Variation and change in Abui : the impact of Alor Malay on an indigenous language of Indonesia

Saad, G.M.

Chapter 3

Methodology

3.1 Introduction

This thesis adopts a mixed-methods approach, integrating a number of different methods in order to understand the processes underlying language variation and change in Abui. The aim of this chapter is to describe and motivate these various methods.

In order to investigate the mechanisms of change, a variationist study of a given linguistic feature must be accompanied by a thorough description of the speech community (Ross, 2013). This resulted in two broad categories of data: 1) sociolinguistic and ethnographic interview data and 2) linguistic data. Sociolinguistic and ethnographic interviews were conducted in order to comprehend the history and current structure of the speech community. More specifically, sociolinguistic survey data was collected from a large number of individuals in order to explore specific variables relevant for explaining linguistic variation. Ethnographic interviews were conducted with a smaller group of speakers to hone into two topics that became relevant for the study of this specific speech community: a) emic notions of life-stages (introduced in §2.4.2 and elaborated on in §3.4) and b) the history of language use at schools (discussed in §2.3). Taken together, data from these interviews was used to construct four age-cohorts based on emic notions of age categories. Linguistic data was collected for each of these age-cohorts and then compared to one another. Three types of linguistic data were col-

lected: conversational data, experimental data, and data from fieldnotes.

Conversational data consisted of various kinds of conversations. Experimental data was collected in the form of both production data and comprehension data. The production data was gathered from a video elicitation task while the comprehension data was gathered from a forced-choice task. The data from fieldnotes includes elicited data and data from participant observation. These latter types of data were important in forming and testing hypotheses in addition to observing speaker's judgments and attitudes towards variation. The diverse types of data are summarized in Table 3.1.

Table 3.1: Breakdown of various types of data

Interview data	Sociolinguistic	Sociolinguistic survey
	Ethnographic	Focus group discussions on age-groups School behaviour interviews
Linguistic data	Conversational	Conversations
	Experimental	Surrey Stimuli production task Forced-choice comprehension task
	Fieldnotes	Elicitation Participant observation
	Other	Put Stimuli, narrative, word list

This chapter is organized as follows: The details of how fieldwork was undertaken are discussed in §3.2. Sociolinguistic and ethnographic interview data are discussed in §3.3. A description of how the speakers were sampled, is presented in §3.4, which is both a result of the interview data and a method of sampling for the linguistic data which is discussed in §3.5. The details of the corpus on the whole are illustrated in §3.6. Section 3.7 describes the Abui corpus assembled by Kratochvíl, which was also consulted for this thesis. Section 3.8 discusses how the various types of data were analyzed and synthesized for the various chapters of this thesis.

3.2 Fieldwork

Fieldwork for this thesis was conducted in the village of Takalelang, *desa* 'vilalge cluster' Lembur Barat, *kecamatan* 'sub-district', Alor Tengah Utara,

kebupaten 'regency' Alor, in the Nusa Tenggara Timur province of Indonesia (see Figures 1.3 and 1.4 in Chapter 1 for maps of Takalelang). In Takalelang, I was introduced into the speech community by Benny Delpada and Simon Lanma who I had met in Singapore and Kupang, respectively, through the help of the linguist, František Kratochvíl. Once in the community, several members introduced me to their friends and family members.

The village of Takalelang was selected as the setting for investigation for two main reasons. First, most of the prior (descriptive and historical linguistic) research done on Abui was done in Takalelang on the Takalelang dialect, by František Kratochvíl. This meant that there was a very steady foundation in place, both in terms of the available published materials and corpora of the language as well as the local community's reception towards foreign researchers.

Secondly, the village of Takalelang lies on the coast, adjacent to Alor's main road. This means that it is an ideal location for the study of ongoing language shift. This is because many community members frequently visit the market in Kalabahi, the regency capital, and thus often come into contact with other ethnic groups with whom they use Alor Malay as a *lingua franca* (see §1.4.2). In addition, there are two primary schools stationed on the western and eastern border of the village, meaning that sociolinguistic work investigating the impact of schools on the local community could also be conducted in a feasible manner.

Three fieldtrips were undertaken for this research in the years 2015, 2016, and 2017. Each fieldtrip lasted for a duration of two to two and half months. The languages used for daily communication included Alor Malay, Indonesian, and Abui. All interviews were always conducted in Alor Malay and Indonesian, as this allowed me to transcribe them myself and analyze them in a more swift manner.

Having studied Standard Indonesian before my first fieldtrip, my first few weeks involved me speaking the standard variety as much as possible. As time went on and I developed more exposure to Alor Malay and to the sociolinguistic situation, I began learning Alor Malay and using it as much as possible during interviews and day-to-day interaction. Using Alor Malay, as opposed to Standard Indonesian, decreased the distance between myself and the speakers, who often associate Standard Indonesian with government and other formal affairs.

During the collection of linguistic data, I typically used Alor Malay to

carry out instructions. Especially in the early stages of fieldwork, my ability to explain instructions clearly was only possible in Alor Malay. Furthermore, it was unclear to what extent younger speakers would fully understand instructions in Abui. This was evidenced by the fact that some of assistants in my research team deliberately chose to address them in Alor Malay when giving them instructions. As such, Alor Malay seemed like the most appropriate option.

In my second and third year of fieldwork, my Abui improved significantly. It allowed me to hold slightly above basic conversations and be able to conduct simple interviews and elicitation sessions. It also enabled me to partake in common etiquette practices within the village.

3.2.1 Research team and local language experts

The work for this project is the product of collaboration from a large research team in Takalelang. The team consisted of at least seven members from the Abui community, Simon Lanma, Benny Delpada, Ansellmus Delpada, Frengki Delpada, Lorens Malbiyeti, Vicary Maufani, and Dori Lanma. All of these individuals, in addition to many others, expressed their willingness to offer their assistance to any tasks involved in this PhD project, including actively teaching me their language.

Ansell Delpada, Simon Lanma, Benny Delpada, Frengki Delpada, Lorens Malbiyeti were responsible for transcribing the data in ELAN and then translating it. Wherever possible, they also added additional notes if they considered the speech to be ungrammatical or infelicitous. Vicary Maufani with her status as a school teacher offered her support in recruiting younger participants for the Surrey Stimuli task. Dori Lanma was of great assistance on a daily basis with transcriptions, translations, and basic vocabulary. She was also an invaluable language coach when necessary, constantly practising Abui and teaching me new expressions. In addition, she too assisted with transcriptions and judgements on other people's speech (see §3.5.3.1).

In line with consultation with several local community members and researchers active in the area, local language experts were reimbursed in a number of different ways. Food and board were reimbursed on a weekly basis or purchased from the local market. Typically, experts transcribing and translating a recording or assisting with interviews were either reimbursed at a rate similar to that of a local school teacher or were offered gifts such as

betel nut and/or cigarettes.

In addition, several of these individuals were actively involved in carrying out their own language documentation and description projects both on an academic and a societal level. This ensured a fruitful relationship that benefited from expertise from both the researcher and members of the local community. For example, for the ethnographic interviews conducted with elders concerning life-stages, we were able to align the goals of this PhD project with the goals of several community members.

The questions were put together with several Abui members with aims that were communally decided upon such as establishing particular life-stages, understanding language behaviour among them, and shedding light on how some of this behaviour can be reversed. These interviews were always conducted by an Abui speaker, Simon Lanma, who used the interviews as an opportunity to ask the interlocutors whether they had specific requests on what could be done in terms of local language teaching, documentation, and revitalization. In addition, many of these recordings also documented social issues in the community and allowed interlocutors to express their views on certain injustices with respect to local government which they felt they were being subjugated to. Simon used this field training and these recordings for his BA thesis which, at the time of writing, had just been completed. These recordings are also being used as proof for efforts by Simon and others to introduce Abui into the local school systems.

3.2.2 Equipment

In the field, a Canon XA-20 video camera was used for the video recordings, along with two shotgun Sennheiser XLR K6P and ME 64 microphones. This set-up was ideal for recording interviews, conversations, and elicitation tasks. Additionally, an Olympus LS-12 audio recorder was used where video was deemed unfit or impractical. Finally, a Panasonic Toughbook CF-53 was used to process the data in the field. The software used for the analysis of the data is discussed in §3.8.

3.3 Sociolinguistic and ethnographic interview data

The interview data comprise two types: sociolinguistic interviews and ethnographic interviews. Dubois and Horvath (1999) discuss the importance

of combining both sociolinguistic methods, following the Labovian tradition, e.g. Labov (1963, 1990) and ethnographic methods, following Eckert and McConnell-Ginet (1992). Instead of being mutually exclusive, Dubois and Horvath (1999) argue that the two are complementary - both offering different levels of granularity to the study of language change in progress. The sociolinguistic survey methodology offers a bird's eye perspective on the data. Focusing on collecting data from a representative sample, it allows for a comparison between a group in question and the broader speech community. In addition, it allows for comparisons across different speech communities (Labov, 1990; Dubois & Horvath, 1999). Where the sociolinguistic approach offers a survey of the community, the ethnographic approach offers an in-depth understanding of the social structure of a particular group in question (Dubois & Horvath, 1999). It allows for a detailed study of locally-defined groups and their relationship to one another. It treats variables such as 'age' and 'gender' as socially constructed, with different manifestations in different places at different times (Eckert and McConnell-Ginet, 1992, 1999; Eckert, 2017).

In the research for this thesis, the sociolinguistic interview was aimed at gathering information from an individual speaker about their own basic information and language use. Such data was used to establish variables, such as 'language use during primary school', for instance. This allowed for quantitative comparisons whose implications are discussed in Chapters, 5, 6, 7. The ethnographic interview, on the other hand, was aimed at eliciting a few individual's views of the makeup of the speech community. The questions were less targeted at the individual's life history but at the history of the community as a whole as well as at its current makeup. Together, both these types of data established four well informed age-cohorts, discussed in §3.4. Sociolinguistic interviews are discussed in §3.3.1 while ethnographic interviews are discussed in §3.3.2.

The interviews were always conducted in a mix of Alor Malay and Indonesian for ease of processing. There were three reasons why the interviews were conducted in Malay Indonesian as opposed to Abui. First, the my ability to conduct interviews in Malay Indonesian was significantly higher than my ability in Abui. Second, in the cases, when the interviews were carried out by an Abui speaker, having them in Malay Indonesian meant that I could follow most of the interview and could intervene if necessary. Third, the interviews would be much less time-consuming in terms of tran-

scription, translation, and coding. All of the interviewees were fluent in Alor Malay, literate, and had some level of schooling. On the rare occasion that an older speaker would fail to fully comprehend the question in Malay Indonesian, the interviewer would switch to Abui briefly.

