i> ‘leguminous plant’, <i>lɛŋ<sup>1</sup></i> ‘be red’
red pepper	<i>nprik</i> /n̩prik/	S. <i>maak<sup>2</sup>p<sup>h</sup>it<sup>5</sup></i> , D. <i>maak<sup>3</sup>phet<sup>3</sup></i> ; Z. <i>pik<sup>5</sup></i> ‘spicy’, <i>si<sup>1</sup>-pik<sup>5</sup></i> ‘chillies’
scutch grass	<i>ya sai</i> /yàsáy/	S. <i>jaa<sup>3</sup>s<sup>h</sup>aj<sup>4</sup></i> ; Lh. <i>ya sai</i> ” (K)
Shorea robusta	<i>mai pau</i> /maypàw/ ~ <i>mai bau</i> /maybàw/	S. <i>maj<sup>5</sup>paw<sup>4</sup></i>
sweet orange	<i>mak wan</i> /màkwán/	S. <i>maak<sup>2</sup>waan<sup>1</sup></i> , D. <i>maak<sup>3</sup>waan<sup>1</sup></i> ; R. <i>mokwán</i> ‘pomelo’
sweet pea	<i>htu hkam</i> /thùkhám/	S. <i>t<sup>h</sup>o<sup>2</sup>k<sup>h</sup>aam<sup>4</sup></i>
teak	<i>mai sak</i> /màysàk/	S. <i>maj<sup>5</sup>s<sup>h</sup>ak<sup>4</sup></i> ; Z. <i>mai<sup>21</sup>sak<sup>21</sup></i> (D), Lac. <i>mai sag</i> , Ng. <i>maisak</i> , R. <i>màysvk</i>
thorny bamboo	<i>mai sang</i> /maysaŋ/	S. <i>maj<sup>5</sup>s<sup>h</sup>aaŋ<sup>4</sup></i> , D. <i>mai<sup>5</sup>saŋ<sup>2</sup></i>
tomato	<i>măhkri sum</i> /məkhri <sup>h</sup> sum/ ~ <i>ba hkri</i> /bàkhri/	S. <i>maak<sup>2</sup>k<sup>h</sup>ɔ<sup>1</sup>s<sup>h</sup>om<sup>3</sup></i> ; D. <i>maak<sup>3</sup>xə<sup>1</sup>som<sup>4</sup></i> ; Z. <i>mē<sup>5</sup>ke<sup>55</sup>sum<sup>11</sup></i> , Lh. <i>pa<sup>F</sup>khyi<sup>L</sup>-ši<sup>L</sup></i> , Lac. <i>ba khyi:</i> , R. <i>baki</i>
tree (kind of) <sup>19</sup>	<i>hpak ha</i> /phákhà/	S. <i>p<sup>h</sup>ak<sup>4</sup>haa<sup>3</sup></i>
vegetable, <i>bf.</i>	<i>hpak</i> /phák/	S. <i>p<sup>h</sup>ak<sup>4</sup></i> ‘pod’, Kh. <i>phak<sup>4</sup></i> ‘vegetable’, D. <i>phak<sup>1</sup></i> ‘vegetable’

<sup>19</sup> ‘a kind of tree, the sprouts of which is used as food’

willow	<i>mai hkai</i> /màykhay/	S. <i>ma<sup>j5</sup>k<sup>h</sup>a<sup>j5</sup></i>
wood, <i>bf.</i>	<i>mai</i> /mày/ ~ /may/	S. <i>ma<sup>j5</sup></i> , Kh. <i>may<sup>2</sup></i> , D. <i>mai<sup>5</sup></i>

#### 4.4. Food and drink

cut fine, <i>v.</i>	<i>soi</i> /sòy/	S. <i>s<sup>h</sup>ɔj<sup>4</sup></i> , Kh. <i>sɔI<sup>3</sup></i> (W)
dish (kind of) <sup>20</sup>	<i>sum wan</i> /sùmwan/	S. <i>s<sup>h</sup>om<sup>3</sup></i> ‘be sour’, <i>waan<sup>1</sup></i> ‘be sweet’
dried fish	<i>ba heng</i> /baheŋ/	S. <i>paa<sup>1</sup>heŋ<sup>3</sup></i>
flour	<i>hkau mun</i> /khàwmún/	S. <i>k<sup>h</sup>aw<sup>3</sup>mun<sup>4</sup></i> , D. <i>xau<sup>4</sup>mun<sup>2</sup></i>
food <sup>21</sup>	<i>hkau dum</i> /khàwdùm/	S. <i>k<sup>h</sup>aw<sup>3</sup>tom<sup>3</sup></i>
food <sup>22</sup>	<i>hkau puk</i> /khàwpúk/	S. <i>k<sup>h</sup>aw<sup>3</sup>puk<sup>4</sup></i> ; B. <i>khóbou?</i> (WB <i>khopup</i> ), R. <i>kòbuk</i>
fresh fish	<i>ba lip</i> /balíp/	S. <i>paa<sup>1</sup>lip<sup>4</sup></i>
green sour curry	<i>hpak chyaw</i> /phákcó/	S. <i>p<sup>h</sup>ak<sup>4</sup>tsɔ<sup>4</sup></i>
glutinous rice	<i>hkau lam</i> /khàwlam/	S. <i>k<sup>h</sup>aw<sup>3</sup>laam<sup>1</sup></i>
jaggery	<i>nam htan</i> /nànthan/ <sup>23</sup>	S. <i>nam<sup>5</sup>t<sup>h</sup>aan<sup>1</sup></i> ; Ng. <i>thanh</i> , Ka. <i>thángāk</i>
liquor	<i>lau hku</i> /làwkhù/	S. <i>law<sup>3</sup>hy<sup>2</sup></i> , D. <i>lau<sup>4</sup></i>
meal	<i>ma</i> /má/	D. <i>maa<sup>2</sup></i> ; Z. <i>ma<sup>51</sup></i> (D), Lan. <i>ma<sup>55</sup></i>
meal, <i>clf.</i>	<i>da</i> /dà/	S. <i>taa<sup>2</sup></i> , D. <i>taa<sup>3</sup></i>
mix	<i>law le</i> /lòʔlèʔ/	S. <i>lɔ<sup>4</sup>lɛ<sup>4</sup></i>
noodles	<i>hkau sin</i> /khàwsìn/ <sup>24</sup>	S. <i>k<sup>h</sup>aw<sup>3</sup>s<sup>h</sup>en<sup>3</sup></i> ; Z. <i>hau<sup>55</sup>seng<sup>55</sup></i>
noodles	<i>hkau soi</i> /khàwsóy/ <sup>25</sup>	S. <i>k<sup>h</sup>aw<sup>3</sup>s<sup>h</sup>ɔj<sup>4</sup></i> ; Z. <i>hau<sup>11</sup>soi<sup>31</sup></i> , Lh. <i>khau soe</i> ” (K), Lac. <i>khau: sue</i> ”
palm sugar	<i>nam oi</i> /nàmʔòy/ ~ <i>nam moi</i> /nàmʔmòy/	S. <i>nam<sup>5</sup>ʔɔj<sup>3</sup></i> , Kh. <i>ʔɔy<sup>5</sup></i> ‘sugar cane’; Z. <i>nam<sup>31</sup>oi<sup>11</sup></i>
parched rice	<i>hkau dek</i> /khàwdèk/	S. <i>k<sup>h</sup>aw<sup>3</sup>tɛk<sup>2</sup></i>
pickled fish	<i>ba sum</i> /basùm/	S. <i>paa<sup>1</sup>s<sup>h</sup>om<sup>3</sup></i>
preparation <sup>26</sup>	<i>hkau dam nga</i> /khàwdamŋa/	S. <i>k<sup>h</sup>aw<sup>3</sup>tam<sup>1</sup>ŋaa<sup>4</sup></i>
preparation <sup>27</sup>	<i>hkau dum gwi</i> /khàwdùmguý/	S. <i>k<sup>h</sup>aw<sup>3</sup>tom<sup>3</sup>koj<sup>3</sup></i>
preparation <sup>28</sup>	<i>hkau hkyep</i> /khàwkhýèp/	S. <i>k<sup>h</sup>aw<sup>3</sup>k<sup>h</sup>ɛp<sup>3</sup></i>

<sup>20</sup> ‘a dish made from finely chopped lean beef, hot rice, limes, and spices mixed together’

<sup>21</sup> ‘a glutinous rice preparation in which the rice, brown sugar, and a species of aromatic banana are wrapped in leaves and steamed’

<sup>22</sup> ‘food item made by pounding steamed glutinous rice’

<sup>23</sup> The second syllable is originally from Burmese *thán* (WB *than<sup>3</sup>*) ‘toddy-palm’.

<sup>24</sup> Possibly from Chinese *xiàn* ‘thread’ through Shan.

<sup>25</sup> Possibly from Burmese *khau?shwé* (WB *khokchwai*) ‘noodles’ through Shan.

<sup>26</sup> ‘a preparation of pounded rice and sesame seeds used in making bread’

<sup>27</sup> ‘a preparation of pounded soaked rice and banana wrapped in a leaf steamed, then eaten’

<sup>28</sup> ‘a preparation of puffed rice, brown sugar in sticky form, and sesame seeds pressed together into a bar and sold or eaten as candy’

rice	<i>hkau</i> /khàw/ ~ /khaw/	S. <i>k<sup>h</sup>aw<sup>3</sup></i> , Kh. <i>khaw<sup>5</sup></i> , D. <i>xau<sup>4</sup></i> ; Hp. <i>khàu shwàm</i>
rice <sup>29</sup>	<i>hkau wun</i> /khàwwun/	S. <i>k<sup>h</sup>aw<sup>3</sup></i> ‘rice’, <i>won<sup>1</sup></i> ‘above’
rice <sup>30</sup>	<i>hkau ya ku</i> /khàwyakúʔ/	S. <i>k<sup>h</sup>aw<sup>3</sup>jaa<sup>2</sup>ku<sup>5</sup></i> ‘dish made from rice, jaggery, peanuts and sesame seeds’
salted fish paste	<i>bǎnau</i> /bənàw/	S. <i>paa<sup>1</sup>naw<sup>3</sup></i> ; Z. <i>be<sup>1</sup>nau<sup>11</sup></i> , R. <i>bvnò</i>
serve a meal	<i>ling</i> /lɪŋ/	S. <i>lep<sup>5</sup></i>
sesame oil	<i>nam man</i> /nàmmán/	S. <i>nam<sup>5</sup>man<sup>4</sup></i> , Kh. <i>nam<sup>2</sup>man<sup>3</sup></i> , D. <i>lam<sup>5</sup>man<sup>2</sup></i> ; R. <i>nṽmmýn</i>
sweet potato leaves	<i>hpǎman</i> /phákman/	S. <i>p<sup>h</sup>ak<sup>4</sup>man<sup>4</sup></i>
taste, v.	<i>chyim</i> /cím/	S. <i>tsim<sup>4</sup></i> , Kh. <i>cim<sup>3</sup></i> , D. <i>tsim<sup>2</sup></i> ; Z. <i>zim<sup>31</sup></i>
tobacco	<i>mǎlut</i> /ʔmələt/	S. <i>lut<sup>2</sup></i> ‘to suck, smoke’; R. <i>mvløt</i>
tobacco leaves <sup>31</sup>	<i>gan ya</i> /gànyà/	S. <i>kaan<sup>3</sup>jaa<sup>3</sup></i>
vinegar	<i>nam sum</i> /ʔnàmsùm/	S. <i>nam<sup>5</sup>s<sup>h</sup>om<sup>3</sup></i> ; Lan. <i>nam<sup>31</sup>sum<sup>31</sup></i>

