

Chapter 7

Documenting Multilingualism in Southwest Sumatra

Bradley McDonnell

Over the past decade, documentary linguistics has seen an increase in the number of studies that both call for and introduce methodologies from sociolinguistics and anthropology to document languages in multilingual contexts (Lüpke 2010, 2013; Childs, Good, and Mitchell 2014; Di Carlo 2016; Hildebrandt, Jany, and Silva 2017; Good 2018). In particular, these studies draw on (linguistic, anthropological, and sociolinguistic) research of language contact, variation, and change (see Meyerhoff and Nagy 2008; Ravindranath 2015; Lim and Ansaldo 2016 for overviews of the extensive literature on these topics) to provide insights into ways of documenting the linguistic practices of multilingual (and multilectal) speech communities. **1**

They have largely been critical of commonly employed approaches to language documentation wherein the lexicogrammatical code of a single language is privileged for the purposes of documentation and description. However, as Lüpke (2010) and others have pointed out, the majority of languages that are the target of documentation projects are spoken in multilingual contexts (see also Vaughan and Singer 2018 and other articles in the same volume for discussion of Indigenous multilingualisms). To better understand how documentary linguistics can document such contexts, Di Carlo (2016) sets up a cline of different approaches to language documentation. At one extreme is the “ancestral code” approach, which, as mentioned above, “filters” out a single language in the interest of producing a grammar and dictionary.

Documentations created with this approach allow linguists to produce such descriptions more easily. At the other extreme is the “phenomenological” approach, which focuses “on the observable communicative practices of a given community” and in which the documenter does not develop an *a priori* plan to document the language (74). Di Carlo (2016) shows how the

former approach can be reductionist because the researcher determines the composition of the corpus, focusing on a single language (see also Dobrin and Berson 2011; Childs, Good, and Mitchell 2014), while the latter approach “promotes the observation of the linguistic practices as they surface in naturally-occurring communicative events” (74). He goes on to provide practical advice for implementing a phenomenological approach, to which I return in the following two sections.

Good (2018), in an article that reflects on the scope of language documentation, divides approaches to language documentation somewhat differently. He observes that the majority of current language documentation projects are oriented towards documenting speech events, which he calls an “event-based” approach. As in Di Carlo’s ancestral code approach, the event-based approach also tends to focus on a single language and often suffers the limitations of a predetermined plan. Good contrasts this approach with a “repertoire-based” approach, which targets the “linguistic behavior and knowledge of individuals” rather than events (17). In Good’s view, orienting documentation to individuals allows us to document both the lexicogrammatical code and what he calls social entailments—that is, “language ideologies, linguistic ecologies, and the sociolinguistic lives of speakers” (17). A repertoire-based approach better captures the complexities of interacting codes and their social entailments in multilingual contexts. In the end, Di Carlo’s and Good’s approaches are similar (not surprising given their collaboration over a number of years): language documentations that are not based on predetermined outcomes and focus not only on the lexicogrammatical code but also on social entailments are better at fully capturing the linguistic practices of speech communities, especially in multilingual contexts. They both warn that the consequences of following the well-trodden path of the event-based/ancestral code approach can lead to many missed opportunities when aiming to document languages, broadly construed to include their social entailments.

The advantages of implementing them within an Indonesian context on the surface seems promising:

multilingualism has long been the norm in Indonesia. And this trend has continued with the rise of Standard and other colloquial varieties of Indonesian (Musgrave 2014; Goebel 2018<AU: **Goebel 2018a or b?** I don’t have

access to the bibliography, so I am not sure if it is a or b. In my submission, I cited the following paper:

2018. “Language Diversity and Language Change in Indonesia.” In *Routledge Handbook of Contemporary Indonesia*, edited by Robert W. Hefner, 378–89. London: Routledge. <https://doi.org/10.4324/9781315628837-31>.>).