3.3.1 Sociolinguistic interviews

The aim of the sociolinguistic interviews was to collect data that could *a priori* be relevant to explain the variation found in the speech community. Data from these interviews was intended to be used alongside data from the experimental studies such that comparisons could be made across age-groups.

The sociolinguistic interviews were based on a questionnaire which was assembled in 2015 before my first field trip, in collaboration with Marian Klamer, Hanna Fricke, and Yunus Sulistyono. The questions were initially drafted in English and then translated into Standard Indonesian. During the first field trip, the questions were read out in Standard Indonesian. This later appeared to create a formal atmosphere; in later field trips, it occurred to me that it would be more appropriate to adjust the questions to Alor Malay, which I felt many speakers would feel more at ease with. This is evident in some of the questions added at a later stage. Often a combination of the two was used within the interview, which is commonplace in Takalelang. As such, on the whole the language of the interviews can be characterized as a mix of Standard Indonesian and Alor Malay. [1]

Much of the data emanating from these interviews may be considered 'self-reports': with speakers giving subjective opinions on language use, language attitudes, and proficiency. The core of the questionnaire included questions about the speaker's name, the names of their immediate family members, their age, gender, migration history, marital status, educational status, language use, language attitudes, and proficiency. Every year the questionnaire was adapted and new questions were added based on new observations and hypotheses that were not available before the first field trip. Additional questions added later included questions such as, "What language did your parents speak to you before primary school" and "Have you every lived with your grandparents?". Table 3.2 presents a small sample

[1] In this thesis, I often refer to this dialect mixing as Malay Indonesian; see §1.4.2 for a discussion of the term Malay Indonesian.

Table 3.2: Sample of sociolinguistic questionnaire

Question (English)	**Question (Malay Indonesian)**
Name	*Nama*
Date of birth	*Tanggal lahir*
Place of birth	*Tempat lahir*
Place of residence	*Tempat tinggal*
Who do you currently live with?	*Sekarang ni, kakak/bapak/mama tinggal/tidur di satu ruma dengan siapa*?
Who are your siblings?	*Kakak/bapak/mama saudara dong tu siapa*?
What language was used to you before primary school?	*Sebelum masuk SD tu, bahasa apa yang orang tua dong pake ko omong dengan kakak/bapak/mama*?
What language was used to you during primary school?	*Waktu masi SD tu, bahasa apa yang orang tua dong pake ko omong dengan kakak/bapak/mama*?
Have you ever lived with your grandparents?	*Perna tinggal dengan nenek dong*?
At what age would you say you felt comfortable speaking Abui?	*Kakak/Bapak/Mama pung umur berapa dulu baru rasa nyaman omong pake bahasa Abui*?

of the questionnaire. For the full questionnaire, readers are referred to Appendix I.

Sociolinguistic data was collected for 66 participants, who were split into four age-cohorts. This is the same number of speakers who took part in the production task (see §3.5.2.1). Of these, 33 were recorded in the form of an interview, while 33 were filled in to the questionnaire manually at a later stage, without producing a recording. The reason not all sessions could be recorded was due to there being some constraints on recording an interview, such as time and the absence of a parent figure in the case of younger speakers. Despite this, relevant sociolinguistic information was recorded for all 66 participants and filled into the excel file containing the questionnaire. The majority of the answers to the questions and variables were not investigated

quantitatively ; however, they were useful in describing the sociolinguistic setting in Chapter 2. At the same time, many of these questions offered a window into factors that might be relevant for future enquiry. They pointed to relevant areas for further investigation with respect to school attitudes and the relevance of marriage in language use - topics later dealt with in the ethnographic interviews in §3.3.2.

3.3.2 Ethnographic interviews

The ethnographic interviews targeted several individuals to discuss the speech community as a whole, instead of giving detailed accounts of their own life history.

Two types of interviews with an ethnographic approach were conducted. The first targeted emic notions of life-stages and age categories (§3.3.2.1). This was done in order to have a better understanding of age as a variable and to split the population into age-cohorts in order to study language variation, following Eckert (2017). The outcome of these interviews is a breakdown of the speech community into age-cohorts and described accordingly in §3.4.

The second type of ethnographic interviews (§3.3.2.2) targeted specific events that led to a shift in schooling and parenting strategies, accounting for the widespread use of Alor Malay in the speech community. The outcome of these interviews is a description of the speech community and its history in Chapter 2. These interviews were conducted in the second and third year respectively, when a sketch of the sociolinguistic setting had been put together through the sociolinguistic survey data.

3.3.2.1 Focus group discussions on life-stages

In order to discuss important life-stages in the Abui community and establish how the population could be split up accordingly, I conducted five focus group discussions which were semi-structured in nature and based on a list of questions already agreed upon beforehand. The goal of these interviews was to adopt an emic approach and gather Abui-specific information on age-related terminology. This included specific events related to age categories, and age-ranges within these categories. This information was synthesized with information from interviews on school history to establish

four age-cohorts, discussed in detail in §3.4.

There were two main motivations for carrying out these interviews in this way. Firstly, in order to robustly test the relationship between age and language change, age must be split into age-cohorts (Eckert, 2017). Because the span of ages is so great, using age as a continuous variable makes it otherwise difficult to achieve statistical significance with fine grained age differentiation. Secondly, an emic approach is often preferred to an etic approach. While an etic approach splits speakers up into pre-defined, arbitrarily determined and equal age spans, an emic approach splits speakers according to a shared experience of time which may be linked to life-stage or history (Eckert, 2017; see also §2.4.2.1).

In preparation for these discussions, a set of questions was drafted in conjunction with two Abui speakers aged 25 and 32, respectively, namely Simon Lanma and Benny Delpada (see Appendix II for full list of questions). We discussed at length what questions would be both relevant and culturally appropriate. In total, five focus group discussions were conducted, interviewing nine speakers. Each interview included either one or two elders, one or two Abui interviewers, and myself. The discussions were mostly led by Simon Lanma, who had often featured in the media and was experienced at interviewing people. Each interview lasted approximately one hour. By having an Abui speaker carry out the interview, there was a higher chance that the responses would be covered in a more comprehensive manner. The Abui interviewer felt at ease clarifying questions that did not immediately elicit the desired response, and the speakers interviewed felt at ease discussing matters in detail.

The interviews had the following structure. First, Simon and I both explained the general goals of our research and the specific goals of the interview. These were outlined as relating to how Abui children are socialized in the community, what language they speak at school, and how today's generation differs from yesterday's generation, as well as how Abui society is composed of various age-groups. Second, Simon asked the speaker(s) to introduce themselves, stating where they grew up, how old they were, and how many children they have. Afterwards, he asked the speakers to discuss what came to mind, in their view, when they thought of a specific age-related term. He started with terms like *moqu kieng* 'infant' and *moqu* 'child'. This went on for the age-groups in question, which included *neeng abet*/ *maayol maak* 'young man/woman' and *kalieta* 'adult'. Speakers were asked to elab-

orate on specific life-stages associated with these events in addition to estimating an age-range for each term, if applicable. He then moved on to ask questions about language use and bride price negotiations. Finally, many speakers also provided information containing comparisons between the situation today and that of their own childhood. Figure 3.1 illustrates the set-up used for one of these interviews.

Figure 3.1: Ethnographic interview on age-groups

The interviews were transcribed and translated using ELAN. For analysis, an Excel sheet with all the common questions was compiled; subsequently, data from all nine participants was added to this questionnaire.

3.3.2.2 Interviews about schooling and language policy

As discussed in Chapter 2, the Abui Takalelang speech community has seen drastic changes in the last 50-60 years. The aim of the interviews discussed in this section was to gather more information on the structural changes underlying a shift in language use and behaviour at school. Since schools played the largest role in promoting Malay Indonesian at the expense of

Abui, interviews were conducted with current and retired school teachers, principals, and supervisors. In some of sociolinguistic surveys, several speakers pinpointed specific teachers who were stark proponents for banning Abui at schools and promoting the use of Malay Indonesian. Some of these individuals were then approached and interviewed.

Specifically, I conducted five interviews with five Abui members associated with schools. The interviews were semi-structured in nature, typically lasting between one and two hours. Unlike the focus group discussions on life-stages, it was not possible to recruit the services of Abui speakers to conduct the interviews; therefore, the interviews were conducted by myself. In these interviews, the interviewees recounted the history of schooling in Alor more generally and Takalelang more specifically. They all gave precise dates about when certain schools moved or opened. They also offered precise information on specific years where some classes spoke better Abui then others, allowing for detailed descriptions of the shift in language use and attitudes.

In addition, they gave in-depth accounts about how they and their colleagues implemented an order that banned the use of Abui in the classroom and in the school playground. They offered their motivations for doing so as well as offering precise information about how various generations responded to this order over the years. This information was synthesized to produce a brief history of schooling in the Takalelang area (presented in §2.3).

The interviews were all audio-recorded. However, since the goal of these interviews was to provide a small sketch of the history of language use in schools, only some fragments were transcribed and translated into English using ELAN. In addition, notes taken during the interviews were also consulted. Appendix III provides a sample question list used during one of these interviews.

3.4 Sampling the speech community

The outcome of the ethnographic interviews led to the formation of four age-cohorts of speakers in the Takalelang speech community, presented in Table 3.3. These age-cohorts are delineated by age and are inspired by Abui age terms denoting certain life-stages and history, which were discussed in §2.4.2.1.

Table 3.3: Age-related terms

Abui term	**English gloss**	**Life-stage; history**	**Age-range (years)**
Moqu	'(Pre)adolescents'	Birth until end of puberty	9-16
Neeng abet / maayol maak	'Young adults'	End of puberty until marriage	17-25
Kalieta	'Adults'	Marriage; grew up speaking Abui and Alor Malay	26-34
Kalieta	'Elders'	Marriage; grew up speaking Abui only	40-75

These age-ranges represent approximations based on numbers given during the ethnographic interviews. They are meant to revolve around people's general notions about certain life-stages, such as the end of puberty marking the start of 'young adulthood' and marriage marking the start of 'adulthood'.

Of course, for any given speaker, it is possible that their age does not correspond to the life-stage described in Table 3.3. For example, the table assumes that any given speaker aged between 26-34 must be married. While this was mostly true, it was not always true. In a situation where there was a possible mismatch between age and life-stage, age took precedence over life-stage in determining group membership. Therefore, if a speaker aged between 26-34 was unmarried, they were still placed in the group of *kalieta* 'adults' (aged 26-34). This was motivated by the fact that it was not always possible to determine whether a particular speaker conformed to a given life-stage, whereas it was always possible to determine their age. In other words, while gathering information on some life-stages, such as marriage was possible, understanding whether someone had reached the end of puberty and was diligent enough to be considered a viable mate (another common trait of 'young adults') posed clear ethical as well as practical limitations. Thus, it was impossible to place a speaker into the group of *moqu*

'(pre)adolescents' or *neeng abet/maayol maak* based on life-stage criteria alone. As such, age was selected as a determining factor, while life-stages are used to characterize the groups in a broad sense.

