Diversifying Science: Insights of Native- and Non-Native-Speaker Collaborations in Linguistic Descriptions

I Wayan Arka
Australian National University/Universitas Udayana

Abstract

This chapter reflects on over two decades of linguistic research in Indonesia, examining collaborations between native- and non-native-speaker linguists and local speech communities in documenting and describing minority and endangered languages. Drawing on case studies from Balinese, Rongga, Marori, and Enggano, it explores how positionality, authority, and co-production shape grammar writing and broader documentation outcomes. Situating these experiences within the shift toward community-led, ethically grounded, and ethnographically informed fieldwork, the chapter underscores the importance of capacity building for native-speaker linguists and sustained engagement with both communities and local institutions. It argues that partnerships grounded in mutual respect and shared epistemic goals can produce richer, more culturally embedded, and socially impactful linguistic descriptions, advancing a more inclusive and decolonial linguistic science.

Key words: *native-speaker linguists, collaborative grammar writing, positionality, authority and co-production, community-led documentation, ethically grounded fieldwork, capacity building, decolonial linguistics, endangered languages, sustained institutional engagement*.

1. Introduction

This chapter offers a critical reflection on my long-standing engagement with linguistic research in Indonesia, focusing on the evolving collaborative dynamics between native- and non-native-speaker linguists as well as speech communities in the documentation and description of languages.[1] Indonesia's extraordinary linguistic diversity, with hundreds of Austronesian and Papuan languages, provides both a rich empirical base and a complex sociocultural terrain in which issues of positionality, authority, and co-production are continually negotiated. In this context, building on Ameka's (2006) co-creative grammar writing, I reflect on and explore the recent paradigm shift in field linguistics (Florey and Himmelmann 2010:122; Dobrin and Berson 2011; Czaykowska-Higgins 2009; Bischoff and Jany 2018, among others), discussing how partnerships between native speakers from local speech communities and non-native linguists—when grounded in mutual respect and shared epistemic goals—can lead to more culturally embedded, ethically sound, and analytically robust linguistic work, extending beyond grammar writing. This aligns with calls to reimagine linguistics as a decolonial, community-led practice that actively challenges the extractive legacies of the field (Tsikewa 2021; Hudley, Mallinson, and Bucholtz 2024).

My reflections are based on over two decades of fieldwork and academic practice, drawing from three intersecting roles that have shaped my perspective. First, as a native-speaker

[1] Dedicated to Nikolaus Himmelmann, whose pioneering work in language documentation and capacity building in Indonesia—and in Austronesian linguistics more broadly—has been a lasting inspiration. I have been fortunate to work with him, including in the UBUD Documentation Workshops (2006–2007) and during my Humboldt Fellowship at the University of Cologne, which he kindly sponsored. I also thank the anonymous reviewers, Sonja Riesberg (editor), and Amrita (Mita) Nugraheni for their helpful comments on this chapter.

linguist working on Balinese and Indonesian, I have experienced the advantages of insider knowledge—along with its limitations—and the transformative impact of collaborative analysis with non-native linguists. Second, as a non-native fieldworker documenting Indonesia's endangered and minority languages such as Rongga, Marori and Enggano. I have navigated the challenges of outsider positionality and developed models for collaborative engagement that centre the agency of local communities. Third, as a supervisor and mentor, I have supported numerous native-speaker students in documenting their own languages, guiding them through different challenges at various stages of their academic trajectories, and witnessing how their training and empowerment reshape the research landscape from within.

This chapter explores the affordances and tensions across these roles, with the aim of identifying practices that support inclusive, pluralistic, and co-productive linguistic science. Drawing on concrete case studies—from working on my native language Balinese to leading documentation projects in *Tanah Papua* and the Barrier Islands regions—I show how linguistic collaboration can thrive when all actors are supported in developing their capacities, and when community goals are integrated with academic objectives. In this view, collaboration is not simply a methodological convenience but an ethical and epistemological commitment.

However, as I have learned through extensive field experience (Arka 2025; Hisa, Mahuze, and Arka 2018; Hisa, Mahuze, and Arka 2017; Arka 2015, 2013, 2010), these collaborations only succeed when linguists' research agendas are developed alongside—and not at the expense of—the interests of the speech communities themselves. As Dobrin and Schwartz (2016) emphasize, collaboration is not a universally appropriate or sufficient model for every context. In many minority settings, linguistic work intersects with broader community aspirations around cultural revitalisation, education, identity, and economic-cultural-social wellbeing. As such, linguistic documentation must be framed not only as a scholarly enterprise but also as a socially responsible practice. This has become a central philosophy in my work: that research is most meaningful when it contributes reciprocally to the communities it engages with, and when it is co-constructed by those who speak, live, and sustain the languages we seek to describe.

In what follows, I outline the conceptual grounding for this co-creative approach by, first of all, discussing the what and how of linguistic research and language description (§2). I then examine the shifting roles of native- and non-native-speaker linguists (§3), reflect on challenges and strategies for equitable collaboration (§4). Along the way, I highlight insights gained through supervising native-speaker students (§5), present case studies—Balinese and Enggano—that illustrate co-creation in practice (§6), and formulate recommendations for the future of linguistic science (§7), with the broader goal of diversifying the scientific enterprise of linguistics through more inclusive and locally rooted forms of knowledge-making. Conclusion is then given in §8.

2. On the What and How of linguistic descriptions and language documentation

A fundamental challenge in linguistic description and language documentation, especially when writing a grammar, lies in determining *what* to document and describe: which phenomena to prioritise, what data to collect, and what kinds of knowledge to foreground.

This task is far from straightforward. Grammar writing, as Evans and Dench (2006:1) remind us, "is a demanding art which calls for the integration of form and meaning, system and use, synchrony and diachrony." It becomes even more complex in the context of language documentation, particularly for minority and endangered languages where linguistic competence among speakers may be uneven, and sociocultural factors heavily shape the ecology of language use. My own fieldwork experience across Indonesia has shown that producing a grammar is only one piece of a much larger puzzle. Alongside what Evans and Dench (2006) call the "grammar-dictionary-text" trinity,— or what Musgrave and Thieberger (2021) term the "documentation quartet," with the addition of media recordings in modern documentary linguistics—there are local expectations, needs, and constraints, including the demand for hypertext and interlinking in digital materials. It is not enough to produce grammars as scholarly artefacts; our work must also respond to what I have elsewhere called the "multi-dimensionality of capacity building" (Arka 2025), which includes fostering local participation, leadership, and community relevance. In addition, in today's digital age, it is increasingly valuable to leverage hypertext and web technologies to enhance accessibility and verifiability of analytical claims—for example, by allowing users to view examples in full context and access supporting evidence.

