

BALANCING LINGUISTIC GOALS AND COMMUNITY RELEVANCE: LESSONS FROM DOCUMENTING THE BADUY DALAM SUNDANESE LANGUAGE

Hernawan¹, Eri Kurniawan², Retty Isnendes³, Rd. Dian Dia-an Muniroh⁴,
Haris Santosa Nugraha⁵, Bradley McDonnell⁶, Michael Ewing⁷
*Universitas Pendidikan Indonesia^{1,2,3,4,5}, University of Hawai'i at Mānoa⁶,
University of Melbourne⁷*

hernawan@upi.edu¹; eri_kurniawan@upi.edu²; retty.isnendes@upi.edu³; ddmuniroh@upi.edu⁴;
harissantosa89@upi.edu⁵; mcdonn@hawaii.edu⁶; m.ewing@unimelb.edu.au⁷

Abstract

Language documentation plays a critical role in preserving linguistic diversity and cultural heritage. However, balancing the academic goals of linguists with the practical needs and interests of indigenous communities remains a challenge. This study reports on the first phase of documenting the relatively isolated indigenous language of Baduy Dalam in Banten, Indonesia. Using a linguist-centered approach, the fieldwork involved eliciting language data and recording daily narratives, culminating in the creation of a small Baduy dictionary. While the project succeeded in capturing targeted word forms and sentence structures, local informants expressed concerns about its relevance to their community. These findings highlight the limitations of linguist-driven documentation efforts, suggesting the need for approaches that prioritize community engagement and shared ownership of language preservation initiatives. A shift towards community-based and multidisciplinary methodologies may foster greater participation, ensuring that language documentation aligns with both academic and local community objectives.

Keywords: Baduy, elicitation, indigeneous community, linguist-focused approach, communitite-based approach

Abstrak

Dokumentasi bahasa sangat penting untuk melestarikan keanekaragaman bahasa dan warisan budaya, tetapi menyeimbangkan antara tujuan akademik para linguis dengan kebutuhan praktis dan kepentingan komunitas adat masih menjadi tantangan. Penelitian ini melaporkan tahap pertama penelitian kami tentang upaya dokumentasi bahasa Sunda Suku Baduy Dalam yang relatif terisolasi di Banten, Indonesia. Penelitian ini melibatkan pengumpulan data bahasa dan pencatatan narasi harian masyarakat Baduy Dalam dengan menggunakan pendekatan yang berfokus pada ahli bahasa. Dengan menggunakan pendekatan yang berpusat pada kepentingan para linguis, penelitian ini berhasil mengumpulkan daftar kata dan struktur kalimat yang ditargetkan, yang menghasilkan sebuah kamus kecil bahasa Baduy. Namun, para informan mempertanyakan relevansi proyek tersebut dengan masyarakat. Temuan kami menyoroti keterbatasan pendekatan yang berfokus pada ahli bahasa dan menyarankan perlunya pendekatan yang lebih melibatkan komunitas adat dan meningkatkan rasa kepemilikan proyek pelestarian bahasa. Pendekatan berbasis komunitas dan multidisiplin dapat meningkatkan partisipasi, serta memastikan bahwa dokumentasi bahasa dapat menyeimbangkan tujuan akaademik dengan kebutuhan komunitas adat.

Kata kunci: *Baduy, elisitasi, masyarakat adat, pendekatan yang berfokus pada ahli bahasa, pendekatan berbasis komunitas*

INTRODUCTION

Languages are an essential part of human culture and identity, and they play a critical role in shaping the way we perceive and interact with the world around us (Gelman & Roberts, 2017). However, many of the world's languages are at risk of disappearing, with estimates suggesting that up to half of the world's 7,000 languages could become extinct in the next century (Brohan, et al., 2022; Simons, 2019). Documenting endangered languages is therefore crucial in preserving linguistic diversity and cultural heritage (de Swaan, 2004; Romaine & Gorenflo, 2017).

Indonesia is one of the world's most linguistically diverse countries, with over 700 living languages spoken across its thousands of islands (Pepinsky et al., 2022). However, many of these languages are endangered, with some facing the threat of extinction within a generation (Bromham et al., 2022; Ewing, 2014). The rapid pace of development, urbanization, and globalization in Indonesia has contributed to the decline of many indigenous languages (Pepinsky et al., 2022), as young people increasingly adopt dominant languages such as Bahasa Indonesia and English. Among the numerous indigenous languages under threat is the Baduy Dalam dialect of Sundanese (language code[bac]), spoken in the Western part of the Java Island (Eberhard et al., 2024).

Having isolated themselves from greater Sundanese society over 500 years ago to avoid increasing foreign influences, the Baduy Dalam 'Inner Baduy' live an intentional and extremely ascetic lifestyle designed to maintain traditional culture and harmony with the natural world (Waluya et al., 2023). The Baduy speak a variety that Ethnologue classifies as Baduy (language code [bac]), a language distinct from Sundanese but which belongs to the Sundanese branch of Western Malayo-Polynesian, a member of the Austronesian family (Hidayati et al., 2017). In our observation, due to a rapidly shifting linguistic and social landscape, Baduy language and culture are quite fragile.

Heretofore, study of the Baduy has principally concentrated on their unique culture, focusing on general descriptions and social customs (Garna, 1989; Bakels, 1988; Makmur, 2001; Permana, 2001), religion (Wessing, 1977, 1988; Barendregt & Wessing, 2003), music (van Zanten, 1995, 2004), agricultural practices (Iskandar, 2007), architecture (Permana, 2006; Barendregt & Wessing, 2008; Marlina, 2009) and the origins of the people (Bakels & Boevink, 1988; Barendregt & Wessing, 2005). There has been relatively little study of their folklore (an oral tradition), the exception being *Beberapa cerita rakyat Baduy* 'Some Baduy folk tales', a largely Indonesian-language volume not widely available (Lembaga Sejarah and Antropologi, 1975).