#### 4.5. Clothing and adornments

blanket	<i>hpa jawng</i> /phàjoŋ/	S. <i>p<sup>h</sup>aa<sup>3</sup></i> ; Z. <i>phɿ<sup>21</sup>tfoŋ<sup>55</sup></i> (D), Lan. <i>phə<sup>31</sup>tfoŋ<sup>35</sup></i> , R. <i>paqzung</i>
cloth	<i>hpa</i> /phà/	S. <i>p<sup>h</sup>aa<sup>3</sup></i> ‘covering, waist cloth’, Kh. <i>pha<sup>5</sup></i> (W), D. <i>phaa<sup>4</sup></i>
cotton cloth	<i>man</i> /màn/	S. <i>man<sup>3</sup></i> , D. <i>man<sup>4</sup></i>
dress	<i>hking</i> /khìŋ/	Kh. <i>khing</i> (P)
dye, v.	<i>mak</i> /màk/ <sup>32</sup>	S. <i>mak<sup>4</sup>ʔ</i> ‘make a mark’; Z. <i>mak<sup>5</sup></i> ‘tattoo’, Hp. <i>məʔ</i> ‘ink’
indigo	<i>nam hkawn</i> /nàmkhon/	S. <i>nam<sup>5</sup></i> ‘water’, <i>hon<sup>4</sup></i> ‘indigo’, D. <i>hɔn<sup>4</sup></i> ‘indigo’
leggings	<i>hpa kau</i> /phàkàw/	S. <i>p<sup>h</sup>aa<sup>3</sup>kaaw<sup>3</sup></i>
shoes	<i>hkyep din</i> /khyépɖin/	S. <i>k<sup>h</sup>ep<sup>4</sup>tin<sup>1</sup></i> ‘sandals’, D. <i>kep<sup>3</sup>tin<sup>6</sup></i> ; R. <i>chøpdòn</i> , Ka. <i>heúttín</i> ‘slipper’
shoes	<i>sawk din</i> /sòkdin/	S. <i>s<sup>h</sup>ɔk<sup>4</sup>tin<sup>1</sup></i>

<sup>29</sup> ‘a species of highland paddy with pearly white grain’

<sup>30</sup> ‘rice mixed with millet’

<sup>31</sup> ‘a mixture of tobacco leaves with the chopped stems of the tobacco plant or other vegetable matter such as hemp, which mixture is milder in flavour than the pure tobacco leaves’

<sup>32</sup> Possibly associated with Chinese *mò* ‘ink’.

silk	<i>lai</i> /lày/ ~ <i>lai</i> /lay/	S. <i>laaj</i> <sup>3</sup> , Kh. <i>naay</i> <sup>5</sup> , D. <i>laai</i> <sup>4</sup> ; Z. <i>lai</i> <sup>11</sup> 'silkworm'
socks	<i>mawng din</i> /moŋdin/	S. <i>moŋ</i> <sup>1</sup> <i>tin</i> <sup>1</sup>
tattoo, v.	<i>sam</i> /sám/	S. <i>s<sup>h</sup>am</i> <sup>4</sup> , Kh. <i>sam</i> <sup>3</sup> , D. <i>sam</i> <sup>2</sup>
towel	<i>hpa jet</i> /phàjèt/	S. <i>p<sup>h</sup>aa<sup>3</sup>tset</i> <sup>5</sup> , D. <i>phaa<sup>4</sup>tset</i> <sup>5</sup> ; Z. <i>phɿ</i> <sup>21</sup> <i>tjet</i> <sup>21</sup> (D), Lh. <i>pha</i> <sup>?</sup> <i>cat</i> <sup>F</sup> , Lac. <i>pha joid</i> , Le. <i>phǎ</i> <sup>31</sup> <i>tjet</i> <sup>31</sup> , R. <i>pazit</i> S. <i>kon</i> <sup>1</sup> <i>hoŋ</i> <sup>2</sup> <i>joŋ</i> <sup>5</sup> , D. <i>kon</i> <sup>6</sup> S. <i>s<sup>h</sup>aan</i> <sup>2</sup> <i>leŋ</i> <sup>1</sup>
trousers	<i>gun hkung</i> /gunkhùŋ/	
turkey-red cloth	<i>san leng</i> /sànleŋ/	
<b>4.6. Dwelling</b>		
Bhamo	<i>Manmaw</i> /manmo/	S. <i>maan</i> <sup>3</sup> <i>mɔ</i> <sup>2</sup> ; Lh. <i>man</i> : <i>mo</i> (K)
camp	<i>dăbang</i> /dəbaŋ/	S. <i>paan</i> <sup>1</sup> 'gathering place'
country	<i>mung</i> /mún/ ~ <i>ming</i> /mín/	S. <i>mɿŋ</i> <sup>4</sup> , Kh. <i>mɿŋ</i> <sup>3</sup> , D. <i>məŋ</i> <sup>2</sup> ; Z. <i>meng</i> <sup>11</sup> , Lh. <i>muk</i> <sup>L</sup> , Lac. <i>moug</i> ; Le. <i>məŋ</i> <sup>33</sup> , R. <i>móng</i> S. <i>mɿŋ</i> <sup>4</sup> 'country', <i>tan</i> <sup>4</sup> ? 'place'; Z. <i>mɿŋ</i> <sup>21</sup> <i>tan</i> <sup>55</sup> (D), Lh. <i>muŋ</i> <sup>H</sup> <i>tan</i> <sup>L</sup> , Lac. <i>mung</i> "dain"; <sup>33</sup> Le. <i>maŋ</i> <sup>55</sup> <i>tan</i> <sup>55</sup> , R. <i>móngdàn</i> S. <i>mɿŋ</i> <sup>4</sup> 'country'; Z. <i>meng</i> <sup>11</sup> - <i>gván</i> <sup>11</sup> , Ng. <i>mungkvn</i> , R. <i>mònggàn</i> S. <i>s<sup>h</sup>en</i> <sup>3</sup> <i>wi</i> <sup>1</sup> ; Lh. <i>sinli</i> : (K) S. <i>keŋ</i> <sup>4</sup> <i>tuŋ</i> <sup>1</sup> S. <i>mɿŋ</i> <sup>4</sup> <i>ku</i> <sup>3</sup> S. <i>nam</i> <sup>5</sup> <i>k<sup>h</sup>am</i> <sup>4</sup> ; Lh. <i>nam kham</i> " (K) S. <i>ho</i> <sup>1</sup> , D. <i>ho</i> <sup>1</sup> ; Lh. <i>ho</i> : (K), Lac. <i>khoo</i> : S. <i>k<sup>h</sup>ɔk</i> <sup>3</sup> ? 'enclosure for keeping animals, cage'; Z. <i>gok</i> <sup>1</sup> , R. <i>kok</i> S. <i>tep</i> <sup>4</sup> <i>k<sup>h</sup>aa</i> <sup>4</sup> , D. <i>tep</i> <sup>3</sup> <i>xaa</i> <sup>2</sup> S. <i>maan</i> <sup>3</sup> , Kh. <i>maan</i> <sup>5</sup> , D. <i>maan</i> <sup>4</sup> S. <i>nam</i> <sup>5</sup> 'water', <i>p<sup>h</sup>aaj</i> <sup>1</sup> 'dam', D. <i>faai</i> <sup>1</sup>
country	<i>mung dan</i> /múnɗan/	
country (world)	<i>mung kan</i> /mùŋkàn/	
Hsenwi	<i>Sinli</i> /sìnli/	
Keng Tung	<i>Kyengdung</i> /kyéŋɗuŋ/	
Mogok	<i>Munggut</i> /múŋgùt/	
Nam Hkam	<i>Namhkam</i> /nàmkhám/	
palace	<i>hkaw</i> /kho/	
room	<i>gawk</i> /gòk/	
thatch	<i>dep hka</i> /dépkhá/	
village	<i>man</i> /màn/	
water gate	<i>nam hpai</i> /nàmphay/	

<sup>33</sup> The first syllable is not *moug*: according to my consultant.