After I had established these age categories, I then used these cohorts to organize the speaker population. For the variationist studies in Chapters 5, 6, and 7, I made sure that the cohorts had a representative amount of speakers fitting into a particular age-category. I recruited appropriate speakers during community gathering such as Church, sports matches, and by simply taking walks in the village and asking people whether they wanted to be recorded. Most speakers were recruited in the village of Takalelang, although some were recorded in the city of Kupang, Timor, where a large diaspora of Takalelang Abui speakers reside. In Takalelang, several members introduced me to their friends and family members, who then took part in the study.

All four age-cohorts are discussed in detail in §§3.4.1-3.4.4. In these sections, four main types of information is presented. First, information is given about the life-stages of each cohort. This is followed by their respective age boundaries. Furthermore, self-report data on language exposure from the sociolinguistic interviews is also presented. This includes i) language used by parents to speaker before primary school, ii) language used by parents to speaker during primary school, and iii) language used with peers in primary school playground. Variables i) and ii) provide insight into parenting strategies. Comparing variable i) across the four groups depicts how parents have altered their language family policy in response to pressure from government and school officials. Comparing variables i) and ii) also shows within the individual's timespan whether their parents have changed their language family policy as the child has settled into primary school. The reason this was selected was that parents often claimed that they were worried that if they wouldn't speak Alor Malay to their children before primary school that their children would risk suffering at school due to abuse from teachers as well as the inability to keep up. They also claimed that once children had safely learned Alor Malay and were doing well at school that they were willing to start using more Abui with them. Finally, variable iii) shed light into what language children were speaking among themselves in the playground. Because the playground was a highly stigmatized setting for play, it gives a conservative account of how often children were speaking Abui to another. Finally, general tendencies about their current linguistic

behaviour are also presented.

3.4.1 *Moqu* '(pre)adolescents'

The '(pre)adolescent' group (9-16 years) is derived on the basis of the Abui *moqu*, a term used to denote children and teenagers. Sexual maturity marks the end of *moqu*hood. In the past, a child would be given a loincloth to ritually mark the passage from *moqu* to sexual maturity (Du Bois, 1944). Today, no such ritual exists anymore, but the concept of *moqu* is still widely used to refer to someone who is not yet fully sexually mature.

Moqu is typically used more broadly in everyday life[2] than it is in this thesis as it typically covers the period from birth up until adolescence. For this thesis, the age of nine was selected as the lower boundary of *moqu*, because running tasks to elicit Abui with children below nine proved to be difficult, as they did not feel comfortable speaking Abui nor conducting the task. Therefore, the lower boundary of nine was selected for practical and experimental reasons; however, the upper boundary of sixteen was selected because it characterizes the end of puberty. It is thus representative of the category itself, as evidenced from excerpt (1).

(1) Q: *Menurut Bayi, kita biasa bilang moqu itu, apa*?
'According to grandpa, what does it mean when we utter the word *moqu*?'
A: *Moqu itu... Anak sejak lahir sampe dengan umur 16, itu dianggap moqu.*
'What we consider a *moqu* is any child from birth up until the age of 16.' [EG.65M.AG]

A socially constructed characteristic feature includes 'knowing how to

[2]The term might be combined with other modifiers to specify more specific periods. For example, *moqu kieng* refers to the first stage of *moqu* and might be loosely translated as 'newborn'. Speakers describe this phase as being the one where the newborn can not be separated from its caregiver. Speakers also refer to this stage using the Malay Indonesian term *anak merah* 'red child'. The next stage of *moqu*hood is *moqu fila* 'small child' which typically refers to the ages of zero to five. Additionally, a term introduced at least in the last half century, is *sakola moqu* 'school child', which typically refers to children from the ages 6-16 years. Crucially enough, Abui does not have a designated word to describe the adolescent stage, typically classifying adolescents in puberty as *moqu* until they are fully sexually mature and become available candidates for marriage [EG.60M.AG].

work'. Knowing how to work typically refers to being able to work in the fields. For men, it also involves knowing how to build a house, while for women it involves being able to undertake domestic chores. *Moqu* typically do not know how to 'work' on their own. As a result, they are not considered viable for marriage, as excerpt (2) shows.

(2) *Moqu belum bisa kerja... moqu berarti itu puber yang memang su begitu* (*15 - 16 tahun*) *tapi belum bisa dikawinkan masi dianggap itu anak*.
'A *moqu* can not fully work yet. *Moqu* does indeed refer to an adolescent in puberty (of around 15-16 years). However, (s)he may not get married.' [EG.65M.AG]

Today, all members of the (pre)adolescents group of *moqu* are Alor Malay-dominant bilinguals. They only speak Alor Malay with their peers and age-mates. The few instances in which they report speaking Abui is when they converse with their grandparents; even then, while grandparents often address their grandchildren in Abui, some respond in Abui while others still respond in Alor Malay. Sociolinguistic data from nineteen members of the *moqu* '(pre)adolescents' group reveals that a great majority of speakers (16/19) report being addressed in Alor Malay by their parents in their early childhood, before entering primary school (see Table 3.4). Six speakers, who report being addressed in Alor Malay before entering primary school, report being addressed in both languages during primary school, indicating that parents adjusted their language patterns. This is in line with two claims that parents make. The first one is about the emphasis of the use of Malay Indonesian with children before they enter primary school so that they can cope with the language of instruction. The second is about children gradually becoming more fluent as they grow older, in this case as they transition into primary school. All nineteen participants report speaking exclusively Alor Malay in the playground. As such, this increased Abui language use by parents has little to do with school as such, but has more to do with the gradual process of language acquisition and socialization that takes place with Abui (see §2.4.2).

There was one outlier in this group who reported being addressed in Abui by his parents both before and during primary school as well as initiating speaking it with peers in the primary school. This was a boy who

was born in the mountain hamlet of Tifol Afeng, who lived in between Tifol Afeng and Takalelang. As discussed in §1.4.2, many children from Tifol Afeng grow up speaking Abui. The two speakers who reported being addressed using both languages both lived with their grandparents (who spoke little Malay) for at least four years.

Table 3.4: Self-reports on early language exposure of (pre)adolescents (N=19)

(Pre)adolescents (age 9 - 16 years)	**Alor Malay**	**Both**	**Abui**
Language used by parents before primary school	16	2	1
Language used by parents during primary school	10	8	1
Language used with peers in primary school playground	18	0	1

3.4.2 *Neeng abet/maayol maak* 'young adults'

Neeng abet 'young male adult' and *maayol maak* 'young female adult' are the terms used to describe sexually mature members who are searching for a spouse. Being a *neeng abet/maayol maak* 'young adult' also means that a person must be able to work independently, which is deemed crucial if one is to be considered as a viable spouse. Today, finishing an education also affects the age in which people get married. Both terms correspond to similar ages in both genders. They typically have a range from about 17 to 25 years, as excerpt (3) shows.

(3) Q: *Terus, bagaimana dengan neeng abet atau maayol maak, menurut nenek*?
'Then, what about *neeng abet* or *maayol maak*, according to grandpa?
A: *Kalo neeng abet dengan maayol maak itu su mulai umur 17 sampe dengan 24-25, paling tinggi 30*
'As for *neeng abet* and *maayol maak*, this begins from age 17 until 24-25, max. 30' [EG.65M.AG]

When asked for the age at which people typically get married, the interviewees typically gave twenty-five as a milestone, arguing that people would start to ask questions if a marriage wasn't planned by that age. The age of

twenty was given as an age that allowed young adults to finish their studies, work for a few years, and accumulate wealth, in preparation for a dowry payment. It was also pointed out by several speakers that while 17-25 is a representative age-range for the category of *neeng abet* 'young adult male', the essential characteristic that the term holds is unmarried status. In other words, if someone within that age-range bears a child or gets married, they will no longer hold the status of *neeng abet*/ *maayol maak* 'young adult' but will move unto *kalieta* 'elder', as evidenced from excerpt (4).

(4) *Umur belasan ju, su kawin berarti su bukan neeng abet lagi, su status bapa tu, status kalieta*
'Even if someone is a teenager, if they were to get married, then they will no longer be a *neeng abet*. Their status will change to *bapa* 'father' (in Alor Malay), or to *kalieta* 'adult/elder' (in Abui).' [EG.57M.AG1]

The reverse is true. If someone in their thirties is not yet married, they are still considered a *neeng abet* or *maayol maak*, as excerpt (5) shows.

(5) *Na, sampe umur tigapuluan juga belum kawin berarti dia masi neeng abet. Umur mulai 35 ke atas itu yang, kadang orang tida bilang neeng abet karena itu suda masuk di usia usia orang tua walaupun dia belum kawin.*
'Further, even if someone has reached their thirties,the fact that they have not yet gotten married yet still means that they are a *neeng abet* 'young adult male'. However, this lasts until 35. From 35 onwards, even if they have not gotten married, they are no longer called a *neeng abet* because they've entered the age of elders.' [EG.45M.AG2]

The young adults are also Alor Malay-dominant bilinguals. Their early exposure as indicated in Table 3.5 resembles that of the group of (pre)adolescents, although the group of young adults have had slightly more exposure to Abui. For example, four young adults report being addressed equally in both languages, while two report mostly being addressed in Abui. Similarly to the (pre)adolescents, we see an increase in exposure to Abui (besides Alor Malay) when the young adults entered primary school.

The two speakers that report being addressed in Abui both before and during primary school also lived with grandparents. The same holds for the speakers who selected both.

Table 3.5: Self-reports on early language exposure of young adults (N=19)

Young Adults (age 17 - 25)	**Alor Malay**	**Both**	**Abui**
Language used by parents before primary school	13	4	2
Language used by parents during primary school	10	7	2
Language used with peers in primary school playground	15	3	1

Today, many young adult speakers who had been living away from the Abui home community for a certain period of time (e.g., for educational and career purposes) report having made conscious efforts to resettle into the community upon their return, and improving their Abui as they take on more important 'adult' roles in the community. Speakers who remained in Takalelang their whole life also report an improvement in their Abui, both in terms of fluency and frequency of use.

3.4.3 *Kalieta* 'adults'

The group of 'adults' consists of speakers aged 26-34 years. It is based on the Abui term *kalieta* which is a term reserved for adults who are married or have children two typically defining features of this age category. Especially if one has a child, this change of status to *kalieta* also bears with it an additional name; the person in question is now called 'child so-and-so's father/mother', as described in excerpt (6).