Therefore, despite a fairly robust literature on cultural issues, the language of the Baduy has received scant attention. As just stated, their language is usually characterized as an archaic form of Sundanese. However, that characterization is not the result of analysis of the language because it has not received organized study and no description is readily available. The only source that has been discovered in our literature search is a brief work of slightly over 60 typescript pages that reports on a few phonological, morphological, and syntactic aspects of the language—*Perbandingan struktur bahasa Sunda Baduy dengan bahasa Sunda Lulugu* 'A comparison of the structure of Baduy with Lulugu', an in-house project report (Universitas Padjadjaran, 1984). This sketch grammar is nearly impossible to find in Indonesia and is only available on an extremely degraded microfiche in a few libraries in the US. Although Ethnologue (Lewis, Simons, & Fennig, 2015) characterizes the Baduy language as a distinct language that is

‘vigorous’, this assessment makes no linguistic distinction between the two groups of Baduy in Banten Province, the *Baduy Dalam* and the more populous *Baduy Luar* ‘Outer Baduy’. The Ethnologue assessment of the vitality of the language takes neither this important distinction nor the potential for language shift into account.

Based on the above descriptions, this paper aims to first outline the methodological approach of the linguistic fieldwork to document the Sundanese variety of Baduy Dalam in order to promote its cultural presence and highlight its importance in preserving linguistic diversity and cultural heritage. Second, the paper reflects on the linguist-focused model of documenting languages and questions whether the model is the right choice for indigenous communities, particularly for the Baduy Dalam communities.

LITERATURE REVIEW

The Baduy

The Baduy, who refer to themselves as the Kanekes (Baduy or Badui is a term derived from the word for *Bedouin*—a term to describe a Sundanese tribe unaffected by modernization or completely isolated from the outside civilization (Alam et al., 2023)), went into seclusion in the 16th century, rejecting increasing foreign influences in order to preserve its view of traditional Sundanese society and values. The Baduy Dalam and Baduy Luar, with a current population estimated at 12,000-13,000 (Kurnia & Sihabudin, 2010), live in a cluster of roughly 50 villages spread over 50 km² in a remote area in the Kendeng Mountains in Banten Province near the city of Rangkasbitung (population of 130,000)—roughly 125 km from the capital city of Jakarta and around 275 km from Bandung.

Traditionally, the Baduy practice Sunda Wiwitan, a belief system that emphasizes preserving harmony and balance with nature and veneration of ancestors. They attempt to preserve life in the manner in which they believe their forebears intended. According to Marlina (2009), there are two overriding principles which guide them. The first is a belief in a single god, Batara Tunggal, who is said to live in the Kendeng Mountains, at the Arca Domas, believed to be point of contact between the one God and the Earth, thus sacred ground forbidden to all but the most holy of humans. The second is to protect the relationship between humans and the natural environment, the purpose being reverence for nature and the preservation of a sustainable society.

As described by Iskandar (2007:115), among the traditional principles guiding Baduy life is “The hill should not be destroyed...The valley should not be damaged...What is long should not be shortened...What is short should not be lengthened...The ancestral injunctions may not be changed.” As a result, the Baduy people are forbidden to reshape the land, in addition to many other proscriptions, including prohibitions against wet rice cultivation (using only dry field cultivation), plowing the soil (planting has traditionally been done with bamboo poles), digging wells, using fertilizers or pesticides. The Baduy uses no inorganic materials in work, clothing or housing. To minimize the impact to the land, houses do not sit directly on the ground but are constructed on pillars that are not anchored in the earth but rest on rocks, which serve to protect the Earth and also level the platforms (Marlina, 2009). There are a large number of other dictates, including prohibitions against wearing sandals or shoes, taking any form of conveyance, using any type of modern technology (such as TV, cameras, computers, and cell phones), all designed to retain harmony with the forest and rivers, avoid violence and confrontation, and uphold traditional values.

In addition to these restrictions, certain activities are proscribed during much of the year—again designed to maintain the balance of nature and honor tradition. For example, hunting and fishing are limited to two months of the year, and are not allowed at other times. Also, marriages may only take place during June and July, the months after the harvest has been completed and the harvest ritual (the *seba*) performed (see Kurnia & Sihabudin, 2010).

The Baduy are divided into two principal groups: the Baduy Dalam and the Baduy Luar. The population of the Baduy Dalam community is approximately 1170 (Kurnia & Sihabudin, 2010) and live in *tangtu telu kampung*, the three villages of Cibeo, Cikertawana, and Cikeusik. The Baduy Dalam adhere strictly to the restrictions and the traditional way of life; they are considered protectors of the balance with nature, caring for the forests, rivers, and mountains so that humans may live in harmony with the Earth. They have historically had very little contact with outsiders. The Baduy Dalam wear usually white shirts and white head coverings (a sign of greater purity) and gray skirts. The Baduy Luar, identifiable by their black clothing and characteristic blue batik, live in some 48 surrounding villages and provide a buffer, protecting the sanctity of the Baduy Dalam from the influences of the outside world. (Each of the *tangtu telu kampung* has between 9 and 20 such outer villages.) The Baduy Luar adhere much less strictly to traditional culture; they do not eschew modern technology (electricity, television, cell phones), they wear shoes, and some own motorcycles or cars. They engage the wider world in limited ways, in many respects designed to help sustain the Baduy Dalam (see Kurnia & Sihabudin, 2010).

The current methods used for documenting endangered languages

The documentation of endangered languages involves the collection, analysis, and preservation of linguistic data (Austin & Sallabank, 1999; Rogers & Campbell, 2015; Seifart et al., 2018). One common method of documenting endangered languages is through the use of fieldwork, which involves linguists visiting communities where the language is spoken and collecting data through various methods such as interviews, recordings, and observations (see Haig, et al., 2011). The data collected through fieldwork can be used to create a grammar, dictionary, or corpus of the language, which can then be analyzed to better understand the linguistic structures and features of the language (Thieberger, 2011). Other methods of documenting endangered languages include using archival materials such as old recordings or written texts, which can provide valuable historical and cultural context for the language (Austin, 2021).