## 4.7. Tools

bag, <i>bf.</i>	<i>htung</i> /thuŋ/	S. <i>t<sup>h</sup>oŋ<sup>1</sup></i> , Kh. <i>thoŋ<sup>4</sup></i> , D. <i>thoŋ<sup>1</sup></i> ; Z. <i>tung<sup>31</sup></i> , Lh. <i>thauŋ<sup>F</sup></i> , Lac. <i>thung:</i> , Ng. <i>thung</i> , R. <i>d̥vnggúŋ</i> ‘shoulder bag’, Ka. <i>shíthaúŋ</i> , G. <i>s<sup>hi</sup>iʔt<sup>h</sup>auŋ</i> (H)
bag (cotton)	<i>htung lai</i> /thuŋlày/	S. <i>t<sup>h</sup>oŋ<sup>1</sup></i> ‘bag’, <i>laaj<sup>3</sup></i> ‘silk’
bag (ordinary)	<i>htung hkin</i> /thuŋkhìn/	S. <i>t<sup>h</sup>oŋ<sup>1</sup></i> ‘bag’; Z. <i>tung<sup>31</sup>-hen<sup>55</sup></i>
bag (shoulder)	<i>htung ba</i> /thuŋbá/	S. <i>t<sup>h</sup>oŋ<sup>1</sup>paa<sup>4</sup></i>
basket (fish)	<i>hkik</i> /khík/	S. <i>k<sup>h</sup>uuk<sup>2</sup></i>
basket (rice) <sup>34</sup>	<i>hkau dawm</i> /khàwdom/	S. <i>k<sup>h</sup>aw<sup>3</sup></i> ‘rice’, <i>təm<sup>2</sup></i> ‘large wicker basket’
basket (wicker)	<i>mung</i> /mùŋ/	S. <i>muŋ<sup>1</sup></i> , D. <i>muŋ<sup>6</sup></i> ; Z. <i>mong<sup>11</sup></i>
button	<i>mak dum</i> /màkdùm/	S. <i>maak<sup>2</sup>tum<sup>2</sup></i> , D. <i>tum<sup>3</sup></i>
chain	<i>sai dawng</i> /saydòn/	S. <i>s<sup>h</sup>aaj<sup>1</sup>təŋ<sup>4</sup></i>
chisel, v.	<i>tawk</i> /tók/	S. <i>tək<sup>2</sup></i> , D. <i>tsək<sup>3</sup></i> ; Z. <i>dvok<sup>5</sup></i>
cleaver	<i>hpa</i> /phà/	S. <i>p<sup>h</sup>aa<sup>5</sup></i> , Kh. <i>phaa<sup>2</sup></i>
cup	<i>gawm</i> /gom/	D. <i>kəm<sup>3</sup></i> ; Z. <i>gom<sup>35</sup></i> , Lh. <i>keim</i> (K), Ng. <i>kom</i>
dipper	<i>ka mai</i> /kamay/	S. <i>moj<sup>1</sup>?</i> ‘water dipper’; Z. <i>gva<sup>55</sup>mai<sup>35</sup></i>
fishhook	<i>myit</i> /myít/ ~ <i>myek</i> /ʔmyék/	S. <i>met<sup>4</sup></i> , Kh. <i>met<sup>4</sup></i> , D. <i>met<sup>3</sup></i> ; Z. <i>mit<sup>1</sup></i> ~ <i>mit<sup>5</sup></i> , Lh. <i>ngo</i> : <i>myhad</i> ’ (K), Lac. <i>wim</i> : <i>myhed</i> , R. <i>ngamit</i> , G. <i>met</i> ‘to fish’ (H)
flint	<i>hpai</i> /phày/	S. <i>p<sup>h</sup>aj<sup>4</sup></i> ‘fire’, Kh. <i>phay<sup>3</sup></i> ‘fire’, D. <i>fai<sup>2</sup></i> ‘fire’; Z. <i>pai<sup>11</sup>-kyet<sup>1</sup></i> ‘match’
girth	<i>sai kyet</i> /saykyét/	S. <i>s<sup>h</sup>aaj<sup>1</sup>ket<sup>4</sup>maa<sup>5</sup></i>
hatchet	<i>hpa tung</i> /phàtúŋ/	S. <i>p<sup>h</sup>aa<sup>5</sup></i> ‘cleaver’, D. <i>phaa<sup>5</sup>tum<sup>3</sup></i> ; Z. <i>pe<sup>1</sup>dvung<sup>31</sup></i>
lamp	<i>pyen ding</i> /pyéndìŋ/	S. <i>ten<sup>4</sup>?</i> ‘candle’; Z. <i>byen<sup>31</sup>-din<sup>11</sup></i>
mat	<i>sat</i> /sát/	S. <i>s<sup>h</sup>aat<sup>2</sup></i> , D. <i>saat<sup>3</sup></i>
mill	<i>lui</i> /luy/	S. <i>loj<sup>1</sup></i> ; Z. <i>lui<sup>35</sup></i> , Lh. <i>lui</i> ’ (K), Le. <i>lui<sup>33</sup></i>

<sup>34</sup> ‘a small basket for storage of rice’



mirror	<i>jam na /jàmna/</i>	S. <i>tsam</i> <sup>3</sup> ‘mirror’, <i>naa</i> <sup>3</sup> ‘face’, Kh. <i>naa</i> <sup>5</sup> ‘face’, D. <i>tsam</i> <sup>6</sup> ‘mirror’, <i>laa</i> <sup>4</sup> ‘face’; Z. <i>man</i> <sup>31</sup> - <i>zyam</i> <sup>31</sup> , (D) <i>mjoʔ<sup>31</sup>tfam<sup>51</sup></i> , Lh. <i>jham na</i> : (K), Lan. <i>mjoʔ<sup>31</sup>tfam<sup>31</sup></i> , R. <i>jv̄mna</i> , Ka. <i>zàn</i>
mosquito net	<i>sut /sút/ ~ əsut /ʔəsút/</i>	S. <i>s<sup>h</sup>ut<sup>4</sup></i> , Kh. <i>sup<sup>4</sup></i> , D. <i>sut<sup>1</sup></i> ; Z. <i>sut<sup>5</sup> ~ sun<sup>31</sup></i> , R. <i>søt</i> , Ka. <i>sūt</i> , G. <i>s<sup>h</sup>ut</i> (H)
nail	<i>mai na /màyná/</i>	S. <i>maaj<sup>2</sup>naa<sup>4</sup></i> ; Z. <i>mai<sup>11</sup>na<sup>31</sup> ~ mai<sup>31</sup>na<sup>11</sup> ~ mai<sup>31</sup>na<sup>31</sup></i> , Lh. <i>mai na</i> ” (K), Ka. <i>maigná</i> (H), G. <i>menná</i> (H)
paddle	<i>dak hpai /dàkpháy/</i>	S. <i>p<sup>h</sup>aaj<sup>4</sup></i> , D. <i>faai<sup>2</sup></i>
paddy mill	<i>hkau lui /khàwluy/</i>	S. <i>k<sup>h</sup>aw<sup>3</sup>loj<sup>1</sup></i>
pierce, v.	<i>chyawk /cók/</i>	S. <i>tsok<sup>5</sup></i>
pin	<i>se /sè/, sumse /sùmsè/ ‘bolt’</i>	S. <i>s<sup>h</sup>ε<sup>3</sup></i> ‘bolt’
ploughshare	<i>na jawn /nàjon/</i>	S. <i>naa<sup>4</sup></i> ‘rice’, <i>tsɔn<sup>5</sup></i> ‘to dig out’
ploughshare	<i>na htai /nàthay/</i>	S. <i>naa<sup>4</sup></i> ‘rice field, t <sup>h</sup> aj <sup>1</sup> ‘plough’, thAI <sup>4</sup> (W), B. thè (WB <i>thay</i> )
pot	<i>maw /mò/</i>	S. <i>mɔ<sup>3</sup></i> , Kh. <i>mo<sup>5</sup></i> , D. <i>mo<sup>4</sup></i>
pot (metallic)	Nu. <i>maw hkang /moo<sup>4</sup>khaaj<sup>5</sup>/</i>	S. <i>mɔ<sup>3</sup>k<sup>h</sup>aaj<sup>1</sup></i>
rope	<i>jik /jìk/</i>	S. <i>tsɿk<sup>3</sup></i> , D. <i>tsək<sup>5</sup></i>
rope, bf.	<i>sai /say/</i>	S. <i>s<sup>h</sup>aaj<sup>1</sup></i> , Kh. <i>saay<sup>4</sup></i> , D. <i>saai<sup>1</sup></i> ; Z. <i>sai<sup>35</sup>haaj<sup>35</sup></i> (M)
rope <sup>35</sup>	<i>sai gang /saygàŋ/</i>	<i>s<sup>h</sup>aaj<sup>1</sup>kaaj<sup>2</sup>?</i> ‘string of a crossbow’
rudder	<i>li hkang /likhánj<sup>36</sup>/</i>	S. <i>haaj<sup>1</sup></i> ‘tail’, Kh. <i>haaj<sup>4</sup></i> ‘tail’, D. <i>haaj<sup>1</sup></i>
scales	<i>yoi /yòy/</i>	S. <i>jɔj<sup>4</sup></i> ; Z. <i>ngun<sup>31</sup>-yoi<sup>11</sup></i>
small bottle	<i>kawk /kók/<sup>37</sup></i>	S. <i>kək<sup>4</sup></i> , D. <i>kək<sup>3</sup></i> ; Z. <i>gok<sup>55</sup></i> , R. <i>gok</i>
small box	<i>yep /ʔyép/</i>	S. <i>ʔep<sup>2</sup></i> , Kh. <i>ʔep<sup>4</sup></i> , D. <i>ʔep<sup>3</sup></i>
spinning wheel	<i>gawng /gòn/</i>	S. <i>koŋ<sup>4</sup></i> , D. <i>koŋ<sup>2</sup></i>
spinning wheel	<i>gawng sai /gònəsày/</i>	S. <i>koŋ<sup>4</sup></i> ‘spinning wheel’, <i>s<sup>h</sup>aaj<sup>1</sup></i>
belt		‘rope’
spoon	<i>chyaw /cò/</i>	S. <i>tsɔ<sup>5</sup></i> , D. <i>tso<sup>5</sup></i> ; Z. <i>zvyo<sup>31</sup></i> , Ka. <i>húnzaú</i> , G. <i>hánstót</i> (H)
sword	<i>lap lang /láplán/</i>	S. <i>laap<sup>2</sup></i> , Kh. <i>naap<sup>4</sup></i> , D. <i>laap<sup>3</sup></i>

<sup>35</sup> ‘ropes holding the baskets, as of a bullock-load, in place’<sup>36</sup> The first syllable *li* expresses ‘boat’ in Jinghpaw.<sup>37</sup> Possibly related to Burmese *khwε?* ‘cup’ (WB *khwak*).

table <sup>38</sup>	<i>ku</i> /kù/	S. <i>ku</i> <sup>2</sup> ‘block, bench, couch, bed’, Kh. <i>ku</i> <sup>4</sup> ‘bed’, D. <i>ku</i> <sup>3</sup> ‘bed’; Z. <i>ku</i> <sup>55</sup> ‘bed’ (D), R. <i>yøpgù</i> ‘bed’
teapot	<i>ngoi</i> /ŋóy/	S. <i>ŋɔj</i> <sup>4</sup> , D. <i>ŋɔi</i> <sup>2</sup> ; Lh. <i>ngoe</i> ” (K), R. <i>ngoy</i>
tobacco pipe	<i>maw ya</i> /mòya/	S. <i>mɔ</i> <sup>3</sup> <i>jaa</i> <sup>3</sup>
tool (kind of) <sup>39</sup>	<i>lang chyo</i> /láŋcý/	D. <i>laŋ</i> <sup>55</sup> <i>tsɔi</i> <sup>31</sup> (DX)
trap	<i>sai</i> /say/	S. <i>s<sup>h</sup>aj</i> <sup>4</sup> , D. <i>sai</i> <sup>2</sup> ; R. <i>se</i>
umbrella	<i>jawng</i> /jɔŋ/	S. <i>tsɔŋ</i> <sup>3</sup> , Kh. <i>tsɔŋ</i> <sup>5</sup> (W), D. <i>tsɔŋ</i> <sup>6</sup> ; Z. <i>zyong</i> <sup>35</sup> , R. <i>zung</i>
umbrella <sup>40</sup>	<i>jawng hkam</i> /jòŋkhám/	S. <i>tsɔŋ</i> <sup>3</sup> <i>k<sup>h</sup>am</i> <sup>4</sup>
water bottle	<i>nam tau</i> /nàmtàw/	S. <i>nam</i> <sup>5</sup> <i>taw</i> <sup>3</sup> , Kh. <i>nAm</i> <sup>2</sup> <i>tAU</i> <sup>5</sup> (W), D. <i>tau</i> <sup>4</sup> ; Z. <i>dvau</i> <sup>55</sup>
wheel	<i>leng</i> /lèŋ/	S. <i>leŋ</i> <sup>1</sup> ‘bicycle’, Kh. <i>leŋ</i> <sup>4</sup> ‘vehicle’ (W), D. <i>leŋ</i> <sup>1</sup> ‘cart’; Z. <i>lyeng</i> <sup>31</sup> , Ng. <i>lhyingh</i> ‘cart’, R. <i>lìng</i> ‘cart’
yoke	<i>ek</i> /ʔék/	S. <i>ʔek</i> <sup>2</sup> , Kh. <i>ʔek</i> <sup>4</sup> , D. <i>ʔek</i> <sup>3</sup>
yoke	<i>gan</i> /gàn/	S. <i>kaan</i> <sup>4</sup> ; Z. <i>xap</i> <sup>21</sup> <i>kān</i> <sup>51</sup>