(6) *Setela dia suda beruma tangga, kawin, dia suda tida neeng abet itu suda 'neeng kalieta'. Apa lagi suda ada anak, itu suda 'neeng kalieta'. Sudah dikatakan kecil punya bapak.*
'When he has married and settled down, he is no longer a *neeng abet* 'young male' but a *neeng kalieta* 'adult male' instead. Especially if he has a child, then he is definitely a *neeng kalieta* 'adult male'. He is (then) referred to as 'child so-and-so's father'.' [EG.45M.AG2]

The lower boundary of twenty five corresponds to the age at which many people typically get married. This age is generally higher today than it was in the past due to the fact that a lot of young Abui speakers pursue tertiary education either in Kalabahi (Alor Island) or in Kupang (Timor Island) [EG.45M.AG2].

Technically speaking, the Abui term *kalieta* is used for any adult who is married and/or has children all the way into old age. For the purposes of this study, however, a distinction is made between *kalieta*, aged between 26-34 and *kalieta*, aged 40 and above (discussed in §3.4.4). There are two important reasons for this distinction, one due to life-stage and the other due to history. In terms of life-stage, there is some evidence to suggest that speakers view 25 to 35 as the ideal window for marriage. This is also related to status. The evidence for this can be seen in how people observe unmarried adults. People below 35 who are not married may still be considered *neeng abet*/*maayol maak* 'young (unmarried) adults'; however, people above 35 who are not yet married will automatically be referred to as *kalieta*, as shown by excerpt (5). This points to the age of 35 as representing some sort of threshold.

The second reason for this is not due to life-stages per se, but due to different history. Speakers aged 26-34 were the first generation to have undergone a shift to Alor Alor Malay, while speakers above forty were all raised in Abui and may be considered true Abui L1 speakers (see §§2.3-2.4.2.1 for discussion on age and a change in schooling patterns respectively).

In terms of language exposure, the adults group was the first to undergo a change in parental linguistic strategies: not all parents unequivocally raised them in Alor Malay; they often mixed Abui and Alor Malay. This is shown in Table 3.6,[3] which presents a more balanced distribution compared to the two younger groups. Before and during primary school, a considerable number of speakers in this age-group is raised either exclusively in Abui, or in both Abui and Alor Malay. Similarly to the (pre)adolescents and young adults, we see a rise in exposure to Abui as the speakers entered primary school. As regards the language used to communicate with friends in the playground, we see that half of the speakers reported using mostly

[3]There was missing data for one speaker with regards to the variable 'Language used with peers in primary school playground' which is why the total number of tokens amounts to 18 instead of 19.

Abui. This number is considerably higher than in the other two groups illustrating that the 1980s and early 1990s was a period when children still spoke Abui to each other.

Here again, the four speakers that report being addressed in Abui before primary school either lived with their grandmothers (N = 3), or grew up in Tifol Afeng (N = 1).

Table 3.6: Self-reports on early language exposure of adults (N=19)

Young Adults (age 26 - 34)	**Alor Malay**	**Both**	**Abui**
Language used by parents before primary school	7	8	4
Language used by parents during primary school	3	9	7
Language used with peers in primary school playground	8	3	7

Today, 'adults' typically speak both Alor Malay and Abui to their peers. They speak mostly Abui to elders and mostly Alor Malay to younger speakers as well as their own children.

3.4.4 *Kalieta* 'elders'

The age-cohort of elders (range: 40 - 75 years) has been designated as the control group in this research. This group is the speaker group from which the patterns and rules of Abui grammar are assumed to be derived (as reported in e.g., Kratochvíl, 2007), and is taken to represent the grammatical norm. Only nine speakers were selected; they showed very little variation in the linguistic variables selected for investigation. There are two crucial metrics on which to split the Abui notion of *kalieta* into two groups of 'adults' and 'elders', one due to them represented a different life-stage and another due to them also having a different history with regards to langauge use and exposure.

With regards to life-stage, the 'elders' group typically represents a later stage in adulthood compared to the group labeled 'adults' (discussed in §3.4.3). In this age category, most members would be suitable to sit on the *meja adat* (Malay Indonesian), roughly translated as 'the table of customs'. This refers to the negotiation table for bride price, land ownership and other

important matters. In order to sit on the *meja adat*, both marriage and proficiency in the Abui ritual language are required. This is not always accessible to speakers in the 'adults' group.

With regards to their language history, while the *kalieta* 'adults' (26-34 years) grew up with a mix of Alor Malay and Abui, *kalieta* 'elders' (40-75 years) grew up speaking Abui, as evidenced by (7).

(7) *Mulai tuju delapan tahun itu kita pung bahasa Abui itu sudah. Kan dulu kan kita hanya omong bahasa Abui saja to*?
'From age 7-8 our Abui would be fluent. I mean, in the past, we only spoke Abui, you know.' [EG.49F.AG]

Elders only received exposure to Alor Malay from teachers when they entered schools, as depicted in excerpt (8). They often dropped out of school after two or three years. This was especially common for women whose parents insisted that they stay home and lend an extra hand in the fields. Most parents of the members of this group spoke Abui and thus raised this group in Abui. The parents did not receive any formal education; however, some of their fathers received some Alor Malay literacy training by foreign (Dutch) administrators, learning to write on stone slates (see §2.3; Du Bois (1944) mentions that barely twenty boys knew how to speak Malay in the entire five-village complex in Atimelang).

(8) *... perkembangan dulu lain. Dulu kami ini tida tau bahasa itu Indonesia yang mana. Bahasa Indonesia saja kita turun sakola baru tau. Hanya bahasa Abui.*
'..in the past, the developments were different. In the past, we did not know what Indonesian even was. We had to go down to school (from the villages in the hills), and then we would learn Indonesian. [It used to be] only Abui.' [EG.60M.AG]

Table 3.7 illustrates that all nine speakers sampled for the sociolinguistic interview report being spoken to by parents in Abui both before and during primary school.[4] In addition, they all claim to have initially used Abui with peers in primary school playground until they were punished for doing so (see §2.3 for more info on school policies).

[4]One speaker did not attend primary school. This is why the field the third row has only 8 speakers.

Table 3.7: Self-reports on early language exposure of elders (N=9)

Young Adults (age 40 - 75)	**Alor Malay**	**Both**	**Abui**
Language used by parents before primary school	0	0	9
Language used by parents during primary school	0	0	8
Language used with peers in primary school playground	0	0	8

Today, the elders use Abui to communicate with other (young) adults and elders, and speak Alor Malay only to children and non-Abui speakers.

3.4.5 Summary of age-groups

Sections 3.4.1-3.4.4 have given descriptions on the four age-cohorts used to categorize speakers for the variationist studies found in this thesis. The age-groups correspond roughly to certain age-ranges which are in turn related to certain life-stages. The cohorts were inspired by the Abui terms and life-stages associated with them. However, speakers are placed into age-cohorts purely based on their age and not on whether they conform to a given life-stage. In addition to age-ranges and life-stages, information was presented on their early language exposure as well as their current language use. The differences between the age-groups in terms of early language exposure and current language use are summarized in Table (3.8). For further information on marriage practices as well as the construct of age as broken down into life-stage and history, see §1.5.4 and §2.4.2.1 respectively.

Table 3.8: Summaries of differences between age-groups

	Moqu **'(Pre)adolescents'**	***Neeng abet/maayol maak*** **'young adults'**	***Kalieta*** **'adults'**	***Kalieta*** **'elders'**
Age-range	9-16	17-25	26-34	40-75
Life-stage	Not yet sexually mature	Sexually mature. Prepare for marriage.	Typically married and/or bear child	Married. Can participate in ritualized negotiations
Early exposure	Raised in Alor Malay by parents. Spoke Alor Malay with peers.	Raised in Alor Malay by parents. Spoke Alor Malay with peers.	Raised by both Abui and Alor Malay by parents. Spoke a mix with peers.	Raised in Abui by parents. Spoke Abui with peers. Learned Alor Malay at school
Current use	Speak Alor Malay with peers and adults.	Speak Alor Malay with peers. Speak Alor Malay and Abui with adults	Speak Alor Malay and Abui with peers. Speak Alor Malay and Abui with adults. Speak Alor Malay with children.	Speak Abui with peers and parents. Speak Alor Malay with children.

3.5 Linguistic data

In order to study on-going language contact in as complete a manner as possible, several threads of linguistic data were collected: conversational data (§3.5.1), experimental data (§3.5.2), data from fieldnotes (§3.5.3). Conversational data includes spontaneous speech involving several speakers. Experimental data includes two types of data: production data gathered by means of the Surrey Stimuli video elicitation task (§3.5.2.1) and comprehension data gathered by means of a forced-choice task (§3.5.2.2). Data from fieldnotes includes elicited data (§3.5.3.1), and data obtained from participant

observation (§3.5.3.2).

The different types of data were used for various purposes across Chapters 4-7. Conversational, experimental, and elicited data were used for the sketch grammar (Chapter 4) and for the study on reduplication (Chapter 7). Conversational and experimental data was used in the two studies on variation and change (Chapter 5, 6).

Taken together, data collection was carried out to capture three main dimensions relevant to the study of language change: a) genre diversity, b) ecological validity, c) comparability across different kinds of participants.

To achieve genre diversity, data from conversational recordings was collected alongside data from experimental recordings. These types were further supplemented by elicited data to test certain hypotheses about the language. Further, the method of participant observation produced data and questions that would not have otherwise arisen in more controlled environments. In addition, even within category of experimental data collection, two modalities were targeted: production and comprehension. This combination allowed for a more complete testing of speaker's competence; it allowed for a testing of speaker's online processing as well as their linguistic knowledge.

The second dimension of ecological validity refers to the extent to which the data here may be representative of how people actually speak and are spoken to. Spontaneous conversational data is typically the closest one can come in terms of collecting ecologically valid data and it is for this reason that this type of data was collected. Furthermore, the age-range of speakers from the conversational data of 1 to 85 years also adds further ecological validity to the data because it is representative of nearly every age category in the community.

The third dimension of comparability across different kinds of participants was met through the collection of experimental data. Therefore, the studies in Chapter 4 and 5, which rely predominantly on experimental data are comparable, as the same speakers were sampled. For the experimental data collected, speakers were split into four groups: three experimental groups and one control group. The experimental groups had similar numbers of participants, spread across gender evenly. While the conversational data was more spontaneous and thus less comparable across speakers, the experimental data involved responses to the same set of stimuli. In addition, the majority of speakers performed both the production and the

comprehension task.

Of course, these types of data do not individually satisfy all three of the criteria outlined above, but taken collectively, they form a varied sample of data that does.

3.5.1 Conversational data

The genre of conversation is often considered highly ecologically valid. Most people use language to converse with one another, so collecting conversations represents language use in its most natural form (Dingemanse & Floyd, 2014). In total, 19 conversational recordings were transcribed and translated. The age-range of speakers was 1 to 85 years. Data from speakers above 40 provided a baseline for how the language is spoken by Abui L1 speakers. Data from younger speakers was representative of how the language is undergoing change.