In recent years, digital technologies have played an increasingly important role in documenting endangered languages. Digital tools such as speech recognition software, transcription software, and online dictionaries have made it easier and more efficient to collect, analyze, and share linguistic data (Jones, 2014). Additionally, digital archives and repositories have become a popular way of preserving and sharing endangered languages with a wider audience (Thieberger, 2011). However, while these technologies have made language documentation more accessible, they also raise ethical concerns around issues such as intellectual property, informed consent, and community participation (Holton, 2011). As such, it is important for researchers to consider the ethical implications of their work when documenting endangered languages.

The linguist-focused model

The linguist-focused model is a method of documenting endangered languages, coined by Czaykowska-Higgins (2009). This model places emphasis on the linguistic structures and features of the language. This approach involves linguists working closely with speakers of the endangered language to collect data, which is then analyzed and documented in detail. A linguist-focused model of research is language research conducted by linguists, for linguists. In this model, the language-speaking community's participation is limited mostly to being the source of fluent speakers, and the level of engagement in the community by a linguist is relatively small. This means that the research is primarily aimed at advancing linguistic knowledge and theory, rather than addressing the needs or concerns of the language-speaking community.

The linguist-focused model has several advantages, according to Czaykowska-Higgins (2009), such as its ability to produce highly detailed and accurate linguistic descriptions of the language. This level of detail can be beneficial for language revitalization efforts, as it can provide a basis for language teaching materials and curriculum development. Additionally, the linguist-focused model can facilitate cross-linguistic comparisons, which can lead to a better understanding of linguistic diversity and the nature of language itself.

However, the linguist-focused model also has some disadvantages. One criticism of this approach is that it can be too focused on the linguistic structures of the language and may overlook other important aspects of language use and cultural context (Czaykowska-Higgins, 2009; Seifart et al., 2018). Additionally, the linguist-focused model can be time-consuming and resource-intensive, which can limit its applicability in certain situations (Czaykowska-Higgins, 2009; Seifart et al., 2018). Furthermore, there is a risk that this approach may not adequately involve the local community, which can lead to issues of ownership and representation of the language (Czaykowska-Higgins, 2009). As such, it is important for researchers to consider the advantages and disadvantages of the linguist-focused model when documenting endangered languages and to be mindful of the ethical considerations of their work (Czaykowska-Higgins, 2009).

Previous studies on documenting endangered languages and their methodologies

Previous studies on documenting endangered languages have focused on creating dictionaries, which present various challenges in satisfying both the interests of linguists and the needs of language communities. One such study was conducted by Mosel (2011), who worked on dictionary projects for endangered languages in Samoa and Bougainville, Papua New Guinea. In her article, Mosel discusses the challenges that arise during lexicographic work in short-term language documentation projects. These include cooperation with the speech community, selecting a dialect, and producing a useful piece of work that meets scientific standards while working with limited resources. She also highlights the differences between lexical databases that result from language documentation projects and the kind of dictionaries that minority speech communities require for educational purposes. Mosel shows how lexical databases can be used to create dictionaries that satisfy both the needs of native speakers and the interests of linguists.

Another similar dictionary-writing project was carried out collaboratively by Anderson, Barth, and Forepiso (2015), working on the Matukar Panau language in Papua New Guinea. They came up with the idea of creating an online dictionary through a community collaborative-based effort, which involved multiple local community speakers in the documentation process. They termed this approach collective elicitation, which allowed groups of speakers of varying

proficiency levels to sit together and contribute words and descriptions to the dictionary. Not only did they discuss the word forms, but they also negotiated the most appropriate definitions or descriptions of the words. Among the major benefits of the initiative was the heightened collective ownership of the language and increased engagement in the project.

Kim's (2019) recent publication on the traditional approach to language documentation underscores the link between morphology and language documentation. Kim discusses how traditional methods of elicitation can mask ingrained assumptions about the ontology of data and the wider context of linguistic research. Critical examination of these assumptions yields a broader range of possible approaches that can be drawn on to produce a corpus theorization that is appropriate to each language situation. In particular, Kim argues that it is sometimes unethical to collect language data in a decontextualized way that prioritizes the linguist's goals above the goals of the language speakers.

METHODOLOGY

This is a qualitative research study, whose principal goal is to describe a naturalistic phenomenon, in this case the descriptive report of the first phase of our Baduy Dalam language documentation project and to reflect on or question the methodological approach we adopted in catering to our needs as linguistics and the interests of the Baduy Dalam community. Soedirgo and Glas (2020) label this methodological practice as 'active reflexivity', where researchers recognize and respond to the complexities of their positionality during the research. Active reflexivity is seen as apt in qualitative research (Fujii, 2017).

The Baduy Dalam research project was initiated in 2014 as a pilot project supported by the Fulbright grant, with William Daniel Davies from the University of Iowa as the principal investigator. Davies was assisted by a team of researchers from Universitas Pendidikan Indonesia who had direct access to the Baduy community, particularly the Baduy Dalam. The primary goal of the five-day exploratory fieldwork conducted in March 2014 was to gather initial sets of data, including word and sentence lists. The first phase of this project adopted a linguist-focused model to collect linguistic data, which involved gathering word lists, conducting interviews, and recording naturalistic language use. These data were analyzed using linguistic tools to identify the grammatical structures in the language. The ultimate aim of this project was to create a language dictionary and grammar sketch of the Baduy Dalam language.

The study on documenting the endangered Baduy Dalam language employed various data collection methods to gather linguistic and cultural information. These methods included word-list elicitation, where 13 male participants provided a list of words related to various themes such as nature, kinship, or culture. This list was adopted from Nurjanah and Solehudin (2010), an adaptation of Swadesh List (1950). It is important to note that Baduy Dalam females were customarily not allowed to participate in any research studies. Furthermore, interviews were conducted with community members to collect information about language use, attitudes towards the language, and perceptions of language loss. Recordings of naturalistic language use were also collected to provide a more comprehensive understanding of the language and its everyday contexts. These recordings were transcribed and analyzed to identify patterns in the language, such as phonetic and grammatical structures. The combination of these data collection methods provided a detailed and nuanced understanding of the endangered Baduy Dalam language and its cultural context.