In fact, Musgrave (2014) states that while there is some evidence of linguistic homogenization, there does not appear to be a reduction of languages or registers. Rather, according to Musgrave, we are seeing changing patterns of multilingualism throughout Indonesia. This finding is supported by a large-scale survey by Abtahian and Cohn (2018), who conclude “that the linguistic situation in Indonesia is better viewed as changing patterns of multilingualism, and not necessarily a loss of the fluid and rich linguistic repertoire of language varieties and styles available to individual speakers and communities” (106). However, multilingual situations in Indonesia are diverse and represent different levels of complexity between interacting languages, lects, and registers. For example, Goebel (2005) shows how Javanese speakers in an urban setting in central Java shift between different registers of Javanese and Indonesian based upon various social factors, while Himmelmann (2010) shows how in certain Tomini-Tolitoli villages the speech community is shifting away from the Tomini-Tolitoli language to a wider use of Indonesian and Bugis in the public domain (see also Errington 1998; Grimes 2010a; Musgrave 2014).

Despite a significant body of research on multilingualism in Indonesia, there has been little discussion of documenting the linguistic practices of multilingual speech communities. The majority of documentation projects in Indonesia isolate languages that are the intended target of documentation, even when a project seeks to document more than one language (see Riesberg, chap. 1, this volume). In response to the gap in the literature on the intersection of multilingualism and language documentation in Indonesia, this chapter presents the challenges and prospects of documenting multilingualism in Sumatra, specifically in two rural communities in southwest Sumatra: a Besemah speech community in the highlands of South Sumatra province and the Nasal speech community in the southernmost reaches of Bengkulu province. While these two projects are located within a half-day car ride from one another, they differ along a number of different dimensions. First, the language ecology of each speech community is significantly different: the Besemah speech community is multilectal and marked by polyglossia while the Nasal speech community is clearly multilingual. Second, the Besemah speech community

represents a fairly stable multilectalism with a relatively small number of non-Besemah-speaking community members living in many Besemah villages, but in the Nasal speech community, multilingualism appears to be far less stable. Because the number of Nasal speakers is much smaller, they acquire or learn other Malay isolects, and while it is likely that children with at least one Nasal parent will learn Nasal, there are still some that do not. Third, the Besemah documentation project has been ongoing for over a decade, while the Nasal documentation project only began in earnest in 2019. Thus, the aim of this chapter is: (i) to look back at the Besemah documentation project, one that started when I was an inexperienced graduate student and primarily focused on documenting speech events that use the ancestral code, and (ii) to look ahead to the Nasal documentation project, one that aims to document the linguistic practices of the Nasal speech community by drawing heavily on a phenomenological/repertoire-based approach. In the course of discussing these projects, I show that incorporating elements from both ancestral-code/event-based and phenomenological/repertoire-based approaches is or was in some cases necessary and in others beneficial.

1. Looking Back at Documenting the Linguistic Practices of the Besemah Speech Community

<AU: CUP prefers to avoid subtitles in subheadings. Can you revise and shorten this, please?>

I initiated the Besemah documentation project as a graduate student in 2008 in the Besemah highlands of southwest Sumatra with the help of Atma Jaya Catholic University of Indonesia in Jakarta, Sriwijaya University in Palembang, the Padang Field Station of the Max Plank Institute for Evolutionary Anthropology, and several members of the Besemah community in the village of Karang Tanding. The goals of this project were to document Besemah, an underdescribed dialect of South Barisan Malay spoken by some 330,000 people.² From the outset, the focus of this project was to document the lexicogrammatical code of Besemah (i.e., the ancestral code)

and to understand how Besemah differs structurally from other Malayic languages of Sumatra and elsewhere. This project came at a time when there was a flurry of research happening on Malayic languages of Sumatra, led primarily by David Gil and Uri Tadmor at the Max Planck Institute for Evolutionary Anthropology as well as by Peter Cole and Gabriella Hermon and their then PhD students Yanti and Timothy McKinnon at the University of Delaware.