There are two types of conversations in my corpus. The first is spontaneous conversations, where I simply recorded conversations already taking place, while the second is directed conversations, where I brought speakers together and directed them to engage in conversation.

In the first type, speakers were approached whilst already sitting together and engaging in conversation. Then, I kindly asked whether I could set up the camera and record their existing conversation, explaining I was interested in recording a *konversasi sehari-hari* 'everyday conversation'. I pointed out that I simply wanted speakers to forget that the camera was there and continue having their everyday conversation, implying that I did not expect a formal register but a casual one. I also highlighted that I wasn't particularly interested in collecting Abui data, making clear that any language or variety that came up would be suitable as long as speakers felt they were conversing naturally. These instructions led to the collection of a number of recordings containing natural speech which often included code-switching of Abui and Alor Malay.[5]

The second type was less spontaneous. It involved directing the recording by asking speakers to come together and converse with one another from a list of topics. This method was implemented particularly for (pre)-adolescent speakers. Typically, it was not always possible to record them

[5] Generally speaking, code-switching is not addressed as it fell outside the scope of this thesis; however, some references are made to it in Chapter 4, example (113) and in §7.6.1.

speaking spontaneously because they were often found running around and playing. However, with the help of a school teacher and/or adult speakers, they often agreed to sit down and have informal conversations about their favourite school subjects or what activities they had been engaging in. A breakdown of all the conversations transcribed and translated in my corpus is listed in Table 3.9.[6]

[6] In addition to the nineteen conversations presented here, around twenty more were recorded but left untranscribed and untranslated.

Table 3.9: Abui corpus of conversational data

Duration	Code	Content	Speaker age and gender
0:01:50	BB	Conversation centred around a baby	1F, 5F, 23M, 40F, 43F, 45F, 50M, 70F
0:16:41	QU	A young male asking an elder man to ask his mother questions about her past	23M, 50M, 85F
0:09:54	SC	2 sisters conversing about school	15F, 25F
0:13:57	MM	A boy reciting the Mon Mot Mot story to his friends	9F, 11F, 12F, 13M
0:08:44	CT	A child telling his friends a traditional story	13M, 16F
0:12:10	DA	A young adult asking 2 adolescents about their daily activities	13M, 16F
0:08:43	LU	Young adult asking questions to 2 adolescents about language use	13M, 16F
0:06:15	AL	Young adult asking questions to adolescent girl about language use	14F, 23M
0:04:15	LP	A (pre)adolescent girl assists a 3 older women in preparing lunch	9F, 40F, 40F, 45F
0:19:49	CP	An (pre)adolescent girl assists a 3 older women in baking a cake	9F, 40F, 40F, 45F
0:41:50	GJ	2 girls making jewelry	23F, 27F
0:22:15	BC	2 boys conversing about school and future plans	24M, 19M
0:19:48	GO	4 girls conversing on Orpa's porch	21F, 22F, 23F
0:24:05	WG	Women gossiping under a tree	22F, 26F, 32M, 34F, 40F, 49F
0:31:31	CS	People of all ages, sitting on a porch and practising for the choir for a Sunday church sermon	7M, 15F, 22F, 40F, 43F, 45F, 46M
0:24:48	AH	A young lady, her baby daughter, her mother, and two other ladies sitting and chatting	0.5F, 22F, 40F, 42F, 43F
0:19:45	JH	3 speakers in their early thirties taking care of their children and conversing with an old lady	4F, 5F, 30F, 32M, 32F, 34M, 75F
0:07:00	LH	A group of young adults sitting at someone's home, conversing with some adults; lots of Abui/Alor Malay code-switching	18M, 22M, 24M, 25M, 26M, 27M, 43M, 47F
0:07:43	HB	3 (pre)adolescents discussing their hobbies and what they did earlier in the week	10F, 12F, 15M
5:01:03	*Total*		

As illustrated in Table 3.10 the conversational data includes data from speakers ranging from age 1-85. In a number of recordings, toddlers were also present. Therefore, these recordings have some instances of child-directed speech as well as utterances from children themselves. In other recordings, some old ladies above the age of 75 were present. All in all, 38 females were recorded, compared to 18 males. There are several possible explanations for this gender bias in some of the age-groups. The first one is coincidence; 6 of the children aged (1-8 years) all happened to be female. Moreover, in the (pre)adolescent group, boys typically felt more shy than girls to take part in a conversation.[7] In the elders group, it appeared to be easier to find women gathered in one space, conversing naturally. Men appeared to be more scattered around.

Table 3.10: Participant table for conversational data

Group	**Age-range (years)**	**M**	**F**	**Total**
Children	1-8	0	6	6
(Pre)adolescents	9-16	2	9	11
Young adults	17-25	7	8	15
Adults	26-34	5	2	7
Elders	40-85	4	13	17
Total	1-85	18	38	56

There were several advantages of using conversations for this research. Generally, when transcribed and translated, conversational data provides the basis for text data which often form the backbone of grammatical analysis. They are also essential in analyzing the more pragmatic, semantic and subtle parts of language (Payne, 1997). Thus they were used extensively for the sketch in Chapter 4. The advantage of using texts is that it becomes possible to study how people actually speak, as opposed to how people think they speak. This is especially important in studies such as the ones included in this thesis, where a lot of statements are made about which variant in the

[7]However, it was observed that girls generally felt more comfortable speaking Abui. This could be related to socialization practices discussed in §2.4.2.2. Gender differences in language use are also discussed in Chapter 5.

speech community is considered felicitous and grammatical.

Another advantage of using conversations is that they may be used for the study of conversational analysis, which is useful as a separate field of study but also proved relevant for the research in this thesis. In the context of studying language change in progress, conversations were insightful in understanding whether or how speakers use other-initiated repair to correct the speech of younger, less fluent speakers. Although not investigated systematically, this type of repair, or 'corrections' by older speakers in conversation proved helpful in pointing out areas of the language that younger speakers struggled with, in addition to highlighting what kind of methods older speakers used to correct their speech, if at all. This thus proved to be a novel means to study the treatment of age-related variation.

This is evident in example (9), where a 22-year-old speaker reduplicates the wrong verb *heyel~heyel-ra* 'go insane' and is then corrected by a 40-year old speaker through the use of *ayoq~ayoq-da* 'split in half'. This phenomenon was one of the methods used in early stages of this research to tag speech that appeared to deviate from community norms. Instances like this, for example, highlighted reduplication as a possible area of investigation, which is why it was later investigated.

(9) a. 22-year-old female

A-pong *heyel~heyel-ra!*
2SG.INAL-face RDP~go.insane-IPFV

'Your face is going insane! (You are drunk)' [CV.22F.AH]

b. 40-year-old female

"A-pong *ayoq~ayoq-da"* *ba* *henil-e.*
2SG.INAL-face RDP~two-IPFV REPORT like.this.IPFV-IPFV

' "Your face is being split in half (You are drunk)!" Like that.' [CV.40F.AH]

However, despite these advantages, conversational data also has some shortcomings. One disadvantage that it is more time-consuming to transcribe than other genres. Furthermore, due to its free nature, it is not always possible to obtain comparable data among speakers. As such, it was not possible to collect conversational data from all the participants I tested to

conduct quantitative analyses. For this reason, I also collected experimental data.

3.5.2 Experimental data

Experimental data was split into production data and comprehension data. To gather production data, the Surrey Stimuli video elicitation task was used (§3.5.2.1). To gather comprehension data, a forced-choice task was used (§3.5.2.2).

3.5.2.1 Production data: Surrey Stimuli video elicitation task

The Surrey Stimuli task (Fedden, Brown, & Corbett, 2010; Fedden & Brown, 2017) is a series of 40 short video clips spanning about five seconds. The clips are listed along with a brief description in Table 3.11. For a list of the 40 clips, plus accompanying screenshots of the videos, see Appendix IV.[8]

One of the main advantages of data from the Surrey Stimuli task is that it provides comparability across various speakers, making it ideal to study language variation across age-groups. It also provides video clips which target many areas relevant for Abui grammar. This is because the task was initially designed by a group of researchers interested in pronominal marking in the Alor-Pantar languages more generally. It contains a large amount of both intransitive events and transitive events involving two participants, making it useful for studying variation in pronominal marking, reflexive possessive marking, lexical semantics, among other topics. Often, semantic factors related to pronominal marking, such as animacy and volitionality were manipulated in the clips, providing an adequate range of experimental conditions. The stimuli set had been tested on Abui and other AP languages (see Fedden et al., 2014; Fedden and Brown, 2017) and did not appear to violate any cultural or pragmatic norms. Finally, it is important to underline that, initially, the elicitation task was not hypothesis-driven, but rather aimed to collect a corpus in which variable grammatical patterns could be identified.

The procedure of the task went as follows: speakers were asked whether they were interested in being recorded as they watch short video clips on a computer and describe what was going on in the clip, using Abui. This was

[8]The initial set had 42 clips, but 2 were not available when downloading the set from http://www.smg.surrey.ac.uk/projects/alor-pantar/pronominal-marking-video-stimuli/.

Table 3.11: The Surrey Stimuli video clips

Code	Description of clip
C01	man pulls other man
C02	girl leans on man
C03	people dancing
C04	boy steps on sleeping man
C05	man sleeping on bed
C06	man sitting against wall and dozing off
C07	woman sitting and laughing
C08	man holds snake, gives it to girl who is afraid
C09	water being poured into glass
C10	man lying down, talking to himself
C11	man sitting against house and eating banana
C12	man is standing, boy runs over to him
C13	one man is standing, another walks into him
C14	man walks over to a wall and sits down
C15	coconut tree; one coconut falls
C16	man walking, bumps into, and then hugs tree
C17	4 logs, 1 large one
C18	girl pulling log
C19	man walks over and sees axe with blood
C20	man walks and steps on banana
C21	man standing and leaning on house
P01	man smells other man
P02	man smells cheese
P03	match goes out
P04	man sitting, leaning against wall and waking up
P05	a short man and tall man are standing
P06	boy hears noise and is startled
P07	boy is sleeping; man comes and wakes him up
P08	bent person on all 4s with rock on back
P09	man walks over and trips on log
P11	banana falls on log
P12	man hears other man
P13	man hugs tree
P14	small log placed on big bent plank
P15	man cradles child
P17	man standing
P18	3 stones of different sizes
P19	banana falls on stomach of man lying down
P20	man running
P21	man leaning on wall, gets up, and walks away

accompanied by the reassurance that they could resort to Alor Malay should they struggle to find the right Abui words to describe the clips. Instructions were always given in Alor Malay; while watching the clips, they were asked *apa yang terjadi?* 'what was going on?'.[9] Participants were video-taped while performing the task. On three occasions where video was inappropriate or unavailable, the session was audio recorded. For the (pre)adolescent group, the instructions were given by a school teacher, as an authoritative yet familiar figure from whom the children were used to following instructions.