In thinking about what kinds of linguistic descriptions we produce, I align with Evans and Dench's call to recognise the interdependence of descriptive, typological, and formal-theoretical approaches. They describe these as "a triadic and mutually complementary relationship" (2006:2). Descriptive linguistics provides grounded accounts of particular languages, typology brings out patterns across languages, and formal linguistics constructs predictive models. In my own documentation projects—whether on Rongga, Marori, or Enggano—I have found it crucial to design the descriptive work so that it not only serves the immediate needs of language maintenance (e.g., orthography development, or school materials), but also informs theoretical questions (e.g., on alignment, voice, or clause structure) and typological debates (e.g., on Austronesian variation). Descriptive work, as I have argued (Arka 2025:32), is the logical entry point before comparative or formal analysis can be carried out, but it must be embedded in a larger strategy that bridges academic inquiry and community benefit.

This multi-stranded goal also requires expanding the very idea of what constitutes a "grammar". As Evans and Dench emphasise, "a grammar is inescapably an interpretive account" (2006:7), shaped by the linguist's choices, theoretical assumptions, and intended readership. From my perspective, working as both a native and non-native researcher in different language contexts, I find that such interpretive work must also be shaped by the language community's aspirations, the available human and institutional resources, and the evolving dynamics of identity and motivation. In an ethnosyntactic or ethnogrammatical sense (Enfield 2002), these priorities influence the content and organisation of the grammar: which categories are described in depth, what examples and varieties are chosen, and how explanations are framed. These considerations are especially critical when grammar writing also serves pedagogical purposes, such as developing teaching materials for local schools. In Enggano, while the actual drafting of the grammar has been primarily undertaken by the non-local linguist members of our project team, the process has been co-creative and evolving, informed by ongoing community consultation through our Community Research Collaborator (CRC). This work has proceeded alongside training the CRC in basic linguistics, ELAN, building a lexicon for school use, and developing literacy resources. In Marori, the descriptive work seeded a broader eco-cultural revitalisation project. These experiences underscore that in the contexts where I work, a grammar is not simply a product of analytical

rigour—it is part of a broader co-creative enterprise involving multiple stakeholders, epistemologies, and goals. Descriptive linguistics, when situated in such multilingual, dynamic, and politically charged environments, is both a scientific and a deeply social act. Grammar writing, at different levels, is thus a co-constructed account in which analytical decisions are shaped and informed by community knowledge, values, and goals.

Understanding *how* to undertake this descriptive task requires equal methodological care. In recent years, linguistic fieldwork has undergone a shift —sometimes referred to as the *documentarist turn*—from decontextualised elicitation toward ethnographically informed, corpus-based, and participatory approaches (Himmelmann 1998; Woodbury 2003, 2011). This shift owes much to the early engagement of Nikolaus P. Himmelmann in the 1990s, whose advocacy for richly contextualised documentation laid a foundation for modern best practices, as well as to other early pioneers such as Anthony C. Woodbury. My own methodological stance builds directly on these foundational ideas, extending them in ways that respond to contemporary challenges of capacity building, community co-authorship, and ethical reciprocity in field linguistics in Indonesia. Traditional elicitation methods—such as sentence translation—often fail to capture real-time interactional dynamics like register alternation, politeness strategies, or epistemic stance. To address this, modern documentation increasingly integrates three key practices: the structured collection of metadata, the development of multimodal corpora, and ethnographic grounding. Recording demographic, genealogical, and contextual metadata about speakers enables researchers to link linguistic variation to socially meaningful variables. Corpus design must include naturalistic and stimulus-based materials that are sensitive to pragmatic domains often underrepresented in spontaneous data. And crucially, these methods must be informed by participant observation and community engagement. This ethnographic lens ensures that linguistic patterns are not misinterpreted, and that grammars reflect the lived sociocultural realities of speakers. As we have argued (Arka & Mullan, to appear), these methodological shifts are essential, for instance, in capturing how language encodes social cognition, and they allow our grammatical descriptions to speak not just to formal and typological concerns, but also to the broader socio-cultural contexts in which language is embedded.

In navigating these complex demands of grammar writing (or language descriptions more generally) serving typological generalisation, formal rigor, and community relevance, I have found that the choice of theoretical framework plays a critical role. It is essential to work with a framework that is not only sufficiently modular and flexible to accommodate the unique structural and sociocognitive patterns of individual languages, but is also capable of revealing cross-linguistic parallels in a principled way. Lexical-Functional Grammar (LFG) offers such a framework (cf., Kroeger 2007). Its parallel and modular architecture enables analysts to tease apart and explicitly model multiple interacting layers of linguistic structure, including constituent structure, grammatical relations, information structure, and social meaning. As we argue in Arka & Mullan (to appear), LFG "provides an effective framework for analysing the intersection of social cognition and typological variation," particularly through its capacity to model context-sensitive grammatical behaviour without flattening language-specific distinctions.

This is especially valuable in multilingual, socially stratified, and contact-heavy ecologies like those in Indonesia, where social cognition is deeply woven into grammar. The discourse pragmatic structure (prag-str) in LFG, for instance, allows us to explicitly encode culturally specific information (e.g. social caste-based relations, *menak* 'high(er) caste' and *jaba* 'commoner') that determines lexical and grammatical choices, such as in Balinese speech

levels or polite pronominal forms. This linguistically relevant information analysed as 'social predicate' can be linked to grammatical modules like f-structure (functional structure) and a-structure (argument structure) through clearly defined correspondences. Such a modular setup lets us account for how variation in language use is shaped by speaker–addressee dynamics, social hierarchy, politeness, and voice constructions—while keeping these components separable for different analytical purposes. This modularity also makes LFG collaborative-friendly: different researchers, even from divergent theoretical traditions, can contribute to the same description by working on distinct but compatible layers (e.g., one on syntax, another on discourse, another on semantics) without requiring full adoption of LFG's formal machinery. In my own work, this flexibility has enabled productive partnerships in grammar writing, both with fellow LFG specialists and with typologists and documentary linguists who use other frameworks. Crucially, this blend of language-specific descriptive adequacy and cross-linguistic comparability, and theoretical openness makes LFG especially well-suited for the kinds of collaborative grammar work I undertake—whether as a native-speaker linguist working on Balinese and Indonesian or a non-native researcher documenting languages like Marori and Enggano.