#### 4.8. Community, occupation and production

barren	<i>hkam</i> /khám/ ~ <i>ham</i> /hám/	S. <i>haam</i> <sup>4</sup> ‘be empty’, Kh. <i>ham</i> <sup>3</sup> (W)
barren field	<i>na hkam</i> /nákhám/	S. <i>naa</i> <sup>4</sup> ‘rice field’, <i>haam</i> <sup>4</sup> ‘be empty’; Z. <i>na</i> <sup>51</sup> <i>xam</i> <sup>51</sup> (D), Lh. <i>na kham</i> (K), R. <i>nàhàm</i>
blacksmith	<i>jang lek</i> /jàŋlék/	S. <i>tsaaŋ</i> <sup>3</sup> <i>lek</i> <sup>4</sup> , D. <i>tsaaŋ</i> <sup>6</sup> <i>lek</i> <sup>3</sup>
blighted rice	<i>hkau pyi</i> /khàwpyì/	S. <i>k<sup>h</sup>aw</i> <sup>3</sup> <i>p<sup>h</sup>i</i> <sup>1</sup>
butcher	<i>jang nwi</i> /jàŋnuy/	S. <i>tsaaŋ</i> <sup>3</sup> <i>nɿ</i> <sup>5</sup>
compartment <sup>41</sup>	<i>lawk</i> /lòk/ ~ <i>lung</i> /lúŋ/	S. <i>lɔk</i> <sup>4</sup> , Kh. <i>lɔk</i> <sup>4</sup> (W), D. <i>lɔk</i> <sup>1</sup> ; Z. <i>lok</i> <sup>1</sup>
compartment <sup>42</sup>	<i>na lawk</i> /nàlòk/	S. <i>lɔk</i> <sup>4</sup> <i>naa</i> <sup>4</sup>
coppersmith	<i>jang tawng</i> /jàŋtón/	S. <i>tsaaŋ</i> <sup>3</sup> ‘be skilled in’, <i>tɔŋ</i> <sup>4</sup> ‘metallic substance’
enclosure	<i>hkawk</i> /khòk/	S. <i>k<sup>h</sup>ɔk</i> <sup>3</sup> , Kh. <i>khɔk</i> <sup>1</sup> ‘room’ (W), D. <i>xɔk</i> <sup>5</sup> ; R. <i>kok</i> ‘room’

<sup>38</sup> ‘a platform, anything in the shape of a table, e.g. *ʔyúp-kù* ‘bed’ (lit. sleep-table), *ɛá-kù* ‘dining table’ (lit. eat-table), *làyka-kù* ‘writing desk’ (lit. letter-table)’

<sup>39</sup> ‘a tool consisting of a bamboo pole about six feet in length with a hook at the end (e.g. as for pitching straw)’

<sup>40</sup> ‘golden umbrella used by royalties or monks for special occasions’

<sup>41</sup> ‘a section, compartment of a lowland field’

<sup>42</sup> ‘a compartment of a rice-field, bounded by bunds’

enclosure	<i>wing</i> /wín/	S. <i>weŋ</i> <sup>4</sup> ‘town’, D. <i>weŋ</i> <sup>2</sup> ‘city’
fence	<i>da hka</i> /dàkhà/	D. <i>ta</i> <sup>33</sup> <i>xa</i> <sup>11</sup> (M); Z. <i>ta</i> <sup>21</sup> <i>kha</i> <sup>55</sup> (D)
garden	<i>sun</i> /sún/	S. <i>s<sup>h</sup>on</i> <sup>1</sup> , Kh. <i>son</i> <sup>4</sup> , D. <i>son</i> <sup>1</sup> ; Z. <i>sun</i> <sup>55</sup> , R. <i>sún</i>
goldsmith	<i>jang hkam</i> /jàŋkhám/	S. <i>tsaaj</i> <sup>3</sup> <i>k<sup>h</sup>am</i> <sup>4</sup>
landing place <sup>43</sup>	<i>da nam</i> /dànám/	S. <i>taa</i> <sup>3</sup> <i>nam</i> <sup>5</sup> , Kh. <i>taa</i> <sup>1</sup> , D. <i>taa</i> <sup>6</sup>
lowland paddy	<i>hkau lung</i> /khàwlun/	S. <i>k<sup>h</sup>aw</i> <sup>3</sup> <i>lon</i> <sup>1</sup>
manure	<i>nam hpun</i> /nàmphun/	S. <i>nam</i> <sup>5</sup> ‘water’, <i>p<sup>h</sup>un</i> <sup>2</sup> ‘manure’, D. <i>fun</i> <sup>3</sup> ; Z. <i>nam</i> <sup>31</sup> - <i>pun</i> <sup>55</sup> , R. <i>nìmpun</i>
martial art <sup>44</sup>	<i>jau hkyen</i> /jàwkhýèn/	S. <i>tsaw</i> <sup>3</sup> ‘master’, <i>k<sup>h</sup>en</i> <sup>1</sup> ‘hand’; Lh. <i>jau khyen</i> , Lac. <i>jau khyen</i>
mine	<i>maw</i> /mo/	S. <i>mɔ</i> <sup>2</sup> , Kh. <i>mo</i> <sup>4</sup> (W), D. <i>mɔ</i> <sup>3</sup> ; Lh. <i>mo</i> : (K), Lac. <i>mo</i> :, Ng. <i>luk mau</i> , R. <i>shémo</i> ‘gold mine’, Ka. <i>maū</i> , G. <i>mɔ</i> ‘Kadu’ (H)
paddy (red) <sup>45</sup>	<i>leng</i> /leŋ/	S. <i>leŋ</i> <sup>1</sup> ‘be red’, Kh. <i>neŋ</i> <sup>1</sup> ‘be red’, D. <i>leŋ</i> <sup>6</sup> ‘be red’; Hp. <i>lɔŋ</i> ‘yellow’
paddy embankment	<i>na dung</i> /nàdùn/	D. <i>toŋ</i> <sup>33</sup> <i>la</i> <sup>55</sup> (DX)
rice field	<i>na</i> /nà/	S. <i>naa</i> <sup>1</sup> , Kh. <i>naa</i> <sup>3</sup> , D. <i>laa</i> <sup>2</sup>
rice field	<i>hkau na</i> /khawnà/	S. <i>k<sup>h</sup>aw</i> <sup>3</sup> ‘rice’, <i>naa</i> <sup>1</sup> ‘rice field’; R. <i>koná</i>
rice nursery	<i>hkau ga</i> /khàwgà/	S. <i>k<sup>h</sup>aw</i> <sup>3</sup> ‘rice’, D. <i>ka</i> <sup>31</sup> <i>xau</i> <sup>31</sup> ‘rice sprouts’ (DX)
ridge <sup>46</sup>	<i>te</i> /té/	S. <i>tɛ</i> <sup>4</sup> , D. <i>tɛ</i> <sup>2</sup> ; Lh. <i>te</i> : (K)
ridge <sup>47</sup>	<i>te na</i> /téná/	S. <i>tɛ</i> <sup>4</sup> <i>naa</i> <sup>4</sup> , Kh. <i>tɛ</i> <sup>3</sup> <i>naa</i> <sup>3</sup> ; R. <i>nádé</i> , Ka. <i>lɔpátəná</i> (H), G. <i>lɔpáténá</i> (H)
seedlings	<i>ka</i> /kà/	S. <i>kaa</i> <sup>3</sup> , Kh. <i>ka</i> <sup>5</sup> (W), D. <i>kaa</i> <sup>4</sup>
silversmith	<i>jang ngin</i> /jàŋŋín/	S. <i>tsaaj</i> <sup>3</sup> <i>ŋum</i> <sup>4</sup> , Kh. <i>ŋum</i> <sup>3</sup> ‘silver’ (W), D. <i>ŋən</i> <sup>2</sup> ‘silver’
smith, bf.	<i>jang</i> /jàŋ/ <sup>48</sup>	S. <i>tsaaj</i> <sup>3</sup> ‘be skilled in’
terraced fields <sup>49</sup>	<i>na kawng</i> /nákòn/	S. <i>naa</i> <sup>4</sup> ‘rice’, <i>kɔŋ</i> <sup>4</sup> ‘ridge, or elevated part of anything’

<sup>43</sup> ‘a ferry or public landing place on the bank of a river or of a lake’

<sup>44</sup> ‘the martial art of self-defence, which is practised in several varieties (including principally a variety using swords or spears, one using wooden staffs, and one using no weapons)’

<sup>45</sup> ‘an insect-caused blight of paddy in which the stalks turn red, lit. red’

<sup>46</sup> ‘a ridge of earth or mound separating the locks or sections of a paddy field’

<sup>47</sup> ‘the terrace of a paddy field’

<sup>48</sup> Possibly from Chinese *jiàng* ‘craftsman’ through Shan.