In more specific, three types of data have been collected over the life of the project:

- a. Natural speech. Various genres of narrative (described elsewhere) and conversations have been video-recorded using a Canon FS200 digital camcorder using an external microphone Sony ECM-MS907 and have been simultaneously recorded using a Marantz PDM 661 MKII handheld solid state voice recorder (to provide back-up audio in case of unforeseen problems). Following best practices as set out by the E-MELD School of Best Practices in Digital Documentation (E-MELD 2005), the Language Archive (Max Planck Institute), and the Pacific and Regional Archive for Digital Sources in Endangered Cultures (PARADISEC 2005), the video media is in .mod format, and audio media is in .wav format recorded at a 48 kHz sampling rate and 24 bit quantization. Videos have been recorded and saved in .mod format, a lossless format that can later be converted (to .mpeg, .avi, .mp4) for general users.
- b. Lexical items. We have directly elicited many lexical items, writing forms in IPA and recording the sessions with the voice recorder.
- c. Morphological and syntactic forms. Primary syntactic and morphological data have been derived from the recordings of natural speech. We have used direct elicitation to complete paradigms, to collect additional examples of the forms, and follow up on discoveries made in analysis.

In terms of data processing, the data were handled using multiple techniques, in part depending on the type of data and in part depending on the particular process.

- a. The video-recorded narratives have been transcribed in ELAN (see <https://archive.mpi.nl/tla/elan>), a software created at the Max Planck Institute, expressly for language documentation. These transcriptions have been checked by at least one senior team member, glossed, and given natural translations in English and Indonesian, creating time-aligned, annotated videos. ELAN is frequently mentioned in discussions of best practices as a recommended software for transcribing and annotating video, such as E-MELD 2005 and PARADISEC 2005. ELAN is an open source software, which ensures accessibility and usability of the data by others (an important criterion discussed in Bird & Simons, 2003).
- b. Audio recorded speech not associated with videos or direct elicitation have been transcribed in ELAN and glossed and translated.
- c. The texts of videos and elicited data have been entered in the Summer Institute of Linguistics Fieldworks (FLEX) and checked, and lexical data have been organized. FLEX (Paxson, 2017) is open source software and recommended as best practices (E-MELD, PARADISEC, and others). It can be exported in .xml format, ensuring interoperability and accessibility.
- d. The videos that are to be used in creating CDs/DVDs for use in Indonesia have been edited using Adobe Premiere Pro. Subtitles in English and Indonesian have been added to the videos.

As for data curation and backup, several steps were involved in preserving the data collected:

- a. All video and audio files have been downloaded onto a laptop and two external hard drives in the field.
- b. Filenames have been assigned following naming conventions that have been proposed in best practices and metadata have been constructed following the best practices set out in the Open Languages Archive Community (Simons & Bird, 2008), the Language Archive, and PARADISEC (Thieberger & Jacobson, 2010).

About archiving, all data files have been archived in PARADISEC (<http://www.language-archives.org/language/bac>), an archiving organization located in Australia. PARADISEC is a member of the Digital Endangered Languages and Music Archives Network (DELANMAN) and carries the Data Seal of Approval, and it has agreed to archive all of the recordings, field notes, and documentary materials that have been produced as part of the proposed project. A letter stating their willingness is included in the supplementary material. PARADISEC requires all materials to be stored in accord with best practices (as set out E-MELD, OLAC, the Language Archive, and others): audio in .wav format, text in .xml, video in .mod, handwritten field notes in .pdf, and still images in .tif.

FINDINGS AND DISCUSSIONS

The results of data elicitation

At the beginning of our project in 2014, like in any typical linguistic field research, the research team initiated the data collection process by employing word-list elicitation from a 576 Sundanese word list developed Nurjanah and Solehudin (2010). This list was constructed based on Swadesh's (1950) List by adding words appropriate in the Sundanese culture. The inventory comprises terms related to address, names of objects within a person's surroundings (animate and inanimate), and grammatical words like conjunctions and prepositions.

sesebutan ka awewe kolot		ete kolot
sesebutan keur budak lalaki		Aceng / ongong
sesebutan keur lalaki dewasa		area (dewasa) adi (salaan Dopen)
amil/lebe/panghulu		Puun
arisan		-
hajat		Salametan
imah-imah		rumah tangga
juru tulis		-
kabayan		-
kamasan		-
kapala desa		Jaro
tua kampung		.kolot kampung
ulu-ulu		-
marebot		-
nu nungguan bale desa		kolot desa
paledang		-
pamong desa		-
polisi desa		-

Figure 1. Sample of word list elicitation instrument

As is seen from Figure 1 above, the lefthand column is a list of target words adapted from Swadesh list that has been translated into prototypical Sundanese variety, making it easier for the

researcher to elicit data and provide hints when informants had no clue what word was being asked. The middle column was reserved for the Indonesian counterparts and the rightmost column for the Baduy variants.

The informants were asked individually or in small groups, depending on their consent and confidence in their ability to respond, as illustrated in Figure 2 below. Data collection occurred entirely through interviews, which varied in length depending on the interview instruments and the situation, such as if the informants appeared tired and needed a break. The researcher(s) typically used the prompt, "*Naon bahasa Baduyana ...*[What is the Baduy equivalent for ...]?" with a follow-up question such as, "*Aya kecap lain deui teu ...* [Is there another word for ...]?" to generate possible synonyms or related terms. As a result of a follow-up question, as apparent in Figure 1 row 3, there are two Baduy words for a single Sundanese word, i.e., *aka* (for adults) and *adi* (for non-adults).



