

Methods in Language Documentation and Description: A Guide to the Kelabit Documentation Project

Charlotte Hemmings

charlotte.hemmings@ling-phil.ox.ac.uk

University of Oxford

Abstract

Kelabit is a Western Austronesian language spoken in Northern Sarawak, Malaysia. This paper provides a guide to the Kelabit documentation project, contextualising the materials collected, and discussing the research methods used. It is hoped that this will make the project outputs more accessible and provide a useful reference for researchers and communities looking to document similar phenomena in related languages.

Keywords: language documentation, research methods, Kelabit, corpus guide

1. Introduction

This paper provides a guide to the Kelabit documentation project, contextualising the corpus of audio and video recordings, and reflecting upon the methods used in collecting data for the project. It discusses the extent to which the different research methods used were effective in achieving the project aims, which were to provide a documentation of the Kelabit language of Northern Sarawak with analysis of symmetrical voice alternations and related morphosyntactic phenomena. It is hoped that this discussion will provide a useful reference for possible methods of documentation and description of similar phenomena in related languages, as well as offering a meta-documentation (Austin, 2013) for the project and its outputs, making the collection more accessible to potential users (Woodbury, 2014).

The paper follows the formats introduced in Caballero (2017) and Salffner (2015) for contextualisation of the contents and design of the corpus. Section 2 introduces the Kelabit language and the documentation project. Section 3 gives an overview of the materials collected and how to access them in ELAR. Section 4 reflects on the methods used in data collection and Section 5 concludes.

2. The Kelabit Language

Kelabit (kzi, 639-3) is a Western Austronesian¹ language of Northern Sarawak, Malaysia (Martin, 1996). It belongs to the Apad Uat subgroup² of Malayo-Polynesian which also includes Lun Bawang/Lundayeh and Sa'ban. Apad Uat languages are spoken in Northern Sarawak as well as across the borders into Brunei, Sabah, and Kalimantan, Indonesia (Kroeger, 1998a). This is indicated in red in Figure 1.

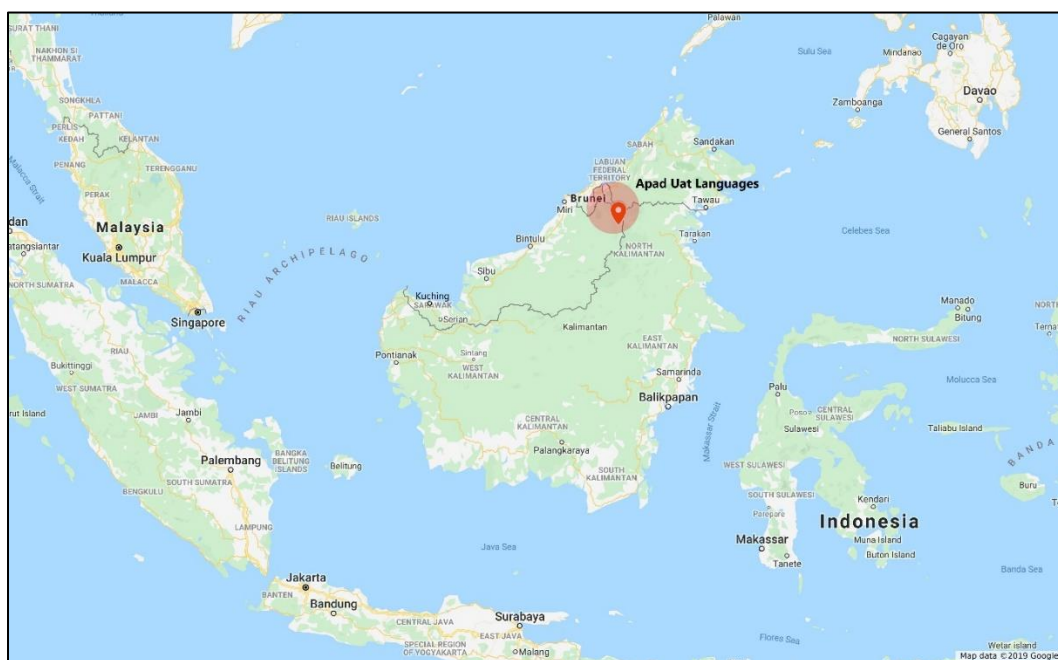


Figure 1: Approximate location of the Apad Uat languages (Map Data: Google, Maxar Technologies)

¹ I use the term Western Austronesian in the typological/geographical sense of Riesberg (2014) and Himmelmann (2005). It includes the languages of Taiwan, the Philippines, Malaysia, Indonesia, and Madagascar that share the property of symmetrical voice (see Section 2.2.1).

² The subgroup is variously referred to as Dayic (Blust, 2013; Smith, 2017), Kelabitic (Clayre, 2005; Kroeger, 1998a) and Apo Duat (Hudson, 1978) in the literature. I use the term Apad Uat as the most neutral cover term. It comes from Hudson's Apo Duat, which Eghenter & Langub (2008) suggest was a mishearing of *apad uat*, the local name for the mountain range on the border between Malaysia and Indonesia. The term 'Dayic' is avoided due to potential confusion with the term 'Dayak', which was previously used to refer indiscriminately to all interior groups in Borneo (King, 1993) and can be associated with negative/pejorative meanings (Schiller, 2007).

Kelabit is traditionally spoken in the Kelabit Highlands, a plateau in central Borneo which lies at the headwaters of the Baram River, and is surrounded by the Apad Uat mountain range to the east and the Tamabu range to the west (Bala, 2002). Today, there are several villages in and around the Kelabit Highlands where the language is spoken. The administrative centre is known as Bario, and is formed of a number of longhouse settlements³ in close-proximity: Bario Asal, Ulung Palang, Arur Dalan, Arur Layun, Pa' Ramapoh Atas and Bawah, Pa' Derung, Padang Pasir and Kampung Baru. Bario is the location of an airport and the only secondary school in the Highlands, and became the de-facto centre of the region following the resettlement of villages during the Confrontation between Malaysia and Indonesia in the 1960s (Bala, 2002). It is marked with a red point in Figure 1. To the north of Bario, there are three villages located along the Debpur river: Pa' Ukat, Pa' Umor and Pa' Lungan. Further downriver, towards the southern end of the Kelabit Highlands are three villages that lie along the Kelapang river: Pa' Mada, Pa' Dalih, and Remudu. There is another village in the Kelapang region, called Batu Patung, which had been uninhabited for some time⁴ but is now in the process of being rebuilt. Between the northern and the southern settlements lies Pa Main, which is now uninhabited,⁵ but was previously an important centre for the Kelabit people. Finally, there are also Kelabit settlements beyond the Kelabit Highlands, that are closer to coastal towns like Miri. The main Kelabit villages outside the Highlands are Long Peluan, Long Lellang, Long Seridan, and Long Napir. Many speakers now live in towns such as Miri and Kuching rather than the traditional villages, as well as other cities in Malaysia and abroad. The rough location of current villages is indicated in Figure 2. Northern (or upriver) villages are indicated in red. Southern (or downriver) villages are indicated in blue, and villages outside the Kelabit Highlands are indicated in green.⁶

³ Longhouses are the traditional dwellings of many indigenous groups in Borneo. It is a communal living space (Janowski, 1995; Saging & Bulan, 1989) and the term refers not only to a physical house but also to the community who collectively live and farm there. As a result of longhouses burning down, or groups seeking new arable land for rice farming, the location of villages in the past was transitory, and new settlements were frequently built. Consequently, the locations of villages today, which are more fixed, may not reflect the river or geographical feature that they were originally named for (*pa'* = river, *long* = confluence, etc.).

⁴ People originally from Batu Patung had relocated to Pa' Dalih.

⁵ Most people from Pa' Main relocated to Ulung Palang in Bario.

⁶ The location of villages is approximate, based on the best of the author's knowledge, and may not be fully accurate. The villages in green do not represent a dialect area.

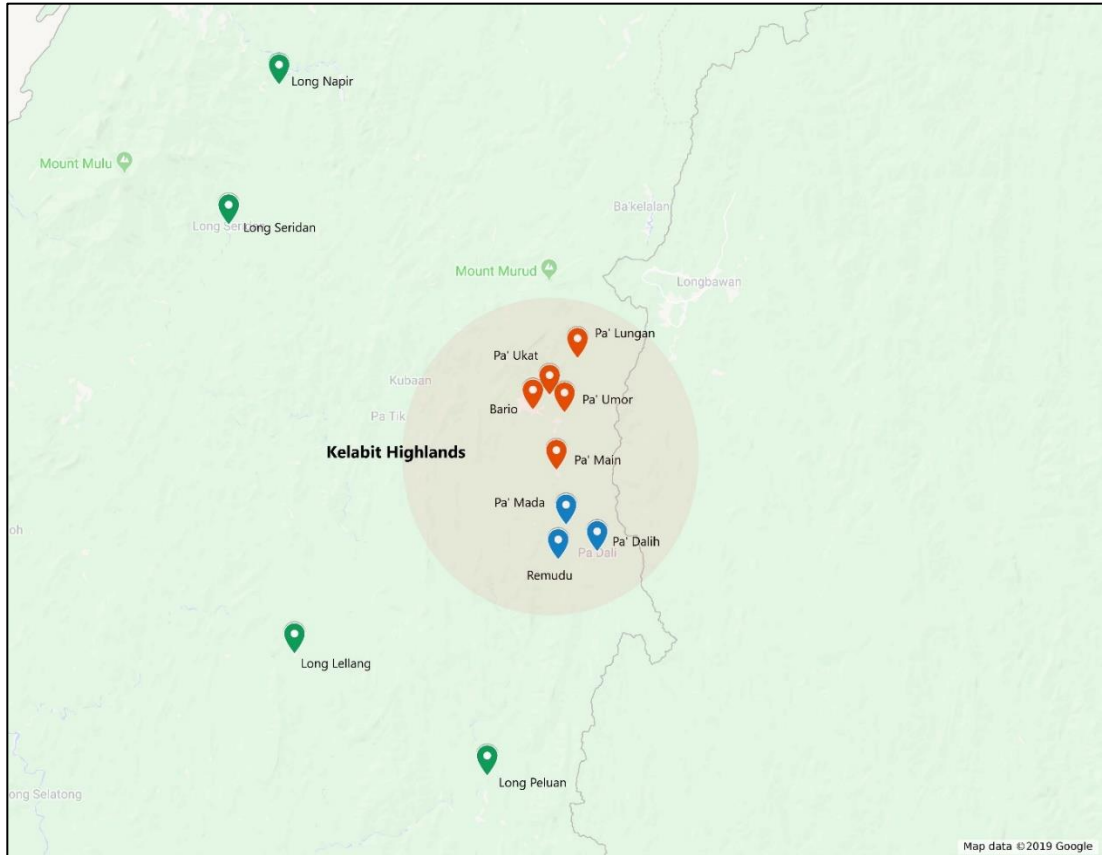


Figure 2: Approximate locations of Kelabit villages (Map data: Google, Maxar Technologies)



Figure 3: Bario during the Bario Food Festival (Pesta Nukenen) 2014

Different villages are associated with different varieties of Kelabit, with some differences being salient to speakers. For example, Gerawat Nulun of Bario Asal⁷ suggests that there are four major dialects of Kelabit distinguished among other factors by the pronunciation of the word ‘day’: *edto* [əd^hɔ:], *echo* [əʃɔ:], *eso* [əsɔ:], and *so* [sɔ:]. The *edto* pronunciation is common in northern (or upriver) villages, like Bario and Pa’ Umor, and the *so* pronunciation is typical of southern (or downriver) villages along the Kelapang river, like Pa’ Dalih. There are other salient phonological and lexical differences between northern and southern varieties. For example, Bario schwa corresponds to Pa’ Dalih /i/, Bario /u/ corresponds to Pa’ Dalih /o/, and Bario /d/ corresponds to Pa Dalih /r/:

(1) *Dialect Differences* (Hemmings, 2016, p. 89)

Bario Kelabit	Kelapang Kelabit	
[ŋadəl]	[ŋadɪl]	<i>ngadel</i> ‘sharp’
[manək]	[manɔk]	<i>manuk</i> ‘bird’
[dadən]	[radən]	<i>dadan</i> ‘long time’

As for lexical differences, the adverb meaning ‘later’ is *na’an* in Bario Kelabit and *ano* in Pa’ Dalih Kelabit. However, the exact number and grouping of dialects has yet to be documented in detail (Blust, 1993). In this documentation project, recordings were collected in Bario, Pa’ Umor, and Pa’ Dalih. Recordings with speakers from the first two locations are tagged as ‘Kelabit – Bario’ to reflect the northern dialect. Recordings with speakers from Pa’ Dalih or other southern villages are tagged as ‘Kelabit – Pa’ Dalih’.

Kelabit is considered to be threatened as measured by the EGIDS scale (see Coluzzi, this volume) since the language is still used among those of child-bearing age but is increasingly not transmitted to the next generations (Simons & Fennig, 2018). This is particularly true in towns, where the majority of speakers now live (Martin & Yen, 1994). Though there are many fluent speakers in urban centres like Miri and Kuching, Martin and Yen (1994) found that the rate of intergenerational transmission was low, particularly in families where only one parent was Kelabit. Similarly, they found that younger speakers who had multi-ethnic friendship groups would often

⁷ Originally from Pa’ Lungan.

use Malay or English or code-switch in group settings (Martin & Yen, 1994). In contrast, in the Kelabit Highlands, the language is still used as a means of daily communication in a range of domains.⁸ However, the population size there is much lower, and the average age is older. Thus, though Kelabit is used as a language of daily communication in the Kelabit Highlands, it can still be considered threatened, given the relatively low number of speakers and the large-scale migration towards towns.

Given the patterns of language shift, it is difficult to estimate the exact number of speakers. The 2010 Sarawak Census listed the Kelabit population as 5,900 – though the total should probably be higher if we take into account population growth outside Sarawak. However, according to the Bario clinic records in 2012, only 1,089 people lived in Bario and the surrounding villages of Pa’ Ukat, Pa’ Umor, and Pa’ Lungan. The total number of speakers is most likely somewhere between 2,000-6,000. All speakers of Kelabit are multilingual (except perhaps the very oldest), and, generally, speakers have proficiency in Kelabit, Malay, and English (Amster, 2003). All the 34 speakers who participated in the documentation project, aged between 40 and 80, spoke Malay and many of them also had proficiency in English and other local languages. This is summarised in Table 1:

Table 1: Linguistic Repertoire of Kelabit Speakers

Language	Number of Speakers	% of population surveyed
Kelabit	34	100%
Malay	34	100%
English	25	74%
Lun Bawang	14	41%
Iban	12	35%
Kayan	7	21%
Penan	6	18%
Kenyah	4	12%
Sa’ban	3	9%
Lun Kerayan/Berian	3	9%
Bidayuh	1	3%
Hakka	1	3%

⁸ Including occasionally in church, in village meetings, at home, and in the village centre but not in schools where Malay is the medium of instruction. Kelabit is traditionally an oral language, but written Kelabit is now used in new domains such as social media.

Moreover, 82% of speakers interviewed spoke three or more languages; 59% spoke four or more languages, and several had proficiency in seven or eight different local and national languages.

Generally, speakers have positive attitudes towards the language, both in the towns and in the Highlands (Martin & Yen, 1994; Rethinasamy, 2014), and maintenance of the language is viewed as important. Consequently, there are ongoing community initiatives to help revitalise it, including Kelabit camps in Miri, and a Kelabit playschool in Bario (Bulan & Labang, 2008). For some years, there was also a Kelabit Radio Station that broadcast in Bario twice a day, and included daily news reports in Kelabit (Harris & Harris, 2011). More recently, there has been a large-scale community project to translate the Bible into Kelabit which has involved discussions around orthography and language standardisation. Hence, language is something that the Kelabit community are keen to preserve and it is hoped that the documentation project outlined in this paper will help with this task.