The vast majority of the documentation took place in a single Besemah village, Karang Tanding, located approximately 15 kilometers from the town of Pagaralam. I took up residence there with my partner and later our daughter for a total of two years from 2008 to 2010 and then another year from 2014 to 2015. During these three years, we participated in daily life in the village and took part in weddings, funerals, naming ceremonies, and holidays like Idul Fitri and Independence Day. We learned to speak Besemah, and we were embedded into the fabric of the Besemah community in Karang Tanding. The documentation of Besemah centered on recordings of narratives, conversations, elicitations of several areas of the phonology and morphosyntax of the language, and recordings of transcription sessions. Over the years, the documentation progressed from simple audio recordings to more complex video and audio recordings. And, as I began to gain skills and experience in documentary linguistics, the documentation of Besemah improved through better quality recordings, metadata, and transcriptions and translations (see McDonnell forthcoming). I worked closely with a handful of people on the transcriptions and eventually provided training to one of them, Hendi Feriza, in Discourse Transcription, translation, and ELAN software (Wittenburg et al. 2006; *ELAN* (version 6.0) 2020).

The Besemah speech community in Karang Tanding and the surrounding villages is multilectal and best characterized as polyglossic. While Pagaralam, the closest market town to Karang Tanding, is seen by many as the cultural center of the Besemah people (W. Collins 1979), it is inhabited by a large number of non-Besemah residents, most of whom use an emerging variety of Malay that features many characteristics of Palembang Malay, Besemah, and even elements of colloquial (Jakarta) Indonesian. People from Karang Tanding commonly go to Pagaralam to sell produce from their farms, buy goods, go to the hospital, or attend festivals and

other large gatherings. Thus, most of them speak some variety of Palembang Malay. Furthermore, when interacting with people from other areas of South Sumatra province, people from Karang Tanding also use Palembang Malay as a means of interethnic communication. There are two exceptions to this pattern. First, people from Karang Tanding spoke Besemah when interacting with others who speak closely related dialects of South Barisan Malay, such as Lintang or Semende. Second, when they interacted with people from Java or other islands, they tended to use a colloquial variety of Indonesian. Weddings, funerals, and Friday sermons at the mosque were mostly conducted in some form of Standard Indonesian. **<AU: is this correct? I'm trying to add some language in here to explain the shifts between present and past. If this is still the case, it might be best to move all of this (after the two exceptions) into present tense** This section is describing general patterns of language use, so I think it could all be kept in the present tense, and the following sentences could also be in present tense. I'd prefer not to use phrases with "my research" in this paragraph.> In only a few instances did anyone speak Besemah during these speeches, and these typically were short asides. Speeches tend to be very formulaic, using the same overall structure and set phrases from Standard Indonesian. Children also learned Standard Indonesian in schools. In many cases, teachers use Besemah in schools, but in some cases teachers are brought in from other provinces in Indonesia and do not speak Besemah.

The polyglossic nature of the Besemah speech community—and the challenges this presents—became apparent early on in the documentation project when I faced difficulties recording simple elicited narratives; the person telling the story would typically use a different Besemah than what I was hearing when the recorders were off. There was a clear "microphone effect," which made sense, given the fact that during formal events Standard Indonesian was often used with a PA system and microphone. Microphones simply represented a higher level of formality. This forced me early on to focus on recording everyday conversations. In some cases, I would come across some people sitting on their porch or in their front yard having a conversation, and I would ask to record them. In other cases, I would ask individuals to make a

recording some time in the future. Typically, there was a point person who would gather a few friends and family together, and they would record at one of their homes or in a hut in their rice paddy. However, based on advice from consultants with whom I worked closely, I was not present for any of these recordings. My presence would have had an effect on the recording. I set up the equipment and came back after an hour or so to gather it.