The experiment was most often conducted at their own home, but sometimes it was also conducted at the household where I was staying. Specific efforts were made to avoid the presence of older speakers who were often tempted to interfere and 'correct' utterances they felt were ungrammatical. When it was unavoidable to be in the presence of older speakers, I made it very clear that I was only interested in the speech of the participant in question, and that the recording could be jeopardized should others get involved. Some older speakers later objected to the idea that I was collecting language data from speakers who are 'not fluent yet'. However, I reiterated my motivation of collecting data from various ages, in order to study the development of the language. I also pointed out that it was important to have an idea of how less than fluent speakers spoke Abui for didactic purposes.

In general, speakers performed the task smoothly and comfortably. On a small number of occasions, I was prompted to initiate repair and intervene. The first of these involved instances where speakers were simply encouraged to elaborate. Speakers sometimes gave too little information or missed an essential part of the target response. This could come in the form of open ended question, such as asking why something happened, if they gave a short response. An example of this would be from a video were a man kneels down, sees an axe in a pool of blood (Clip C19 in Table 3.11). If a response merely yielded, 'the man was scared', I would probe further and ask 'why was scared?' or 'what did he see?'.

Other instances that prompted me to interject included utterances that

[9] As mentioned in §3.2, speakers were given instructions in Alor Malay for two main reasons. First, especially in the early stages of fieldwork, explaining instructions clearly was only possible in Alor Malay. Second, it was unclear to what extent younger speakers would fully understand instructions in Abui. This was evidenced by the fact that some of my assistants deliberately chose to address them in Alor Malay when giving them instructions. As such, Alor Malay seemed like the most appropriate option.

weren't clearly audible due to poor articulation or background noise. In asking speakers to repeat their utterance, some speakers shuffled around their utterance and used a different choice of words. In all of these cases of researcher-initiated repair, the second utterance was always considered. The second utterance was also considered in instances of self-repair, where the participant would repeat their responses in favour of a more accurate and better constructed sentence. Having said that, there is still a large amount of variation in the responses given.

In total, the Surrey Stimuli task was collected from 66 speakers, spread across the four age-cohorts, as illustrated in Table 3.12.[10] For each the three experimental groups, (pre)adolescents, young adults, adults, 19 speakers were recorded.[11] However, for elders (age 40-75), the control group, only nine speakers were recorded because their speech showed little variation from one another. Furthermore, most of the published resources on Abui are based on data from speakers in this age-cohort (e.g. Kratochvíl, 2007, 2011a; Kratochvíl and Delpada, 2014). Nonetheless, the sum of nine still appeared to be large enough to run certain statistical analyses, such as the Kruskall-Wallis test, reliably.

Table 3.12: Participant table for Surrey Stimuli Production Task

Group	**Age-range (years)**	**M**	**F**	**Total**	**Mean age**
(Pre)adolescents	9-16	9	10	19	13.47
Young adults	17-25	10	9	19	21.42
Adults	26-34	10	9	19	30.29
Elders	40-75	4	5	9	50.44
Total	9-75	33	33	66	25.51

3.5.2.2 Comprehension data: Forced-choice task

While it is essential to examine speaker's language production, it is also important to investigate speakers' comprehension in order to explore to what

[10] For details on how this type of data was transcribed and translated, see §3.8.

[11] In the adults group, the recording of one of the speakers was not recorded firsthand, but was taken from the Kratochvíl corpus; see §3.7.

extent a given feature or simplification pattern is entrenched in a speaker's competence (Onar Valk, 2015; see also §1.8.5). In order to test for speakers' comprehension, a forced-choice task was conducted.

In a forced-choice task, speakers are typically presented with more than one similar constructions in either writing or speech and must select which of the sentences is most felicitous, grammatical, or appropriate for a given stimulus. They are thus 'forced' to select which construction fits best. This is especially useful in testing constructions with similar underlying meanings. Thus, it proved useful in testing variation within a given linguistic category, such as reflexivity in third person possession, for example.

In this research, the task was devised to test four linguistic variables (see Table 3.13). The first step involved selecting video clips which could elicit constructions suitable to these variables. A total of 30 clips were selected, most of which were taken from the Surrey Stimuli production task, while five were recorded with Abui speakers for the purpose of this task. For every video clip shown to participants (also labeled a 'trial'), two conflicting sentences were recorded, one containing a match and the other a mismatch. Thus, a total of 30 videos were shown accompanied by 60 audio stimuli.

In order to find the appropriate sentence pair for every trial, I sat down with native elder speakers and discussed sentences that could be considered matches. Once they had been selected, they were then uttered by a native speaker (25-year-old male speaker) and audio recorded. Subsequently, the task was piloted with two other native speakers above the age of 40 to test whether the target sentences were fully grammatical and/or felicitous to them and also whether the nontarget sentences were not fully grammatical and/or felicitous.

The task was split into two experimental sets, A and B. The 30 video clips were then randomly shuffled for each set. Half of the participants were assigned to set A while the other half were assigned to set B. The participants watched the video clips and listened to the recording of the two sentences accompanying the clip, which had also been randomized, so that the target was sometimes first, and sometimes second.

The procedure went as follows. After agreeing to the task, a speaker sat down next to me, facing a laptop. Using a jack splitter, both the speaker and myself each had our own set of headphones. This ensured that we could both listen to the audio-recorded stimuli. I would then play a video, and follow it up with sentence 1, and then sentence 2. I then asked the speaker in

Alor Malay, *mana yang pas*? 'which one fits [with the video]?'. Speakers then uttered ' one' or 'two'. They were always presented with two trial rounds in the beginning, involving simple sentences such as 'the man is sitting' vs. 'the man is standing', along with a video of one of these postures. This afforded them time to acclimatize to the task. Everyone passed the trial round. All sessions were audio-recorded. After the sessions, the sessions were played back and the answers were then filled into an excel sheet. Figure 3.2 illustrates the set-up of the task.

Figure 3.2: Forced choice set-up

An example of a trial involving a match and a mismatch is presented in (10a-b). The video stimulus shown here involves 'a man pulling his (own) friend' (Clip C01 in Table 3.11). The variable in question is the use of the reflexive possessive (discussed in Chapter 5). Option (10a) represents a match because it contains the reflexive possessive prefix *de-* while option (10b) represents a mismatch because it contains the non-reflexive possessive prefix *he-*.

(10) Target: Reflexive possessive (*dV*-) Video clip C01 [man$_x$ pulling his$_x$ friend]

a. **Reflexive match: (*dV*- prefix)**

Neeng	*nuku*	***de-feela***	*ha-fik-e.*
man	one	3.REFL.AL-friend	3.PAT-pull-IPFV

'A man$_x$ is pulling his$_x$ friend.' [FC.POSS1.MATCH]

b. **Non-reflexive mismatch (*hV*- prefix)**

Neeng	*nuku*	***he-feela***	*ha-fik-e.*
man	one	3.NREFL.AL-friend	3.PAT-pull-IPFV

'A man$_x$ is pulling his$_y$ friend.' [FC.POSS1.MIS]

The four linguistic variables addressed by the 30 trials include: 1) reflexive possessive marking, 2) verb usage, 3) pronominal indexation, and 4) clause linkers. For the full list of trials for the forced-choice task, see Appendix V. The order of trials was randomly assigned, meaning that since they belonged to different variables, they acted as distractors for one another. Despite four variables being tested, this thesis only reports on two of them: Reflexive possessive marking in Chapter 5 and verb usage in Chapter 6.

To ensure that speakers were not guessing and could tease apart a match from a mismatch, every clip that targeted a given form was counterbalanced by another clip which targeted the mismatch of that form. For example, there were three videos such as (10), which targeted the reflexive possessive *dV*-, where the non-reflexive possessive represented a mismatch. To counterbalance this, there were also three trials targeting the non-reflexive *hV*-, where the non-reflexive possessive *hV*- represented a match and the reflexive *dV*- represented a mismatch. This was done for every variable; see Table 3.13.

Table 3.13: Description of trials for forced-choice task

Variable	Condition	Trials
Possessive marking	Reflexive possessive target	3
	non-reflexive possessive target	3
Verb usage	Visual perception 1: 'see'	2
	Visual perception 2: 'look at'	2
	Motion 1: 'fall over'	2
	Motion 2: 'fall from above'	2
	Change of state 1: 'get up'	2
	Change of state 2: 'wake up'	2
Pronominal indexation	GOAL prefix target	3
	PATIENT prefix target	3
Clause linkers	SEQUENTIAL *ya* target	3
	LINKER *ba* target	3
Total		30

A total of 60 participants took part in the forced-choice task during a two-month fieldtrip in 2017. Most of the participants who took part in the production task also participated in the comprehension task. In total, 9/66 participants from the Surrey Stimuli production task were not available for the forced-choice task. To compensate, three new speakers were added for the forced-choice task. Information about the participants is laid out in Table 3.14. All 60 recordings may be found in The Language Archive (click here).

Table 3.14: Participant lists for experimental tasks

		Production			Comprehension		
Groups	**Age-range**	**M**	**F**	**Total**	**M**	**F**	**Total**
(Pre)adolescents	9-16	9	10	19	9	9	18
Young adults	17-25	10	9	19	9	5	14
Adults	26-34	10	9	19	9	8	17
Elders	40-75	4	5	9	5	6	11
Total	9-75	33	33	66	32	28	60

3.5.3 Fieldnotes

In this thesis, I make a distinction between data that has been recorded by audio or video (such as conversational and experimental data) and data which has been jotted down in a notebook or on a laptop (fieldnotes). Data from my fieldnotes generally comes in two forms: elicited data (§3.5.3.1) and data obtained from participant observation (§3.5.3.2).

3.5.3.1 Elicited data

Both elicitation and textual data (conversations, narratives) are crucial for a comprehensive descriptive account of a language (Payne, 1997, p. 366). Payne defines elicitation as 'samples of language that accomplish hypothetical communicative tasks'. The reason they are considered hypothetical is that typically a intermediary language is used to elicit a translation equivalent utterance. This type of elicitation can be especially useful in the early parts of fieldwork (Bowern, 2008, p. 77). Elicitation can also involve back translation, meaning that the researcher offers a sentence in the target language and then asks the speaker to translate it into the intermediary language. Duranti (1997, p. 98) discusses the importance of elicitation as allowing the linguist to uncover regularities in the linguistic system and get access to forms that might not be very accessible in every day speech. Another form of elicitation includes asking for grammaticality judgements. This involves presenting an utterance to a speaker and asking them whether the utterance is a good sentence (Bowern, 2008, p. 78). This can be particularly tricky, be-

cause speakers can have all sorts of different opinions on what constitutes a good or a bad sentence, and it is not entirely clear which part of the sentence they might be at odds with. All three types, namely translation, back translation, and judgements were used throughout this thesis.