Figure 2. Interview setting for elicitation sessions

We also employed Davies (2010) sentence list, comprising 60 sentences, which he compiled from an array of sentence structures extracted from his many fieldwork experiences in Java and Madura. The syntactic constructions under examination range from declaratives with different types of transitivity status, existential structure, equational structure, locative and causative structures, relative clauses to interrogatives.

Locative

24. Mobil itu di belakang rumah. *Rev fustha may sibe di ~~teundeun~~ di tucaya
 anak may sibe uin di harepan imah*
25. Ibu ada di kantor. *Ames aya di harepan parako*

Causative

26. Bambang membuat ayahnya marah. *karmain geus ngalampalikeun maraheun
 ka ayah*
27. Koki itu menajamkan pisaunya. *Eta budak ngasal kolokna
 Eta budak ngaputah jamangna*
28. Ibu menyuruh anak-anak untuk berjalan kaki ke rumah. *Ames aya nitah aralna leumpang ka huma*

Figure 3. Sample of base sentence instrument

Through that work, which consisted of video recordings of four narratives and two music/poetry performances, word lists and some basic morphosyntactic elicitation, we have uncovered differences between the Baduy Dalam language and modern Sundanese. Lexical differences are present throughout the language, but two semantic areas of early interest that were uncovered are terms relating to kitchen technology and architecture.

Table 1. A list of distinct words used in the standard Sundanese and Baduy Sundanese

Sundanese	Baduy	
<i>katél</i>	<i>kékéncéng</i>	‘frying pan/wok’
<i>leukeur</i>	<i>kuluwung</i>	‘place to store pots’
<i>susuk</i>	<i>kalaci</i>	‘spatula for stirring’
<i>téko</i>	<i>cérét</i>	‘tea kettle’
<i>gelas awi</i>	<i>somong</i>	‘bamboo drinking cup’
<i>bobotolan</i>	<i>colobong</i>	‘food bowl’
<i>palupuh</i>	<i>surung</i>	‘bamboo floor of house’
<i>kusén</i>	<i>lawang</i>	‘door frame’
<i>tepas</i>	<i>sosoro</i>	‘porch like part of house’
<i>jandéla</i>	<i>lolongok</i>	‘window’
<i>taneuh</i>	<i>lemah</i>	‘ground where house is built’
<i>tatapakan</i>	<i>umpak</i>	‘stone pillars supporting house’

Morphosyntactically, the voice markers so prevalent in Indonesian-type languages have different manifestations in Baduy. A prefix *si-*, with a quite different distribution in standard Sundanese, is used quite productively with some intransitive predicates in the Baduy Dalam Sundanese variant, meaning ‘wearing of’.

<i>jamang</i>	‘shirt’
<i>sijamang</i>	‘wear a shirt’
<i>dudukuy</i>	‘hat’
<i>sidudukuy</i>	‘wear a hat’

The same prefix is also attested in the standard Sundanese, with multiple meanings. Noteworthy is the fact that this prefix is found only in a very limited set of predicates. In other words, the prefix is no longer productive in the standard Sundanese.

<i>sibanyo</i>	‘wash one’s face and feet’
<i>sidadep</i>	‘cross one’s arms across chest as a sign of distancing oneself’
<i>siduru</i>	‘sit in front of a fire to warm one’s body in cold weather’

In addition, there are differences in the use of the active nasal morpheme. While both Baduy and Sundanese take the root of ‘must’ to be *kudu* (which is common among Indonesian-type languages), in Baduy the nasal prefix is used, *ngudu*, whereas Sundanese and other languages use the form without the prefix. This same distinction arises with some other predicates such as *ngajadi* ‘become’, and *ngalahir* ‘give birth’ (where the mainstream Sundanese variety uses forms with no prefix), as well as *mantog* ‘go’ (Sun. *indit*), *nganing* ‘eat’ (Sun. *dahar*), *nyaah* ‘love’ (Sun. *bogoh*).

The final product of the first phase of our project was a small dictionary of Baduy language (consisting of 534 lemmas) complemented by transcriptions and transliterations of four Baduy stories (Davies et al., 2014).

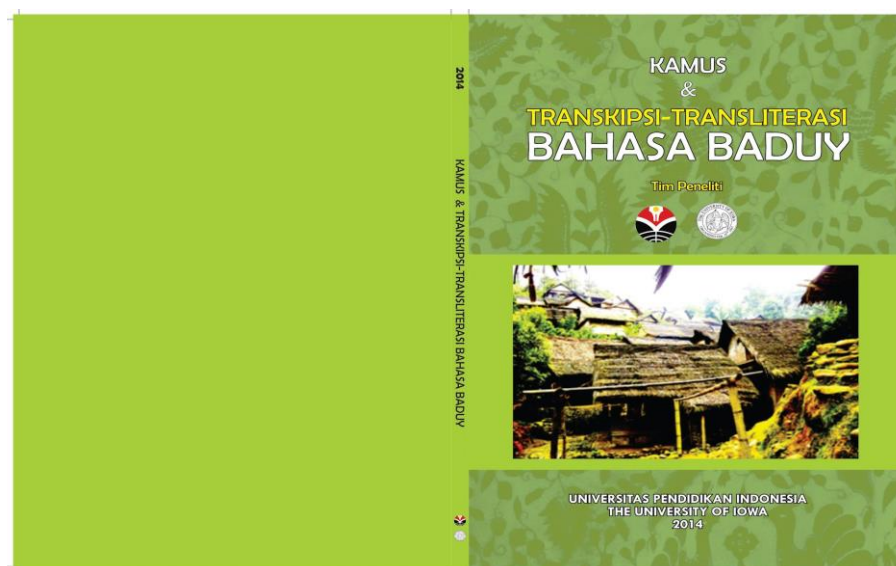


Figure 4. Sundanese Baduy dictionary

Figure 4 shows the cover of the dictionary, which is the final product of the first phase of our project. The dictionary was printed in a limited number and distributed to various libraries in Indonesia. Its significance lies in the fact that it was also presented to the representatives of the Baduy Dalam and Luar as evidence of our project and an early effort to preserve their language and culture. However, it remains to be seen how the people of Baduy Dalam will make use of the dictionary, given the absence of formal education and literacy in the community.