2.1 Previous Documentation and Description

There are very few resources available for Kelabit or any of the closely related Apad Uat languages (cf. Martin, 1996). The earliest sources include short wordlists collected by missionaries and government officers during the period of Brooke rule from 1841-1946, including de Crespigny (1896), Douglas (1911), Ray (1913), Roth (1896), and Rutter (1929). Ray (1913) is considered the most important resource and it includes 200 lexical items in Kelabit. The first descriptive work on the Apad Uat languages is Southwell (1949), who published notes on verbal morphology of Kemaloh Lundayeh. More recently, there have been some descriptive works, mainly on morphology and phonology (Asmah, 1983; Blust, 1974, 2006, 2016). There are also two short dictionaries (Amster, 1995; Blust, 1993), and some online resources (The Borneo Dictionary⁹, Kelabit wiki,¹⁰ and the Kelabit Portal¹¹). Moreover, there are extensive ethnographic materials on Kelabit written by members of the community (e.g. Bala, 2002; Saging, 1976/77; Saging & Bulan, 1989; Talla, 1979) and outside anthropologists (e.g. Amster, 1998, 1999, 2003, 2006; Harrisson, 1954, 1959a, 1959b; Janowski, 1988, 1991, 2003, 2012; Rubenstein, 1973, and Schneeberger,

⁹ <https://borneodictionary.com/kelabit/>

¹⁰ https://www.bario.info/index.php?title=Main_Page

¹¹ <http://kelabitportal.com/>

1979). Many of these texts include vocabulary, formulaic greetings, and/or transcripts of traditional songs, such as *laku* which document oral history and personal life stories. Finally, the Summer Institute of Linguistics (SIL) are in the process of producing materials for children within the community; there is an unpublished Toolbox lexicon¹² produced by Peter Martin (some of which appears on the Borneo Dictionary website), and an ongoing dictionary project led by Gerawat Nulun and assisted by Alexander Smith. Nonetheless, documentation is limited, particularly in relation to syntax and discourse features, which has led to the particular design and development of the current documentation project.

2.2 The Kelabit Documentation Project

2.2.1 Motivations

As is often the case with language documentation and description, this project arose from my PhD research. The academic focus was an analysis of symmetrical voice alternations and related phenomena, such as case-marking and word order. Symmetrical voice systems are a characteristic feature of Western Austronesian languages (Himmelman, 2005) but are otherwise rare in the world's languages.¹³ Like other voice systems, they involve an alternation in the mapping of semantic arguments to syntactic functions that is morphologically encoded (Kulikov, 2011). However, unlike active-passive and ergative-antipassive alternations, they do not involve the demotion of arguments or detransitivisation. Consequently, languages with symmetrical voice systems have more than one transitive clause type (Riesberg, 2014). An example from Malay is:

(2) *Malay*

a. **Actor Voice (AV)**

Anak	saya	me-lihat	orang	itu
child	1SG	AV-see	person	DIST

‘My child saw that person’

¹² Toolbox is a software tool used by linguistics to store lexical data and help with interlinear glossing.

¹³ Some inverse systems have been said to share commonalities with symmetrical voice, e.g. Movima from Bolivia (Haude & Zúñiga, 2016). Similarly, Dinka, a Western Nilotic language from Sudan, has been claimed to have a symmetrical voice system (Erlewine, Levin, & van Urk, 2014).

b. **Undergoer Voice (UV)**

Orang itu di-lihat anak saya

Person DIST UV-see child 1SG

‘My child saw that person’ (Himmelmann, 2005, p. 112)

Both constructions in (2) are syntactically transitive with two core arguments: the actor and the undergoer. However, they differ in terms of which argument is mapped to subject and realised pre-verbally; this difference is reflected in the verbal morphology. Given that symmetrical voice systems are cross-linguistically rare, and differ from other voice systems such as active/passive, there has been longstanding debate over how best to analyse them (see e.g. Adelaar, 2013; Chen & McDonnell, 2019 for discussion). This has generated further debate on the nature of grammatical functions (Bickel, 2010; Dryer, 1997; Kroeger, 1993; Schachter, 1976, 1996) and alignment in Austronesian (Aldridge, 2004, 2012; Foley, 2008; Kroeger, 1993; Riesberg, 2014).

There is significant variation within Western Austronesian voice systems. Consequently, languages are often subcategorised as being ‘Philippine-type’ or ‘Indonesian-type’ (Arka & Ross, 2005; Himmelmann, 2005). Among other things, Philippine-type and Indonesian-type languages differ on:

- the number of voice alternations;
- whether the language additionally has a true passive construction;
- whether the language has applicatives;
- whether voice morphology also encodes tense-aspect-mood (TAM) distinctions;
- whether the language has case-marking; and
- whether the basic word order is subject-verb-object (SVO) or verb-initial (Arka, 2002; Hemmings, 2015).

It is generally understood that Philippine-type languages are more conservative and preserve structural properties that they inherited from Proto-Austronesian (Blust, 2013). In contrast, Indonesian-type languages, such as Malay, are believed to represent historical innovations (Adelaar, 2005). What makes Kelabit so interesting is that the languages of Northern Sarawak, and the Apad Uat subgroup in particular, fall at an important point of transition, both genetically and geographically, between the languages of the Philippines and the languages of Indonesia (Adelaar, 1995; Hemmings, 2016; Hudson, 1978). Kelabit appears to be intermediate between Lun

Bawang/Lundayeh, which evinces Philippine-type verbal morphology and a three-way system of voice alternations, and Sa'ban which has a two-way system of voice alternations and seems to have lost much of the Philippine-type verbal morphology (Clayre, 2005, 2014). Consequently, Kelabit is an interesting case study to better understand variation in Western Austronesian voice systems, and to explore possible diachronic changes.

2.2.2 Project Design

Given the lack of existing documentation, the context of language endangerment, and the typological interest in Kelabit as a means to explore wider theoretical debates surrounding symmetrical voice, the Kelabit documentation project was devised with the dual aims of producing a representative documentation of the language alongside in-depth analysis of the voice system. The PhD project that was the foundation for documentation was funded by a Wolfson Scholarship at SOAS, University of London. As an outside researcher from the UK, I was originally introduced to the Kelabit community by other researchers working in the area, and for making contacts, obtaining informed consent, negotiating the project goals, and recruiting participants, I received significant assistance from community members in the UK and Malaysia who I was in contact with before my first field trip.¹⁴ With the help of community members, I collected audio and video recordings of Kelabit over a period of roughly three months in 2013 (from 20 September to 20 December) and three months in 2014 (from 30 June to 19 September). During this time, I was based in Bario but also collected recordings in Pa' Dalih and Pa' Umor. In order to create as representative, complete, and comprehensive a corpus as possible, a diverse range of texts were collected that varied according to speaker, topic, and genre (see Seifart, 2008 for discussion of representativeness in language documentation). These were made accessible to a wider audience via transcription into a provisional orthography (see Section 3.1.3), and translation into English, using the ELAN annotation software (Version 5.2, 2019). Comments on individual words, morphemes, or constructions of interest were also given as annotations in the ELAN files. Finally, metadata concerning each recording session and each speaker was recorded. Standard formats for annotation, grammatical description, and metadata were followed (cf. Bird & Simons, 2003;

¹⁴ With special thanks to Monica Janowski, Beatrice Clayre, Val Mashman, Julia Raja, Lynette Smith, Lucy Bulan, Poline Bala, Garnette Jalla, and Florance Lapu Apu.

Bowern, 2015; Himmelmann, 1998; Himmelmann, 2006; Schultze-Berndt, 2006; Woodbury, 2003, 2011, among others).

I received funding to continue the Kelabit documentation project through a Leverhulme Early Career Fellowship (ECF-2016-425) exploring ‘Information Structure in the Languages of Northern Sarawak’ at the University of Oxford. This involved additional fieldwork from 16 October to 8 December 2017 and from 28 January to 22 March 2019, though not exclusively in Kelabit-speaking villages.¹⁵ During this time, I worked with community members to collect additional recordings as well as undertaking detailed elicitation relating to grammatical functions and information structure. Again, recordings were annotated with transcription, and translation into English; associated metadata was also collected.

At the start of both research projects, I sought approval from the head of *Rurum Kelabit Sarawak* (The Kelabit Association) in Bario, and the headmen in the villages that I visited. I also spoke to as many people as possible about the nature of my research in order to recruit participants and ensure informed consent. To do this, I worked with community members to translate key messages into Kelabit at the start of fieldwork. To date, 34 speakers have contributed to the documentation project by recording naturalistic texts and responding to experimental stimuli. Florance Lapu Apu has been centrally involved as the primary language consultant during elicitation sessions in Bario, and annotator for the majority of recordings. In Pa’ Dalih, Jeffrey Malang was the language consultant for elicitation, and the annotator of recordings. Lucy Bulan also acted as a language consultant and assisted in the annotation and collection of materials from Pa’ Umor.¹⁶ Many more community members have been involved, from discussing analyses to

¹⁵ The project also involves documentation of the Lun Bawang variety spoken in Ba Kelalan and the Sa’ban variety spoken in Long Banga.

¹⁶ Florance Lapu Apu is well-known in the community and has helped enormously in terms of identifying people to record texts or participate in experiments, providing her knowledge of Kelabit in elicitation sessions and annotating texts. I began working with her after a mutual contact from the community put us in touch as she had previously assisted a number of researchers. Florance has excellent knowledge of both Kelabit and English and, over several years of working together, is now familiar with my style of asking elicitation questions, as well as the processes for transcription and translation. This makes working together very rewarding. Jeffrey Malang was introduced to me by a researcher who he had previously assisted as someone who could help with introducing me to people in Pa Dalih, finding people to record texts, and translating from Kelabit to English. Finally, Lucy Bulan is my adopted mother in Bario and was one of the first people that I contacted at the start of my research. She was previously the head teacher at the school in Bario and the head of the Kelabit Association (*Rurum Kelabit*) Education Unit, which seeks to support the Kelabit language, among other things. As such, Lucy played a central role in many community-led language revitalisation projects. Lucy provided considerable assistance to the project by introducing me to Kelabit speakers in the UK, important community stakeholders in Kuching and Miri, and speakers in Pa’ Umor, the longhouse of her ancestors. Alongside Florance, she has discussed community hopes for the Kelabit language with me and helped to

suggesting speakers to work with and topics to record. I am extremely grateful to everyone who provided assistance and I acknowledge their collective contribution to the work.



Figure 4: From left, Charlotte Hemmings (Linguist), David Lian @ Maran Talla @ Ribuh Tepun¹⁷ (Consultant), Florance Lapu Apu (Annotator and Research Assistant) after a recording session.



Figure 5: Stanley Isaac Ibh @ Mikat Balang @ Paren Kera'i' recording at the Radio Bario studio.

ensure that the documentation project supports these plans as much as possible. An important goal of the project has been to maintain good relations with the Kelabit community.

¹⁷ The Kelabit have a practice of name-changing to reflect significant life events. Typically, people will change their name upon the birth of their first child, and then again after the birth of their first grandchild (see e.g. Amster, 1999; Saging & Bulan, 1989). The @ sign is used by convention to separate the different names.

Given the high levels of proficiency among the Kelabit community, English was chosen as the main language of communication when conducting elicitation and discussing the transcription and translation of recordings.¹⁸ However, as my proficiency in Kelabit improved, I also attempted to converse with participants in Kelabit in both elicitation and text collection contexts.

The corpus of audio and video recordings collected during this project is in the process of being archived with the Endangered Languages Archive at SOAS (Hemmings, 2019). Copies of all recordings have also been left with the community in the care of Florance Lapu Apu, the primary language consultant. The project resulted in my PhD thesis (Hemmings, 2016), and several publications relating to Kelabit grammar, including Hemmings (2015, forthcoming) They highlight that Kelabit, and the languages of Borneo, do not fit neatly into the two-way typology of ‘Philippine-type’ and ‘Indonesian-type’ but rather show a mixture of properties. Hence, a more fine-grained approach is needed to capture morphosyntactic variation within Western Austronesian languages.

3. The Documentary Corpus

3.1 Corpus Conventions

3.1.1 Recording Formats

Audio recordings were collected using a digital recorder (Zoom H4N in 2013/2014 and Zoom Q8 in 2017/2019), a lavalier microphone (Audio Technica PRO70) for individuals, and a stereo microphone (Superlux E523/D) for groups and pairs in elicitation. The recordings were stored in WAV format with a 44.1kHz sampling rate, 16-bit depth. Video recordings were originally collected using a Canon EOS 60D DSLR camera with Canon EF 28mm f/1.8 USM lens. The DSLR camera is primarily intended for still photography but can film up to 12-minute videos with very high image quality. This has the advantage of the videos being aesthetically pleasing, but the disadvantage is that they are limited in length. The recording set-up was relatively easy and involved connecting the camera to the Zoom H4N in order to record high quality audio using external microphones. From 2017, I began using the Zoom Q8 for both video and audio recording,

¹⁸ I did not use Malay for elicitation purposes in order to avoid prompting the choice of AV or UV in translation tasks.

reducing the amount of equipment necessary to take to the field.¹⁹ Using Handbrake software videos were converted to MP4 for use with ELAN, and for archiving.

3.1.2 File Naming and Metadata

Files from the PhD project are named according to a code schema PPPDDMMYYYYAA_XX, where PPP indicates the place of recording (e.g. BAR = Bario, PDA = Pa' Dalih, PUM = Pa' Umor), DDMMYYYY is the date of recording (recordings made after 2014 used the ISO order YYYYMMDD), and AA represents the data collector (e.g. CH = Charlotte Hemmings), and XX distinguishes multiple recordings made on the same date. Files that are related to each other, e.g. audio, video, and ELAN files for the same recording session, have the same name (with different extensions).

The following metadata was collected for each recording following the ISLE Meta Data Initiative (IMDI) standard (Gippert, Himmelmann, & Mosel, 2006): filename; recording title; description; date; location; genre; topic; keywords; participants; languages; access; and recording equipment. This information is accessible and searchable on the ELAR deposit page. Topics, genres, and participants can be filtered from the deposit homepage, as discussed further in Section 3.2. The search bar can be employed to search via keywords or filename. Subgenre information is included at the beginning of each recording title, and as a keyword. Information on the content of recordings and recording equipment used is found under the metadata category 'description'. For elicitation recordings, descriptions also include information on morphemes or constructions of interest discussed in the recording. In addition, sociolinguistic metadata was collected for each speaker, including: name; date of birth; gender; birthplace; primary language; other languages; ethnic group; occupation; mother's name; mother's primary language; mother's other languages; mother's birthplace; mother's ethnic group; father's name; father's primary language; father's other languages; father's birthplace; and father's ethnic group. This is stored in a passworded Excel file to protect personal information and is not currently accessible via the archival deposit, but can be requested from the researcher in an anonymised format.

¹⁹ The image quality is lower than the DSLR camera but there is less chance of problems with audio due to cables becoming detached etc. This was ultimately felt to be a more important consideration.

3.1.3 Transcription and Translation

In order to add value, each recording is transcribed into a provisional orthography and translated into English. The procedure for annotation was as follows. First, I prepared the files for annotation by converting video files and creating an ELAN file with annotations roughly aligned on the basis of pauses. I then filled in as much transcription as possible, and subsequently worked with a native speaker annotator to check the transcription, fill in any gaps, and translate each utterance into English. Finally, the annotator and I would discuss and add comments on points of interest, including notes on semantics, paradigmatically related forms of verbs, and ethnographic context. For elicitation sessions where sentences were elicited with a direct translation, I added transcription and translation directly without the assistance of a native speaker consultant.

As Kelabit is traditionally an oral language, there is no standardised spelling system and discussions relating to orthography are ongoing. The most controversial issue is how to represent the glottal stop. Following Labang (2012), I began the project using ‘q’ to represent glottal stop word-finally and an apostrophe word-medially (see Hemmings 2016). However, the use of ‘q’ is unpopular with some speakers and it was subsequently decided that the apostrophe would be used for the glottal stop in the community-led Bible translation project.²⁰ For this reason, I am in the process of converting the transcriptions to the apostrophe everywhere. The next stage for the project is to provide further annotation in the form of morpheme-by-morpheme glossing using SIL’s FLEx software, as well as annotations relating to grammatical functions and information structure, following the guidelines in the Grammatical Relations and Animacy in Discourse (GRAID) Manual (Haig & Schnell, 2014) and the Information Structure in Cross-Linguistic Corpora Manual (Dipper, Götze, & Skopeteas, 2007). These more elaborated annotations will become available upon completion.