In these conversations, I discovered a number of lexicogrammatical patterns that have been otherwise unattested in other Malayic languages (see McDonnell 2016), such as the nature of Generalized Noun-Modifying Clause Constructions (McDonnell 2020) and universal quantification (McDonnell 2022). After living in the village for more than a year, I also came across a genre of storytelling called *andai-andai*. This genre of fairytales was known primarily by women over fifty years old and told to children at night. (Apparently, the custom was for parents and grandparents to exchange a massage from their children or grandchildren for a story.) However, *andai-andai* became less and less common. Today, people under fifty years old cannot recite these stories, and the younger generation has never heard one. Thus, I also began to document *andai-andai* from a handful of speakers who felt confident enough to perform them. From early on in the documentation, I elicited lexical and grammatical information compiled in a Fieldworks Language Explorer (FLEX) database (SIL International 2020). More recent documentation contains a prosody experiment (see McDonnell and Turnbull 2018) and targeted elicitation of grammatical phenomena.

Circling back to the approaches to documentation outlined by Di Carlo (2016) and Good (2018) in section 1 of this chapter, we see that the Besemah documentation described above would primarily fall under an ancestral-code/event-based approach. The documentation was organized around speech events, and I explicitly targeted as many different native speakers of Besemah (the ancestral code) as possible. And, as both Di Carlo and Good pointed out, these decisions led to a number of missed opportunities in the documentation of the linguistic practices of a Besemah speech community. I list the three most obvious missed opportunities below:

1. There is no documentation of weddings or other formal events. I avoided recording these events for several reasons, but the most salient ones are that these events involved many people and they were conducted entirely in Standard Indonesian. The prior is a logistical concern. Will any significant portion of the recording be able to be transcribed? Or, will the sea of people talking drown out any potential speech? When so many people are involved, how does the researcher obtain informed consent? Looking back, I see reasonable solutions to both issues, but at the time these were real concerns. The latter, however, reveals a clear ancestral code bias.
2. There is no documentation of Besemah speakers using Palembang Malay or talking to others who speak a dialect of South Barisan Malay. Again, I privileged Besemah as *the* object of study and simply did not document people from Karang Tanding visiting Pagaralam or talking to people who speak other dialects.
3. There are few records of individual Besemah speakers across different contexts. The documentation of Besemah has a diversity of Besemah speakers in the same general contexts (e.g., everyday conversations) but does not include much in the way of the same speakers across different contexts (e.g., speaking to non-Besemah friends in Pagaralam, giving a speech at a wedding, etc.), a hallmark of Good's repertoire-based approaches that focus on the sociolinguistic lives of speakers.

By not following these practices, I missed the opportunity to capture the full range of the linguistic repertoires of Besemah people. I am not necessarily proposing that the documentary record be chock full of recordings of Besemah people using Indonesian, but some representation in the corpus would give us a fuller picture of the linguistic practices of the Besemah speech community. It is, of course, not too late to collect these sorts of recordings and fill in gaps in the documentation of Besemah. Indeed, there are plans to do just this in the coming years. Admittedly, it would have been much better for me to document the full repertoire of some

Besemah speakers during these extended times of fieldwork. However, given the central role that Besemah plays in daily life, the Besemah conversations should make up the core and majority of the documentation. Including the use of Standard Indonesian and Palembang Malay by Besemah speakers should play a relatively minor role. Furthermore, while adding such documentation better demonstrates the use of different codes and the linguistic ecology of the speech community, it may not get at language ideologies. Thus, following Di Carlo (2016) and others, I believe that conducting ethnographic interviews would be an extremely valuable addition to the documentary record in order to better understand the language ideologies of Besemah speakers (see below for a description of sociolinguistic questionnaires in the Nasal documentation).