3. Native-Speaker and Speech Communities in Linguistic Description and documentation

My early linguistic training and later academic development have been shaped by a bilingual upbringing in Balinese and Indonesian. As a native-speaker linguist, I bring to my analyses an intuitive grasp of sociocultural nuances and communicative practices that often elude external analysts. This insider position enables more in-depth interpretations of subtle pragmatics and lexicogrammatical features, offering unique insights into language description and into specific areas prioritised during documentation. However, the benefits of this insider role can only be fully realised—and its potential pitfalls mitigated—through quality linguistic training and sustained collaborative research experience. Such pitfalls include taking implicit knowledge for granted or overlooking deeper structural patterns due to over-familiarity, even among trained linguists (§3.2). Before discussing the importance of linguistic training for native speakers, I first examine the diverse roles that native speakers can play in linguistic description (§3.1).

3.1. Diverse roles and engagements

As a linguist working across multiple roles—both as a native speaker describing my own languages and as a non-native researcher documenting others—I have come to appreciate that linguistic description is most meaningful when conceived as a collaborative, socially grounded enterprise. Native speakers and local communities contribute to our research in a variety of ways, each role offering different kinds of insights and support. At one end of the spectrum are relatively passive language consultants who participate by providing elicited responses or contributing to recordings of natural speech. Their involvement is often shaped by the researcher's agenda, and while crucial, it may be limited to the provision of linguistic forms rather than interpretive or contextual information.

More engaged are those whom I consider relatively active language consultants—often local elders, teachers, or community leaders with a deep understanding of their language and its

sociocultural foundations. In my experience, this group also includes younger, well-educated native speakers who, despite lacking formal linguistic training, offer informed perspectives on what linguistic or cultural domains are important to document. Their input has guided decisions on documentation priorities, corpus planning, and even methodological choices. These collaborators play an important role in shaping the research trajectory, serving not just as data sources but as cultural interlocutors and knowledge custodians. This aligns with Ameka's (2006:70–71) call for native speakers to be seen as agents who "help shape the record" and whose contributions are essential to generating "optimal descriptions."

A third and especially vital role is that of the local research assistant, or Community Research Collaborator (CRC)—an individual who not only functions as a language consultant, combining aspects of both passive and active involvement, but also serves as a liaison and coordinator, managing field logistics, building trust within the community, and facilitating interactions between the external researcher and the broader speech community. The importance of this role cannot be overstated. In my fieldwork across Indonesia, the success and sustainability of language documentation efforts have often depended on the recruitment and training of at least one committed and capable local CRC. Such individuals become the anchor of long-term, locally embedded linguistic research, ensuring that the project's impact extends beyond the typical 2–3 year lifespan of an outsider's funded initiative. They help institutionalise documentation processes, contribute to the community's linguistic and cultural resilience, incorporate indigenous epistemologies and research methodologies, as well as support the emergence of future native-speaker linguists—thus embodying the kind of co-creative model advocated by Ameka (2006:95) and Tsikewa (2021), also echoed in my own work (Arka 2025).

In sum, meaningful linguistic documentation must be attuned to the diverse roles that native speakers and local communities can play. Each form of involvement, from the quiet contribution of a consultant in a recording session to the strategic guidance of a local CRC, enriches the descriptive record and strengthens its relevance for both academic and community goals. Our task as linguists is to recognise, respect, and cultivate these roles, ensuring that our work is grounded in reciprocal relationships and shared epistemic authority.

3.2. On the need of native speaker training in linguistics

The growing recognition of native speakers as epistemic collaborators in linguistic research must be accompanied by a serious commitment to training. While native speakers bring irreplaceable cultural knowledge, lived experience, and intuitive linguistic insight to the table, they often lack access to the analytical tools, methodological rigour, and theoretical frameworks that structure academic linguistics. As Felix Ameka (2006:75) notes, "the most powerful research machine imaginable is the trained linguist who is also a native speaker of the object language." Yet he also reminds us that being a native speaker linguist is a "mixed blessing" (2006:76): while the depth of intuitive knowledge may yield "flashes of insight," such insight can be compromised by theoretical bias or a narrow view shaped by the speaker's own idiolect or social positioning.

There are a growing number of native speakers of Indonesian languages (with myself as an example) who are committed to serious training in linguistic theory and method. As I have argued with Mullan (Arka & Mullan, to appear), grammars are not neutral inventories of structure—they are interpretive models that benefit from both linguistic theory and sensitivity to speaker-based social cognition. The trained native speaker, especially when equipped with

an understanding of grammatical architectures like LFG, can bring to light subtle yet systematic dimensions of meaning encoded in everyday language use, such as sociopragmatic distinctions and interactional motivations. These are frequently obscured in purely formal or typologically driven accounts.

In my own practice—working as a native-speaker linguist on Balinese and Indonesian, and as a non-native researcher on languages like Marori and Enggano—I have seen first-hand how training native speakers in basic linguistics transforms the documentation enterprise. The basic training in linguistics for local RAs enables them not only to contribute as informed consultants, but to take on active roles in data collection, transcription, analysis, and publication. It also creates a more equitable collaborative space, where descriptive and analytical decisions are co-owned. This training need not always take place within the traditional university setting. In the projects that I have led, we have pursued context-sensitive, field-based training—through hands-on workshops on ELAN and Praat, guided corpus annotation, collaborative text glossing, and participatory grammar-writing sessions. These activities build confidence, competence, and a sense of ownership, while demystifying the analytical process.