<sup>49</sup> ‘terraced fields on the side of a hill watered by irrigation’

timber log	<i>mai lung</i> /maylùŋ/	S. <i>maj</i> <sup>5</sup> ‘wood’, <i>log</i> <sup>1</sup> ‘forest’; Lh. <i>mai lung</i> : (K), Lac. <i>mai lung</i> :
------------	--------------------------	--

#### 4.9. Commerce and trade

bazaar	<i>hkau gat</i> /khàwɡát/	S. <i>k<sup>h</sup>aw</i> <sup>3</sup> ‘rice’, <i>kaat</i> <sup>2</sup> ‘market’
market <sup>50</sup>	<i>gat</i> /ɡát/	S. <i>kaat</i> <sup>2</sup> , D. <i>kaat</i> <sup>3</sup> ; Lh. <i>kāt</i> <sup>H</sup> , R. <i>gvt</i>
measure <sup>51</sup>	<i>jaw</i> /jó/	S. <i>tsɔ</i> <sup>4</sup>
measure <sup>52</sup>	<i>hkan</i> /khán/	S. <i>k<sup>h</sup>an</i> <sup>1</sup>
measure <sup>53</sup>	<i>hpak</i> /phák/	S. <i>p<sup>h</sup>aak</i> <sup>2</sup> ? ‘part, side (as of a body of water), division or portion of anything divided lengthwise into halves or fourths’; Z. <i>pak</i> <sup>1</sup> ? ‘half a pound’, Lh. <i>phag</i> ! (K)
measure <sup>54</sup>	<i>joi</i> /joy/	S. <i>tsɔj</i> <sup>5</sup> ; Lh. <i>joe</i> : (K), Lan. <i>tʃoi</i> <sup>35</sup> ‘steelyard’, Lac. <i>joi</i> :, R. <i>joy</i>
price	<i>ka</i> /kà/	S. <i>kaa</i> <sup>3</sup> , D. <i>kaa</i> <sup>6</sup> ; R. <i>ká</i> ‘debt’
tax	<i>hkan</i> /khàn/ <sup>55</sup> ~ <i>hkan se</i> /khanse/ ~ <i>hkan si</i> /khansi/	S. <i>k<sup>h</sup>an</i> <sup>1</sup> ‘price’, D. <i>xan</i> <sup>35</sup> <sub>se</sub> <sup>55</sup> (DX); Z. <i>kāŋ</i> <sup>21</sup> (D), Lh. <i>khaŋ</i> <sup>L</sup> , Lac. <i>khaung</i> ” doug:
trade	<i>hpăga</i> /phəga/	S. <i>kaa</i> <sup>5</sup> ‘to trade’; Z. <i>pe</i> <sup>5</sup> <i>ga</i> <sup>35</sup> ‘trade’, Lh. <i>phăka</i> <sup>L</sup> <i>pyu</i> <sup>F</sup> (trade-person), Lac. <i>pha</i> : <i>ga</i> :, R. <i>pvga</i>
trade, v.	<i>ga</i> /ga/	S. <i>kaa</i> <sup>5</sup> , D. <i>kaa</i> <sup>5</sup> ; R. <i>ga</i>

#### 4.10. Communication and transportation

bridle	<i>gak</i> /gàk/	S. <i>kak</i> <sup>5</sup>
cart	<i>law</i> /lò/	S. <i>lɔ</i> <sup>5</sup> , Kh. <i>lɔ</i> <sup>2</sup> , D. <i>lo</i> <sup>5</sup> ; Hp. <i>lɔ</i>
journey <sup>56</sup>	<i>hkau</i> /khàw/	S. <i>k<sup>h</sup>aaw</i> <sup>4</sup>
package	<i>chyawk</i> /cók/	S. <i>tsɔk</i> <sup>4</sup> ; Z. <i>zyok</i> <sup>5</sup>
way <sup>57</sup>	<i>tang</i> /táj/	S. <i>taaj</i> <sup>4</sup> , Kh. <i>taŋ</i> <sup>3</sup> (W), D. <i>taaj</i> <sup>2</sup>

<sup>50</sup> Zaiwa *gai*<sup>31</sup> ‘market’ is from Chinese local dialect *gai* (Lustig 2010: 135), which is also found in Lhaovo spoken in Shan State *gai* (Yabu 2000: 34), Langsu *kai*<sup>31</sup>, Leqi *kei*<sup>55</sup> and Ngochang *kaih*.

<sup>51</sup> ‘a measure of volume or capacity equal to ten baskets’

<sup>52</sup> ‘a unit of weight equal to the weight of ten silver Rupees, one tenth of a viss’

<sup>53</sup> ‘measure of capacity equal to two *järe*, about four cups by volume’

<sup>54</sup> ‘one viss, a unit of weight equal to 1.63 kg’

<sup>55</sup> Possibly from Burmese *ʔăkhùn* (WB *akhwan*) through Shan.

<sup>56</sup> ‘stage of a journey, a day’s journey, generally with a pack-bullock’

<sup>57</sup> ‘a way, a road; an antecedent, a cause’

## 4.11. Culture and entertainment

book	<i>laika</i> /làyka/ <sup>58</sup>	S. <i>lik</i> <sup>3</sup> <i>laaj</i> <sup>4</sup> , D. <i>lik</i> <sup>5</sup> <i>laai</i> <sup>2</sup> ; Z. <i>lai</i> <sup>11</sup> <i>gva</i> <sup>55</sup> , R. <i>lèga</i>
book	<i>lik</i> /lík/	S. <i>lik</i> <sup>3</sup> <i>laaj</i> <sup>4</sup> ‘book’, D. <i>lik</i> <sup>5</sup> <i>laai</i> <sup>2</sup> ‘book’
book, <i>clf.</i>	<i>bap</i> /bàp/	S. <i>pap</i> <sup>5</sup> ‘book’, Kh. <i>pap</i> <sup>1</sup> ‘book’, D. <i>pap</i> <sup>5</sup> ‘book’
culture	<i>laili laika</i> /làyilàyka/	S. <i>lik</i> <sup>3</sup> <i>laaj</i> <sup>4</sup> ‘book’ Z. <i>lai</i> <sup>21</sup> <i>li</i> <sup>21</sup> <i>pha</i> <sup>21</sup> <i>tji</i> <sup>51</sup> (D)
flute	<i>pyi</i> /pyi/, <i>sumpyi</i> /sumpyi/ ~ <i>sampyi</i> /sampyi/	S. <i>pi</i> <sup>2</sup> , Kh. <i>pi</i> <sup>4</sup> (W), D. <i>pi</i> <sup>3</sup> ; Z. <i>bvi</i> <sup>55</sup> , <i>sam</i> <sup>55</sup> <i>pji</i> <sup>55</sup> (D), R. <i>bilóm</i> , <i>biman</i> ‘flute’, <i>bishun</i> ‘long flute’
gamble	<i>taw</i> /tò/ <sup>59</sup>	S. <i>to</i> <sup>2</sup> , D. <i>to</i> <sup>3</sup> ; Lh. <i>to</i> ”, R. <i>dó</i>
handiwork	<i>lai</i> /lày/	S. <i>laaj</i> <sup>3</sup> ? ‘silk’
harp	<i>tingse</i> /tiŋsè/	S. <i>tiŋ</i> <sup>2</sup> ‘harp’, <i>se</i> <sup>2</sup> ‘to play’; Z. <i>tīŋ</i> <sup>55</sup> <i>je</i> <sup>51</sup> (D)
manner	<i>lai</i> /lày/	S. <i>laaj</i> <sup>4</sup>
manner <sup>60</sup>	<i>tang het</i> /tánghét/	S. <i>taaj</i> <sup>4</sup> <i>het</i> <sup>4</sup>
model	<i>lai len</i> /làylen/	S. <i>laaj</i> <sup>4</sup> <i>len</i> <sup>4</sup>
paper	<i>je</i> /jè/ <sup>61</sup>	S. <i>tse</i> <sup>3</sup> , D. <i>tse</i> <sup>4</sup>
paper	<i>mai sau</i> /màysàw/ ~ <i>mau sau</i> /màwsàw/	S. <i>maj</i> <sup>5</sup> <i>s<sup>h</sup>aw</i> <sup>3</sup> ‘slender piece of wood’; Z. <i>mau</i> <sup>11</sup> <i>sau</i> <sup>11</sup> , Lh. <i>muk</i> <sup>F</sup> <i>suk</i> <sup>H</sup> , Lac. <i>moug</i> : <i>soug</i> ”, Le. <i>mou</i> <sup>53</sup> <i>sou</i> <sup>55</sup> , Ng. <i>maussauh</i> , R. <i>mèsò</i>

## 4.12. Cults, customs and socializing

day of fasting	<i>wan gam</i> /wángam/	S. <i>wan</i> <sup>4</sup> <i>kam</i> <sup>1</sup> <i>s<sup>h</sup>in</i> <sup>1</sup>
pagoda	<i>gawng ngu</i> /gòŋŋù/ <sup>62</sup>	S. <i>kəŋ</i> <sup>4</sup> <i>mu</i> <sup>4</sup> , D. <i>kəŋ</i> <sup>2</sup> <i>mu</i> <sup>2</sup> ; Z. <i>koŋ</i> <sup>21</sup> <i>ŋu</i> <sup>21</sup> (D), Lh. <i>gug ngung</i> (K)

<sup>58</sup> The second syllable *ka* is a Jinghpaw word which means ‘to write’.

<sup>59</sup> Possibly from Chinese *dū* ‘to gamble’ through Shan.

<sup>60</sup> ‘manner of doing, one’s gait, posture, deportment’

<sup>61</sup> This word is mostly used in northern Jinghpaw dialects.

<sup>62</sup> Possibly from Burmese *káunhmū* ‘good deed’ (WB *koí<sup>3</sup>mhu*) through Shan (Professor Mathias Jenny, p.c., 2016). The mismatch of the initial consonant of the second syllable between Shan and Jinghpaw can be accounted for in terms of progressive assimilation occurred in Jinghpaw. This feature shared among Kachin languages suggests that this item was introduced into non-Jinghpaw Kachin languages through Jinghpaw.

spirit	<i>hpyi</i> /phyi/	Lac. <i>goung ngoo</i> , Ng. <i>gongs-ngu</i> , Li. <i>kong<sup>5</sup>mu<sup>5</sup></i> S. <i>p<sup>h</sup>i<sup>1</sup></i> , Kh. <i>phi<sup>4</sup></i> , D. <i>phi<sup>1</sup></i> ; Z. <i>pi<sup>11</sup></i> , G. <i>p<sup>h</sup>itun</i> ‘evil spirit’ (H)
spirit country	<i>mung hpyi</i> /mùŋphyi/	S. <i>mɿŋ<sup>4</sup>p<sup>h</sup>i<sup>1</sup></i>
vanish <sup>63</sup>	<i>hpoi</i> /phòy/	D. <i>phoi<sup>6</sup></i> ‘be fragile’; Z. <i>poi<sup>11</sup></i> , R. <i>pòy</i>

## 4.13. Human body

body	Nu. <i>hking</i> /khiŋ/	S. <i>k<sup>h</sup>iŋ<sup>4</sup></i> , Kh. <i>khIŋ<sup>3</sup></i> (W), D. <i>xij<sup>2</sup></i>
semen	<i>nam ngan</i> /nàmŋan/	S. <i>nam<sup>5</sup></i> ‘water’, <i>ŋaan<sup>1</sup></i> ‘be fertile, not castrated’, D. <i>lam<sup>5</sup>ŋaan<sup>1</sup></i>
weight	<i>nam nak</i> /nàmnák/	S. <i>nam<sup>5</sup>nak<sup>4</sup></i>