Looking back, then, there appear to be two major roadblocks to creating a fuller documentation of the linguistic practices of the Besemah speech community. The first roadblock, mentioned above, was a focus on the ancestral code that unnecessarily limited the scope of the documentation. The second issue is related and is one that many documentary linguists may struggle with: the documentation of linguistic repertoires is likely to be messy, uneven, and generally unpredictable. That is, when I recorded everyday conversations, I was able to look out for interfering factors, such as avoiding recording next to a noisy street or a location where a large crowd is gathered. However, documenting repertoires by recording a set of individuals in different environments is likely to result in some recordings occurring in noisy environments or where there is little conversation for an extended period of time. While I was prepared to deal with the messiness of conversational data between Besemah speakers, I was apparently not prepared to deal with the messiness of recording formal situations that involved large groups of people or recording the same speaker in diverse and unpredictable settings. These recordings would likely be full of noisy data that may or may not be useful, but I failed to see at the time that a portion of these recordings would likely contain not just usable but important information about the repertoires of Besemah speakers.

While the Besemah documentation project adopted the ancestral code/event-based approach, there are also numerous elements of Di Carlo's phenomenological approach in it. For

one, the focus on everyday conversation is certainly in line with Childs, Good, and Mitchell (2014), and my commitment to recording everyday conversations was not a decision made *a priori*. Rather, it was a response to the types of unnatural monologic recordings I was gathering in the beginning. Second, I chose to provide rich transcriptions in the documentation of Besemah conversations that contain information on context, language use, sociocultural factors, and language ideologies (see McDonnell forthcoming). These insights are part of the documentary record and combine the knowledge I gained during the transcription process with that gained from participant observation. Much of the documentation—even aspects of the lexicogrammatical code—came about during the course of my fieldwork. This includes the documentation of *andai-andai* fairytales and experiments on word prosody, which struck me as markedly different from other Malay languages I had heard. These aspects of the Besemah documentation project fall in line with the phenomenological approach, one that was generally responsive to communicative events I was observing while living with the community.

2. Looking Forward at Documenting the Linguistic Practices of the Nasal Speech Community

<AU: CUP prefers to avoid subtitles in subheadings. Can you revise and shorten this, please?>

The Nasal speech community differs significantly from the Besemah speech community described in the previous section. The community is decidedly multilingual. They speak the Nasal language, an underdocumented Sumatran language (Billings and McDonnell 2024), alongside one or more varieties of Malay in daily interactions. And, much like the members of the Besemah speech community, they use Standard Indonesian in formal ceremonies and varieties of colloquial Indonesian with some visitors and newcomers to the village. Nasal is spoken by an estimated three thousand people in four villages on the border of Bengkulu and Lampung provinces, making this community much smaller than both Besemah and the

surrounding speech communities of Kaur and Semende. Kaur, a Malay language that is closely related to Besemah, is the language most widely spoken in the region, with approximately forty thousand speakers. Semende, a dialect of South Barisan Malay and very closely related to Besemah, is spoken by fairly large numbers of people who have migrated from the highlands to the coast. The four villages in which Nasal is spoken represent a cross-section of people from these and other groups, and it is commonplace to hear Nasal, Kaur, and Semende spoken throughout these villages. In some cases, Kaur, Semende, and even Batak or Javanese people have taken up residence in these villages to live and work. In this case, people do not appear to speak or even attempt to learn Nasal. Instead, they continue to use Kaur or Semende, or if they come from Java or elsewhere in Sumatra, they use some variety of Indonesian or learn Kaur. In other cases, people from these same groups have married a Nasal person and moved to their spouse's village. In this case, very few spouses learn to speak Nasal but may have a passive understanding. However, Nasal speakers invariably speak at least Kaur and likely also Semende, apparently from very early ages. Thus, as Nasal people have commonly commented to me, to speak Nasal is to speak Kaur and Semende. That is, the multilingualism in the Nasal speech community is asymmetrical: Nasal speakers learn Kaur and/or Semende while Semende and Kaur speakers rarely learn Nasal. As in the Besemah speech community, weddings, funerals, and other formal events are conducted in Standard Indonesian, and children learn Standard Indonesian in schools. When traveling to market towns like Merpas or Bintuhan (the capital of the regency), Nasal people apparently use Kaur.