Reflection on the linguist-focused model in documenting Baduy Dalam

The study documenting the endangered Sundanese language of Baduy Dalam highlights limitations in the linguist-focused model used to document endangered languages, especially in the context of Baduy Dalam. The study found that this model may not effectively document Baduy Dalam because it relies heavily on linguistic analysis and may not take into account the cultural context in which the language is spoken. As a result, this approach may not be sensitive to the needs, interests and perspectives of the Baduy community, evidently leading to a lack of engagement and ownership in the project. Although we had established personal relationships with Baduy informants and consents from the community have been obtained, concerns were raised by the *wakil jaro* (deputy chief), representing the community's concerns about the purpose of the research and the follow-up. This illustrates the level of engagement and sense of ownership that the community has in the project. Furthermore, the Baduy did not yet see the usefulness of the language project for their community, as reflected in the deputy chief questions after the elicitation sessions, “What is this research for? What’s in there for us?”.

The relatively low level of engagement of the Baduy community was evident in the number of Baduy Dalam people involved, especially those who hold not only the knowledge and wisdom but also the authority to share their life narratives with outsiders. In the first year of the project, only a limited few (the deputy chief, *Ayah Mursid*, his colleagues or close neighbors from the same village) were willing to participate, despite our established relationships with the community. Even when a larger group was present, only a few were poised to respond or provide answers, typically represented by the chiefs. This raises questions about the effectiveness of the linguist-focused model in documenting indigenous language varieties like Baduy Dalam.

Furthermore, the study found that the linguist-focused model may not be effective in promoting language revitalization efforts, as it may not involve a much wider group of community members in language documentation and may not take into account the potential impact of language loss on the community. This highlights the need for a community-based approach to language documentation that involves the local community in all stages of the process, from data collection to analysis and dissemination (inter alia Akumbu, 2022; Fitzgerald, 2018; Langley et al., 2018; McCarty, 2018). Overall, the study underscores the limitations of the linguist-focused model in documenting endangered languages (see Kim, 2019) and highlights the need for some modification such as collective elicitation ala Anderson et al. (2015) or educational dictionaries as alternative products ala Mosel (2011).

Comparing the results of the study on documenting the endangered language of Baduy Dalam with previous studies on documenting endangered languages, some similarities and differences can be identified. Like previous studies, the present study emphasizes the importance of alternative approaches such as a multidisciplinary approach (e.g. Di Carlo, 2016), drawing insights from multiple fields, e.g., anthropology, verbal arts, information science, media and recording arts, and ethnoscience (Esch et al., 2019; Austin, 2017), and community involvement in language documentation efforts (e.g. Akumbu, 2022). Additionally, the present study raises the need for cultural sensitivity, and ethical considerations in language documentation so that there is a growing sense of ownership on the part of the community to the language and the project (Anderson et al., 2015).

However, there are also some differences between the present study and previous studies. For example, the present study specifically examines the use of the linguist-focused model for

documenting an endangered language, whereas previous studies have explored a variety of approaches, including participatory methods and collaborative approaches (Anderson et al., 2015; Kim, 2019; Mosel, 2011). Additionally, the present study focuses on the endangered language of Baduy Dalam in Indonesia, whereas previous studies have examined a range of endangered languages in different geographic and cultural contexts, such as Mosel (2011) and Anderson et al. (2015) in Papua New Guinea, Akumbu (2022) in Africa, or Langley et al. (2018) in Louisiana. Overall, the present study contributes to the growing body of literature on language documentation and underscores the importance of a multidisciplinary and community-based approach to documenting endangered languages.

CONCLUSION

In conclusion, the study on documenting the endangered language of Baduy Dalam and the use of the linguist-focused model highlights the importance of a community-based approach that takes into account linguistic, cultural, and ethical considerations when documenting endangered languages. The study provides insights into the advantages and disadvantages of the linguist-focused model and underscores the importance of providing more engagement of the local community in the documentation process. The study also highlights the need for researchers to consider the long-term impact of their work on the community and the potential for language revitalization efforts. Overall, the study provides valuable guidance for researchers working in the field of language documentation and emphasizes the importance of preserving linguistic diversity and cultural heritage.

The results of the present study certainly have several implications for documenting other endangered languages. The study highlights the adoption of a multidisciplinary approach that takes into account linguistic, cultural, and ethical considerations when documenting endangered languages. Furthermore, the study emphasizes the need for researchers to involve the local community in the documentation process and to consider the long-term impact of their work on the community. A potential challenge for field research studies especially the ones focusing on indigenous languages is strategic data collection techniques that could capture real language use and contexts on the one hand, but adhere to the restrictions of not doing electronic recording in the villages where the speakers reside.

Several limitations do emerge. One limitation is that the study focused solely on the linguist-focused model and did not explore other approaches to language documentation. Additionally, the study was conducted in a specific cultural and linguistic context and may not be generalizable to other contexts. Future research could address these limitations by examining other approaches to language documentation and exploring how these approaches can be adapted to different cultural and linguistic contexts. Furthermore, future research could focus on the impact of language documentation on language revitalization efforts and how language documentation can be used to promote linguistic diversity and cultural heritage. Overall, the study provides a valuable contribution to the field of language documentation, but further research is needed to build upon these findings and address the limitations of the current study.

NOTE

The authors would like to thank the anonymous reviewers who have given constructive feedback to the earlier version of this paper.

ACKNOWLEDGMENT

This research was generously supported by funding from the National Science Foundation (USA) under Award Number 1562130. The support provided for international activities associated with this project is gratefully acknowledged.