In the research for this thesis, Alor Malay was used as a language of elicitation, especially in the early stages. This had several advantages and disadvantages. The main advantage was it was the language that most speakers are fluent in, and some even native. This created one less hurdle in terms of understanding one another. In addition, its reputation as a *bahasa sehari-hari* 'colloquial day-to-day language', meant that speakers felt at ease speaking it during elicitation sessions. This brought about a more casual atmosphere, which perhaps might have been different through the use of the more formal register of Standard Indonesian. However, when compared to Abui, Alor Malay lacks much of the complexity both lexically and morpho-syntactically. This meant that it was often difficult to capture some of the subtleties using Alor Malay. Towards the later stages, Abui was used as the language of elicitation.

Elicitation served two main functions. First, it was used to gather grammaticality judgements on and modifications of the speech of younger speakers, by older speakers. Second, it allowed for the filling of gaps in sentences involving minimal pairs, used primarily for the sketch grammar (Chapter 4).

Presenting older speakers sentences uttered by young speakers and asking them for their judgements made it possible to detect areas of the grammar that appeared to show variation across age-groups. This method brought to my attention that the topic of reflexivity in possession, verb usage, and reduplication warranted further investigation. In Chapter 7 on reduplication, judgements by older speakers, in the form of corrections, were presented as converging evidence to illustrate that many of the forms used by younger speakers differed from those used by older speakers. This (along with actual utterances by older speakers) was taken to show that younger speakers were innovating the system (see §7.5 for more on this methodology).

This is illustrated in example (11a-b). (11a) presents an utterance by a 13-year-old female who uses a reduplicated form *hayok~ha-yok* 'RDP~3.PAT-shake'; (11b) presents an elicited sentence (a correction) of (11a) by a 40-year-old female, showing the use of a serial verb construction *ha-yok ha-tel* '3.PAT-shake 3.PAT-tie.together'.

(11) a. 13-year-old female

Di *de-raala* ***hayok~ha-yok*** *ba* *di*
3.AGT 3.REFL.AL-throat RDP~3.PAT-shake LNK 3.AGT
taa.
sleep.IPFV

Intended: 'He shakes his neck back and forth and then sleeps. [SS.13F.55]

b. Correction of (11a) by a 40-year-old female

Di *da-wata=ng* ***ha-yok*** ***ha-tel.***
3.AGT 3.REFL.INAL-neck=ALL 3.PAT-shake 3.PAT-tie.together

'He shakes his neck back and forth.' [FN.40F]

These 'corrections' were elicited in the field, while going over a recorded text with an older speaker. The elicited utterances were often noted down in the notes tier of the ELAN file, which accompanied a recording.

The second purpose of elicitation was to test hypotheses and fill in gaps for the sketch grammar. This was done in the field by using fieldnotebooks. After I had returned from the field, digital media was also used extensively to gather further elicited data. Skype calls facilitated elicitation. The most salient medium used was Facebook, a platform that has become increasingly more widespread in the elicitation of linguistic data. Indonesians are generally active users of Facebook and this includes the Abui as well. Certain groups on Facebook, such as *Abui Tanga Hetahai* 'Learning Abui', proved to be useful settings to elicit grammaticality judgments in addition to gathering translations of certain Abui phrases (see also §1.5.5 for more information on digital media use). Furthermore, some utterances were elicited using WhatsApp, which allows for instant voice messages to be sent out back and forth. These methods are not without pitfalls, however, with the main concern being the fact that the researcher is not only eliciting data (which in itself is out of context) but he is also far removed from the actual speakers. For these reasons, where possible, data elicited using these media was kept to a minimum, such as when illustrating example sentences involving minimal pairs.

To sum up, the elicited data used in this research comprises: translations, back translations, and judgements. Translations and back translations are scattered across notebooks, and conversations among digital me-

dia (Facebook, Whatsapp). Judgement data is all recorded in the notes tier of the ELAN files accompanying the recordings. In total, it is estimated that there are at least 300 elicited phrases and/or clauses scattered across these categories.

3.5.3.2 Data from participant observation

Having spent over six months in the village of Takalelang over the span of three years, I developed a level of Abui proficiency that was sufficient for basic conversations. Abui speakers constantly made active efforts to speak to me in Abui. As such, about 10% of fieldnote data used in this thesis, come from active participation in the community. The type of participant observation can be characterized as a combination of *passive participation* and *complete participation* (Duranti, 1997). When recording conversational data, I engaged mostly in *passive participation*, trying to be as unintrusive as possible, and walking away from the scene, such that speakers could converse more naturally. During day-to-day affairs, I tried to engage as much as possible in *complete participation*, attempting to speak Abui whenever appropriate.

As such, some of the examples tagged as fieldnote data comes from my own intuitions about the language, which stem from conversational phrases used in everyday interactions. In addition, some examples also come from phrases that have been overheard and jotted down. While participant observation is an essential method for conducting fieldwork, it may be less reliable than audio and video recordings because there is no raw data to fall back upon. Examples derived from participant observation were generally kept to a minimum.

3.5.4 Other

In addition to conversational, experimental, and data from fieldnotes, there were a few other miscellaneous recordings that were used as sources of data in this thesis. These include a short narrative about the story of Ateng Ateng, a recording of a video elicitation set, called the Put Stimuli (Bowerman, Gullberg, Majid, & Narasimhan, 2004),[12] and a word list of around 600 items

[12]The Put stimuli can be downloaded from the following url: http://fieldmanuals.mpi.nl/volumes/2004/put-project/.

(Saad, 2019b). The narrative and Put Stimuli were collected in the early phases of the research with the simple aim of collecting as many types of data as possible. The word list was collected for the Lexirumah database (Kaiping et al., 2019).

Figure 3.3: Collection of word list with three Abui elders

3.6 The corpus

This section provides an overview of all the recordings belonging to the categories discussed in §3.3 and §3.5.[13] Section 3.6.1 provides details on the recordings as organized by type of data, while §3.6.2 provides a breakdown of the four age-groups and the type of data collected for each age-group and speaker.

[13]The corpus may be accessed via the following link:
https://archive.mpi.nl/islandora/object/lat%3A1839_0ae02515_e62a_47f4_835d_fb1b4bd43f23

3.6.1 Organized by type of data

A breakdown of the corpus of sociolinguistic and ethnographic interviews is presented in Table 3.15.

Table 3.15: Breakdown of corpus of interview data

Type	Code	Subtype	N recordings	Total Duration	Age-range of speakers
Sociolinguistic interviews	[SL]	-	33	∼25:00:00	9-75
Ethnographic interviews	[EG]	Age-group	5	04:06:00	40-65
		School history	5	09:43:00	48-56

Note: Duration corresponds to HH:MM:SS

A breakdown of the linguistic data is presented in Table 3.16.[14] All of the types of data, except for 'fieldnotes', have recordings and transcripts, which have been archived and can be freely accessed.

[14] This duration includes the total number of transcribed and translated text and not the total duration of the actual recording. Thus, the total duration of transcribed and translated conversational text is 05:01:03. There are at least double the amount of untranscribed and untranslated conversations.

Table 3.16: Breakdown of corpus of linguistic data

Type	Code	Subtype	N recordings	Total duration	age-range of speakers
Conversational	[CV]	-	19	05:01:03	1-85
Experimental	[SS]	Surrey Stimuli Production task	66	∼22:00:00	9-75
	[FC]	forced-choice Comprehension task	60	∼15:00:00	9-75
Fieldnotes	[FN]	Elicited data	∼300 phrases	-	23-75
		Data from participant observation	∼30 phrases	-	1-85
Other	[NR]	Narrative	1	00:01:25	67
	[PT]	Put stimuli	1	00:22:46	43
	[WL]	Wordlist	1 (∼600 words)	02:25:05	40

Note: Duration corresponds to HH:MM:SS

Every linguistic example, excerpt, or claim extracted from my corpus and presented in this thesis contains a code which offers information on a) the genre of the recording, b) the age and gender of the speaker, and c) the unique identifier of the recording (see Tables 3.15 & 3.16 for abbreviations). For instance, example (12a) is cited as [CV.50M.QU]. The CV code implies the genre of the recording is 'conversational', the 50M implies the speaker is a 50-year-old male, and the QU corresponds to the unique identifier of the recording which links to a recording of a man asking his mother 'QUestions' about her past.

[SS.59F.33]

Example (12b) is taken from a response to the Surrey Stimuli elicitation task and thus has the following citation [SS.59F.33]. The SS corresponds to 'Surrey Stimuli'. The 59F denotes that the speaker is a 59-year-old female, and the 33 points to her being coded as the 33rd participant of the Surrey Stimuli task. Example (12c) is taken from my fieldnotes (either through elicitation or participant observation) and is thus given the code FN.[15] Ex-

[15]Typically, when an example was elicited from a speaker, a code for age and gender is

amples from the recordings labeled 'other' each contain the genre code relating to their specific source recording (e.g. NR 'narrative').

(12) a. *maayol*
'woman' [CV.50M.QU]

b. *neeng*
'man' [SS.59F.33]

c. *kaai*
'dog' [FN]

All examples, excerpts, and claims taken from recordings contain a link to the video/audio files in the archive. Clicking on any code will take the reader directly to the session. The only exception is for conversational recordings. Clicking a conversational code (CV) will take the reader to Table 3.9 which contains a list of all the conversational recordings. The list itself contains links to the recordings in the archive. It must be noted that the codes issued in this thesis differ slightly from the file names used for the actual recordings. Furthermore, data from Abui wordlists, which is stored in the LexiRumah database (Kaiping et al., 2019) is cited using references that take the reader to the database (e.g. Saad, 2019a, 2019b, 2019c).

In addition, I collected a small corpus of Alor Malay, with five Surrey Stimuli tasks and one conversation, as presented in Table 3.17. All but one of these recordings were conducted with Abui-Alor Malay bilinguals; one recording was done with an Adang-Alor Malay speaker. In total, seven speakers were sampled. This was used for the Alor Malay examples used throughout this thesis as well as to have a baseline of the contact variety in question.

Table 3.17: Breakdown of Alor Malay corpus of linguistic data

Type	Code	Subtype	N recordings	Total Duration	age-range of speakers
Conversational	[CV.AM]	-	1	00:20:17	23-27
Experimental	[SS.AM]	Surrey Stimuli	5	∼01:40:00	15-40

given. If the example is taken from participant observation, no code for age and gender is given.