The documentation of the linguistic practices of the Nasal speech community was initiated in 2019 with a team of collaborators, including fellow linguists Yanti (Head, Center for Language and Culture Study, Atma Jaya Catholic University of Indonesia) and Jacob Hakim (PhD student, Department of Linguistics, University of Hawai'i at Mānoa) as well as three members from the community, Johan Safri, Anton Supriyadi, and Wawan Sahrozi. Other linguists joined the project after it was already underway: Blaine Billings (PhD student, Department of Linguistics, University of Hawai'i at Mānoa) joined in 2021, and Dendi Wijaya

(Researcher, National Research and Innovation Agency; BRIN) joined in October 2023. My interest in initiating the Nasal documentation project was due to a number of factors. First, Nasal had not been known to linguists until 2007 (Anderbeck and Aprilani 2013), and basic questions about its structure (phonology, morphology, syntax) and its relation to other Austronesian languages were largely unanswered. Second, given the precarious position of Nasal, the language is likely endangered. Thus, I wanted to document the language while it is still actively spoken with the hope of supporting language maintenance. Furthermore, I desired to engage in a documentation project that placed community goals and involvement at the center of the project (see McDonnell 2017 for a discussion of the prospects and challenges in reaching this goal). Finally, having developed expertise in this region, I was eager to build upon it in a new project. And given the multilingual speech practices in the Nasal speech community, I planned to document the complexities surrounding these practices by drawing on the approaches that document the lexicogrammatical code alongside the social entailments (Childs, Good, Mitchell 2014; Di Carlo 2016; Good 2018).

In order to meet these different research- and community-oriented goals, the project design includes four main components: (i) documenting everyday conversations, (ii) conducting interviews based on sociolinguistic questionnaires, (iii) documenting culturally important events and the lexicon for community outcomes, and (iv) eliciting aspects of the lexicogrammatical code for grammatical descriptions. As we will see below, for several reasons, we have not limited ourselves to a repertoire-based approach but incorporate elements of both event-based and repertoire-based approaches to meet both research and community goals. The Nasal documentation project began in earnest in October 2019 when Jacob Hakim and I visited the Nasal community working closely with Johan Safri, Wawan Sahrozi, and Anton Supriyadi. We also met with community leaders and began to record speech events and elicit basic phonological, lexical, and grammatical properties of Nasal and Kaur. Training and capacity building for the three team members of the Nasal community began with a training workshop in transcription, translation, and specialized language documentation software in February 2020 at

Atma Jaya Catholic University of Indonesia, and sustained support for this training has been maintained remotely and in person ever since. Many of the project activities continue today.

As in the case of Besemah, documenting everyday conversations has taken center stage in this project, but for Nasal it draws on two different methodologies for collection. The first follows the methods I developed in documenting Besemah conversation (see the previous section for details). It aims to bring together small groups of either Nasal or Kaur speakers to record conversations. As Nasal speakers will readily switch to Kaur or Semende even if a single person does not speak Nasal, we wanted to create a space where Nasal would likely be used. These conversations better ensure that we have documentation of Nasal. These documentations also relied on more complex recording setups that capture high-quality video using multiple cameras and high-quality audio using several audio recorders with headset microphones. In October 2019, we recorded a handful of conversations in both Nasal and Kaur. The second methodology more closely follows recommendations in Good's (2018) repertoire-based approach, which is based on well-established traditions in sociolinguistic research.³ Beginning in late 2020, Nasal collaborators on the project Johan Safri and Wawan Sahrozi were provided recording equipment to begin recording everyday interactions that they observed. Recording setups for these situations were minimal (i.e., a small video recorder and/or audio recorder) to ensure that they could easily record in as many different contexts as possible. This allowed them to record formal meetings at the village meeting hall, the types of speech events I did not record in the Besemah documentation project. As a part of this effort, we also expanded the metadata for each speech event and each participant in the events that we recorded, following recommendations by Childs, Good, and Mitchell (2014). This now includes basic speaker traits, linguistic "life histories," multilingual competencies and contexts of acquisition, basic event information, the context of how the event came about, and speaker relationships within the speech event.