REFERENCES

- Akumbu, P. W. (2022, April). A community approach to language documentation in Africa. In *53rd Annual Conference on African Linguistics (ACAL 53)*.
- Alam, N. R., Handoyo, S., Suhermanto, & Prihatini, L. (2023). Legal protection of ulayat rights in the traditional communities of the Baduy tribe. *Novateur Publications*, 2, 173-182.
- Anderson, G. D., Barth, D., & Forepiso, K. R. (2015). The Matukar Panau online talking dictionary: Collective elicitation and collaborative documentation. In I. W. Arka, N. L. S. Malini, & I. A. M. Puspani (Eds.), *Language Documentation and Cultural Practices in the Austronesian World* (pp. 111-126). Asia-Pacific Linguistics.
- Austin, P. K., & Sallabank, J. (1999). *The Cambridge Handbook of Endangered Languages*. Cambridge University Press.
- Austin, P. K. (2017). Language documentation and legacy text materials. *Asian and African Languages and Linguistics*, 11, 23-44.
- Austin, P. K. (2021). Language documentation and language revitalization. In J. Olko & J. Sallabank (Eds.), *Revitalizing Endangered Languages: A practical Guide* (pp. 199-219). Cambridge University Press. <https://doi.org/10.1017/9781108641142.014>
- Bakels, J. (1988). The Baduy of Banten: The hidden people. In R. Schefold, V. Dekker, & N. de Jonge (Eds.), *Indonesia in Focus* (pp. 39-45). Meppel: Edu'Actief.
- Bakels, J., & Boevink, W. (1988). *De Baduy van West-Java*, Amsterdam: CAS A. [Werkdocument 2.]
- Barendregt, B., & Wessing, R. (2003). Encountering the authentic: The creation of the Baduy of Banten, West Java. In A. D. Syukur & A. Suryana (Eds.), *Tulak Bala: Sistim Pertahanan Tradisional Masyarakat Sunda dan Kajian Lainnya Mengenai Budaya Sunda* (pp. 85-116). Yayasan Pusat Studi Sunda.
- Barendregt, B., & Wessing, R. (2005). Tending the spirit's shrine: Kanekes and Pajajaran. *Moussons* 8: 3-26.
- Barendregt, B., & Wessing, R. (2008). Centred on the source: Hamlets and houses of Kanekes (Baduy). In Reimar Schefold, Peter J.M. Nas, Gaudenz Domenig and Robert Wessing (eds.), *Indonesian Houses vol 2 Survey of Vernacular Architecture in Western Indonesia*. Leiden: KITLV Press.
- Bird, S., & Simons, G. (2003). Seven dimensions of portability for language documentation and description. *Language*, 557-582.
- Bromham, L., Dinnage, R., Skirgård, H., Ritchie, A., Cardillo, M., Meakins, F., Greenhill, S., & Hua, X. (2022). Global predictors of language endangerment and the future of linguistic diversity. *Nature Ecology & Evolution*, 6(1), 163-173. <https://doi.org/10.1038/s41559-021-01604-y>
- Davies, W. D. (2010). *A grammar of Madurese*. De Gruyter Mouton.

- Davies, W. D., Kurniawan, E., Isnendes, R., Hernawan, Dallyono, R., & Muniroh, R. D. D. (2014). *Kamus & Transkripsi-transliterasi Bahasa Baduy*. Universitas Pendidikan Indonesia & Iowa University.
- de Swaan, A. (2004). Endangered languages, sociolinguistics, and linguistic sentimentalism. *European Review*, 12(4), 567-580. <https://doi.org/10.1017/S1062798704000481>
- Eberhard, D. M., Simons, G. F. & Fennig, C.D. (2024). *Ethnologue: Languages of the World*. 27th edn. SIL International.
- ELAN (Version 6.8) [Computer software]. (2024). Nijmegen: Max Planck Institute for Psycholinguistics, The Language Archive. Retrieved from <https://archive.mpi.nl/tla/elan>
- Esch, D., Foley, B., & San, N. (2019). Future directions in technological support for language documentation. *Proceedings of the Workshop on Computational Methods for Endangered Languages*. <https://doi.org/10.33011/COMPUTEL.VII.341>.
- Ewing, M. C. (2014). Language endangerment in Indonesia. *International Journal of Education*, 8(1). <https://ejournal.upi.edu/index.php/ije/article/view/1764>
- Fitzgerald, C. M. (2018). Creating sustainable models of language documentation and revitalization. *Insights from Practices in Community-based Research: From Theory to Practice around the Globe*, 94-111.
- Fujii, L. A. (2017). *Relational Interviewing: An Interpretive Approach to Social Science Research*. Routledge.
- Garna, J.K. (1989). *The Baduy of West Java: Adaptation and Change to Development*. Bandung: Fakultas Pascasarjana Universitas Padjajaran.
- Gelman, S.A., & Roberts, S.O. (2017). How language shapes the cultural inheritance of categories. *Proceedings of the National Academy of Sciences of the United States of America*, 114(30), 7900-7907. <https://doi.org/10.1073/pnas.1621073114>
- Hidayati, S., Iman, N., Samin, & Franco, F.M. (2017). Using ethnotaxonomy to assess traditional knowledge and language vitality: A case study with the Urang Kanekes (Baduy) of Banten, Indonesia. *Indian Journal of Traditional Knowledge*, 16(4), 576-582.
- Holton, G. (2011). The role of information technology in supporting small and endangered languages. In P. K. Austin & J. Sallabank (Eds.), *The Cambridge Handbook of Endangered Languages* (pp. 371-399). Cambridge University Press.
- Iskandar, J. (2007). Responses to environmental stress in the Baduy swidden system, South Banten, Java. In Roy Ellen (ed.), *Modern Crises and Traditional Strategies: Local Ecological Knowledge in Island Southeast Asia*, pp. 112-132. Berghahn Books.
- Jones, M. C. (Ed.). (2014). *Endangered Languages and New Technologies*. Cambridge University Press. <https://doi.org/10.1017/CBO9781107279063>
- Kim, Y. (2019). Morphology and language documentation. In R. Lieber (Ed.), *Oxford Encyclopedia of Morphology* (pp. 1-23). Oxford University Press.
- Kurnia, A. D., & Sihabudin, A. (2010). *Saatnya Baduy Bicara [It's time for Baduy to speak]*. Bumi Aksara.
- Langley, B., Langley, L., Martin, J. B., & Hasselbacher, S. (2018). The Koasati Language Project: A collaborative, community-based language documentation and revitalization model. *Insights from Practices in Community-based Research: From Theory to Practice around the Globe*, 132-50.
- Lembaga Sejarah dan Antropologi. (1975). *Beberapa Cerita Rakyat Baduy*. Lembaga Sejarah dan Antropologi, Direktorat Jendral Kebudayaan, Departemen Pendidikan dan Kebudayaan.