3.2 An Overview of the Corpus

The corpus includes roughly 55 hours of audio and 10 hours of video recordings with associated metadata and annotations in ELAN. The files are organised into bundles, where each bundle contains a recording from a particular session (either audio or audio + video), and the associated ELAN annotation file. Occasionally, bundles also include a PDF document containing typed field

²⁰ It was decided, however, that individuals are free to use ‘q’ or apostrophe as they see fit in personal communication and both spellings are recognised.

notes. The bundles are organised into three major genres reflecting the methods used for data collection: elicitation, stimuli, and naturalistic texts. Materials classed as ‘elicitation’ involve discussion of linguistic elements, including particular word classes, morphemes, and constructions. These sessions made use of questionnaires, direct translation, and grammaticality judgements, and are detailed in Section 3.2.1. Materials classed as ‘stimuli’ arise from data collection via picture or video stimuli, and are detailed in Section 3.2.2. Finally, materials classed as ‘naturalistic text’ were collected without linguistic or non-linguistic stimuli. They cover a range of sub-genres and are detailed in Section 3.2.3. There are currently 103 bundles deposited in the ELAR collection (Hemmings, 2019).

3.2.1 Elicitation

The corpus includes approximately 39 hours of recordings of elicitation and several notebooks with field notes. Elicitation sessions were typically held with a single consultant, though judgements were subsequently confirmed with other speakers. The majority of recordings relate to the Bario dialect of Kelabit. However, I also collected a basic word list and cardinal numbers in Pa’ Dalih Kelabit. Initially, most sessions were conducted in English. However, during later trips, I sometimes attempted monolingual elicitation in Kelabit, at least for the purposes of identifying grammaticality judgements.

I started the project with a Swadesh 200-word list, translating from English to Kelabit. Subsequently, I elicited additional verbs from different semantic classes using the list in Haspelmath (1993) and the Leipzig Valency Project. Having elicited the verbal root, I attempted to elicit related forms using the verbal affixes documented for Kelabit in Asmah (1983). I would offer a form, based on the stem plus an affix, and elicit a judgement about whether this was a possible word in Kelabit. Where related forms were judged to be acceptable, example sentences were then elicited to illustrate the difference. Finally, I used the Austronesian Elicitation Schedule, originally designed for Oceanic Languages (Johnston, 1989), to elicit information about tense, aspect, and mood via translation from English to Kelabit. These initial recordings were helpful for basic vocabulary, language learning, and illustration of the roles of different morphemes and constructions.

Building on this initial analysis, I used grammaticality judgements to collect negative evidence. An example was the elicitation of pronominal paradigms: having elicited the pronoun

forms using the Austronesian Elicitation Schedule, I created sentences that contained the different forms in different functions, different positions, and different clause-types. Some of these I understood to be grammatically incorrect but wanted to check speaker intuitions. I presented these sentences in written form, without English translation, and then asked for a grammaticality judgement for each. This was useful for identifying the range of possibilities.

Finally, I collected some written examples by asking Florance Lapu Apu to write Kelabit language tests for me. This served the dual purpose of improving my language skills, and creating sentences which had not been unduly influenced by me. Tests typically involved passages for me to translate from Kelabit to English, which could be used as the basis for discussion of particular constructions, and fill-the-gap exercises in order to better understand paradigmatic relations within clauses.

During subsequent trips, elicitation sessions focused on specific grammatical topics, sometimes on the basis of draft documents produced during the breaks between field trips. For example, there are several elicitation sessions that involve discussions on phonology, morphology and syntax on the basis of documents that I had drafted summarising the descriptive findings. I also used targeted elicitation to identify which word orders were possible in Kelabit. To do this, I used colour-coded word cards (with red for verbs, blue for nouns etc.).²¹ I would ask the consultant to build a sentence from the word blocks and subsequently change the order of the cards, asking the consultant to judge if the sentence was still acceptable and, if so, to provide a possible context for its use. This has the advantage of making visually clear what is being changed and avoiding a situation where a structure is judged to be bad on the basis of other factors (including the researcher's pronunciation or prosody).

I used the same method to investigate grammatical functions. I would initially ask for a translation (either of a verb or a whole sentence) and write this out on word cards. Then, I would rearrange the word cards and elicit a judgement as to whether the sentence was still possible and what differences in meaning arose. Again, this makes visually clear what is being changed, which is helpful in the context of complex constructions. To provide further evidence for the analysis of

²¹ As in many Western Austronesian languages, Kelabit roots can often derive either verbal or nominal forms under productive derivational processes. Similarly, adjectives have much in common with stative verbs. This has led to a debate in the literature as to whether Austronesian roots are pre-categorial (Foley, 1998). I follow Kroeger (1998b) and Himmelmann (2008) in arguing that word classes can be distinguished in Kelabit on account of their functional and distributional properties. For the purposes of this elicitation task only clear examples of nouns (e.g. humans, pronouns) and verbs (inflected, transitive verbs) were used.

Austronesian voice systems outlined in (2), I looked for syntactic constructions associated with subjects and objects cross-linguistically. This included binding, raising, control, quantifier float, secondary predicates, and the position of adverbs (Falk, 2006; Keenan, 1976). I then used the above method to elicit examples of each construction and negative evidence to show the difference in behaviour between the argument privileged in the verbal morphology (i.e. the actor in actor voice and the undergoer in undergoer voice) and other core arguments.

To investigate the role of information structure in syntactic choices, I used a similar word card method to explore the possible position of focus markers, such as *tupu* ‘only’ and *meto* ‘as well’. I asked the consultant to produce a possible sentence including the marker and then varied its position and elicited grammaticality judgements. I also created various information structure contexts using questions or contrast (see Lambrecht, 1994; Van der Wal, 2016), and asked speakers to judge/rank the grammaticality of different theoretically possible structures. Last but not least, I used the translation tasks in the Questionnaire on Information Structure (QUIS) (Skopeteas *et al.*, 2006) and the Questionnaire on Focus Semantics (QUIS_{SEM}) (Renans, Zimmermann, & Greif, 2011).

Elicitation sessions can be located in the deposit by selecting ‘elicitation’ as the genre. They are subcategorised as structured elicitation and semi-structured elicitation in the session titles. Structured elicitation includes anything that was elicited using a questionnaire or word list or closely following a draft document. Semi-structured elicitation includes all other sessions where the speakers had more control over the examples that they gave. These classifications can be accessed via the search bar on the deposit page. Recordings can also be filtered according to topic. The main topics for elicitation are: ‘information structure’; ‘morphology’; ‘phonology’; ‘syntax’ and ‘word list’. Within the main topic of morphology, recordings are also tagged with the sub-topics ‘verbs & verbal morphology’ and ‘word classes’. Within the main topic of syntax, recordings are tagged with the sub-topic ‘grammatical functions’, ‘pronouns’, ‘voice’ and ‘word order’. A full list of recordings is found in the Appendix.

3.2.2 Stimuli

There are three main resources collected using stimuli: the prosody experiment reading, the Pear Story collection, and the Unhappy Rats translation. These are all tasks where the same types of material were collected from multiple speakers. The prosody experiment set out to explore whether

pronouns in Kelabit are clitics and whether the two case forms (NOM and GEN) behave differently. I worked with Florance Lapu Apu to design a set of test sentences in Kelabit that varied according to the different contexts in which the pronouns occurred. This included different case forms, different verbs (i.e. both actor voice, undergoer voice and non-voice marked/lower transitivity verbs), different functions (i.e. subject or not), and different positions (i.e. pre-verbal, immediately post-verbal and clause-final). We identified several predicates that could take pronominal subjects and objects and designed sentences accordingly that were as close as possible to each other. We also included a context paragraph so that the test sentence was neither the first nor the last intonation unit of the utterance. In total 26 such paragraphs were created, and subsequently verified for naturalness with one male and one female speaker before approaching participants to record.

To collect the data, the paragraphs were presented to five participants - two men and three women - in a randomised order as a written document, with instructions to read each paragraph aloud with a short pause between them. Following Himmelmann and Ladd (2008), reading was seen as a good method of eliciting intonation without influencing the participants by having them repeat after the fieldworker or a native speaker research assistant. In order to prevent disfluency, the examples were printed in a large, clear font and speakers were given time to read through the sentences and familiarise themselves with the spelling system used. Each of the paragraphs was repeated twice. This amounts to 40 minutes of audio recording that is potentially useful for other forms of prosodic analysis where it is necessary to have the same words or sentences repeated by multiple speakers.

The Pear Story collection includes several retellings of the Pear Story from a video stimulus (Chafe, 1980) which I had piloted during the first field trip. In each case, two or more speakers were involved. One speaker would watch the video clip, which depicts a man picking pears, a boy who takes some of the pears and several ensuing incidents, all without audio. That speaker would then tell the story to the other speaker(s), who had not seen the clip. This enabled me to collect multiple narratives with similar content without using direct translation and creating any linguistic bias. The result is approximately 30 minutes of audio (and in one case video) recording.

The Unhappy Rats translation task was developed by Latrouite and Riester (2018) as a means of collecting comparable data on the role of information structure context on syntactic choices. It involves two sets of six short paragraphs, each containing the same target sentence but differing in terms of the context. In the first series, the target sentence has an indefinite/non-specific

undergoer ('cats chase rats') and in the second series, the target sentence has a definite undergoer ('my sister kicked my dog'). The paragraphs differ as to the status of information as topic, focus or background. For example, one paragraph establishes the dog as topic (e.g. 'my dog is the poorest dog in the world. He got abandoned as a puppy [...] also my sister kicked him'). Another establishes the sister as a topic (e.g. 'my sister is so mean. She must be the meanest person in the world. She plays tricks on people [...] also she kicks my dog'). A third is about violence in general, where both the sister and the dog represent new information (e.g. 'people are so aggressive nowadays. Take yesterday, someone drilled a hole into a car to steal gasoline [...] also my sister kicks my dog'). The paragraphs were given to six participants in English written form, who then had as much time as they wanted to translate them into Kelabit. Four of these translations were audio recorded, resulting in approximately 20 minutes of recording.

Finally, I piloted several picture and video stimuli with Florance Lapu Apu. Firstly, the Maus film clip – a video stimulus which the consultant was asked to narrate. Secondly, the Topological Relations Picture Series (Bowerman & Pederson, 1992) – an exercise in which the consultant is asked to describe the location of an object in a series of pictures. Thirdly, the Circle of Dirt picture story (Eisenbeiss, McGregor & Schmidt, 1999) in which the consultant is shown a series of pictures and asked to create a narrative around them. I first asked the consultant to describe each picture in turn, which we wrote out on paper, and then to recite the story from memory, which I subsequently recorded. Lastly, we attempted several exercises from the Questionnaire on Information Structure (QUIS) (Skopeteas *et al.*, 2006) and the Questionnaire on Focus Semantics (QUIS_{SEM}) (Renans *et al.*, 2011).

During the PhD research trip, we piloted QUIS experimental task 3 in which the consultant sees two pictures and is asked to describe them in turn (Skopeteas *et al.*, 2006, p. 39-73). The first picture is intended to provide some context for the second and differed as to whether the actor and undergoer were animate or inanimate and which of the two was given. During later research trips, we also piloted several experimental tasks:

1. QUIS experimental task 19 in which the consultant sees a storyboard of pictures entitled the Tomato Story and has the narrative explained to them in English (Skopeteas *et al.*, 2006, p. 149-155). The consultant then narrates the story from the perspective of (a) an external narrator, (b) the youngest child, and (c) the mother. They also answer a series of related questions;

2. QUIS experimental task 12 in which the consultant sees two pictures and is asked to describe each in turn (Skopeteas *et al.*, 2006, p. 107-110). This is designed to elicit contrastive topics;
3. QUIS experimental task 24 which is designed to elicit partial topics (Skopeteas *et al.*, 2006, p. 169-173). The consultant is shown a picture and asked a question that applies to only some of the people or items visible in the picture;
4. the picture stimuli from QUIS_{SEM} (Renans *et al.*, 2011, p. 33-35) was used for an experiment in which I asked target questions of the pictures that differed in their focus domain (e.g. all new, predicate focus, argument focus on actor, argument focus on undergoer) to see if this context would affect the structure of the answer. There are five pictures and I chose a different focus domain question for each picture, with two undergoer focus questions (one in actor voice and one in undergoer voice). Over five consecutive days, I repeated the experiment, varying which picture was introduced with which type of question.

The materials collected using stimuli can be subdivided into reading task, translation task, picture stimulus, and video stimulus, depending on the nature of the stimulus provided to speakers. These sub-genres appear in the titles of sessions and can be searched in the deposit. Users can also find materials by selecting ‘stimuli’ as the genre and then filtering by topic, namely ‘prosody experiment’, ‘pear story’, ‘unhappy rats’, and ‘information structure’. The other stimuli are listed by name.

3.2.3 Naturalistic Text

The corpus includes roughly 14 hours of naturalistic text or recordings of longer discourse that were not prompted by any linguistic or non-linguistic stimuli but simply recorded on request. In some cases, topics were suggested to the speaker based on their known interests or expertise and/or topics that had not previously been documented. In other cases, speakers had previously mentioned an interest in telling a particular story or describing a particular activity and an occasion was identified on which to record it. In Seifart’s (2008) terms, the selection of texts was largely opportunistic. I identified speakers both on the basis of recommendations and those who I knew well and had time to work with me. Whenever speakers were willing, I aimed to video record texts,

in order to document the extra-linguistic context of the recording. I took time to familiarise participants with the recording equipment in order to limit its potential intrusiveness.

Bowern (2015) recommends documenting audio, video, and written materials in as many genres as possible in order to ensure maximal representativeness. In this project, I tried to include genres that are culturally relevant and differ along the so-called ‘spontaneity parameter’ (Himmelman, 1998, p. 117) in terms of how ‘planned’ the recording is (Ochs, 1979). The sub-genres included under naturalistic text are: conversation, description, folk story, formal speech, joke, personal narrative, procedural text, radio broadcast, and song. These appear in the titles of each session and can be found using the search bar. Conversation is a cover term for any (typically unplanned) recording in which more than one speaker is conversing (including interview-like recordings). Description involves discussion and description of cultural activities, and is also unplanned but may have been discussed before and include domain-specific vocabulary. Folk stories are planned in the sense that the storyline is well-known and recited rather than spontaneously created. Formal speech is planned and written in advance. Personal narratives are texts with narrative structure that detail the narrator’s own experiences in the past, and are typically less planned than folk stories. Procedural texts are similar to descriptions but involve instructions on how to complete a task. Radio broadcasts include several news reports on Radio Bario that were translated in advance from the Borneo Post newspaper. They also include the flight schedule which followed a set structure every day. Finally, songs include several culture-specific categories, including *laku* (oral history/life stories), *ulin* (love song), *ri lekuah* (song of praise), *kuab* (war song), *sikih* (dream song), *sido* (mourning song), and children’s rhymes (see Saging & Bulan, 1989, p. 94-95 for further discussion of songs and oral traditions in Kelabit). Further information on each of these categories is provided in the description for individual recordings. Recording a variety of genres is not only important to ensure a more representative documentation but also useful for analysis of complex syntactic structures, which are known to be used in different ways in planned versus spontaneous speech.