In order to better understand language ideologies and the sociolinguistic lives of speakers, my collaborators and I have also carried out recorded interviews based on a sociolinguistic questionnaire conducted by Dendi Wijaya, which covers four primary areas:

1. Biographical details, linguistic life history
2. Language repertoire (lects one speaks, how well, and motivations for learning/using them)
3. Language attitudes, ideologies
4. Language acquisition, socialization

These four areas are based on questionnaires developed by Abtahian and Cohn (2018) and Di Carlo (2016). I, along with the other members of the team, am currently processing the data and plan to publish results of the interviews to gain a better perspective on linguistic ideologies, language use, and the linguistic ecology in the Nasal speech community.

In the beginning stages of the project, conversations with Nasal leaders revealed that there was no language work among members of the Nasal speech community. However, early discussions with community members and leaders have shown that there is interest in a multilingual dictionary of Nasal that at the very least contains Indonesian and English (see McDonnell 2017). Community leaders have also shown some interest in documentation of culturally important events, such as Sengkukho, wherein Nasal people dress in elaborate costumes made from woven palm leaves, sugar palm fibers, and the charcoal of burned palm branches and dance from one end of the village of Gedung Menung to the Nasal River. In order to meet these goals, Blaine Billings and I have worked closely with Johan Safri, Wawan Sahrozi, and other members of the Nasal community to produce the desired documentation. Billings is also leading efforts to document the lexicon. For the multilingual dictionary, aside from drawing on words from the transcribed conversations, he has also followed Mosel's (2004) "active elicitation" model by collecting words by semantic domain in small groups of three to four people. These small group sessions are being recorded and make up part of the documentary record of Nasal. In documenting Sengkukho, we have not only documented the event itself but

also preparations for the event. The final product will depend upon the community's input on how they would like to edit and present these recordings.

Finally, since Nasal and Kaur have received very little linguistic description, my project collaborators and I plan to produce various descriptions of these languages, including a reference grammar of Nasal and sketch grammars of both Kaur and Semende. These descriptions will be based on the conversations as well as elicitation of different areas of the grammar and will take into account issues of language use, language contact, and diachrony. One area that required specialized elicitation and description was Nasal prosody, which is detailed in Hakim (2024).

The documentation of the linguistic practices of the Nasal speech community involves elements from each of the approaches discussed above but in many ways follows the spirit of the phenomenological approach to document the linguistic practices of the community as they naturally emerge. There are certainly elements of event-based/ancestral-code approaches in this project. Recordings of conversations that ensure that participants use either Nasal or Kaur follow these approaches, as do the multilingual dictionary, recordings of culturally important events, and elicitation of the lexicogrammatical code for descriptions. The need to draw on event-based/ancestral-code approaches comes from (i) a lack of even basic documentation/description of Nasal and Kaur and (ii) the desire of the community to create resources that valorize Nasal, a language whose vitality is in question. As Di Carlo (2016) notes, it would not be impossible for a project to include elements of both approaches (91–96). And indeed, unlike the Besemah documentation project, the Nasal documentation project includes a number of elements based upon the repertoire-based approach<AU: I changed this because it seems like this sentence is a confirmation of the previous one (i.e., you have a project that uses both approaches) rather than a contrast (requiring “however”). Am I misunderstanding? This is correct. Thank you.>. Recordings of a limited number of individuals in different contexts as well as the use of sociolinguistic questionnaires aim to document social entailments and better understand the full scope of the linguistic practices of the Nasal speech community. Combining these approaches helps us to avoid the pitfalls of following the well-trodden paths of previous

documentation projects that only document the lexicogrammatical code of a single language while also providing needed descriptions and support for community goals. As Childs, Good, and Mitchell (2014) and Di Carlo (2016) maintain, this sort of approach is only possible with a team, which is true of the Nasal documentation project but not the Besemah documentation project. I would also add that experience allows documenters to try phenomenological/repertoire-based approaches that have far less structure than an ancestral code/event-based approach.