As for advanced training in linguistics**,** based on both my own academic training and extensive supervisory experience, I have seen that such training in linguistics—at undergraduate and postgraduate levels—enables native speakers to contribute meaningfully across multiple stages of research, significantly improving the efficiency of documentation and the depth and quality of analysis. This is particularly evident in my supervision of PhD students who are native speakers working on their own languages: their insider knowledge, combined with formal linguistic training, has led to richly informed and theoretically grounded descriptive work. However, such outcomes are only possible with sustained and targeted support**,** as advanced training—especially at the PhD level—is time-intensive, costly, and often takes several years to yield tangible results. Moreover, it is a major personal and academic undertaking, and in many communities, there may be few individuals who have both the capacity and willingness to commit to such a demanding path.

In general, through a combination of context-sensitive, field-based mentoring and structured analytic exposure, I have seen native speaker collaborators grow from data providers into confident co-analysts. For example, the development of the *pragmatic structure* (prag-str) module within LFG (Arka & Mullan, to appear, §5.2) owes much to my native-speaker intuitions and life-experience about culturally salient predicates—like the Balinese concepts of *menak* and *jaba*—which map directly onto syntactic and information-structural behaviour. Likewise, the multilingual Enggano dictionary (Rajeg et al. 2025) is an outcome of years of collaboration with local Enggano native speaker experts. Without a native and shared analytic vocabulary and metalinguistic awareness, these intuitions might never have found formal expression in grammar and dictionary.

For long-term benefits, training native speakers must go beyond technical skills in linguistics, and ideally include some training in related disciplines and leadership skills, with the hope that this will cultivate their capacity to work on their language and culture through broader interdisciplinary perspectives with practical benefits (e.g. standardisation and needs analysis for teaching material development and local literacy, as well as initiating/managing language maintenance or revitalisation programs). It should include critical reflection on linguistic ideology, dialect diversity, and the politics of representation: whose variety is being described and used as a “standard” to be taught at local schools? Who defines what counts as "core

grammar"? How do social categories like gender, caste, or kinship inform grammatical descriptions and their use? Is there a need to establish a local language committee to oversee the development of textbooks and a dictionary? Who decides membership and leadership in such a committee? As Ameka (2006:83–84) cautions, native speaker linguists can also fall into the trap of prescriptive bias or fail to reflect on the social variation within their language. This risk may be mitigated in the modern Indonesian context, where internet access is now available even in remote areas, enabling community members to access rich, interlinked resources as described by Musgrave & Thieberger (2021). Such community access and engagement allows users to move seamlessly among resources and to become aware of the linguistic variability in their language, which can then be included in examples, dictionary entries, and teaching materials. Thus, robust training programs must therefore engage students with theoretical pluralism, sociolinguistic sensitivity, and an awareness of the historical, political, and ecological dimensions of language work. This aligns with our argument (cf., Arka 2025; Arka & Mullan, to appear) that linguistic analysis must remain accountable to the community's perspectives and communicative norms, not just to the demands of typology or theory.

Ultimately, training native speakers in linguistics is not only about producing better grammars. It is about restructuring the ecology of linguistic science itself—expanding who gets to speak authoritatively about language, and what kinds of knowledge count. It is also key to sustaining long-term engagement: many of my most successful documentation projects owe their longevity not to the periodic visits of outside researchers, but to the persistence and leadership of trained local collaborators, including community-based research assistants who have grown into independent researchers. These individuals do not merely extend the reach of external researchers; they become institutional bridges and cultural stewards, managing corpora, coordinating fieldwork, and serving as mediators between local language ideologies and academic research goals. In short, training native speakers in linguistics is not only a methodological necessity—it is an ethical and epistemological investment in a more inclusive and co-creative model of linguistic science.

4. Non-Native Linguists and (Minority) Indonesian Languages

While the previous section focused on the contributions, capacities, and evolving roles of native-speaker linguists and community members in documentation, this section shifts perspective to my work as a non-native linguist documenting minority endangered languages across Indonesia. In this role, I am an outsider navigating unfamiliar linguistic ecologies and complex sociocultural terrains. The challenges in this role are considerable, ranging from the need to build trust with community members to managing expectations (particularly around financial compensation), and contending with residual suspicions associated with the term *proyek* in the Indonesian context. In this subsection, I first provide some historical background in §4.1, outlining the foundational contributions of non-native linguists—primarily missionaries—during both the pre- and post-Indonesian independence periods. In §4.2, I examine the challenges encountered by non-local linguists, whether Indonesian or foreign, in undertaking documentary research within present-day Indonesia.

4.1. Contribution of Non-Native Linguists in Indonesia: A Historical Overview

In reflecting on the history of linguistic research in Indonesia, I acknowledge that much of the foundational work on both dominant and minority languages has been carried out by non-native linguists, particularly foreign researchers, since the Dutch colonial period. These early efforts, beginning in the 19th century, were led primarily by Christian missionaries who had the training, resources, and institutional support that local communities at the time could not access. Figures such as H.N. van der Tuuk, who compiled a grammar and bilingual dictionary of Batak, and J.A. Verheijen, who worked extensively on Manggarai—including ethnobiological and ethnographic studies—set the stage for future linguistic work. Similarly, Petrus Drabbe contributed valuable grammatical descriptions and texts on the languages of Tanimbar and southern New Guinea. These early works—though shaped by religious and colonial motivations—provided a lasting record that continues to inform current linguistic and typological research (Arka 2025: 5).

This tradition of non-native linguistic engagement continued into the post-independence era. While Indonesia gained political autonomy in 1945, the linguistic research landscape remained heavily influenced by foreign institutions and researchers. In the second half of the 20th century, organisations such as the Summer Institute of Linguistics (SIL) played a major role in documenting local languages across the archipelago, particularly in remote regions. In some areas, SIL's presence lasted for decades before their activities were restricted due to political and religious sensitivities (Arka 2025: 6–7). Their contributions, however, were undeniably significant—not only in producing descriptive materials, but also in introducing methods of data collection and analysis that shaped future generations of linguists, including my own.