3.6.2 Organized by speaker

Tables 3.18-3.21 list all the speakers from the four age-groups group who participated in the research for this thesis, along with the tasks that they performed.[16]

Table 3.18: Speaker and recording information for *moqu* '(pre)adolescents' group

Speaker code	Age	G	Surrey Stimuli	Forced-choice	Conversational	Sociolinguistic questionnaire
Ma	9	F	✓	✓		✓
La	10	M	✓			✓
Lo	11	M	✓			✓
Ma	10	F	✓	✓	✓	✓
Od	12	F			✓	
Je	12	M	✓			✓
Es	12	F	✓	✓	✓	✓
La	13	M	✓	✓		✓
At	13	M	✓	✓	✓	✓
Le	15	M	✓	✓	✓	✓
Te	14	F	✓			✓
Ri	14	M	✓	✓		✓
Ap	14	F	✓	✓	✓	✓
Ar	14	F	✓	✓		✓
At	15	M		✓		
Ro	15	M	✓	✓		✓
Ma	15	F	✓	✓		✓
Ri	15	F	✓	✓	✓	✓
As	15	M	✓	✓		✓
Na	16	M	✓	✓		✓
Ag	16	F	✓	✓	✓	✓
Ri	16	F	✓	✓		✓

[16] A checkmark under sociolinguistic questionnaire implies that the basic amount of sociolinguistic data has been filled in for a given speaker. This includes: age, gender, place of birth, information on siblings, information on (grand)parents, and early language exposure.

Table 3.19: Speaker and recording information for *neeng abet/maayol maak* 'young adults' group

Speaker code	**Age**	**G**	**Surrey Stimuli**	**Forced-choice**	**Conversational**	**Sociolinguistic questionnaire**
Ar	17	M	✓	✓		✓
Ti	17	F	✓			✓
Ri	18	M	✓			✓
Ti	19	M	✓	✓	✓	✓
Na	18	F	✓			✓
De	19	F	✓			✓
Ma	20	F	✓	✓		✓
Vi	20	F	✓	✓		✓
We	21	F			✓	
Lu	21	F			✓	
An	22	F	✓	✓	✓	✓
Be	22	M	✓	✓		✓
Ad	22	F	✓	✓	✓	✓
An	23	M	✓	✓	✓	✓
Si	23	M	✓	✓	✓	✓
Di	23	F	✓	✓		✓
Yu	24	M	✓	✓	✓	✓
Vi	24	M	✓	✓		✓
Ya	25	M	✓	✓		✓
Lo	25	M	✓	✓		✓
Sa	25	F	✓	✓		✓

Table 3.20: Speaker and recording information for *kalieta* 'adults' group

Speaker code	**Age**	**G**	**Surrey Stimuli**	**Forced-choice**	**Conversational**	**Sociolinguistic questionnaire**
Yo	26	M	✓	✓		✓
Sa	26	F	✓	✓	✓	✓
Fr	26	M	✓	✓		✓
Ko	27	F	✓	✓		✓
Fi	27	M	✓	✓		✓
Ya	28	F	✓	✓		✓
Le	28	F			✓	
Bo	30	M	✓	✓		✓
Pe	30	M	✓	✓		✓
An	30	M	✓	✓		✓
Lo	30	F	✓	✓		✓
Me	31	M	✓	✓		✓
Ma	31	F	✓	✓		✓
Lu	32	F	✓	✓		✓
Hi	32	F	✓		✓	✓
Be	32	M	✓	✓	✓	✓
Do	33	F	✓	✓		✓
Ar	32	M	✓	✓		✓
An	34	F	✓	✓	✓	✓
Je	34	M	✓		✓	✓

Table 3.21: Speaker and recording information for *kalieta* 'elders' group

Speaker code	Age	G	Surrey Stimuli	Forced-choice	Conv	Socioling	age-group discussions	School history interview
Sa	40	F			✓			
Lo	40	F	✓	✓		✓		
Do	40	F	✓	✓		✓		
Or	43	F	✓	✓		✓		
La	43	M	✓	✓		✓		
Su	40s	F			✓			
Ma	40s	M					✓	
Ye	45	M					✓	
Ar	46	M	✓	✓		✓		✓
Ni	47	F			✓		✓	
Si	48	M						✓
Ju	49	F					✓	
Ju	50	M						✓
Ab	50	M	✓	✓	✓	✓		
Ma	50?	M			✓			
Ag	50+	F		✓				
Ma	53	M						✓
So	55	F					✓	
Ar	56	M						✓
Da	56	M	✓	✓		✓		
Ka	57	M					✓	
Ni	59	F	✓	✓		✓		
Ma	60	M					✓	
Ay	65	M					✓	
Ne	70	M		✓				
Am	70+	F				✓		
Ni	70+	F			✓			
Lo	75	F	✓	✓		✓		
Lo	80+	F			✓			

Note: Conv: conversation, Socioling: sociolinguistic questionnaire

3.7 Kratochvíl corpus

In addition to the corpus collected by the author, another corpus which was used for the research in this thesis is the Kratochvíl corpus. The Kra-

tochvíl corpus, generously made available by František Kratochvíl, is the largest Abui corpus available. It consists mainly of Abui data collected in the years 2003-2016 in Takalelang as well as Tifol Afeng, Mainang, and some other villages. The corpus was made available using the software SIL Toolbox and consists of text files. The genres include conversations, narratives, elicitation stimuli responses, elicitation, text messages, as well as transcripts from Nicolspeyer (1940). The corpus has data from speakers who are Abui L1 speakers, with the exception of data from speakers who were working directly with Kratochvíl on transcriptions. The names of the speakers were always available meaning that it was possible to also derive their age. Many examples in Chapters 4 and 5 are borrowed from the Kratochvíl corpus and are cited generically as follows: (Kratochvíl corpus).

3.8 Data analysis

The data used in this thesis went through several analytical stages before it could be used for Chapters 1 to 7 of this thesis. What follows is an explanation of these stages, from a general level of how the recordings were analyzed using software such as ELAN and FLeX to the specific level of how each of the chapters implemented certain types of analyses.

One of the first levels of analysis involved transcribing and translating the video and audio recordings. This was done using ELAN, version 5.2 (Sloetjes, 2018). The advantage of using ELAN is that it allows for simultaneous playback of video and transcription. Transcriptions are also time aligned. Every recording corresponds to one ELAN file, meaning that there is no centralized database where all the data can be browsed at once. However, it is possible to conduct concordance searches across all ELAN files, a feature that I used extensively.

As for the ethnographic and sociolinguistic interviews, I transcribed most of them on my own. This was possible because the language being used was Malay Indonesian, which I felt comfortable transcribing on my own. Plus, I myself had been participating in the interviews so I was aware of the content. An Excel sheet database containing broad questions pertaining to the ethnographic interviews was created whereby all the answers could be presented in a centralized fashion. This was also the case for the sociolinguistic interviews, which contained data on the majority of speakers recor-

ded in the corpus.

As far as the Abui linguistic data recordings are concerned, every recording which was transcribed and translated was done so under the guidance of Abui native speakers, aged between 22 and 50 years. On several occasions, Abui speakers who were trained in the use of ELAN transcribed and translated recordings on their own and sent them to me. This was done to maximize the amount of transcription and translation. Native speakers assisting in transcription and translation were always instructed to also flag utterances that appeared ungrammatical or unfelicitous and propose modified sentences in a separate tier in ELAN (see §3.2.1 for more information on the team of Abui speakers that assisted in transcription and translation).

The 66 Surrey Stimuli recordings were also imported into the programme FieldWorks SIL (FLEx), version 8.2. The advantage of using FLEx is that it offers a centralized platform where all texts can be viewed at once. The disadvantages of using it are that it is not possible to link the texts to multimedia files and that it is most suitable for texts involving one speaker only. This means that it must be used in tandem with ELAN.

For the three chapters on language variation and change, Chapters 5, 6, 7, three excel databases were created, documenting the linguistic variable in question.

The next paragraphs describe how each type of data was used in the various chapters. They are intended to be used by a reader who has preferably already read the chapters and is seeking further information on the process of data analysis. In Chapters 1 and 2, both the sociolinguistic and ethnographic interview data was used to describe the sociolinguistic setting and history. For the current chapter, the ethnographic interviews focussing on age categories were used to establish the four age-groups discussed in §3.4. For Chapter 4: A sketch grammar of Abui, the Flex corpus, the ELAN files of all of the linguistic data, and fieldnotes were used. For Chapter 5: Variation and change in the reflexive possessive, Chapter 6: Variation and change in verb usage, and Chapter 7: Variation and change in reduplication, Surrey Stimuli data from all 66 participants was copied into an Excel file. Each speaker was listed chronologically based on their age. In addition, every one of the 40 Surrey Stimuli responses was listed as a separate column. This proved to be the most visually pleasing way of maintaining a central database, where I could browse variation for every speaker.

For Chapter 5: Variation and change in the reflexive possessive, all of

the nouns were coded for as either being possessed or unpossessed. Further, they were then coded based on which possessive was used, indicating whether the possessive was used in a standard way or not, for example, whether a reflexive possessive prefix was used in a reflexive possessive environment. A total of fourteen coding options were implemented. However, for the actual analysis, only 4 were applied. This data was then entered into SPSS version 25 and a Kruskall-Wallis test was performed. In addition, the conversational data was used to gather frequency counts on the desired constructions. Finally, the forced-choice data was coded for the reflexive possessive in SPSS version 25; a Chi-squared test was used to analyze the data.

For Chapter 6: Variation and change in verb usage, the clips containing the relevant events were isolated and then coded for based on the verb being used. The visual perception event domain contained five clips, the falling event domain contained five clips, while the change of state event domain contained three clips. These were then coded for based on which verb the speaker used to describe the event. This data was then entered into SPSS version 25 and a Kruskall-Wallis test was performed on the production data. The Kratochvil corpus was used to gather frequency counts on the desired constructions. As for comprehension data, the forced-choice data was coded for the selected verbs in SPSS version 25 and then a Chi-squared test was used to analyze the data.

For Chapter 7: Variation and change in reduplication, all instances of reduplication in my corpus were selected and then added to an excel sheet. Each utterance was then coded for age of speaker and type of reduplication. In addition, translations were given in Alor Malay along with any judgements and corrections that older speakers had with regards to the original utterance. The same was done for a database of parallel serializations as well as for Alor Malay reduplications.

3.9 Summary and conclusion

This chapter has described and motivated the methods used in the research for this thesis. It covered the two broad types of data: sociolinguistic and ethnographic interview data as well as linguistic data. It discussed how these interviews were carried out and how they were crucial in establishing a sketch of the social setting and history of the speech community. It also presented

the four age-groups that were constructed based on the interview data. Four features of each group were discussed: their age-range, their life-stage, their early exposure, and their current usage.

Furthermore, the various types of linguistic data were elaborated upon: conversational, experimental, and fieldnotes. Experimental data included both a production and comprehension task, while fieldnotes included elicited data and data from participant observation. These various types of data were then summarized in a description of the corpus collected over a span of three years. Finally, a discussion was given on how the data was analyzed in addition to how it fit into each chapter of this thesis.