3. Conclusion

Since the reemergence of documentary linguistics around the turn of the twenty-first century, the field has seen a shift away from the perspective that language documentations are ancillary to language descriptions and towards one that values the creation, curation, and dissemination of multipurpose records of the linguistic practices of speech communities (Austin 2016; McDonnell, Berez-Kroeker, and Holton 2018). This conception of documentary linguistics has seen an increased focus on documenting language use within its sociocultural context, forging collaborations with the speech communities or even initiating community-based models, and developing interdisciplinary team-based approaches, all in an effort to create rich documentations that have the ability to serve multiple purposes now and in the future. As Good (2018) aptly points out, language documentation has been primarily oriented towards documenting events and in many cases is particularly good at capturing the use of the lexicogrammatical code of a target language. It is Good's contention, however, that such event-based documentations in many cases—and especially in multilingual contexts—fail to capture many aspects of language, including the social entailments. In response to Good (2018) and earlier proposals by Childs, Good, and Mitchell (2014) and Di Carlo (2016), this chapter has reflected on the documentation of the linguistic practices of the Besemah speech community. It has shown that while the project generally took a phenomenological approach, it still focused too heavily on the ancestral code by documenting events where only this code was used. I missed out

on opportunities to document the full linguistic repertoires of members of the Besemah speech community. Thus, there is currently a gap in the record of social entailments. In response to these missed opportunities and the multilingual context of the Nasal speech community, this chapter laid out how I, along with my collaborators, made a plan to better document multilingualism by incorporating elements of a repertoire-based approach, which includes documentation of individuals' linguistic practices in different contexts as well as sociolinguistic questionnaires, to better understand linguistic ideologies.

7. Documenting Multilingualism in Southwest Sumatra

I would like to thank Maya Abtahian, Joseph Errington, and an anonymous reviewer for comments on this paper. I would also like to thank people from the Nasal and Besemah villages, who took part in these documentation projects. I would especially like to thank Hendi Feriza, Johan Safri, Anton Supriyadi, and Wawan Sahrozi, with whom I have worked closely on these projects. I am grateful to Asfan Fikri Sanaf and Kencana Dewi for kindly allowing me stay at their home in Karang Tanding. I am also grateful to collaborators on the Nasal project, including Yanti, Anggy Denok Sukmawati, Jacob Hakim, Blaine Billings, and Dendi Wijaya, as well as the Language Institute at Sriwiaya University and the Center for Culture and Language Studies at Atma Jaya Catholic University of Indonesia for sponsoring this research. I am grateful to the Ministry of Research and Technology and the National Research and Innovation Agency (BRIN) in Indonesia for allowing me to conduct research in both Besemah and Nasal communities. This material is based upon work supported by the National Science Foundation under

Grant No. (1911641). Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author and do not necessarily reflect the views of the National Science Foundation.

1 I use multilingual(ism) for both multilingual(ism) and multilectal(ism). In cases where I mean to distinguish the two, I make the distinction explicit with the use of multilectal(ism).

2 South Barisan Malay has also been referred to as Central Malay or Middle Malay, which is translated from the Dutch *Midden-Maleisch*.

3 An anonymous reviewer points out that longitudinal recordings of a single speaker have long been a part of sociolinguistic research, including Labov's data collection of Hawai'i Pidgin in the 1970s (see Labov 1990) and Bickerton's data collection in Guyana (see Bickerton 1971). The same reviewer points out that this type of data collection is referred to as "freestyle" recording in a more recent dissertation by Daleszynska (2011), who discusses many of the challenges associated with these recordings (113–16).