Since the turn of the 21st century, and particularly following the *documentarist turn* (cf. Himmelmann 1998; Woodbury 2003), a new phase of non-native linguists' engagement in Indonesia emerged—one shaped by modern, corpus-based, community-engaged language documentation. A milestone in this development, especially in Tanah Papua, was the establishment of the Center for Endangered Languages Documentation (CELD) at the University of Papua in Manokwari. This centre was launched alongside the Wooi Documentation Project, funded by DoBeS (*Dokumentation Bedrohter Sprachen*), under the supervision of Nikolaus Himmelmann. More than a single project, this initiative set a precedent for building both documentary corpora and local capacity. CELD went on to coordinate major documentation efforts, including the "highlands summit" program in 2012 (targeting Yali, Walak, and Western Dani), the 2014 projects on Iha and Mpur, and the 2019–2022 ELDP-funded documentation of Konda (Yaben) and Kais.

Himmelmann's role was pivotal not only in supervising individual projects but also in fostering a national network of trained Indonesian linguists. His leadership extended to capacity-building initiatives such as the landmark national language documentation workshops held in Bali in 2006–2007 (Florey & Himmelmann 2010), in which I participated as an instructor. This model—bringing together participants from across Indonesia, including both minority and non-minority language speakers—proved transformative: several participants went on to complete PhDs and become leaders in their own right, among them Yusuf Sawaki, now Director of CELD. The ripple effects of this model can still be seen today in similar workshops run by other institutions, including the *Pusat Riset Preservasi Bahasa dan Sastra* (Centre for the Preservation of Language and Literature) BRIN (National Agency for Research and Innovation). These workshops have combined technical training in tools

such as ELAN and FLEx with guidance on grant writing and project design, a critical combination for enabling Indonesian linguists to secure competitive funding and lead independent documentation projects on the international stage.

At the same time in more recent years, I have seen a gradual but important shift: more Indonesian linguists—especially those working for Badan Bahasa or as academics at universities—have taken up the task of documenting and describing the languages of the archipelago. However, these researchers are often still non-native to the languages they study, especially in the case of minority languages spoken in remote or marginalised regions. This reflects a persistent structural limitation: the communities where many of these languages are spoken often lack access to higher education and have limited pathways for their members to become trained linguists. In many of the places where I have worked—such as Merauke, central Flores, and the Barrier Islands—there simply were no local individuals from the minority speech communities with the linguistic training and experience necessary to lead research projects.[2]

Because of this, modern language documentation in Indonesia—including many of my own projects—has continued to depend on external support, often through international collaborations funded by bodies such as ELDP, DoBeS, and the AHRC. These collaborations have resulted in rich, publicly available multimedia corpora and have contributed to typological and theoretical inquiry, as well as to language maintenance. Yet they also highlight the ongoing need for capacity building and the continued relevance of non-native linguists in supporting this work (Arka 2008a, 2025; Florey and Himmelmann 2010). From a historical and practical standpoint, the role of non-native linguists remains central—not as replacements for local leadership, but as collaborators and facilitators who can help fill critical gaps while working towards a more inclusive future for linguistic research in Indonesia.

4.2. Challenges of Non-Local Linguists in Linguistic and Documentation Research

While non-local linguists—especially foreign scholars—have made essential contributions to the documentation of minority languages in Indonesia, their role is not without complexity. As outsiders, they face a range of methodological, ethical, and practical challenges that shape the effectiveness, sustainability, and reception of their work. These challenges must be understood not as individual shortcomings, but as structural and relational dynamics

[2] In recent decades, an increasing number of Indonesian linguists—including those from non-minority speech communities such as Nias—have completed advanced degrees (MA and PhD) in linguistics, particularly from Indonesian universities . This growing cohort represents an important potential force for advancing language documentation within Indonesia. However, based on my experience serving on the international ELDP panel reviewing grant applications from Indonesia and Southeast Asia more broadly, it is evident that many of these emerging scholars still require greater exposure to collaborative research practices and training in the full range of documentation methodologies in order to compete effectively on the international stage. To date, only a small number of Indonesian linguists have successfully secured highly competitive international grants such as ELDP and ELF (Endangered Language Funds) funding independently—that is, without the partnership or mentorship of foreign collaborators. Recognizing this gap, initiatives were launched to support local capacity building, including the ELDP- and Volkswagen-sponsored language documentation workshops held in Indonesia (Florey and Himmelmann 2010), which, as mentioned above, inspired similar workshops organized by (local) Indonesian institution. Recently (2019-2025), the OCSEAN, an MSC RISE exchange program funded by the European Commission (https://www.ocsean.eu/), also includes capacity building for Indonesian scholars through hands-on training in data collection and analysis, developing transferable skills, and enabling community-led research with lasting impact.

embedded in sociocultural, educational, and historical contexts (Arka 2018; Sawaki & Arka 2018:259–260).

4.2.1. The Role of the Outsider

Non-local linguists often operate with positionality shaped by their status as outsiders—linguistically, socially, and institutionally. In my experience, the success of documentation projects is often significantly influenced by how non-local researchers manage this position of difference. Gaining access to communities, building trust, and negotiating one's role often take considerable time and sensitivity. In remote and marginalised communities across eastern Indonesia, where linguistic diversity is high but educational and infrastructural support is low, researchers may be met with suspicion, indifference, or misplaced expectations—especially in places historically conditioned by missionary or government development projects (Sawaki & Arka 2018:263–264).

This positionality also extends to practical fieldwork issues. In many areas, foreign linguists are better trained, resourced, and institutionally supported, making them more capable of initiating and sustaining documentation projects. However, this advantage can also deepen dependency and limit opportunities for local scholars and communities to assume leadership roles. As I note in Sawaki & Arka (2018:260), "foreign linguists are better equipped for fieldwork… [and] are in a better position to generate publications, organize seminars, and facilitate training activities." This asymmetry underscores the importance of recognising and mitigating the implications of outsider-led documentation.

4.2.2. Methodological and Ethical Challenges

One of the most persistent challenges for non-local linguists is negotiating expectations. Communities often associate the term *proyek* 'project' with government-initiated programs that distribute funds with little accountability, leading to the assumption that any outsider project is a cash-disbursement activity. This has led to misunderstandings and frustration when documentation projects do not align with these expectations. As a strategy, I and colleagues in the Center for Endangered Languages Documentation (CELD), for example, have found it more effective to use the terms *program* or *aktivitas* 'activity' to manage expectations and communicate goals more clearly (Sawaki & Arka 2018:263).