## 4.14. Life, sickness and death

bullet	<i>mak</i> /màk/	S. <i>maak<sup>2</sup></i> ‘round thing’ <sup>64</sup>
bullet	<i>mak lung</i> /màklùŋ/	S. <i>maak<sup>2</sup>loŋ<sup>1</sup></i>
cannon	<i>gawng dang</i> /gòŋdàŋ/	S. <i>kəŋ<sup>3</sup></i> ‘gun’; Z. <i>gong<sup>11</sup>dang<sup>11</sup></i> ‘large underground fireworks’
cholera	<i>sawng hkun</i> /soŋkhun/	S. <i>s<sup>h</sup>əŋ<sup>1</sup>hon<sup>1</sup></i> ; Z. <i>soŋ<sup>55</sup>xun<sup>55</sup>mo<sup>55</sup></i> (D), Lh. <i>saung: khun:</i> (K), Lac. <i>seing: khein:</i> , R. <i>songkun</i>
dropsy	<i>nam ling</i> /nàmlíŋ/	S. <i>nam<sup>5</sup>lɿŋ<sup>1</sup></i>
epilepsy	<i>ma mu</i> /màmu/	S. <i>maa<sup>3</sup>mu<sup>1</sup></i> , Kh. <i>maa<sup>5</sup></i> ‘be crazy’, <i>mu<sup>4</sup></i> ‘pig’, D. <i>maa<sup>4</sup>mu<sup>1</sup></i>
foster	<i>bau</i> /baw/	S. <i>paur<sup>5</sup></i> , Kh. <i>pAur<sup>2</sup></i> (W); Z. <i>bau<sup>11</sup></i>
gun	<i>gawng</i> /gòŋ/	S. <i>kəŋ<sup>3</sup></i> ‘gun’, D. <i>kəŋ<sup>4</sup></i>
gun (kind of) <sup>65</sup>	<i>gawng hpai</i> /gòŋphay/	S. <i>kəŋ<sup>3</sup></i> ‘gun’, <i>p<sup>h</sup>aj<sup>4</sup></i> ‘fire’
herbal medicine	<i>ya ya</i> /yàya/	S. <i>jaa<sup>3</sup>jaa<sup>1</sup></i> ; Z. <i>ya<sup>11</sup>ya<sup>35</sup></i>
leprosy	<i>dut</i> /dút/	S. <i>tut<sup>3</sup></i> , D. <i>tut<sup>5</sup></i>
machine gun	<i>gawng jawk</i> /gòŋjòk/	S. <i>kəŋ<sup>3</sup>tsaak<sup>3</sup></i>
measles	<i>mǎling</i> /mǎliŋ/	S. <i>maak<sup>2</sup>leŋ<sup>1</sup></i> ; R. <i>mǎrlòng</i>
medicine, bf.	<i>ya</i> /yà/	S. <i>jaa<sup>1</sup></i> , Kh. <i>yaa<sup>5</sup></i> , D. <i>yaa<sup>4</sup></i> ; Z. <i>ya<sup>11</sup></i> ‘to cure’
overcome	<i>pye</i> /pyè/	S. <i>pɛ<sup>5</sup></i>

<sup>63</sup> ‘to vanish, disappear, to lose potency, become inane, to have vanished or been lost (e.g. as luck, glory, honour, flavour)’

<sup>64</sup> ‘fruit; also applied to anything round’

<sup>65</sup> ‘a primitive kind of gun that requires a firebrand for discharging’

percussion cap <sup>66</sup>	<i>mak hpai</i> /màkpháy/	S. <i>maak</i> <sup>2</sup> ‘round thing’, <i>p<sup>h</sup>aj</i> <sup>4</sup> ‘fire’
poison	<i>gung</i> /guŋ/	S. <i>kəŋ</i> <sup>5</sup> , Kh. <i>kəŋ</i> <sup>2</sup> (W), D. <i>kəŋ</i> <sup>5</sup>
powder flask	<i>yam kaw</i> /ʔyàmkók/	S. <i>jaam</i> <sup>4</sup> ‘gunpowder’, <i>kək</i> <sup>4</sup> ‘cup’; R. <i>yàm</i> ‘gunpowder’
syphilis	<i>kālang gyi</i> /kəlaŋ gyi/	S. <i>ka<sup>1</sup>laaŋ</i> <sup>2</sup>

#### 4.15. Types of people

abbot	<i>mun jau</i> /munjàw/	S. <i>mun<sup>1</sup>tsaw</i> <sup>3</sup>
beggar	<i>kun yawn</i> /kúnyón/	S. <i>kon</i> <sup>4</sup> <i>jən</i> <sup>4</sup> , D. <i>yən</i> <sup>2</sup> ‘to beg’
beggar	<i>maw hpyi</i> /mophyi/	D. <i>mo<sup>1</sup>phi</i> <sup>1</sup> ; Z. <i>mo</i> <sup>35</sup> <i>pi</i> <sup>55</sup> , Lan. <i>mo</i> <sup>35</sup> <i>phji</i> <sup>31</sup> , Le. <i>mo</i> <sup>33</sup> <i>phi</i> <sup>33</sup> , R. <i>mopi</i>
Burman	<i>man</i> /màn/	S. <i>maan</i> <sup>3</sup> , Kh. <i>man</i> <sup>1</sup> (W), D. <i>maan</i> <sup>6</sup>
Burman	<i>myen</i> /myèn/ <sup>67</sup>	S. <i>maan</i> <sup>3</sup> <i>mɛŋ</i> <sup>4</sup> ; Z. <i>myen</i> <sup>31</sup> , Lac. <i>lamyen</i> , Ng. <i>lvmyens</i> , R. <i>mìn</i>
child	<i>luk</i> /lúk/	S. <i>luk</i> <sup>3</sup> , Kh. <i>luk</i> <sup>1</sup> , D. <i>luk</i> <sup>5</sup>
Chin	<i>hkang</i> /kháŋ/ <sup>68</sup>	S. <i>k<sup>h</sup>aaŋ</i> <sup>1</sup> ‘Kachin’; Lh. <i>khaung</i> ” (K), Ng. <i>khang</i> , Hp. <i>kǎxáŋ</i> ‘Kachin’, Ka. <i>hàháj</i> (H), G. <i>hàháj</i> (H)
elder	<i>jau lung</i> /jàwluŋ/	S. <i>tsaw</i> <sup>3</sup> ‘master’, <i>luŋ</i> <sup>4</sup> ‘father’s or mother’s elder brother’
headman	<i>htǎmung</i> /thəmūŋ/	S. <i>t<sup>h</sup>a<sup>1</sup>moŋ</i> <sup>1</sup>
hunter	<i>jau gawng</i> /jàwɡòŋ/	S. <i>tsaw</i> <sup>3</sup> <i>kəŋ</i> <sup>3</sup> ‘gunner’; Z. <i>zyau</i> <sup>11</sup> <i>gong</i> <sup>11</sup> , Lac. <i>jau goung</i>
king	<i>hkaw hkam</i> /khokhám/	S. <i>hə<sup>1</sup>k<sup>h</sup>am</i> <sup>4</sup> ‘royal palace’; Z. <i>xo</i> <sup>55</sup> <i>kham</i> <sup>51</sup> (D), Lh. <i>khə<sup>L</sup>kham</i> <sup>F</sup> , Lan. <i>khə<sup>35</sup>kham</i> <sup>55</sup> , Lac. <i>khoo</i> : <i>kham</i> ”, R. <i>koqkám</i>
king	<i>jau wawng</i> /jàwwòŋ/	S. <i>tsaw</i> <sup>3</sup> ‘master’, C. <i>huáng</i> ‘emperor’
king	<i>mun jau hkaw hkam</i> /munjawkhokhám/	S. <i>mun<sup>1</sup>tsaw</i> <sup>3</sup> ‘Buddhist monk’, <i>hə<sup>1</sup>k<sup>h</sup>am</i> <sup>4</sup> ‘royal palace’
lazy person	<i>kun hkan</i> /kúnkhan/	S. <i>kon</i> <sup>4</sup> <i>k<sup>h</sup>aan</i> <sup>5</sup>
loafer	<i>kun le</i> /kùnlè/	S. <i>kon</i> <sup>4</sup> <i>lən</i> <sup>1</sup> ? ‘dishonest man’
master	<i>jau</i> /jàw/ <sup>69</sup>	S. <i>tsaw</i> <sup>3</sup> , Kh. <i>caw</i> <sup>5</sup> , D. <i>tsau</i> <sup>4</sup> ; Lh. <i>jau</i> (K), Lac. <i>jau</i> , G. <i>sə</i> ‘God, lord’ (H)

<sup>66</sup> ‘a percussion cap used with a rifle’

<sup>67</sup> Possibly from Chinese *mǎn* ‘Burma’ through Shan.

<sup>68</sup> “uncivilized, a term the Shan applies to Kachin, and the Kachin passes on to the Chin” (Maran 1964: 42)

<sup>69</sup> Shan *tsaw*<sup>3</sup> ‘master’ was borrowed into Jinghpaw in two different forms with different meanings: *jàw* ‘master’ and *dzàw* ‘prince’ (see ‘prince’ below).