The recordings can be located by selecting ‘naturalistic text’ as the genre, and then filtered according to topics. Some topics, including ‘Folk Story’, ‘Joke’, ‘Radio Bario Broadcast’, ‘Speeches’ and ‘Songs’ correspond directly to the sub-genre and allow for easy filtering. Within the topic of folk stories, I tagged recordings with the sub-topics ‘Animal Stories’, ‘Legends’ and ‘Palug Rayeh & Palug I’it’. Animal stories include traditional stories about the turtle, the

mousedeer, the monkey, and other stories that are common in Borneo. Legends include oral traditions, such as the story of Batuh Lawih and Batuh Apui who fell out over Batuh Lawih’s wife. Finally, the *Palug Rayeh & Palug I’it* (lit. Older Fool and Younger Fool) stories are a collection of tales about two brothers and the tricks that the younger brother plays on the elder brother.

In addition to the topics listed above, there are also recordings tagged with the topics: ‘Culture and Traditions’, ‘Oral History’, and ‘The Kelabit Language’. These recordings include different genres (e.g. description, personal narrative, conversation) but are linked by an overall theme. Oral History covers any recording relating to important events in the past, including discussions of the first school in the Kelabit Highlands, the Bario Revival in 1973, the development of the eBario project to bring IT resources to the Highlands, and memories of the Confrontation and the Second World War. The Kelabit Language covers any recording relating to the status of the Kelabit language and community revitalisation efforts. Finally, Culture and Traditions is a cover topic for a series of sub-topics relating to customs and practices in the Kelabit Highlands, as set out in Table 2 (see also the Appendix)

Table 2: Sub-topics of Culture and Traditions

Topic	Content
Arts & Crafts	material culture and entertainment, including Kelabit dances and traditional weaving
Farming	hill farming and wet paddy farming; harvesting and processing rice; activities related to farming, such as scaring birds away from ripe paddy
Food and Feasts	preparation of food, such as the traditional way of serving rice wrapped in a leaf (<i>nuba’ laya’</i>); traditional Kelabit weddings; important feasts such as the name changing ceremony
Hunting, Gathering & Fishing	practices of hunting in the jungle; different methods of fishing; collecting wild honey; gathering wild yams
Making Salt	traditional method of making salt from local salt springs
Traditional Beliefs	taboos and omens observed in the past; the giant Pun Tumid
Traditional Games	childhood games and hobbies
Traditional Knowledge	traditional medicinal plants
Travel and Transport	walking through and communicating in the jungle
Village Life	longhouses and life in Kelabit villages

3.3 Accessing and Using the Deposit

Everyone recorded as part of the project consented to the recording and specified the level of access using the current ELAR categories of open access (O), accessible to all registered users of ELAR (U), and restricted to subscribers, who must first approach the depositor (S). The vast majority of deposit recordings are U, however, some speakers requested that their recordings remain for the community and/or approved users only (access level S).²²

Please cite the collection as follows: Hemmings, Charlotte. 2019. Documentation of the Kelabit Language, Sarawak, Malaysia. London: SOAS, Endangered Languages Archive. URL: <https://elar.soas.ac.uk/Collection/MPI1029735>. Accessed on [insert date here]. To cite data from individual bundles, please reference the bundle ID (e.g. BAR22102013CH_05). Please also acknowledge the contribution of individuals who have collected, transcribed, or translated recordings, as given in the metadata for each recording.

4. Evaluation of Research Methods in the Kelabit Documentation Project

4.1 Elicitation

The elicitation methods used in the project included translation from English to Kelabit, and grammaticality judgements on word formation, word order, and syntactic constructions. Elicitation is sometimes criticised as a method of collecting linguistic data since the materials are by definition influenced by the researcher, and translation is known to introduce bias (Dimmeiulaal, 2001; Himmelmann, 2006; Lüpke, 2009; Mithun, 2001). However, it can be used to complement naturalistic text data in providing full paradigms of forms that may appear only rarely in a corpus (Seifart, 2008). Moreover, it allows the discovery of metalinguistic awareness of speakers and is useful in providing negative examples of structures that are not grammatical (Lüpke, 2009). Consequently, it is important for providing linguistic evidence that can be further supported on the basis of non-elicited data.

The major limitation of elicitation is that the researcher only collects data relating to the topics that they decide to research – and may therefore miss other structures of potential interest. Moreover, in the case of grammaticality judgements, when a structure is judged to be unacceptable it is not always clear what this judgement really means. For example, it could mean that a structure

²² There are currently only three bundles with the access description S.

is not part of the grammar of the language, or that it is semantically odd or pragmatically weird. Even if the structure is judged to be ungrammatical in its technical sense, it may not be for the reason that the linguist expects. For example, in my research into word order, I was interested in determining the order of arguments (i.e. actor & undergoer) in the different voice constructions and how this related to the grammatical function that they fulfilled. However, word order is not only affected by grammatical function but also by the definiteness of arguments, among other factors. There is a risk, therefore, that a linguist who does not know the grammar well may interpret an ungrammatical word order as stemming from the function of an argument, when it actually stems from its definiteness. To minimise the risk of misinterpreting grammaticality judgements, I have worked with the same language consultants over a number of years and explicitly discussed what it means to be ungrammatical. In more recent work, I attempt to elicit judgements in context and to confirm judgements with more than one speaker and/or on more than one occasion.²³

Elicitation has been extremely useful in the project as the basis of the grammatical analysis presented in Hemmings (2016) and in identifying the syntactic choices available to speakers. It provided data that was immediately incorporated into analyses and directly addressed the research questions of the linguist. There is some debate as to whether description of this sort actively contributes to documentation. Indeed, in his seminal work defining the subfield of language documentation, Himmelmann (1998, 2006) clearly distinguished between the goals and methods of documentation and description. However, I would argue, following Austin and Grenoble (2007) and Austin (2016), that descriptive analysis is an important part of making the documentary corpus accessible. It can be vital for understanding the function of constructions that are attested in naturalistic texts, as well as for documenting less frequent and more complex constructions. Moreover, it informs our annotation of the primary data, from translation and glossing to more detailed theoretical notes. Consequently, elicitation can be seen as directly contributing to the knowledge of the linguist, which in turn allows them to achieve the stated aims of a documentation that has added-value and becomes useable to a range of potential future users (Austin, 2016). Nonetheless, it is clear that elicitation alone cannot produce a representative documentation and

²³ In some cases, of course, different speakers give different judgements. Moreover, for some marginal constructions different judgements may be given on different occasions. This reflects the fact that grammaticality is not always binary.

should be supplemented with other materials that may be of more interest to the community and their efforts to revitalise the language.²⁴

4.2 Stimuli

Using picture and video stimuli is a means of eliciting data with some control over content, form, or context, whilst avoiding the pitfalls of translation bias typically associated with structured elicitation. As discussed above, it has the advantage that one can use the same stimuli with multiple speakers to collect more than one version of similar texts. This can be useful when exploring variation, and ensures that analyses are not based purely on the idiolect of a single speaker. In the case of established stimuli in language documentation, such as the Pear Story (Chafe, 1980), it can also add to parallel texts across different languages to enable cross-linguistic or typological comparison (Lüpke, 2009).

In this project, the Pear Story collection and the reading task allowed me to collect useful data for the analysis of voice, word order, and pronoun status (Hemmings, 2016), and the Unhappy Rats translation task provided data on how information structure affects morpho-syntactic choices in Kelabit. This fed into the analysis in providing evidence for how structures are really used in context, and how usage frequency can be compared across speakers. Moreover, the context of the recording, and the vocabulary choices, are more controlled than in the naturalistic text corpus.

However, the texts are best considered staged communicative events (Himmelmann, 1998; Lüpke, 2009) as they were created for the purposes of the research project rather than reflecting naturalistic language use, arguably making them of less value for creating a representative documentation. Moreover, in contrast to elicitation, the texts required further preparation and coding before they could be analysed. In the case of the reading task, developing the sentences also took quite a long time as we attempted to find sentences that varied according to the test variables but still sounded natural.²⁵ Consequently, stimuli can take more preparation than

²⁴ As a reviewer pointed out, it is not necessarily the case that elicited materials are less useful in revitalisation than naturalistic texts. For example, though naturalistic text may be rich in culturally relevant information, it may be inaccessible for new speakers, and may not include the types of phrases that are useful for language learners. In this case, elicitation can help to record the materials that learners might need (e.g. basic greetings, asking for clarification etc.). In this project, recording BAR15102013CH_01 includes some useful phrases for language learners taken from a Malay textbook (Othman, 2012). However, the vast majority of elicitation relates to grammatical description.

²⁵ Creating sentences that sound natural but also contain the relevant variables and no other confounding factors is not always easy. The time constraints of fieldwork mean that you may not be able to repeat the experiment and hence preparation is important.

questionnaires or translation elicitation, particularly if they are created to address a particular research question or specialised for the cultural context of the field site. When pre-existing stimuli are used, such as the Pear Story, it can be much quicker to run the experiment but it may not always be clear to the participant what they are supposed to say, and certain vocabulary items (e.g. pear!) may not exist in the language. Given that cultural contexts differ, it is not clear to what extent one can assume that stimuli are interpreted in the same way and hence that cross-linguistically comparable data has really been collected. Moreover, the materials collected are unlikely to be culturally-relevant, and hence, much like elicitation, stimuli data may be a less desirable outcome for the speaker community.

I have found that retelling a story from a video stimulus is generally clearer for speakers than picture tasks, where the narrative outline may be less obvious. In the case of the QUIS manual tasks, for example, it was often not clear to either the researcher or the consultant exactly how to conduct the task, or what was required as an answer, creating a source of frustration. To use stimuli appropriately requires some training time for both the researcher and the participant with a test set. The reading task worked well as a source of data on prosody, however, it was harder to conduct in the context of a language without a written tradition as the naturalness of the reading is affected by the speaker's familiarity with the orthography used in the stimuli. Finally, the translation task was a useful means of collecting comparable data on the effect of information structure on morphosyntax in Kelabit. However, it did not avoid translation bias and one has to make the assumption that the speaker interpreted the information structure contexts in the way intended in the stimuli. Consequently, one has to interpret the outputs of such experiments with an understanding that this may not accurately represent everyday language use.

4.3 Naturalistic Text

Naturalistic text collection is often considered the most important component of a documentation project (Lüpke, 2009), allowing for the collection of texts that are culturally relevant and reflect the local customs and practices of the speaker community. As such, the materials may be of wider use than elicitation or staged communicative events to the community themselves and also to anthropologists, historians, and other people interested in the content of recordings rather than the form of the language. It has the advantage of limiting the influence of the researcher and any working languages, since the person being recorded has control over what they say, rather than

translating. The constructions and lexical items used in such texts may also be unknown to the researcher and prompt additional research questions. This can help the researcher to better understand the language, rather than simply looking for patterns that they already expect to find. However, as with texts produced using stimuli, the recordings need to be processed before they can be useful for analysis. At the very least, the assistance of native speakers is needed to transcribe, translate and gloss texts – tasks which all require time, training and expertise. Last but not least, naturalistic text corpora cannot provide negative evidence that an example is ungrammatical. and may not provide many examples of rarer constructions. In such cases, elicitation is necessary in combination with naturalistic data.

In this project, I collected a range of naturalistic texts in different genres and have used this to show the impact that text genre can have on the frequency of different constructions, comparing news reports, folk stories, and narratives collected using the Pear Story stimulus (Hemmings, 2016, forthcoming). I also found the transcription and translation of texts very useful as a means of identifying constructions and forms that were relevant to the analysis of voice and information structure. Indeed, the discovery of examples in the naturalistic text corpus often led to the development of a hypothesis that could be further tested using elicitation or stimuli. For documentation purposes, the text collection was also the most representative of naturalistic language use and documents cultural practices as well as culture-specific vocabulary.

As to whether the naturalistic text corpus is fully representative and comprehensive, unfortunately the answer is probably not. As is often the case in documentation projects (see Austin, 2016), certain genres – such as narrative and description – are over-represented in the corpus, whilst others – such as day-to-day conversation²⁶ – are under-represented. Moreover, many recordings were made with only myself and the speaker present, which means that communicative events were removed from the context in which they would normally occur. For example, folk stories and formal speeches were produced without the context of an audience, making the performance not entirely naturalistic. Furthermore, the stories and descriptions were often narrated to me, a non-native speaker, which undoubtedly had an effect on the language used. To limit the effects of the observer's paradox (Labov, 1972), I spent time trying to increase my fluency in

²⁶ This is partly a logistical issue in that it is easier to arrange for one speaker to be free at a given time than multiple speakers. Moreover, being recorded having an everyday conversation may feel much more unnatural than recording a description of a cultural practice or a traditional story. Newer forms of language use, such as Facebook posts or text messages, are completely absent from the deposit.

Kelabit so that I could understand what speakers were saying to me and thereby making the communicative context more natural. I also found that having another native speaker present, or in the recording event itself, makes participants feel more comfortable and at ease.

Finally, it is sometimes argued that language documenters focus on what Woodbury (2011) terms ‘the ancestral code’, documenting an idealised version of the language as it may once have been spoken, rather than the documenting the linguistic ecology that endangered languages exist in today (see Austin, 2016 and references therein). To a certain extent, this project is guilty of this critique – focusing on Kelabit without reference to the multilingual repertoires of speakers beyond including them as metadata. Recordings do in some cases include examples of code-mixing and other contact-phenomena that are common in the context of endangered languages (O’Shannessy, 2011). However, there has not yet been a systematic effort to document differences between Kelabit as spoken in the villages, and Kelabit as spoken in towns and cities, despite the sociolinguistic differences discussed above. Neither has there been systematic effort to describe differences between the Kelabit spoken by older generations and that spoken by younger generations. Thus, there are many ways in which the comprehensiveness and the representativeness of the documentation could be improved. As a first step towards this, I now work closely with key project members, such as Florance Lapu Apu, to agree on aspects of language and culture that could benefit from being documented, and approach speakers to record on these topics or in these genres. This ultimately results in a richer documentation, greater community involvement and control, and helps to ensure that the documentation is useful to a wide range of users.

In summary, the project has benefitted from using a range of research methods, including elicitation, stimuli, and naturalistic text collection. Different methods produce outputs with different uses (and potential users) and allow the researcher to address different research questions. The methods differ in terms of how natural the resulting data is and how much control the linguist has over the structure of the data collected. Elicitation gives the most control, but results in the least natural form of data, since examples are given out of context and open to translation bias. The use of stimuli allows the linguist to collect multiple versions of the same text, which is invaluable for comparison. However, depending on the nature of the task, the data may not be entirely naturalistic and the linguist has less direct control over the structures that are produced, meaning that the data may be harder to interpret. Finally, naturalistic text collection results in data

that has the least influence from the researcher and is consequently the most natural. Nonetheless, the researcher still has an effect on the data by being present at the recording, and, of course, by what they happen to record and what they do not. The important point is to collect data using a variety of methods in order to triangulate any analyses or description, and to create a documentation that is as representative and multipurpose as possible (Himmelman, 2006).

5. Conclusion

This paper is a response to the need to provide meta-documentation for language documentation projects (Austin, 2013) and, in so doing, make the outputs of documentation more accessible to users (Woodbury, 2014). It has contextualised the Kelabit documentation project with its dual aims of providing a representative documentation and basic description of the grammar alongside analysis of voice alternations. It has also provided an overview of the materials collected and the different methods used in data collection, evaluating each in terms of the data collected and the specific uses that it has been put to in the project. As such, it is hoped that this will be a useful resource for anyone accessing the archival deposit of Kelabit materials in ELAR (Hemmings, 2019), as well as for linguists working with similar phenomena in related languages.

References

- Adelaar, A. (1995). Borneo as a cross-roads for comparative Austronesian linguistics. In P. Bellwood, J. J. Fox & D. Tyron (Eds.), *The Austronesians: Historical and comparative perspectives* (pp. 75-95). Canberra, Australia: Australian National University.
- Adelaar, A. (2005). The Austronesian languages of Asia and Madagascar: A historical perspective. In A. Adelaar & N. P. Himmelmann (Eds.), *The Austronesian languages of Asia and Madagascar* (pp. 350-376). London, England: Routledge.
- Adelaar, A. (2013). Voice variation in Austronesian languages of Indonesia: Introduction. In A. Adelaar (Ed.), *Voice variation in Austronesian languages of Indonesia* (Vol. 54) (pp. 1-3). Jakarta, Indonesia: NUSA.
- Aldridge, E. (2004). *Ergativity and word order in Austronesian languages*. (Unpublished doctoral dissertation). Cornell University, Ithaca, New York, NY.
- Aldridge, E. (2012). Antipassive and ergativity in Tagalog. *Lingua*, 122(3), 192-203.