- Lewis, M. P., Simons, G. F., & Fennig, C. D. (Eds.). (2015). *Ethnologue: Languages of the World (18th ed.)*. SIL International
- Makmur, A. (2001). Pamarentahan Kanekes Bedouin village: Perspectives of kinship. *Jurnal Sosiohumaniora*, 4(2), 104-115.
- Marlina, E. (2009). Traditional community of Baduy, West Jawa: Toward architecture sustainability. Ms., Yogyakarta University of Technology.
- McCarty, T. L. (2018). Community-based language planning: Perspectives from Indigenous language revitalization. In *The Routledge Handbook of Language Revitalization* (pp. 22-35). Routledge.
- Mosel, U. (2011). Lexicography in endangered language communities. In P. K. Austin & J. Sallabank (Eds.), *The Cambridge Handbook of Endangered Languages* (pp. 337-353). Cambridge University Press.
- Nurjanah, N. & Solehudin, O. (2010). *Linguistik Historis Komparatif*. CV. Wahana Karya Grafika
- Paxson, V. (2017). *Flex*. Retrieved from <https://github.com/westes/flex>
- Pepinsky, T.B., Abtahian, M.R., & Cohn, A.C. (2022). Urbanization, ethnic diversity, and language shift in Indonesia. *Journal of Multilingual and Multicultural Development*, 43(4), 317-334. <https://doi.org/10.1080/01434632.2022.2055761>
- Permana, R. C. E. (2001). *Gender Equality in the Core Universe Bedouin Custom*. Wedatama Widya Sastra.
- Permana, R. C. E. (2006). *Tata Ruang Masyarakat Baduy*. Jakarta: Wedata Widya Sastra.
- Rogers, C., & Campbell, L. (2015, December 03). Endangered Languages. *Oxford Research Encyclopedia of Linguistics*. Retrieved 27 Nov. 2024, from <https://oxfordre.com/linguistics/view/10.1093/acrefore/9780199384655.001.0001/acrefore-9780199384655-e-21>.
- Romaine, S., & Gorenflo, L. J. (2017). Linguistic diversity of natural UNESCO world heritage sites: bridging the gap between nature and culture. *Biodiversity and Conservation*, 26(8), 1973-1988. <https://doi.org/10.1007/s10531-017-1340-x>
- Seifart, F., Evans, N., Hammarström, H., & Levinson, S. C. (2018). Language documentation twenty-five years on. In F. Seifart, N. Evans, H. Hammarström, & S. C. Levinson (Eds.), *Language* (Vol. 94, Issue 4). Linguistic Society of America. <https://doi.org/10.1353/lan.2018.0070>
- Simons, G. F. (2019). Two centuries of spreading language loss. *Proceedings of the Linguistic Society of America*, 4(27), 1-12.
- Simons, G., & Bird, S. (2008, November). Toward a global infrastructure for the sustainability of language resources. *Proceedings of the 22nd Pacific Asia Conference on Language, Information and Computation* (pp. 87-100).
- Soedirgo, J., & Glas, A. (2020). Toward active reflexivity: Positionality and practice in the production of knowledge. *PS: Political Science & Politics*, 53(3), 527-531. doi:10.1017/S1049096519002233
- Swadesh, M. (1950). Salish internal relationships. *International Journal of American Linguistics*, 16: 157-167.
- Thieberger, N. (Ed.). (2011). *The Oxford Handbook of Linguistic Fieldwork*. Oxford University Press.

- Thieberger, N., & Jacobson, M. (2010). Sharing data in small and endangered languages: Cataloging and metadata, formats, and encodings. In L.A. Grenoble & N. L. Furbee (Eds.), *Language Documentation* (pp. 147-158). John Benjamins.
- Universitas Padjadjaran. (1984). *Perbandingan Struktur Bahasa Sunda Baduy dengan Bahasa Sunda Lulugu [A comparison of the Structure of Baduy with Lulugu]*. Faculty of Humanities, Universitas Padjadjaran.
- van Zanten, W. (1995). Aspects of Baduy music in its sociocultural context, with special reference to singing and angklung. *Bijdragen tot de Taal-, Land- en Volkenkunde* 151: 516-544.
- van Zanten, W. (2004). Les tentations des ascètes: Patrimoine culturel immatériel des Baduy dans le Banten du sud (Indonésie). In Jean Duvignaud and Chérif Chaznadar (eds.), *Le patrimoine culturel immatériel; Les enjeux, les problématiques, les pratiques*, pp. 124-150. Paris: Babel/Maison des Cultures du Monde.
- Waluya, B., Malihah, E., Ruhimat, M., & Wiyanarti, E. (2023). Cultural ecology and environmental education: Lesson learned from Baduy indigenous community. *Indonesian Journal of Geography*, 55(1). <https://doi.org/10.22146/ijg.77203>
- Wessing, R. (1977). The position of the Baduy in the larger West Javanese society. *Man*, 12(2), 293-303.
- Wessing, R. (1988). Spirits of the earth and spirits of the water: Chthonic forces in the mountains of West Java. *Asian Folklore Studies*, 47, 43-61.