merchant	<i>jau poi /jàwpòy/</i>	S. <i>tsaw</i> <sup>3</sup> ‘master’, <i>pəj</i> <sup>4</sup> <i>tsaa</i> <sup>4</sup> ‘broker’ <sup>70</sup> ; Lh. <i>jhau</i> : boe (K)
military officer	<i>jau bu /jàwbù/</i>	S. <i>tsaw</i> <sup>3</sup> ‘master’, <i>po</i> <sup>2</sup> ‘military officer’ <sup>71</sup>
monk	<i>jau mun /jàwmun/</i>	S. <i>tsaw</i> <sup>3</sup> <i>mun</i> <sup>1</sup> ; Ka. <i>səmon</i> (H), G. <i>səmuŋ</i> (H)
nun	<i>ya hkau /yàkhaw/</i>	S. <i>naa</i> <sup>4</sup> <i>k<sup>h</sup>aaw</i> <sup>1</sup> , Kh. <i>khaaw</i> <sup>4</sup> ‘white’, D. <i>yaa</i> <sup>6</sup> ‘lady’, <i>laaŋ</i> <sup>2</sup> <i>xaau</i> <sup>1</sup> ; Z. <i>ja</i> <sup>21</sup> <i>khau</i> <sup>55</sup> (D)
paddy dealer prince	<i>jau hkaw /jàwkhàw/</i> <i>zau /zàw/</i> <sup>72</sup>	S. <i>tsaw</i> <sup>3</sup> <i>k<sup>h</sup>aw</i> <sup>3</sup> S. <i>tsaw</i> <sup>3</sup> ‘master’, D. <i>tsau</i> <sup>4</sup> ‘master’; Z. <i>zvau</i> <sup>11</sup> ‘officer’, Lh. <i>zug</i> : (K), Lac. <i>zau</i> :, Ng. <i>zau</i> ‘officer’
princess	<i>nang /nàŋ/</i> <sup>73</sup>	S. <i>naaŋ</i> <sup>4</sup> , Kh. <i>naŋ</i> <sup>3</sup> (W), D. <i>laaŋ</i> <sup>2</sup> ; R. <i>nvang</i> ~ <i>vnavang</i> ‘the name of the first daughter’, Ka. <i>naūngsaúng</i> ‘wife’s younger sister’, G. <i>nàuŋs<sup>h</sup>ɔ</i> ‘wife’s younger sister’ (H)
rich man	<i>kun mi /kúnmi/</i>	S. <i>kon</i> <sup>4</sup> <i>maak</i> <sup>3</sup> <i>kon</i> <sup>4</sup> <i>mi</i> <sup>4</sup> , Kh. <i>mi</i> <sup>3</sup> ‘be rich’ (W), D. <i>mi</i> <sup>2</sup> ‘be rich’
robber	<i>jun /jun/</i>	S. <i>tson</i> <sup>1</sup> , Kh. <i>tsun</i> <sup>1</sup> (W), D. <i>tson</i> <sup>6</sup>
robber	<i>kun jun /kúnjun/</i>	S. <i>kon</i> <sup>4</sup> <i>tson</i> <sup>1</sup> ; Z. <i>gvun</i> <sup>31</sup> <i>zyun</i> <sup>35</sup>
servant	<i>ningchyang /nìŋcaŋ/</i> ~ <i>nchyang /ñcaŋ/</i>	S. <i>tsaaŋ</i> <sup>3</sup> ‘to hire’, Kh. <i>tsaŋ</i> <sup>5</sup> ‘to hire’ (W), D. <i>tsaaŋ</i> <sup>4</sup> ‘to hire’; R. <i>lìngzvang</i> S. <i>luk</i> <sup>3</sup> <i>s<sup>h</sup>uk</i> <sup>4</sup>
soldier	<i>luk suk /lùksùk/</i>	S. <i>luk</i> <sup>3</sup> <i>s<sup>h</sup>uk</i> <sup>4</sup>
Tai	<i>tai /táy/</i>	S. <i>taj</i> <sup>4</sup> , Kh. <i>tay</i> <sup>3</sup> , D. <i>tai</i> <sup>2</sup>
young girl	<i>sau /saw/</i>	S. <i>s<sup>h</sup>aaw</i> <sup>1</sup> , Kh. <i>saaw</i> <sup>4</sup> , D. <i>saau</i> <sup>1</sup>

#### 4.16. Activity and mental activity

compete	<i>hkying /khyìŋ/</i>	S. <i>k<sup>h</sup>et</i> <sup>2</sup>
hate	<i>chyang /cáy/</i>	S. <i>tsaŋ</i> <sup>4</sup> , D. <i>tsaŋ</i> <sup>2</sup>

<sup>70</sup> From Burmese *pwézá* (WB *pwaicā*<sup>3</sup>) ‘broker’.

<sup>71</sup> From Burmese *bò* ‘commander’ (WB *buil*) < Pali.

<sup>72</sup> “The male members of a chief’s family carry the title *Zau* (*jau*), a term borrowed from the Shan, meaning Lord.” (Hanson 1913: 174–175) “first names indicating chiefly status. This name will replace an extraneous affix the Kachins put before the sibling position. Hence, a Ja Naw is a chiefly name for a second son in traditional fashion. This will assume a Zau from the Shan and become Zau Naw.” (Maran 1964: 38)

<sup>73</sup> “The daughters of a chief carry the title *Nang*, also a Shan term.” (Hanson 1913: 175)



hit <sup>74</sup>	<i>mak</i> /ʔmák/	S. <i>maak</i> <sup>3</sup> ‘blade’, <i>maak</i> <sup>2</sup> ‘be in confusion’
knead	<i>nut</i> /nùt/	S. <i>not</i> <sup>3</sup>
liberate	<i>boi</i> /boy/	S. <i>pɔj</i> <sup>2</sup> , D. <i>pɔi</i> <sup>3</sup>
pass by	<i>pun</i> /pun/	S. <i>pon</i> <sup>5</sup>
prepare	<i>hkyen</i> /khyén/	S. <i>hen</i> <sup>4</sup>
search	<i>sawk</i> /sòk/	S. <i>s<sup>h</sup>ɔk</i> <sup>3</sup> , Kh. <i>sɔk</i> <sup>1</sup> (W); Z. <i>zok</i> <sup>1</sup>
stir <sup>75</sup>	<i>wai</i> /way/	S. <i>waaj</i> <sup>3</sup>
wait for	<i>mawng</i> /mónŋ/	S. <i>mɔŋ</i> <sup>4</sup>

## 4.17. State and quality

be alert	<i>let</i> /lét/	S. <i>let</i> <sup>4</sup> ; Z. <i>lyet</i> <sup>1</sup> , R. <i>lak</i> <sup>?</sup>
be deep	<i>sung</i> /sùŋ/	S. <i>s<sup>h</sup>uŋ</i> <sup>1</sup> ‘be high’, Kh. <i>suŋ</i> <sup>4</sup> ‘be high’, D. <i>suŋ</i> <sup>1</sup> ‘be high’
be different	<i>lak</i> /ʔlák/, <i>lak lai</i> /làkláy/	S. <i>laak</i> <sup>2</sup> , D. <i>lak</i> <sup>1</sup>
be difficult	<i>yak</i> /yàk/	S. <i>jaak</i> <sup>3</sup> , D. <i>yaak</i> <sup>5</sup> ; Z. <i>yak</i> <sup>1</sup> , R. <i>yvk</i>
be educated	<i>kat</i> /kàt/	S. <i>kat</i> <sup>5</sup> ; Z. <i>kat</i> <sup>21</sup> ‘be able’
be foolish	<i>ngawk</i> /ŋòk/	S. <i>ŋuuk</i> <sup>4</sup>
be fragrant	<i>hawm</i> /hom/	S. <i>hɔm</i> <sup>1</sup> , Kh. <i>hɔm</i> <sup>4</sup> , D. <i>hɔm</i> <sup>1</sup> ; Z. <i>kom</i> <sup>55</sup> ~ <i>hom</i> <sup>55</sup>
be glad	T. <i>hum hum</i> /hum <sup>2</sup> hum <sup>2</sup> /	S. <i>hom</i> <sup>1</sup>
be grateful	<i>chyum</i> /cum/	S. <i>tsom</i> <sup>4</sup>
be great	<i>hkik hkam</i> /khìkkhàm/	D. <i>xək<sup>5</sup>xam<sup>3</sup></i> ; Z. <i>khɿk<sup>21</sup>kham<sup>21</sup></i> (D)
be heavy	<i>nak</i> /nák/	S. <i>nak</i> <sup>4</sup> , Kh. <i>nak</i> <sup>4</sup> , D. <i>lak</i> <sup>1</sup>
be insatiable	<i>mak</i> /màk/	S. <i>maak</i> <sup>3</sup> ? ‘be abundant’; Z. <i>mak</i> <sup>1</sup>
be jelled	<i>dung</i> /duŋ/	S. <i>tuŋ</i> <sup>1</sup>
be lazy	<i>hkan</i> /khan/	S. <i>k<sup>h</sup>aan</i> <sup>5</sup> , Kh. <i>khaan</i> <sup>2</sup> , D. <i>xaan</i> <sup>5</sup>
be long	<i>yau</i> /yàw/	S. <i>jaaw</i> <sup>4</sup> , Kh. <i>yaaw</i> <sup>3</sup> , D. <i>yaau</i> <sup>2</sup>
be new	<i>mau</i> /màw/	S. <i>maur</i> <sup>2</sup> , Kh. <i>maU</i> <sup>4</sup> (W), D. <i>maɪ</i> <sup>3</sup>
be old	<i>htau</i> /thaw/	S. <i>t<sup>h</sup>aw</i> <sup>3</sup> , Kh. <i>thaw</i> <sup>5</sup> , D. <i>thau</i> <sup>4</sup>
be potent	<i>hkin</i> /khín/	S. <i>k<sup>h</sup>in</i> <sup>1</sup>
be pretty	<i>kya</i> /kyáʔ/	S. <i>kjaa</i> <sup>5</sup> ? ‘be excellent’
be restless	<i>lu</i> /lu/	S. <i>lu</i> <sup>4</sup> , Kh. <i>lu</i> <sup>3</sup> (W); R. <i>bønlu</i>
be short	<i>bawt</i> /bót/	S. <i>pɔt</i> <sup>4</sup> , Kh. <i>pɔt</i> <sup>4</sup> (W), D. <i>pɔt</i> <sup>1</sup>
be strong	T. <i>heng</i> /heŋ <sup>2</sup> /	S. <i>heŋ</i> <sup>4</sup>
be superior	<i>lawng</i> /loŋ/	S. <i>loŋ</i> <sup>1</sup> , D. <i>loŋ</i> <sup>1</sup>
be suspicious <sup>76</sup>	<i>lawm lem</i> /lomlem/	S. <i>lɔm</i> <sup>1</sup> <i>lem</i> <sup>1</sup> ‘slyly, stealthily’

<sup>74</sup> ‘to hit, strike, as with a sword, to be hit and thus confused, distracted, in disorder’<sup>75</sup> ‘to stir with a ladle’<sup>76</sup> ‘to be suspicious of a person and attempt to find tell-tale signs of intrigue and foul-play, if any, in the person’

be well-finished	<i>hkyem</i> /khyém/	S. <i>k<sup>h</sup>em<sup>3</sup></i>
be wide	<i>gang</i> /gàŋ/ <sup>77</sup>	S. <i>kwaarj<sup>3</sup></i> , Kh. <i>kaarj<sup>5</sup></i> , D. <i>kaarj<sup>4</sup></i> ; R. <i>gang</i>
grow worse	<i>awk</i> /ʔók/	S. ʔək <sup>2</sup> ? ‘come out, as the eruption of a disease’
in vain	<i>li la</i> /lilà/	S. <i>laa<sup>3</sup>li<sup>1</sup></i> ? ‘thoughtlessly’

## 4.18. Time

every year	<i>gu byi</i> /gùbyì/	S. <i>ku<sup>3</sup></i> ‘every’, <i>pi<sup>1</sup></i> ‘year’; Ka. <i>kùpì</i> (H), G. <i>kùpì</i> (H)
lunar month	<i>lun</i> /lun/	S. <i>lɔn<sup>1</sup></i> , Kh. <i>nɔn<sup>1</sup></i> , D. <i>lən<sup>6</sup></i>