- Amster, M. H. (1995). *Kelabit/English, English/Kelabit glossary: A concise guide to the Kelabit language*. Kuching, Malaysia: Rurum Kelabit Sarawak.
- Amster, M. H. (1998). *Community, ethnicity, and modes of association among the Kelabit of Sarawak, East Malaysia*. (Unpublished doctoral dissertation). Brandeis University, Waltham, MA.
- Amster, M. H. (1999). Tradition, ethnicity, and change: Kelabit practices of name changing. *Sarawak Museum Journal*, 54(75), 183-200.
- Amster, M. H. (2003). Gender complementarity and death among the Kelabit. In W. D. Wilder (Ed.), *Journeys of the soul: Anthropological studies of death, burial, and reburial practices in Borneo* (pp. 251-307). Kuching, Malaysia: Borneo Research Council.
- Amster, M. H. (2006). Narrating the border: Perspectives from the Kelabit Highlands of Borneo. In A. Horstmann & R. L. Wadley (Eds.), *Centering the margin: Agency and narrative in Southeast Asian borderlands* (pp. 209-230). New York, NY: Berghahn Books.
- Arka, I. W. (2002, July). *Voice systems in the Austronesian languages of Nusantara: Typology, symmetry and undergoer orientation*. Paper presented at the 10th National Symposium of the Indonesia Linguistics Society, Bali, Indonesia.
- Arka, I. W., & Ross, M. (Eds.). (2005). *The many faces of Austronesian voice systems: Some new empirical studies*. Canberra, Australia: Pacific Linguistics.
- Asmah, H. O. (1983). *The Malay peoples of Malaysia and their languages*. Kuala Lumpur, Malaysia: Dewan Bahasa dan Pustaka.
- Austin, P. K. (2013). Language documentation and meta-documentation. In S. Ogilvie & M. Jones (Eds.), *Keeping languages alive: Documentation, pedagogy and revitalization* (pp. 3-15). Cambridge, England: Cambridge University Press.
- Austin, P. K. (2016). Language documentation 20 years on. In L. Filipović & M. Pütz (Eds.), *Endangered languages and languages in danger: Issues of documentation, policy, and language rights* (pp. 147-170). Amsterdam, Netherlands: John Benjamins.
- Austin, P. K., & Grenoble, L. (2007). Current trends in language documentation. *Language Documentation and Description*, 4, 12-25.
- Bala, P. (2002). *Changing borders and identities in the Kelabit Highlands: Anthropological reflections on growing up near an international border*. Kota Samarahan, Malaysia: UNIMAS Publisher.

- Bickel, B. (2010). Grammatical relations typology. In J. J. Song (Ed.), *The Oxford handbook of linguistic typology* (pp. 399-444). Oxford, England: Oxford University Press.
- Bird, S., & Simons, G. (2003). Seven dimensions of portability for language documentation and description. *Language*, 79(3), 557-582.
- Blust, R. (1974). *The Proto-North Sarawak vowel deletion hypothesis*. (Unpublished doctoral dissertation). University of Hawai'i at Mānoa, Honolulu, HI.
- Blust, R. (1993). Kelabit-English vocabulary. *Sarawak Museum Journal*, 44(65), 141-226.
- Blust, R. (2006). The origin of the Kelabit voiced aspirates: A historical hypothesis revisited. *Oceanic Linguistics*, 45(2), 311-338.
- Blust, R. (2013). *The Austronesian languages*. Canberra, Australia: Pacific Linguistics.
- Blust, R. (2016). Kelabit-Lun Dayeh phonology, with special reference to the voiced aspirates. *Oceanic Linguistics*, 55(1), 246-277.
- Bowerman, M., & Pederson, E. (1992). Topological relations picture series. In *Space stimuli kit 1.2* (pp. 51): Max Planck Institute for Psycholinguistics.
- Bowern, C. (2015). *Linguistic fieldwork: A practical guide*. New York, NY: Palgrave MacMillan.
- Bulan, L., & Labang, D. L. (2008, July). *Community efforts in developing and financing an MLE project: The Kelabit language of Sarawak*. Paper presented at the 2nd International Conference on Language Development, Language Revitalisation, and Multilingual Education in Ethnolinguistic Communities, Bangkok, Thailand.
- Caballero, G. (2017). Choguita Rarámuri (Tarahumara) language description and documentation: A guide to the deposited collection and associated materials. *Language Documentation & Conservation*, 11, 224-255.
- Chafe, W. L. (1980). *The pear stories: Cognitive, cultural, and linguistic aspects of narrative production*. Norwood, NJ: Ablex.
- Chen, V., & McDonnell, B. (2019). Western Austronesian voice. *Annual Review of Linguistics*, 5, 173-195.
- Clayre, B. (2005). Kelabitic languages and the fate of 'focus': Evidence from the Kerayan. In I. W. Arka & M. Ross (Eds.), *The many faces of Austronesian voice systems: Some new empirical studies* (pp. 17-57). Canberra, Australia: Pacific Linguistics.

- Clayre, B. (2014). A preliminary typology of the languages of Middle Borneo. In P. Sercombe, M. Boutin & A. Clynes (Eds.), *Advances in research on cultural and linguistic practices in Borneo* (pp. 123-151). Phillips, ME: Borneo Research Council.
- de Crespigny, C. C. (1896). *Memoirs of Sir Claude Champion de Crespigny, Bart.* London, England: Lawrence and Bullen.
- Dimmeiulaal, G. J. (2001). Places and people: Field sites and informants. In P. Newman & M. Ratliff (Eds.), *Linguistic Fieldwork* (pp. 54-75). Cambridge, England: Cambridge University Press.
- Dipper, S., Götze, M., & Skopeteas, S. (2007). *Information structure in cross-linguistic corpora: Annotation guidelines for phonology, morphology, syntax.* Potsdam, Germany: Universitätsverlag Potsdam.
- Douglas, R. S. (1911). A comparative vocabulary of the Kayan, Kenyan and Kelabit dialects. *Sarawak Museum Journal*, 1(1), 75-119.
- Dryer, M. S. (1997). Are grammatical relations universal. In J. Bybee, J. Haiman, & S. A. Thompson (Eds.), *Essays on language function and language type dedicated to T. Givón* (pp. 115-143). Amsterdam, Netherlands: John Benjamins.
- Eghenter, C., & Langub, J. (2008). Past meets future: A trans-border forum for a sustainable future for the highlands of Borneo. *Borneo Research Bulletin*, 39, 286–294.
- Eisenbeiss, S., McGregor, B., & Schmidt, C. M. (1999). Story book stimulus for the elicitation of external possessor constructions and dative constructions ('the circle of dirt'). In D. Wilkins (Ed.), *Manual for the 1999 field season* (pp. 140-144). Nijmegen, Netherlands: Max Planck Institute for Psycholinguistics.
- ELAN [Computer Software]. (2019). Retrieved from <https://tla.mpi.nl/tools/tla-tools/elan/>
- Erlewine, M. Y., Levin, T., & van Urk, C. (2014, May). *What makes a voice system? On the relationship between voice marking and case.* Paper presented at the Proceedings of the 21st Meeting of the Austronesian Formal Linguistics Association (AFLA), the University of Hawai'i at Mānoa, Honolulu, HI.
- Falk, Y. N. (2006). *Subjects and universal grammar: An explanatory theory.* Cambridge, England: Cambridge University Press.

- Foley, W. A. (1998, June/July). *Symmetrical voice systems and precategoriality in Philippine languages*. Paper presented at the Workshop on Voice and Grammatical Functions in Austronesian Languages, LFG98, Brisbane, Australia.
- Foley, W. A. (2008). The place of Philippine languages in a typology of voice systems. In P. K. Austin & S. Musgrave (Eds.), *Voice and grammatical relations in Austronesian languages* (pp. 22-44). Stanford, CA: CSLI Publications.
- Gippert, J., Himmelmann, N., & Mosel, U. (2006). *Essentials of language documentation*. Berlin, Germany: Mouton de Gruyter.
- Haig, G., & Schnell, S. (2014). Annotations using GRAID (Grammatical Relations and Animacy in Discourse): Introduction and guidelines for annotators (Version 7.0). Retrieved from <https://lac.uni-koeln.de/en/multicast/>
- Harris, C. A., & Harris, R. W. (2011). Information and communication technologies for cultural transmission among indigenous peoples. *The Electronic Journal of Information Systems in Developing Countries*, 45(1), 1-19.
- Harrison, T. (1954). Outside influences on the culture of the Kelabits of North Central Borneo. *Sarawak Museum Journal*, 6(4), 104-125.
- Harrison, T. (1959a). The Kelabits and Muruts. In T. Harrison (Ed.), *The peoples of Sarawak* (pp. 57-72). Kuching, Malaysia: Sarawak Museum.
- Harrison, T. (1959b). *World within: A Borneo story*. London, England: The Cresset Press.
- Haspelmath, M. (1993). More on the typology of inchoative/causative verb alternations. In B. Comrie & M. Polinsky (Eds.), *Causatives and transitivity* (Vol. 23) (pp. 87-121). Amsterdam, Netherlands: John Benjamins.
- Haude, K., & Zúñiga, F. (2016). Inverse and symmetrical voice: On languages with two transitive constructions. *Linguistics*, 54(3), 443-481.
- Hemmings, C. (2015). Kelabit voice: Philippine-type, Indonesian-type or something a bit different? *Transactions of the Philological Society*, 113(3), 383-405.
- Hemmings, C. (2016). *The Kelabit language, Austronesian voice and syntactic typology*. (Unpublished doctoral dissertation). SOAS, University of London, England.
- Hemmings, C. (2019). Documentation of the Kelabit Language, Sarawak, Malaysia. Retrieved from <https://elar.soas.ac.uk/Collection/MPI1029735>

- Hemmings, C. (forthcoming). When an antipassive isn't an antipassive anymore: The actor voice construction in Kelabit. In K. Janic & A. Witzlack-Makarevich (Eds.), *The multifaceted aspects of antipassive*. Amsterdam, Netherlands: John Benjamins.
- Himmelman, N. P. (1998). Documentary and descriptive linguistics. *Linguistics*, 36, 161-195.
- Himmelman, N. P. (2005). The Austronesian languages of Asia and Madagascar: Typological characteristics. In A. Adelaar & N. P. Himmelman (Eds.), *The Austronesian languages of Asia and Madagascar* (pp. 110-181). London, England: Routledge.
- Himmelman, N. P. (2006). Language documentation: What is it and what is it good for. *Essentials of Language Documentation*, 178, 1-30.
- Himmelman, N. P. (2008). Lexical categories and voice in Tagalog. In P. K. Austin & S. Musgrave (Eds.), *Voice and grammatical relations in Austronesian languages* (pp. 247-293). Stanford, CA: CLSI.
- Himmelman, N. P., & Ladd, D. R. (2008). Prosodic description: An introduction for fieldworkers. *Language Documentation & Conservation*, 2(2), 244-274.
- Hudson, A. B. (1978). Linguistic relations among Bornean peoples with special reference to Sarawak: An interim report. *Studies in Third World Societies*, 3, 1-44.
- Janowski, M. (1988). The motivating forces behind changes in the wet rice agricultural system in the Kelabit Highlands. *Sarawak Gazette*, 114(1504), 9-21.
- Janowski, M. (1991). *Rice, work and community among the Kelabit of Sarawak, East Malaysia*. (Unpublished doctoral dissertation). London School of Economics, England.
- Janowski, M. (1995). The hearth-group, the conjugal couple and the symbolism of the rice meal among the Kelabit of Sarawak. In J. Carsten & S. Hugh-Jones (Eds.), *About the house: Levi-Strauss and beyond* (pp. 84-104). Cambridge, England: Cambridge University Press.
- Janowski, M. (2003). *The forest, source of life: The Kelabit of Sarawak*. Kuching, Malaysia: Sarawak Museum.
- Janowski, M. (2012). Imagining the forces of life and the cosmos in the Kelabit Highlands. In M. Janowski & T. Ingold (Eds.), *Imagining landscapes: Past, present and future* (pp. 143-163). London, England: Ashgate.
- Johnston, R. L. (1989). *Grammar and basic vocabulary in Oceanic Austronesian languages: A standard elicitation schedule*. Ukarumpa, Papua New Guinea: Summer Institute of Linguistics.

- Keenan, E. L. (1976). Towards a universal definition of 'subject'. In C. N. Li (Ed.), *Subject and topic* (pp. 303-333). New York, NY: Academic Press.
- King, V. T. (1993). *The peoples of Borneo*. Oxford, England: Blackwell.
- Kroeger, P. R. (1993). *Phrase structure and grammatical relations in Tagalog*. Stanford, CA: CSLI Publications.
- Kroeger, P. R. (1998a). Language classification in Sarawak: A status report. *The Sarawak Museum Journal*, 53(74), 137-173.
- Kroeger, P. R. (1998b, June/July). *Nouns and verbs in Tagalog: A reply to Foley*. Paper presented at the Workshop on Voice and Grammatical Functions in Austronesian Languages, LFG98, Brisbane, Australia.
- Kulikov, L. (2011). Voice typology. In J. J. Song (Ed.), *The Oxford handbook of linguistic typology* (pp. 368-398). Oxford, England: Oxford University Press.
- Labang, D. L. (2012). *Karuh Tauh – Bahasa Kelabit: Simple orthography*. In Rurum Kelabit Sarawak Education Unit, 11th Annual Education Excellence Award. Miri, Sarawak.
- Labov, W. (1972). *Sociolinguistic patterns*. Oxford, England: Blackwell.
- Lambrecht, K. (1994). *Information structure and sentence form: Topic, focus, and the mental representations of discourse referents*. Cambridge, England: Cambridge University Press.
- Latrouite, A., & Riester, A. (2018). The role of information structure for morphosyntactic choices in Tagalog. In S. Riesberg, A. Shiohara & A. Utsumi (Eds.), *Information structure in Austronesian languages* (pp. 247-284). Berlin, Germany: Language Science Press.
- Lüpke, F. (2009). Data collection methods for field-based language documentation. *Language Documentation and Description*, 6, 53-100.
- Martin, P. W. (1996). A comparative ethnolinguistic survey of the Murut (Lun Bawang) with special reference to Brunei. In P. W. Martin, C. Ožóg, & G. Poedjosoedarmo (Eds.), *Language use and language change in Brunei Darussalam* (pp. 268-279). Athens, OH: Ohio University Press.
- Martin, P. W., & Yen, E. (1994). Language use among the Kelabit living in urban centres. In P. W. Martin (Ed.), *Shifting patterns of language use in Borneo* (pp. 147-163). Williamsburg, VA: Borneo Research Council.