Ethical challenges also arise in ensuring that documentation is not extractive, highlighting a clear call for decolonial, community-engaged practice.[3,4] These challenges echo the critiques raised by Tsikewa (2021) and Hudley et al.(2024), who argue that dismantling these entrenched imbalances requires systemic reform in how linguistics is taught, funded, and institutionally rewarded. Misinterpretation of linguistic data due to unfamiliarity with local norms, culturally inappropriate recordings, or exclusion of relevant community voices can all

[3] The move away from extractive practices in Indonesia, as in many other post-colonial contexts, reflects a broader reaction against the exploitative resource extraction characteristic of European colonial rule. In linguistics, this shift aligns with the wider agenda of decolonizing the field. In contemporary Indonesia, it also entails resisting extractive approaches by central authorities or Jakarta-based institutions toward local communities, echoing broader debates on local and regional autonomy (Arka 2008a).

[4] However, it should be noted that linguistic fieldwork or documentary linguistics exists on a continuum from traditional, linguist-focused models—where community members are mainly sources of data—to fully collaborative, community-based models that center community agency, expertise, and priorities (Czaykowska-Higgins 2009). Methodologically, research relationships must be tailored to the specific goals, needs, and aspirations of each community; what is collaborative and appropriate in one context may not be so in another.

compromise the integrity of the documentation. Sawaki & Arka (2018:262) also report how local rivalries and clan-based tensions may complicate the recruitment of consultants or the sharing of cultural knowledge. Furthermore, some community members may engage in projects primarily for financial compensation, contributing data that lacks contextual or narrative depth. Together, these challenges point to the need for dialogue-based, community-participatory methods that foreground reciprocal understanding and collective goals. In any case, when navigating challenging local dynamics (e.g., internal community divisions, shifting allegiances or unpredictable politics; cf., Dobrin and Schwartz 2016) , it is wise to remain flexible and guided by ongoing reflexivity and sensitivity.

4.2.3. Learning Through Ethnographic Engagement

Despite these challenges, non-local linguists can contribute meaningfully when they engage in ethnographic listening and long-term collaboration—particularly to gain a deeper understanding of the social meanings behind linguistic choices (Briggs 1986: 3–5), especially in contexts of multilingualism and identity negotiation (Dobrin and Berson 2011). Many valuable insights emerge not from formal elicitation sessions but from participating in rituals, attending communal events, or sitting with elders who share linguistic and ecological knowledge—the very goals of language documentation and documentary linguistics (cf., Himmelmann 1998; Woodbury 2011, 2003). For example, in our ethnobiological work with Marori and Wooi speakers, we documented not only the language itself but also how it encoded social-ecological knowledge about sago and mangrove ecosystems in traditional and contemporary contexts —insights that could only be accessed through culturally immersive fieldwork (Hisa, Mahuze, and Arka 2017:258; Sawaki and Arka 2018).

A key lesson is the importance of working with and through local leaders and intermediaries—individuals who understand both community values and research goals. Involving respected local figures early on in the project and inviting them to help shape participation strategies mitigates conflicts and enhances buy-in (Sawaki & Arka 2018:263–264). Moreover, when trained locals are empowered to carry out documentation themselves, the work becomes more sustainable and more deeply rooted in community perspectives.

5. Supervising Indonesian Native-Speaker Students: A Transformative Process

One of the most promising developments in Indonesian linguistics in recent decades is the growing number of native-speaker students engaging in the documentation and analysis of their own (heritage) languages. I use *heritage* here in the sense of Polinsky & Kagan (2007:369) and Montrul (2016:16)—that is, languages acquired in the home during childhood but which, due to a later shift in language dominance, are no longer the individual's strongest or primary language. Some of these students could be described as *semi-native* speakers: while they possess valuable insider cultural knowledge and at least partial competence in the language spoken by their parents, they are now dominant in Indonesian, the national language, whether because they moved with their families to urban centres such as Java or because Indonesian became the lingua franca among their peer group even within their home communities. In such a situation, full fluency is often confined to the older generations. When equipped with robust analytical tools and theoretical frameworks**,** such students are uniquely positioned to contribute to a form of scholarship that bridges local epistemologies and academic rigour in their research. In my supervisory role over the past two decades—primarily through Udayana University and the Australian National University (ANU)—I

have found the supervision and long-term mentoring of these students to be a deeply transformative process**,** both for the students and for myself as a supervisor.

The needs of native-speaker students vary depending on whether they pursue graduate education domestically in Indonesia or overseas. For those studying abroad, particularly in English-speaking countries[5] with high-ranking institutions (e.g., Australia, the UK, and the US), additional preparation is often required. This includes securing competitive scholarships, meeting English proficiency standards, and adjusting to different academic cultures and expectations. This preparatory stage, based on my experiences helping Indonesian students be accepted at the ANU, takes years. In contrast, these requirements are often less pressing for students undertaking postgraduate studies domestically. However, even within Indonesia, the rising emphasis on international academic standards—such as the requirement to publish in peer-reviewed journals for thesis completion—has increased pressure on students and, by extension, on supervisors, to provide more structured academic support.

In today's globalised academic landscape, students are increasingly expected to demonstrate scholarly productivity and competitiveness on the international stage, even for local academic and research positions in Indonesia. While top-ranked international universities may not formally require publications during doctoral study, publishing in reputable journals has become common practice as it signals a strong academic trajectory and enhances employability in a competitive job market. For many native-speaker students, particularly first-generation scholars, guidance in navigating the publication process is essential and becomes a core part of the supervisory relationship.

From a supervisor's perspective, supporting PhD students—especially those working on their own languages—has been a rewarding and intellectually generative experience. My large research projects have always been conceived with local capacity building in mind; hence involving Indonesian PhD students (e.g., the *Barrier Islands Project*), or Indonesian Postdoctoral fellows (e.g. the *Enggano Project*) in the team. These Indonesian scholars are not only pursuing independent PhD goals, but are also contributing meaningfully to collaborative, team-based research with shared objectives**.** Their projects serve as integral components of broader research and documentation efforts, and they benefit from participating in collective activities such as joint fieldwork, data processing, analytic workshops, and co-authored publications.