## 4.19. Number

one hundred	<i>sen</i> /sèn/ <sup>78</sup>	S. <i>s<sup>h</sup>en<sup>1</sup></i> , D. <i>sen<sup>1</sup></i> ; Ng. <i>dv<sup>s</sup>ens</i> , R. <i>tiqsèn</i>
thousand		
one thousand	<i>hkying</i> /khyiŋ/ <sup>79</sup>	S. <i>heŋ<sup>1</sup></i> , D. <i>heŋ<sup>1</sup></i> ; Z. <i>hing<sup>55</sup></i> ~ <i>king<sup>55</sup></i> , Lh. <i>khyiŋ<sup>H</sup></i> , Lac. <i>khyeing</i> ”, Ng. <i>dvkhyoengh</i> , R. <i>kíng</i> , Hp. <i>héŋ</i>
ten thousand	<i>mun</i> /mùn/	S. <i>mun<sup>2</sup></i> , D. <i>mun<sup>3</sup></i> ; Z. <i>mun<sup>11</sup></i> , Lh. <i>mun<sup>F</sup></i> , Lac. <i>mun</i> , Ng. <i>mons</i> , R. <i>mùn</i> , Hp. <i>món</i> ‘hundred thousand’
two	<i>sawng</i> /soŋ/	S. <i>s<sup>h</sup>ɔŋ<sup>1</sup></i> , Kh. <i>sɔŋ<sup>4</sup></i> , D. <i>sɔŋ<sup>1</sup></i> ; Ka. <i>saūng</i> , G. <i>s<sup>h</sup>auŋ</i> (H)

## References

- Aung Kyaw, Caay Hân Maü and Khun Aay. 2001. *Shanbunkaken Gengochousahyou* (*Linguistic Questionnaire for the Shan Cultural Area*). Tokyo: Institut Liêu-Am d’Études Panphysiques.
- Benedict, Paul K. 1972. *Sino-Tibetan: A Conspectus*. New York: Cambridge University Press.
- Bradley, David. 1996. “Kachin”. In Stephen A. Wurm, Peter Mühlhäusler and Darrell T. Tryon (eds.) *Atlas of Languages of Intercultural Communication in the Pacific, Asia, and the Americas*, vol. 2.1. Berlin: Mouton de Gruyter. pp.749–751.
- Cushing, Josiah N. 1881 [1914]. *A Shan and English Dictionary*. Rangoon: American Baptist Mission Press.
- Dai, Qingxia. 2005. *Langsuyu Yanjiu* (*A Study of the Langsu Language*). Beijing: Nationality Press.

<sup>77</sup> Possibly from Chinese *guǎng* ‘be wide’ through Shan.

<sup>78</sup> Possibly from Burmese *théin* ‘hundred thousand’ (WB *sin<sup>3</sup>*) through Shan.

<sup>79</sup> Possibly from Chinese *qiān* ‘thousand’ through Shan.

- Dai, Qingxia and Xu Xijian. 1995. *Jingpoyu Cihuixue (Jingpho Lexicology)*. Beijing: Central University for Nationalities Press.
- Dai, Qingxia and Li Jie. 2007. *Leqiyu Yanjiu (A Study of the Leqi Language)*. Beijing: Central University for Nationalities Press.
- Diehl, Lon G. 1988. "AB B as noun plus verb: A second look". Paper presented at the 21st International Conference on Sino-Tibetan Languages and Linguistics, Kunming Institute of Technology, Yunnan, China.
- Duoshi, Yongtang, Xijian Xu and Duan Mutu. (eds.) 1992. *Han-Zai Cidian (Chinese-Zaiwa Dictionary)*. Chengdu: Sichuan People's Press.
- Duroiselle, Charles. 1916. "Literal transliteration of the Burmese alphabet". *Journal of the Burma Research Society* 6. pp.81–90.
- Edmondson, Jerold A. 2008. "Shan and other Northern Tier Southeast Tai languages of Myanmar and China: Themes and variations". In Anthony V. N. Diller, Jerold A. Edmondson and Yongxian Luo (eds.) *The Tai-Kadai Languages*. London & New York: Routledge. pp.184–206.
- Fraser, James. 1922. *Handbook of the Lisu (Yawyin) Language*. Rangoon: Superintendent, Government Printing, Burma.
- Grierson, Geroge A. (ed.) 1928. *Linguistic Survey of India*, vol. 2. pt. 2. Calcutta: Government of India, Central Publication Calcutta Branch.
- Haas, Mary. 1964. *Thai-English Student's Dictionary*. Stanford, California: Stanford University Press.
- Hanson, Ola. 1896. *A Grammar of the Kachin Language*. Rangoon: American Baptist Mission Press.
- . 1906. *A Dictionary of the Kachin Language*. Rangoon: American Baptist Mission Press.
- . 1913. *The Kachins: Their Customs and Traditions*. Rangoon: American Baptist Mission Press.
- Harris, Jimmy G. 1976. "Notes on Khamti Shan". In Thomas W. Gething, Jimmy G. Harris and Pranee Kullavanijaya (eds.) *Tai Linguistics in Honor of Fang-Kuei Li*. Bangkok: Chulalongkorn University Press. pp.113–141.
- Henderson, Eugénie J. A. 1986. "Some hitherto unpublished material on Northern (Megyaw) Hpun". In John McCoy and Timothy Light (eds.) *Contributions to Sino-Tibetan Studies*. Leiden: Brill. pp.101–134.
- Hudak, Thomas J. 2000. *Cushing's Shan-English Dictionary: A Phonetic Version*. Arizona: Arizona State University.
- Huziwara, Keisuke. 2012a. "Ganango on-inron (Ganan phonology)". *Journal of the Research Institute for World Languages* 7. pp.121–144.
- . 2012b. "Luisogo no saikou ni mukete (Towards a reconstruction of Proto-Luish)". *Kyoto University Linguistic Research* 31. pp.25–131.
- . 2014. "Luisogo no saikou (Proto-Luish reconsidered)". *Kyoto University Linguistic Research* 33. pp.1–32.
- Kato, Atsuhiko. 2008. "Birumago hatsuonhyouki no ichirei (An example of transcription of Burmese pronunciation)". ms.
- Kurabe, Keita. 2014. "The reflexes of the Proto-Jingpho glides in modern dialects". *Linguistics of the Tibeto-Burman Area* 37.2. pp.181–197.
- . 2016. "A grammar of Jinghpaw, from northern Burma". Ph.D. dissertation, Kyoto University.
- LaPolla, Randy J. and David Sangdong. 2015. *Rawang-English-Burmese Dictionary*. Privately published for limited circulation.
- Leach, Edmund R. 1954. *Political Systems of Highland Burma: A Study of Kachin Social Structure*. London: G. Bell and Sons.
- Lengtai, Aggasena. 2009. "Shan phonology and morphology". MA thesis, Mahidol University.
- Li, Fang-Kuei. 1977. *A Handbook of Comparative Tai*. Honolulu: The University of Hawaii Press.
- Luo, Yongxian. 1999. *A Dictionary of Dehong, Southwest China*. Canberra: Pacific Linguistics, Australian National University.
- Lustig, Anton. 2010. *A Grammar and Dictionary of Zaiwa*. Leiden: Brill.
- Macdonell, Arthur A. 1929. *A Practical Sanskrit Dictionary with Transliteration, Accentuation, and Etymological Analysis Throughout*. London: Oxford University Press.

- Maran, La Raw. 1964. "Shan loan words in Kachin bilingualism in acculturation". MA thesis, The University of Arizona.
- . 1978. "A dictionary of modern spoken Jingpho". ms.
- Matisoff, James A. 1974. "The tones of Jinghpaw and Lolo-Burmese: common origin vs. independent development". *Acta Linguistica Hafniensia* (Copenhagen) 15.2. pp.153–212.
- . 2003. *Handbook of Proto-Tibeto-Burman: System and Philosophy of Sino-Tibetan Reconstruction*. Berkeley, Los Angeles and London: University of California Press.
- . 2013. "Re-examining the genetic position of Jingpho: Putting flesh on the bones of the Jingpho/Luish relationship". *Linguistics of the Tibeto-Burman Area* 36.2. pp.15–95.
- Meng, Zunxian. 2007. "Zaiwayu de Daiyu jieci (Dai loanwords in the Zaiwa language)". *Minzu Yuwen* 2007.3. pp.59–65.
- Morey, Stephen. 2007a. "Turung-English dictionary". ms.
- . 2007b. "Draft dictionary, Singpho (Numhpuk Hkawng) - English". ms.
- Myanmar Language Commission (ed.) 2009. *Myanmar-English Dictionary*. Yangon: Department of the Myanmar Language Commission.
- Nasaw Sampu, Wilai Jaseng, Thocha Jana and Douglas Inglis. 2005. "A preliminary Ngochang-Kachin-English lexicon". Research paper #107. Chiangmai: Payap University.
- Pulu, Jimi. 1997. *Khamti Phrase Book*. Itanagar: Directorate of Research, Government of Arunachal Pradesh.
- Rhys Davids, T.W. and William Stede (eds.) 1921–1925. *The Pali Text Society's Pali-English Dictionary*. Chipstead: Pali Text Society.
- Sangdong, David. 2012. "A grammar of the Kadu (Asak) language". Ph.D. dissertation, La Trobe University.
- Sao Tern Moeng. 1995. *Shan-English Dictionary*. Kensington, Maryland: Dunwoody Press.
- Sawada, Hideo. 2003. "Lhaovogo no gairashakuyougo (Loanwords in Lhaovo)". ms.
- . 2004. "A tentative etymological word-list of Lhaovo (Maru) language". In Setsu Fujishiro (ed.) *Approaches to Eurasian Linguistic Areas*. Kobe: Kobe City College of Nursing. pp.61–122.
- . 2006. "Lhaovogo seishohou (Lhaovo orthography)". In Asako Shiohara and Shigeaki Kodama (eds.) *Writing Unwritten Languages*. Tokyo: ILCAA, Tokyo University of Foreign Studies. pp.81–88.
- . 2011. "Kachinshuu no Taigochimei oboegaki (A note on Tai place names in Kachin State)". In Setsu Fujishiro (ed.) *Dynamics of Eurasian Languages II: Studies on Languages in Multilingual Areas*. Kobe: Kobe City College of Nursing. pp.127–152.
- Scott, George J. 1900. *Gazetteer of Upper Burma and the Shan States*, part 1, vol. 1. Rangoon: Superintendent of Government Printing and Stationery.
- Wannemacher, Mark. 2011. *A Phonological Overview of the Lacid Language*. Chiang Mai: Linguistics Institute, Payap University.
- Weidert, Alfons. 1977. *Tai-Khamti Phonology and Vocabulary*. Wiesbaden: Franz Steiner Verlag.
- Xu, Xijian. 1987. "Jingpoyu de liangci (Classifiers of Jingpo)". *Minzu Yuwen* 1987.5. pp.27–35.
- Yabu, Shiro. 1982. *Atsigo Goishuu (An Atsi Vocabulary)*. Tokyo: ILCAA, Tokyo University of Foreign Studies.
- . 2000. *Marugo Goishuu (A Maru Vocabulary)*. Report of the Grant-in-Aid for Scientific Research (no. 09610535) by Ministry of Education, Science, Sports and Culture.