- Mithun, M. (2001). Who shapes the record: The speaker and the linguist. In P. Newman & M. Ratliff (Eds.), *Linguistic fieldwork* (pp. 34-54). Cambridge, England: Cambridge University Press.
- O'Shannessy, C. (2011). Language contact and change in endangered languages. In P. K. Austin & J. Sallabank (Eds.), *The Cambridge handbook of endangered languages* (pp. 78-99). Cambridge, England: Cambridge University Press.
- Ochs, E. (1979). Planned and unplanned discourse in discourse and syntax. *Syntax and Semantics Ann Arbor, Mich.*, 12, 51-80.
- Othman, Z. (2012). *Colloquial Malay: The complete course for beginners*. Abingdon, England: Routledge.
- Ray, S. H. (1913). The languages of Borneo. *Sarawak Museum Journal*, 1(4), 1-196.
- Renans, A., Zimmermann, M., & Greif, M. (2011). *Interdisciplinary studies on information structure: Questionnaire on focus semantics* (Vol. 15) (2nd edn.). Potsdam, Germany: Universitätsverlag Potsdam.
- Rethinasamy, S. (2014). *Language use among the Kelabit in Bario, Sarawak*. (Unpublished ms). UNIMAS, Sarawak, Malaysia.
- Riesberg, S. (2014). *Symmetrical voice and linking in western Austronesian languages*. Berlin, Germany: De Gruyter Mouton.
- Roth, H. L. (1896). *The natives of Sarawak and British North Borneo* (Vols 1 & 2). London, England: Truslove and Hanson.
- Rubenstein, C. (1973). Poems of indigenous peoples of Sarawak: Some of the songs and chants. *Sarawak Museum Journal, Special Monograph 2*.
- Rutter, O. (1929). *The pagans of North Borneo*. London, England: Hutchinson.
- Saging, R. L. R. (1976/77). *An ethno-history of the Kelabit tribe of Sarawak. A brief look at the Kelabit tribe before World War II and after*. (Graduation Exercise submitted in partial fulfilment of the requirements for the Degree of Bachelor of Arts, Hons., Jabatan Sejarah, University of Malaya, Kuala Lumpur).
- Saging, R. L. R., & Bulan, L. (1989). Kelabit ethnography: A brief report. *Sarawak Museum Journal*, 40(61), 89-118.
- Salffner, S. (2015). A guide to the Ikaan language and culture documentation. *Language Documentation & Conservation*, 9, 237-267.

- Schachter, P. (1976). The subject in Philippine languages: Topic, actor, actor-topic, or none of the above? In C. N. Li (Ed.), *Subject and topic* (pp. 491-518). New York, NY: Academic Press.
- Schachter, P. (1996). The subject in Tagalog: Still none of the above. *UCLA Occasional Papers in Linguistics*, 15, 1-61.
- Schiller, A. (2007). Activism and identities in an East Kalimantan Dayak organization. *The Journal of Asian Studies*, 66(1), 63-95.
- Schneeberger, W. F. (1979). *Contributions to the ethnology of central northeast Borneo: Parts of Kalimantan, Sarawak and Sabah*. Berne, Switzerland: Institute of Ethnology, University of Berne.
- Schultze-Berndt, E. (2006). Linguistic annotation. In J. Gippert, N. P. Himmelmann & U. Mosel (Eds.), *Essentials of language documentation* (pp. 213-251). Berlin, Germany: Mouton de Gruyter.
- Seifart, F. (2008). On the representativeness of language documentations. *Language Documentation and Description*, 5, 60-76.
- Simons, G. F., & Fennig, C. D. (2018). *Ethnologue: Languages of the world* (21st edn.). Dallas, TX: SIL International.
- Skopeteas, S., Fiedler, I., Hellmuth, S., Schwarz, A., Stoel, R., Fanselow, G., . . . Krifka, M. (2006). *Questionnaire on information structure (QUIS): Reference manual* (Vol. 4). Potsdam, Germany: Universitätsverlag Potsdam.
- Smith, A. D. (2017). *The languages of Borneo: A comprehensive classification*. (Unpublished doctoral dissertation). University of Hawai'i at Mānoa, Honolulu, HI.
- Southwell, C. H. (1949). Structure of the Murut language. *Sarawak Museum Journal*, 5(1949), 104-115.
- Talla, Y. (1979). *The Kelabit of the Kelabit Highlands, Sarawak*. Penang, Malaysia: School of Comparative Social Sciences, Universiti Sains Malaysia.
- Van der Wal, J. (2016). Diagnosing focus. *Studies in Language*, 40(2), 259-301.
- Woodbury, A. C. (2003). Defining documentary linguistics. *Language Documentation and Description*, 1, 35-51.
- Woodbury, A. C. (2011). Language documentation. In P. K. Austin & J. Sallabank (Eds.), *The Cambridge handbook of endangered languages* (pp. 159-186). Cambridge, England: Cambridge University Press.

Woodbury, A. C. (2014). Archives and audiences: Toward making endangered language documentations people can read, use, understand, and admire. *Language Documentation and Description*, 12, 19-36.

Appendix

*A.1 Elicitation Topics*²⁷

1. Information Structure

BAR20082014CH_02 Structured Elicitation – Information Structure 01

Charlotte & Florance use the translation task from QUIS to discuss the effect of information structure on word order in Bario Kelabit.

BAR20190204CH_01 Structured Elicitation – QUIS Manual Translation Task 01

Charlotte & Florance work through the translation exercise from the QUIS manual to discuss information structure in Bario Kelabit.

BAR20190204CH_02 Structured Elicitation – QUIS Manual Translation Task 02

Charlotte & Florance work through the translation exercise from the QUIS manual to discuss information structure in Bario Kelabit.

BAR20190204CH_03 Structured Elicitation – QUIS Manual Translation Task 03

Charlotte & Florance work through the translation exercise from the QUIS manual to discuss information structure in Bario Kelabit.

BAR20190204CH_04 Structured Elicitation – QUIS Manual Translation Task 04

Charlotte & Florance work through the translation exercise from the QUIS manual to discuss information structure in Bario Kelabit.

²⁷ Please note that a more detailed description is provided for each recording as part of the ELAR deposit.

BAR20190204CH_05 **Structured Elicitation – QUIS Manual Translation Task 05**
Charlotte & Florance work through the translation exercise from the QUIS manual to discuss information structure in Bario Kelabit.

BAR20190205CH_03 **Semi-Structured Elicitation – Focus Markers 01**
Charlotte & Florance use word cards to explore the position of *tupu* ‘only’ in actor voice and undergoer voice clauses in Bario Kelabit.

BAR20190205CH_03 **Semi-Structured Elicitation – Focus Markers 02**
Charlotte & Florance use word cards to explore the position of *tupu* ‘only’ and *meto* ‘as well’ in actor voice and undergoer voice clauses in Bario Kelabit.

BAR20190206CH_02 **Structured Elicitation – Focus Translation Questionnaire**
Charlotte & Florance work through the translations exercises from the QUIS_{SEM} questionnaire in order to discuss how focus is expressed in Bario Kelabit.

BAR20190215CH_02 **Structured Elicitation – Information Structure 02**
Charlotte & Florance discuss a summary of field notes on information structure in Bario Kelabit.

2. Phonology

BAR30072014CH_01 **Semi-Structured Elicitation – Phonology 01**
Charlotte & Florance record lexical items and example sentences in Bario Kelabit for the purposes of phonological analysis.

BAR30072014CH_02 **Semi-Structured Elicitation – Phonology 02**
Charlotte and Florance record lexical items and example sentences in Bario Kelabit for the purposes of phonological analysis.

3. Morphology

3.1 Verbs & Verbal Morphology

BAR17102013CH_01 Semi-Structured Elicitation - Verbs and Verbal Morphology 01

Charlotte and Florance discuss verbs and verbal morphology in Bario Kelabit. Includes discussion of transitivity and detransitivising morphology and a list of verbs from Haspelmath (1993)

BAR28102013CH_01 Semi-Structured Elicitation – Verbs and Verbal Morphology 02

Charlotte and Florance discuss verbs and verbal morphology in Bario Kelabit. Includes vocabulary on illness and health and discussion of transitivity and detransitivising morphology.

BAR28102013CH_02 Semi-Structured Elicitation – Verbs and Verbal Morphology 03

Charlotte and Florance discuss verbs and verbal morphology in Bario Kelabit. Includes discussion of transitivity and detransitivising morphology.

BAR28102013CH_02 Semi-Structured Elicitation – Verbs and Verbal Morphology 04

Charlotte and Florance discuss verbs and verbal morphology in Bario Kelabit. Includes discussion of transitivity and detransitivising morphology.

3.2 Word Classes

BAR30072014CH_03 Semi-Structured Elicitation - Word Classes 01

Charlotte & Florance discuss words and example sentences for the purpose of morphological analysis in Bario Kelabit.

BAR30072014CH_04 Semi-Structured Elicitation – Word Classes 02

Charlotte & Florance discuss words and example sentences for the purpose of morphological analysis in Bario Kelabit.

4. Syntax

BAR15102013CH_01 **Semi-Structured Elicitation - Locations, Positions and Directions**

Charlotte & Florance elicit sentences in Bario Kelabit. Includes some basic phrases for learners and discussion of locations, positions and directions.

BAR21102013CH_01 **Structured Elicitation - Austronesian Standard Elicitation Schedule 01**

Charlotte & Florance discuss basic clause and noun structure in Bario Kelabit using the Austronesian Standard Elicitation Schedule (Johnston, 1989). Includes discussion of tense, aspect, mood and noun phrase structure.

BAR21102013CH_02 **Structured Elicitation - Austronesian Standard Elicitation Schedule 02**

Charlotte & Florance discuss basic clause and noun structure in Bario Kelabit using the Austronesian Standard Elicitation Schedule (Johnston, 1989). Includes discussion of adverb placement, co-ordination and subordination, numerals and comparison.

4.1 Grammatical Functions

BAR31072014CH_05 **Structured Elicitation - Relative Clauses 01**

Charlotte & Florance discuss relative clauses in Bario Kelabit.

BAR20171103CH_01 **Semi-Structured Elicitation – Position of Time Adverbials**

Charlotte & Florance discuss the position of time adverbials in Bario Kelabit using word cards. This is a test for object status.

BAR20171106CH_01 **Semi-Structured Elicitation – Raising Verbs 01**

Charlotte & Florance discuss raising verbs, such as believe and expect, and how this interacts with voice alternations/grammatical functions in Bario Kelabit.

BAR20171106CH_02 **Semi-Structured Elicitation – Raising Verbs 02**

Charlotte & Florance discuss raising verbs, such as believe and expect, and how this interacts with voice alternations/grammatical functions in Bario Kelabit.

BAR20171106CH_03 **Semi-Structured Elicitation – Raising Verbs 03**

Charlotte & Florance discuss raising verbs, such as believe and expect, and how this interacts with voice alternations/grammatical functions in Bario Kelabit.

BAR20171107CH_01 **Semi-Structured Elicitation – Control Verbs**

Charlotte & Florance discuss control verbs, such as *nuru* ‘order’, and how this interacts with voice alternations/grammatical functions in Bario Kelabit.

BAR20171107CH_02 **Semi-Structured Elicitation – Binding**

Charlotte & Florance discuss binding and how this interacts with voice alternations/grammatical functions in Bario Kelabit.

BAR20171107CH_03 **Semi-Structured Elicitation – Quantifier Float 01**

Charlotte & Florance discuss quantifier float and how this interacts with voice alternations/grammatical functions in Bario Kelabit.

BAR20171107CH_04 **Semi-Structured Elicitation – Quantifier Float 02**

Charlotte & Florance discuss quantifier float and how this interacts with voice alternations/grammatical functions in Bario Kelabit.

BAR20171108CH_01 **Semi-Structured Elicitation – Descriptive Predicates**

Charlotte & Florance discuss descriptive predicates and how they interact with voice alternations/grammatical functions in Bario Kelabit. Sentences are translated from Schultze-Berndt and Himmelmann (2004).

BAR20171120CH_01 **Structured Elicitation – Grammatical Functions**

Charlotte & Florance discuss a summary of field notes on grammatical functions in Bario Kelabit.

BAR20171120CH_02 **Structured Elicitation – Descriptive Predicates**

Charlotte & Florance discuss a summary of field notes on descriptive predicates in Bario Kelabit.

4.2 Pronouns

BAR31072014CH_04 **Structured Elicitation – Pronouns 01**

Charlotte & Florance discuss the pronoun system in Bario Kelabit.

BAR20190205CH_02 **Semi-Structured Elicitation – Case Marking**

Charlotte & Florance discuss differential actor marking in Kelabit undergoer voice clauses using grammaticality judgements on NOM vs GEN forms in different information structure contexts.

4.3 Voice

BAR31072014CH_01 **Structured Elicitation – Voice 01**

Charlotte & Florance discuss the voice system and voice morphology in Bario Kelabit.

BAR31072014CH_02 **Structured Elicitation – Voice 02**

Charlotte & Florance discuss the voice system and voice morphology in Bario Kelabit.

BAR31072014CH_03 **Structured Elicitation – Voice 03**

Charlotte & Florance discuss the voice system and voice morphology in Bario Kelabit.

4.4 Word Order

BAR18082014CH_01 **Semi-Structured Elicitation - Word Order 01**

Charlotte & Florance discuss basic word order in Bario Kelabit. Includes discussion of emotions and basic order with intransitive predicates.

BAR18082014CH_02 **Semi-Structured Elicitation - Word Order 02**

Charlotte & Florance discuss basic word order in Bario Kelabit using word cards. Includes discussion of basic order with actor voice and undergoer voice predicates.

BAR19082014CH_03 **Semi-Structured Elicitation - Word Order 03**

Charlotte & Florance discuss basic word order in Bario Kelabit using word cards. Includes discussion of basic order with actor voice and undergoer voice predicates, as well as focus domains.

5. Word List

BAR14102013CH_01 **Structured Elicitation - The Swadesh List 01**

Charlotte & Florance use the 200 word Swadesh list to elicit lexical items and sentences in Bario Kelabit.

BAR14102013CH_02 **Structured Elicitation - The Swadesh List 02**

Charlotte & Florance use the 200 word Swadesh list to elicit lexical items and sentences in Bario Kelabit.

PDA05112013CH_01 **Structured Elicitation - The Swadesh List 03**

Charlotte & Jeffrey use the 200 word Swadesh list to elicit lexical items and sentences in Pa' Dalih Kelabit.

PDA05112013CH_02 **Structured Elicitation - The Swadesh List 04**

Charlotte & Jeffrey use the 200 word Swadesh list to elicit lexical items and sentences in Pa' Dalih Kelabit.

PDA05112013CH_03 **Structured Elicitation - The Swadesh List 05**

Charlotte & Jeffrey use the 200 word Swadesh list to elicit lexical items and sentences in Pa' Dalih Kelabit.

PDA06112013CH_01 **Structured Elicitation – Numerals 01**

Jeffrey recites the numbers in Pa' Dalih Kelabit.

A.2 Stimuli

1. Prosody Experiment

BAR12082014CH_01 **Reading Task - Pilot Prosody Experiment 01**

Florance Lapu Apu reads an initial set of 26 paragraphs developed by Charlotte and Florance for the purpose of studying prosody in Bario Kelabit.

BAR12082014CH_02 **Reading Task - Pilot Prosody Experiment 02**

Florance Lapu Apu rereads the initial set of 26 paragraphs twice.

BAR12082014CH_03 **Reading Task - Pilot Prosody Experiment 03**

Florance Lapu Apu reads a revised set of 52 paragraphs developed by Charlotte and Florance for the purpose of studying prosody in Bario Kelabit.

BAR18082014CH_03 **Reading Task - Final Prosody Experiment 01**

Connie Aping reads the set of 52 paragraphs developed by Charlotte and Florance for the purpose of studying prosody in Bario Kelabit.

BAR19082014CH_01 **Reading Task - Final Prosody Experiment 02**

Lian Terawe reads the set of 52 paragraphs developed by Charlotte and Florance for the purpose of studying prosody in Bario Kelabit.

BAR19082014CH_02 **Reading Task - Final Prosody Experiment 03**

Florance Lapu Apu reads the set of 52 paragraphs developed by Charlotte and Florance for the purpose of studying prosody in Bario Kelabit.

BAR21082014CH_10 **Reading Task - Final Prosody Experiment 04**

Lucy Bulan reads the set of 52 paragraphs developed by Charlotte and Florance for the purpose of studying prosody in Bario Kelabit.

BAR09092014CH_03 **Reading Task - Final Prosody Experiment 05**

Joe Frazier Kaya reads the set of 52 paragraphs developed by Charlotte and Florance for the purpose of studying prosody in Bario Kelabit.

2. Pear Story²⁸

BAR30102013CH_01 **Video Stimulus - Pear Story 01**

²⁸ The speakers in version 4 and 6 of the Pear Story retelling chose that their recording should not be shared online.

Florance Lapu Apu narrates the pear story whilst watching the video stimulus. This was a pilot study.