Such collaborative models provide a dynamic research environment in which students are mentored to work both independently and as team players. They are trained across all stages of research—from planning and fieldwork to data analysis and dissemination. This experience equips them with critical academic and professional skills, preparing them to become local agents and future leaders in their fields, at least in Indonesia. From my

[5] While this discussion focuses on English-speaking countries because English proficiency is a universal requirement for publication in leading international journals (Curry and Lillis 2018:3-4), the point applies more broadly. In countries where the medium of instruction is not English, students generally need English in addition to the national language. For example, in Germany or France, graduate students must typically demonstrate competence in the local language for everyday life and administrative purposes, while also maintaining a high level of academic English. In contrast, in some European contexts such as the Netherlands, many graduate programs—especially in the humanities and social sciences—are fully taught in English, so Dutch is not an admission requirement. Similarly, in Japan, while most universities require Japanese proficiency, a number of institutions offer postgraduate programs entirely in English, admitting students without prior Japanese language skills.

standpoint as a supervisor, the benefits are equally significant: these collaborations have generated insights that would not have emerged through individual effort alone. Regular individual supervision and group meetings have fostered a vibrant intellectual exchange, enriching the research and strengthening the outcomes of both student and team-led work.

These experiences underscore the vital role that capacity building through graduate supervision plays in the future of linguistic scholarship in Indonesia. Supporting Indonesian native-speaker students through structured mentorship, international exposure, and collaborative research opportunities not only empowers individuals but also contributes to the broader goal of sustaining and revitalising Indonesia's rich linguistic heritage. By investing in this next generation of scholars—who are deeply rooted in their communities and increasingly connected to global research networks—we lay the groundwork for a more inclusive, resilient, and locally driven model of language documentation and analysis. This approach, I believe, is essential for ensuring the long-term impact of our collective work, both within academia and in the communities whose languages we seek to understand and support.

6. Co-Creation in Practice, Case Studies: Balinese, Marori and Enggano

In this section, I reflect on my collaborative experiences as both a native-speaker linguist and a non-native researcher, to illustrate how co-creation between native and non-native linguists can lead to richer, more nuanced, and culturally grounded language descriptions. Drawing on three case-study projects—Balinese, Marori and Enggano—I highlight how collaborative, interdisciplinary, and cross-positional approaches to language documentation and analysis have shaped my research practice over the years. The first case (§6.1) describes my experience as a native speaker of Balinese working with both Balinese and non-Balinese linguists. The second case (§6.2) turns to my role as a non-native researcher working with the native Marori and Enggano communities and speakers.

6.1. Working on Balinese as a Native Speaker: From Apprenticeship to Mentorship

My PhD research at the University of Sydney, later published as Arka (2003), marked the beginning of a four-year intensive co-creation process with non-Balinese linguists—particularly my supervisors, Professor William A. Foley and Professor Jane Simpson. This formative period was not just about completing a dissertation; it was a deeply collaborative academic apprenticeship that shaped me into an independent researcher. Beyond the dissertation, I co-authored several self-contained papers addressing key issues from the project. These included joint works with supervisors and other linguists, such as Weschler & Arka (1998), Arka & Simpson (1998, 2008), and Arka & Manning (2008). Each collaboration brought new insights, methodologies, and disciplinary cross-fertilisation, pushing the boundaries of how Balinese grammar could be analysed and theorised.

Two critical lessons emerged during the early phase of my academic career. First, I became acutely aware of the limitations of native-speaker intuitions—a point also underscored by Ameka (2006). As native speakers, we often carry unconscious biases: we may privilege our own dialects, fail to account for variation, or overlook linguistic phenomena that seem self-evident and thus remain unarticulated. This experience taught me the importance of data representativeness, particularly the need to test linguistic judgments across speakers of different Balinese dialects. Additionally, working collaboratively with non-native linguists

helped surface culturally significant distinctions and grammatical patterns that I, and other Balinese linguists, had previously under-analysed or overlooked.

This issue becomes particularly evident when examining traditional Balinese grammars written by fellow native-speaker linguists. For example, Oka Granoka et al. (1985) illustrates many of the limitations mentioned above, especially a tendency to overlook language-specific peculiarities, such as the existence of an undergoer voice distinct from the passive voice, which would likely be recognised if the linguist had a solid understanding of Balinese's typological profile within the Austronesian family. The treatment in Oka Granoka et al. is notably cursory—a feature common to many grammars produced under the auspices of Indonesian *Badan Bahasa*, which often follow a uniform, structurally top-down template rooted in basic structuralist linguistics. For example, chapters on syntax typically present basic constituent or grammatical structures—clauses, phrases, subjects, predicates, complements—accompanied by a single, simple example per structure. These examples appear to be self-elicited, lacking contextual grounding in actual language use, and notably, they are not drawn from any corpus of natural language use.

Although the grammar is written in Indonesian and aimed at both Balinese and non-Balinese Indonesian readers, the examples are presented without interlinear glosses, which hampers grammatical insight—only free translations are provided. Crucially, salient grammatical phenomena are often either under-described or completely absent. Consider the following contrasting examples, which the grammar uses to illustrate active and passive voice:

1 a. *Gegendong-e* *maek-in* *I* *Made Diarsa*
beggar-DEF AV.close-APPL ART name
'The beggar approached I Made Diarsa'.

b. *I* *Made Diarsa* *paek-in-a* *baan* *gegendong-e*.
ART name close-APPL-PASS by beggar-DEF
'I Made Diarsa was approached by the beggar.'
(Oka Granoka et al. 1985:86; glosses added)

While these examples do reflect a voice contrast, the grammatical analysis provided is superficial. The discussion does not address the fact that the suffix *-a*, glossed here as PASS, can also appear in structures that lack an overt PP. In such cases, *-a* is ambiguous, functioning as passive marker as in (1b) above, or as a bound pronominal clitic (*=a*), forming a morphosyntactically distinct construction known in Balinese as the Undergoer Voice or Objective Voice (UV/OV) (Arka 2003, 2008b)—a voice type commonly found in Indonesian-type languages. Furthermore, the grammar omits any discussion of the *ka-* passive, which I (Arka 2003) have described as the "high passive."