BAR30102013CH_02 **Video Stimulus - Pear Story 02**

Florance Lapu Apu retells the pear story after watching the video stimulus. This was a pilot study.

BAR31072014CH_06 **Video Stimulus - Pear Story 03**

Lucy Bulan tells the pear story to Zac Lawai Labang after watching the video stimulus.

BAR01082014CH_02 **Video Stimulus - Pear Story 05**

Dara Tigan tells the pear story to Muda Tigan after watching the video stimulus.

BAR03082014CH_01 **Video Stimulus - Pear Story 07**

Stanley Isaac Ibuh tells the pear story to Florance Lapu Apu and Alaw Tungen after watching the video stimulus.

BAR03082014CH_02 **Video Stimulus - The Meaning of the Pear Story**

Stanley Isaac Ibuh and Alaw Tungen discuss the meaning of the pear story video.

BAR02092014CH_01 **Video Stimulus - Pear Story 08**

Joe Frazier Kaya tells the pear story to Daud Ibuh after watching the video stimulus.

3. Unhappy Rats

BAR20190207CH_02 **Translation Task – Unhappy Rats 01**

Muring Apu completes the unhappy rats translation task.
Subsequently audio recorded by Florance Lapu Apu.

BAR20190214CH_01 **Translation Task – Unhappy Rats 02**

Florance Lapu Apu completes the unhappy rats translation task.

BAR20190215CH_01 **Translation Task – Unhappy Rats 03**

Lucy Bulan completes the unhappy rats translation task.

BAR20190216CH_01 **Translation Task – Unhappy Rats 04**

Supang Terawe completes the unhappy rats translation task.

4. Information Structure

BAR20082014CH_01 **Picture Stimulus - QUIIS Manual Task 3**

Florance completes task 3 from the QUIIS Manual (Skopeteas *et al.*, 2006). This task involves looking at two pictures in a series and narrating the two pictures as though they represented a story: the first picture providing the context for the second.

BAR20190205CH_01 **Picture Stimulus – QUIIS_{SEM} Picture Task 01**

Florance is shown pictures A-E from QUIIS_{SEM} and asked questions with different focus domains to explore how information structure is expressed in Bario Kelabit.

BAR20190206CH_01 **Picture Stimulus – QUIIS_{SEM} Picture Task 02**

Florance is shown pictures A-E from QUIIS_{SEM} and asked questions with different focus domains to explore how information structure is expressed in Bario Kelabit.

BAR20190206CH_03 **Picture Stimulus – QUIIS_{SEM} Picture Task 03**

Florance is shown pictures A-E from QUI_{SEM} and asked questions with different focus domains to explore how information structure is expressed in Bario Kelabit.

BAR20190208CH_01 **Picture Stimulus – QUI_{SEM} Picture Task 04**

Florance is shown pictures A-E from QUI_{SEM} and asked questions with different focus domains to explore how information structure is expressed in Bario Kelabit.

BAR20190208CH_02 **Picture Stimulus – QUI_{SEM} Picture Task 05**

Florance is shown pictures A-E from QUI_{SEM} and asked questions with different focus domains to explore how information structure is expressed in Bario Kelabit.

BAR20190205CH_05 **Picture Stimulus – Tomato Story 01**

Florance narrates the tomato story in QUI manual task 19 from the perspective of an external narrator.

BAR20190205CH_06 **Picture Stimulus – Tomato Story Questions 01**

Florance answers questions on the tomato story in QUI manual task 19 on the basis of her retelling in Tomato Story 01.

BAR20190205CH_07 **Picture Stimulus – Tomato Story 02**

Florance narrates the tomato story in QUI manual task 19 from the perspective of the youngest child.

BAR20190205CH_08 **Picture Stimulus – Tomato Story Questions 02**

Florance answers questions on the tomato story in QUI manual task 19 on the basis of her retelling in Tomato Story 02.

BAR20190205CH_09 **Picture Stimulus – Tomato Story 03**

Florance narrates the tomato story in QUIIS manual task 19 from the perspective of the mother.

BAR20190205CH_10 **Picture Stimulus – Tomato Story Questions 03**

Florance answers questions on the tomato story in QUIIS manual task 19 on the basis of her retelling in Tomato Story 03.

BAR20190206CH_04 **Picture Stimulus - QUIIS Manual Task 12**

Florance completes task 12 from the QUIIS manual, which explores how contrastive topics are expressed.

BAR20190206CH_05 **Picture Stimulus - QUIIS Manual Task 24**

Florance completes task 24 from the QUIIS manual, which explores how partial topics are expressed.

5. Other Stimuli

BAR30102013CH_03 **Picture Stimulus - Topological Relations Picture Series**

Florance completes the Topological Relations Picture Series, describing the location of a series of objects in the visual stimulus materials created by Bowerman and Pederson (1992)

BAR30102013CH_04 **Video Stimulus - Maus Film**

Florance retells the story shown in the Maus Film (Maus-Original-10) stimulus.

BAR30102013CH_05 **Picture Stimulus - The Circle of Dirt**

Florance narrates a story using the picture stimulus set ‘The Circle of Dirt’ developed by (Eisenbeiss *et al.*, 1999)

A.3 Naturalistic Text Topics

1. Culture and Traditions

1.1. Arts & Crafts

BAR20171117CH_02 **Description – Kelabit Dances**

Connie Aping talks about Kelabit dances and her memories of running away as a child in fear of being made to dance.

BAR20190211CH_01 **Procedural Text - Traditional Weaving, Part 1**

Sineh Nu'uh Ulun demonstrates the traditional process of weaving a mat from *kaber* leaves. This was recorded in Bario but is in Pa' Dalih Kelabit (Remudu).

BAR20190211CH_02 **Procedural Text - Traditional Weaving, Part 2**

Sineh Nu'uh Ulun demonstrates the traditional process of weaving a mat from *kaber* leaves. This was recorded in Bario but is in Pa' Dalih Kelabit (Remudu).

BAR20190211CH_03 **Procedural Text - Traditional Weaving, Part 3**

Sineh Nu'uh Ulun demonstrates the traditional process of weaving a mat from *kaber* leaves. This was recorded in Bario but is in Pa' Dalih Kelabit (Remudu).

BAR20190211CH_06 **Procedural Text - Traditional Weaving, Part 4**

Sineh Nu'uh Ulun demonstrates the traditional process of weaving a mat from *kaber* leaves. This was recorded in Bario but is in Pa' Dalih Kelabit (Remudu).

BAR20190211CH_07 **Procedural Text - Traditional Weaving, Part 5**

Sineh Nu'uh Ulun demonstrates the traditional process of weaving a mat from *kaber* leaves. This was recorded in Bario but is in Pa' Dalih Kelabit (Remudu).

BAR20190211CH_08 **Procedural Text - Traditional Weaving, Part 6**

Sineh Nu'uh Ulun demonstrates the traditional process of weaving a mat from *kaber* leaves. This was recorded in Bario but is in Pa' Dalih Kelabit (Remudu). Continued from above.

1.2 Farming

BAR27102013CH_01 **Description - Fish Traps and Scaring Birds**

Gerawat Nulun discusses Kelabit traditions relating to fishing and farming. This includes how to set a *mering* fish trap and several methods of scaring birds away from the paddy in the rice fields. This is in the Bario Kelabit dialect. It is not open access.

BAR27102013CH_02 **Description - How to Pound Rice**

Gerawat Nulun describes how Kelabit women used to pound rice in the past to produce rice that can be cooked (*bera*) from the paddy that is harvested (*padey*). This is in the Bario Kelabit dialect. It is not open access.

PDA06112013CH_04 **Conversation - Bargaining over a Rice Mortar**

Jeffrey Malang and Melamud Tepun bargain over the cost of purchasing a rice mortar (*iung*). This is in the Pa' Dalih dialect of Kelabit.

PDA06112013CH_05 **Conversation - Farming and Omens, Part 1**

Jeffrey Malang and Melamud Tepun discuss the Kelabit farming process and some of the omens observed in the past. This is in the Pa' Dalih dialect of Kelabit.

PDA06112013CH_06 **Conversation - Farming and Omens, Part 2**

Jeffrey Malang and Melamud Tepun continue to discuss the Kelabit farming process and some of the omens observed in the past. This is in the Pa' Dalih dialect of Kelabit.

BAR20171110CH_01 **Personal Narrative - Catching Rats**

Lillian Lipang Bulan talks about childhood adventures going to catch rats in the farm and the different methods they used. This is in the Bario Kelabit dialect.

BAR20171110CH_02 **Personal Narrative - Eating Rats**

Lillian Lipang Bulan talks about how some people used to eat farm rats in the past. This is in the Bario Kelabit dialect.

BAR20171117CH_01 **Description - Hill Farming**

Connie Aping describes how people in Pa' Main used to practise hill farming in the past and talks through the various stages. This is in the Bario Kelabit dialect

BAR20190210CH_02 **Description - Harvesting Rice**

Sineh Ngimet Ayu explains the rice harvest at different times of the year when practising wet rice farming. This is in the Bario Kelabit dialect.

1.3 Food & Feasts

BAR17082014CH_01 **Procedural Text - How to Make Nuba' Laya'**

Florance Lapu Apu explains how to make *nuba' laya'* (lit. soft rice) – the traditional way of preparing and serving rice by boiling, mashing and then wrapping the rice in leaves. This is in the Bario Kelabit dialect.

BAR20171109CH_04 **Description – Kelabit Traditional Weddings**

Lucy Bulan describes Kelabit traditional weddings. This is in the Bario Kelabit dialect.

BAR20171109CH_05 **Description - Marking the Landscape**

Lucy Bulan describes the practice of making a *kawang* or gaps in the mountain range to honour your parents as an important feast. This is in the Bario Kelabit dialect.

BAR20190210CH_04 **Description - The Name Changing Ceremony**

Caleena Lallang Sakai discusses the name changing ceremony and how people travelled from Long Lellang to Pa' Main for ceremonies like that in the past. This is in the Bario Kelabit dialect.

1.4 Hunting, Gathering & Fishing

PDA10112013CH_02 **Personal Narrative - Hunting for Wild Boar**

Garawat Riboh talks about his memories of hunting for wild boar in Pa' Dalih with the former headteacher Bayeh Ribuh. This was recorded in Pa' Dalih but is in the Bario Kelabit dialect.

BAR01082014CH_04 **Conversation - Picking Fruit in Pa Umor**

Dara Tigan, Muda Tigan and Charlotte discuss possible weekend plans, including picking fruit in Pa' Umor. This is in the Bario Kelabit dialect.

BAR21082014CH_04 **Description – Fishing**

Stanley Isaac Ibh describes some traditional methods of fishing among the Kelabit. This is in the Bario Kelabit dialect.

BAR10092014CH_02 **Personal Narrative - Gathering Wild Yams**

Muring Apu talks about her memories of gathering wild yams and other vegetables as a child. This is in the Bario Kelabit dialect.

BAR20171118CH_01 **Description - Collecting Beehives**

Dara Tigan talks about the process of collecting honey from beehives in Pa' Umor. This involves climbing to the top of very tall *dangar* trees. This is in Bario Kelabit.

BAR20190207CH_01 **Conversation – Collecting Honey from Stingless Bees**

Jerome Giak Ipa talks to Florance Lapu Apu about how to collect honey from stingless bees (*anget*). This is in Bario Kelabit.

1.5 Making Salt

BAR10092014CH_01 **Personal Narrative - Making Salt**

Muring Apu talks about her parents making salt in Pa' Main in the past and describes the salt making process. This is in the Bario Kelabit dialect.

PDA06112013CH_10 **Conversation - Making Salt in Pa Dalih**

Jeffrey Malang and Melamud Tepun discuss the process of making salt in Pa' Dalih. This is in the Pa' Dalih dialect of Kelabit.

1.6 Traditional Beliefs

PDA07112013CH_05 **Conversation – The Legend of Pun Tumid**

Jeffrey Malang, Riboh Ayu and Belaan Paren discuss the legend of Pun Tumid, a giant who lives in the jungle and is known to many of the indigenous groups in Northern Sarawak by different names. This is in the Pa' Dalih dialect of Kelabit.

BAR20171119CH_01 **Description – Taboos**

Penghulu Robertson Bala talks about three taboos (*ali*) observed by previous generations of Kelabit before they became Christian. This is in the Bario Kelabit dialect.

1.7 Traditional Games

BAR01082014CH_03 Conversation - Pa Umor in the past

Dara Tigan and Muda Tigan discuss life in Pa' Umor in the past, including childhood games and activities. This is in Bario Kelabit.

BAR08092014CH_05 Conversation - Childhood Memories and Games, Part 1

Dara Tigan and Florance Lapu Apu discuss their memories of growing up in the Kelabit Highlands. They talk about childhood games and stories, comparing Pa' Umor and Ulung Palang. This is in Bario Kelabit.

BAR08092014CH_06 Conversation - Childhood Memories and Games, Part 2

Dara Tigan and Florance Lapu Apu continue to discuss their memories of growing up in the Kelabit Highlands. They talk about childhood games and stories, comparing Pa' Umor and Ulung Palang. This is in Bario Kelabit.

1.8 Traditional Knowledge

BAR20190205CH_11 Description - Medicinal Plants

Jerome Giak Ipa talks about some of the plants that grow in the Kelabit Highlands that have medicinal properties and how they can be cultivated. This is in the Bario Kelabit dialect.

1.9 Travel & Transport

BAR25102013CH_04 Description - Camping in the Jungle

David Labang explains how to set up camp in the jungle. This is in the Bario Kelabit dialect.

BAR17082014CH_10 Conversation - Old Words Relating to Moving between Villages

David Labang and Florance Lapu Apu discuss the meaning of old Kelabit words for traditional customs when walking through the jungle between villages. This is in the Bario Kelabit dialect.

BAR20190210CH_05 **Description - Walking from Long Lellang to Bario**

Caleena Lallang Sakai explains how children walked from Long Lellang to Bario in the past to go to school. This is in Bario Kelabit.

BAR20190213CH_01 **Description - Calls in the Jungle**

David Labang talks to Florance Lapu Apu about Kelabit calls or oral signals in the jungle and other sound words. This is in Bario Kelabit.

1.10 Village Life

BAR22102013CH_01 **Description – Weather in Bario**

Florance Lapu Apu gives a short description of the weather in Bario. This is in the Bario Kelabit dialect.

BAR22102013CH_02 **Description – Living in Bario**

Florance Lapu Apu talks about living in Bario. This is in the Bario Kelabit dialect.

BAR22102013CH_04 **Personal Narrative – Growing up in Pa Umor**

Dara Tigan talks about her memories of growing up in Pa' Umor. This is in the Bario Kelabit dialect.

BAR08092014CH_04 **Description - Scolding Children**

Dara Tigan talks about what parents say in Kelabit to scold their children. This is in the Bario Kelabit dialect.

BAR09092014CH_01 **Description – Bulan’s Homestay, Part 1**

Bulan Ipang invites guests to visit her homestay in Bario and describes activities that visitors can do. This is in Bario Kelabit.

BAR09092014CH_02 **Description – Bulan’s Homestay, Part 2**

Bulan Ipang invites guests to visit her homestay in Bario and describes activities that visitors can do. This is in Bario Kelabit.

BAR20171109CH_01 **Description - Sitting with the Elders**

Florance Lapu Apu talks about her memories of sitting with the village elders (*lun merar*). This is in the Bario Kelabit dialect.

BAR20190210CH_01 **Description – The Longhouse**

Sineh Ngimet Ayu talks about how people made longhouses and why people wanted to live in longhouses in the past. This is in Bario Kelabit.

BAR20190210CH_03 **Description - The Kelabit Custom of Greeting Guests**

Sineh Aren Tuan describes the Kelabit custom of greeting guests and other Kelabit traditions. This is in the Bario Kelabit dialect.

2. Folk Story

2.1 Animal Stories

BAR25102013CH_01 **Folk Story - The Mouse Deer**

David Labang tells a story about how the mouse deer tricks humans and other animals. This is in Bario Kelabit.

BAR25102013CH_02 **Folk Story - The Animal Council 01**

David Labang tells a story about when the animals hold a meeting to decide who is the biggest and strongest in the world. This is in Bario Kelabit.

PDA10112013CH_01 **Folk Story - The Tortoise and The Monkey**

Garawat Riboh tells the story of the tortoise and the monkey that steals her prized musical instrument. The story was told in Pa' Dalih but using the dialect of Bario Kelabit.

BAR17082014CH_04 **Folk Story - The Animal Council 02**

David Labang retells the story of when the animals hold a meeting to decide who is the biggest in the world. This is in Bario Kelabit.

BAR17082014CH_06 **Folk Story - The Animal Council 03**

David Labang retells the story of when the animals hold a meeting to decide who is the strongest in the world. This is in Bario Kelabit.

BAR17082014CH_09 **Folk Story - The Buffalo and The Cow**

David Labang tells the story of how the buffalo and the cow swapped shirts. This is in Bario Kelabit

BAR08092014CH_07 **Folk Story - The Tortoise & The Mouse Deer**

Florance Lapu Apu tells Dara Tigan the story of the tortoise and the mouse deer and how the tortoise escapes from enemies after they are caught stealing fruit. This is in Bario Kelabit.

BAR20171109CH_03 **Folk Story - The Mouse Deer and The Crocodiles**

Lucy Bulan reads a traditional folk story that she translated from Malay. The story describes how the mouse deer tricks the crocodiles into letting him cross the river by pretending to count them as he jumps across. This is in Bario Kelabit.

2.2 Legend

BAR17082014CH_02 Folk Story - Batuh Lawih & Batuh Apui

Florance Lapu Apu tells the story of Batuh Lawih and Batuh Apui. In the story, Batuh Apui is jealous of Batuh Lawih's wife and decides to cut her head off with his machete. He then throws the head down from the mountain where the two stone peaks of Batuh Lawih and his wife can be seen today (at the edge of the Kelabit Highlands). This is in Bario Kelabit.

2.3 Palug Rayeh & Palug I'it

BAR27102013CH_03 Folk Story - Palug Rayeh & Palug I'it 01

Gerawat Nulun tells a folk story about two brothers and how the older brother (Palug Rayeh) is tricked by the younger brother (Palug I'it). In this story, they catch crows. The story is told in Bario Kelabit. The recording is not open access.

BAR17082014CH_03 Folk Story - Palug Rayeh & Palug I'it 02

Florance Lapu Apu tells a folk story about two brothers and how the older brother (Palug Rayeh) is tricked by the younger brother (Palug I'it). In this story, they go fishing and transport rice from the rice hut. The story is told in Bario Kelabit.

BAR17082014CH_07 Folk Story - Palug Rayeh & Palug I'it 03

David Labang tells a folk story about two brothers and how the older brother (Palug Rayeh) is tricked by the younger brother (Palug I'it). In this story, they collect fish in bamboo. The story is told in Bario Kelabit.

BAR17082014CH_08 Folk Story - Palug Rayeh & Palug I'it 04

David Labang continues telling the folk story about two brothers and how the older brother (Palug Rayeh) is tricked by the younger brother (Palug I'it). In this story, they collect fish in bamboo. The story is told in Bario Kelabit.

BAR20171119CH_02 Folk Story - Palug Rayeh & Palug I'it 05

Penghulu Robertson Bala tells a folk story about two brothers and how the older brother (Palug Rayeh) is tricked by the younger brother (Palug I'it). In this story, they are fishing with fish traps and poison. The story is told in Bario Kelabit.

BAR20190211CH_04 Folk Story - Palug Rayeh & Palug I'it 06

Sineh Nu'uh Ulun tells a folk story about two brothers and how the older brother (Palug Rayeh) is tricked by the younger brother (Palug I'it). In this story, they go hunting for wild boar. The story was recorded in Bario but told in the Pa' Dalih dialect (Remudu).

BAR20190211CH_05 Folk Story - Palug Rayeh & Palug I'it 07

Florance Lapu Apu tells Sineh Nu'uh Ulun a folk story about two brothers and how the older brother (Palug Rayeh) is tricked by the younger brother (Palug I'it). In this story, they collect honey. The story is told in Bario Kelabit.

3. Joke

BAR17082014CH_11 Joke – A Practical Joke

David Labang plays a practical joke on Florance Lapu Apu. This is in the Bario Kelabit dialect.

4. Oral History

BAR22102013CH_05 **Personal Narrative - Memories of the Confrontation**

Dara Tigan talks about her memories of the Confrontation in 1960s and the soldiers who were stationed in Bario and Pa' Umor. This is in the Bario Kelabit Dialect

BAR25102013CH_03 **Personal Narrative - Memories of Tom Harrisson and the Z force**

David Labang talks about his memories of the Z Force (a special unit of the Allied forces) parachuting into the Kelabit Highlands during the Second World War. Tom Harrisson, who spent a long time among the Kelabit and later became the director of the Sarawak Museum, was part of this group of soldiers. This is in the Bario Kelabit dialect.

BAR17082014CH_05 **Personal Narrative - When Cats were Parachuted into Bario**

David Labang shares his memories of when cats were parachuted into Bario in the 1950s to solve the problem of too many rats after DDT spray was used by the World Health Organisation to control malaria in Sarawak. This is in the Bario Kelabit dialect.

BAR21082014CH_05 **Personal Narrative - The eBario Project, Part 1**

Lucy Bulan explains how the eBario project got started and brought the internet, telephones and better telecommunications to Bario. This is in the Bario Kelabit Dialect.

BAR21082014CH_06 **Personal Narrative - The eBario Project, Part 2**

Lucy Bulan continues her story of how the eBario project got started and brought the internet, telephones and better telecommunications to Bario. This is in the Bario Kelabit Dialect.

BAR21082014CH_07 **Personal Narrative - The eBario Project, Part 3**

Lucy Bulan continues her story of how the eBario project got started and brought the internet, telephones and better telecommunications to Bario. This is in the Bario Kelabit Dialect.

BAR04092014CH_01 **Personal Narrative - The School in Pa Main, Part 1**

David Lian describes how Tom Harrison helped to set up the first school in the Kelabit Highlands in Pa' Main. He describes his role as one of the first teachers. This is in the Bario Kelabit dialect.

BAR04092014CH_02 **Personal Narrative - The School in Pa Main, Part 2**

David Lian continues his story about how Tom Harrison helped to set up the first school in the Kelabit Highlands in Pa' Main. He describes his role as one of the first teachers. This is in the Bario Kelabit dialect.

BAR20171118CH_02 **Personal Narrative - The Bario Revival**

Dara Tigan talks about her memories of the Bario Revival in 1973. This is in the Bario Kelabit dialect.

5. Radio Bario Broadcast

BAR29112013CH_01 **Radio Broadcast - News Report 2013-11-29**

Connie Aping presents the news on Radio Bario, which she had prepared and translated from the Borneo Post edition of 29th November 2013. This is in the Bario Kelabit dialect.

BAR29112013CH_02 **Radio Broadcast - Introductions 2013-11-29**

Lucie Palan Trang presents the introduction segment of her Radio Bario Broadcast with Connie Aping. This is in Bario Kelabit.

BAR29112013CH_03 **Radio Broadcast - Flight Announcement 2013-11-29**

Lucie Palan Trang presents the flight announcements as part of her Radio Bario Broadcast with Connie Aping. This is in Bario Kelabit.

BAR21082014CH_01 **Radio Broadcast - News Report 2014-08-21 Part 1**

Stanley Isaac Ibhuh presents the news on Radio Bario, having prepared and translated from the Borneo Post edition of the 20th August 2014. This is in the Bario Kelabit dialect.

BAR21082014CH_02 **Radio Broadcast - News Report 2014-08-21 Part 2**

Stanley Isaac Ibhuh continues with the news report on Radio Bario that he prepared and translated from the Borneo Post edition of 20th August 2014. This is in the Bario Kelabit dialect.

BAR21082014CH_03 **Radio Broadcast - Flight Announcement 2014-08-21**

Stanley Isaac Ibhuh announces the flight schedule via Radio Bario. This is in the Bario Kelabit dialect.

BAR02092014CH_03 **Radio Broadcast - News Report 2014-09-02**

Connie Aping presents the news on Radio Bario, which she had prepared and translated from the Borneo Post edition of 2nd September 2014. This is in the Bario Kelabit dialect.

6. Song

- PUM18102013CH_03 **Song - Unih Lega Lem Tegu'ang**
A dancing song by Sineh Pasang, Sineh Maren Lugun and Sineh Tingang
- PUM18102013CH_04 **Song – Titih**
A song used to welcome and honour important visitors to the Kelabit Highlands by Sineh Pasang, Sineh Maren Lugun and Sineh Tingang
- PUM18102013CH_05 **Song – Nuba' Umur-Umur**
A love song sung for entertainment by Sineh Pasang, Sineh Maren Lugun and Sineh Tingang
- PUM18102013CH_06 **Song – Sia'-Sia' Iko Langit**
A children's rhyme by Sineh Pasang, Sineh Maren Lugun and Sineh Tingang
- PUM18102013CH_07 **Song - Uih Emu' Kiung**
A children's rhyme by Sineh Pasang, Sineh Maren Lugun and Sineh Tingang. This may originate from the Indonesian side of the border
- PUM18102013CH_08 **Song – A'it-A'it Ko Idang**
A children's rhyme by Sineh Pasang, Sineh Maren Lugun and Sineh Tingang
- PUM18102013CH_09 **Song – Iu'-Iu' Alung-Alung**

A children's rhyme by Sineh Pasang, Sineh Maren Lugun and Sineh Tingang. Children clap along to the rhythm.

PUM18102013CH_10 **Song - Ruyuh-Ruyuh Batang Ubung**

A children's rhyme by Sineh Pasang, Sineh Maren Lugun and Sineh Tingang

PUM18102013CH_11 **Song – Siki**

Two *siki* or dream songs by Sineh Pasang, Sineh Maren Lugun and Sineh Tingang. The first is *la'al iek* and the second *labo lati*. These songs involve a lead and a chorus and encourage love for nature among children.

PUM18102013CH_12- **Song - Ulin Legku' Ai'-Ai'**

01

An *ulin* or love song by Sineh Pasang, Sineh Maren Lugun and Sineh Tingang. These would have been sung at night by the fire or in big groups whilst harvesting rice.

PUM18102013CH_12- **Song - Ulin Legku' Teno-Teno**

02

An *ulin* or love song by Sineh Pasang, Sineh Maren Lugun and Sineh Tingang. These would have been sung at night by the fire or in big groups whilst harvesting rice.

PUM18102013CH_12- **Song - Ulin Reteb Padan Mayung**

03

An *ulin* or love song by Sineh Pasang, Sineh Maren Lugun and Sineh Tingang. These would have been sung at night by the fire or in big groups whilst harvesting rice.

PUM18102013CH_12- **Song - Ulin Reteb Padan Balang**

04

An *ulin* or love song by Sineh Pasang, Sineh Maren Lugun and Sineh Tingang. These would have been sung at night by the fire or in big groups whilst harvesting rice.

- PUM18102013CH_12-05 **Song - Ulin Lyrics**
Sineh Tingang recites the lyrics to each of the four *ulin* songs in turn.
- PUM18102013CH_13 **Song - Ri Lekuah Bawang Elung**
A pair of *ri lekuah* songs by Sineh Pasang, Sineh Maren Lugun and Sineh Tingang. The first is Bawang Elung to celebrate an important woman. The second is Bawang Anid to celebrate an important man.
- PUM18102013CH_14 **Song - Ri Lekuah Lyrics**
Sineh Maren Lugun recites the lyrics to the two *ri lekuah* songs above.
- PUM18102013CH_15 **Song - Lakuh Anak Lun Pa Umor**
Sineh Maren Lugun sings a *laku*h song that she composed herself. *Laku*h songs normally tell of real events or the history of an individual.
- PUM18102013CH_16 **Song - Lakuh Tenga'an Suruh Batang**
Sineh Tingang sings the *laku*h of Tenga'an Suruh Batang. It tells of two ladies who fought over the same man.
- PUM18102013CH_17 **Conversation - The Lakuh of Dayang Alud, Part 1**
Sineh Pasang, Sineh Maren Lugun, Sineh Tingang and Lucy Bulan discuss the meaning of the *Laku*h Tenga'an Suruh Batang and another *laku*h composed by Dayang Alud, telling her life story.
- PUM18102013CH_18 **Conversation - The Lakuh of Dayang Alud, Part 2**
Sineh Pasang, Sineh Maren Lugun, Sineh Tingang and Lucy Bulan continue to discuss the meaning of the *Laku*h Tenga'an Suruh

Batang and another lakuh composed by Dayang Alud, telling her life story.

PUM18102013CH_19 **Song - Lakuh Dayang Alud, Part 1**

Sineh Maren Lugun recites the lyrics to the first part of the lakuh of Dayang Alud.

PUM18102013CH_20 **Song - Lakuh Dayang Alud, Part 2**

Sineh Maren Lugun recites the lyrics to the second part of the lakuh of Dayang Alud.

PUM18102013CH_21 **Song - Lakuh Sineh Maren Lugun**

Sineh Maren Lugun sings her own lakuh that tells of her life story and her family.

PUM18102013CH_22 **Song - Kuab Rebu'-Rebu' Dedtem Kinih**

Sineh Pasang, Sineh Maren Lugun and Sineh Tingang sing a pair of kuab songs. The first is *rebu'-rebu' dedtem kinih* and the second is *rebu'-rebu' uih kinanak*. Kuab songs were composed during the headhunting era and sung to welcome men home from headhunting trips.

PUM18102013CH_23 **Song - Kuab Lun Pa Umor**

The first in a pair of kuab song by Sineh Pasang, Sineh Maren Lugun and Sineh Tingang. Kuab songs were composed during the headhunting era and sung to welcome men home from headhunting trips.

PUM18102013CH_24 **Song - Kuab Lun Peniteb**

The second in a pair of kuab song by Sineh Pasang, Sineh Maren Lugun and Sineh Tingang. Kuab songs were composed during the

headhunting era and sung to welcome men home from headhunting trips.

PUM18102013CH_25 **Song - Siki Elet Kiret**

Sineh Pasang, Sineh Maren Lugun and Sineh Tingang sing a pair of siki songs, Elet Kiret and Elet Bayung. The song is about a girl named Dayang who sadly died in a flood. Legend has it that the song was sung by her ghost.

PUM18102013CH_26 **Song – Tutu’ Udan Nepera**

A celebration song by Sineh Pasang, Sineh Maren Lugun and Sineh Tingang. The song has verses and repeated refrain. The lyrics of the verses are often composed on the spot to welcome new guests.

BAR04092014CH_04 **Song – Mourning Song**

David Lian sings a *sido* or mourning song, composed by a husband after his wife passed away to express his grief.

BAR20190210CH_06 **Song - Nursery Rhymes**

Caleena Lallang Sakai sings several children’s rhymes, including *iu’-iu’ alung-alung*.

7. Speeches

BAR08092014CH_03 **Formal Speech – Passing Exams**

Dara Tigan repeats a version of a motivational speech that she had delivered to the school children at SK Bario a few days before to encourage them to do well in their exams.

8. The Kelabit Language

BAR21082014CH_08 **Description - Rurum Kelabit Sarawak Education Unit, Part 1**

Lucy Bulan describes how the Education Unit of the Kelabit Association (Rurum Kelabit Sarawak) help to promote and preserve the Kelabit language. This is in the Bario Kelabit dialect.

BAR21082014CH_09 **Description - Rurum Kelabit Sarawak Education Unit, Part 2**

Lucy Bulan continues describing how the Education Unit of the Kelabit Association (Rurum Kelabit Sarawak) help to promote and preserve the Kelabit language. This is in the Bario Kelabit dialect.

BAR20171109CH_02 **Description - Revitalising the Kelabit Language**

Lian Raben talks about his involvement with Kelabit language revitalisation. This is in the Bario Kelabit dialect.