Recent studies reveal that these *-a* and *ka-* passives coexist as integral components of the Balinese voice system, reflecting complementary functions with subtle distinctions shaped by semantics and social cognition (Arka and Mullan to appear). For instance, the *-a* passive is generally preferred in contexts requiring agentivity or control (e.g., purposive clauses, negative imperatives), whereas the *ka-* passive, especially in its non-agentive or stative usage, aligns with less volitional, achieved-state readings (Udayana, Rajeg, and Puspani 2025).

Another important lesson I recognised in my early academic career is the essential role of rigorous training across multiple strands of linguistics—including theoretical linguistics—for

native-speaker linguists who wish to contribute meaningfully to the field. The ability to formulate precise research questions, engage critically with data, structure coherent arguments, and produce scholarly work—whether in presentations, publications, or theses—requires far more than intuitive knowledge. The earlier discussion of passive constructions in Balinese illustrates this point well: early grammatical descriptions by native-speaker linguists failed to capture the full complexity and typological significance of the language's passive constructions. In contrast, more recent work by fellow Balinese linguists has begun to address these gaps—progress that has been enabled by advanced linguistic training. These developments clearly underscore the importance of such training for native-speaker linguists (see also §3.2).

Importantly, this kind of training involves more than theoretical instruction: it demands exposure to diverse linguistic models, academic writing conventions, and hands-on analytical skills, with computer-based technical proficiency. In today's digital context, this includes familiarity with widely used tools in language documentation and analysis, such as ELAN, FLEx, and Praat. In addition, for archiving purposes and following the "quartet" model proposed by Musgrave & Thieberger (2021), robust language documentation increasingly depends on interoperable, machine-readable formats that facilitate navigation and re-use across multiple documentation components. My own experience as a PhD student and as a team leader in ELDP-funded major documentation projects in Indonesia laid the foundation for this awareness and has since shaped how I mentor students and junior colleagues—particularly native speakers documenting their own languages, as discussed further in §5.

Now, as a senior scholar with over two decades of research experience on Balinese and other languages of Indonesia, I often engage in co-creation as a mentor, collaborator, and academic partner. Whether working with fellow Balinese linguists or non-Balinese colleagues, the co-creation process typically involves joint efforts in project conceptualisation, securing funding, developing methodologies, supervising data collection and analysis, and co-authoring publications and presentations. These collaborations are not hierarchical but reciprocal, drawing on our respective strengths and perspectives to produce work that is both theoretically robust and locally grounded. The lessons learned from working on Balinese continue to inform my approach in other research contexts, including my work with the Enggano community, which I turn to in the following subsection.

6.2. Working as an Indonesian Balinese Linguist with Enggano and Marori Speakers

In contrast to my experience as a native speaker working on Balinese, my engagement with Enggano (Austronesian) and Marori (Papuan) has placed me in the role of a non-native linguist, facing both familiar and unique challenges common to outsider researchers (cf. Ameka 2006). These challenges have been both linguistic and non-linguistic in nature, requiring thoughtful strategies, collaborative structures, and a flexible, ethnographically informed approach to ensure a meaningful and respectful documentation process.

Linguistically, the most immediate hurdle was the absence of native-speaker intuitions, which non-native linguists naturally lack. Unlike Balinese, where I could draw on deeply internalised knowledge of structural variation and subtle semantic distinctions, in Enggano and Marori, I have had to learn the language from scratch—decoding forms, functions, and usage patterns through intensive collaboration with local speakers. A cornerstone of this

process has been the involvement of dedicated local research assistants who then became CRCs, Mr. Engga Z. Sangian (for Enggano)**,** Mr Agus Mahuze and Mr Maximus Ndiken (for Marori). Their contributions have been essential. Over time, Engga in particular, has acquired basic linguistic training, enabling him to play a central role in interpreting grammatical nuances, identifying variation across dialects, and refining analytical distinctions. Together with other project team members, we have worked to develop a large spoken corpus, representing different villages and dialects of Enggano. This corpus—comprising naturalistic, transcribed, and translated texts—has become an indispensable resource for detecting structural patterns and meaning contrasts that neither outsider nor insider intuitions alone could fully capture.

One of the most rewarding aspects of this work has been the co-creative relationship between non-native and native researchers. Our team is deliberately structured to be complementary: my theoretical and methodological expertise is paired with local CRCs' deep cultural and linguistic knowledge. This collaboration helps mitigate the risks of both native-speaker blind spots (cf. §6.1) and outsider misinterpretation. For the Enggano project in particular, Engga's expanding skills—ranging from data recording and metadata management to transcription, translation, and even academic writing—have allowed him to co-author papers and contribute to publications that reflect both scholarly standards and community-based perspectives; Rajeg et al (2024) and Sangian et al (2024). His involvement exemplifies the transformative potential of training and capacity building, echoing the models discussed in §5.

For the Marori Project, the impact is promising and beyond linguistics. Long-term mentoring in my Marori documentation projects has enabled local CRC, in particular Agus Mahuze, to transform technical skills in recording, transcription, and data management into broader leadership, project management, and community engagement capacities. These skills have sparked creative initiatives beyond linguistics, as seen in Agus Mahuze's establishment of *Mahuze Mandiri* and *Tunas Mandiri*, NGOs that combine cultural documentation with ecotourism and agricultural ventures in Kampung Wasur, securing funding, government support, and inspiring similar projects in the region (Arka 2025). For example, as seen in Figure 1, Kampung Wasur's local nursery showcases native flora, including wild orchids, reflecting the community's deep connection to the land and its traditions while drawing visitors from near and far. This trajectory—from skills transfer to innovation—illustrates how capacity building in documentation can generate lasting cultural and economic benefits for local communities.

Figure 1. Local garden featuring native orchids as a tourist attraction, with Agus pictured serving as a tour guide.

My approach has also been ethnographic, recognising that grammatical forms and cultural practices are tightly interwoven. As Ameka (2006) argues, outsider linguists working in unfamiliar cultural settings are prone to missing culturally embedded meanings unless guided by community members. Local CRCs' participation has been crucial in bridging this gap: they not only explain linguistic usage but also situate it within local belief systems, rituals, and social norms. This kind of grounded interpretation cannot be gleaned from forms alone—

it requires lived cultural experience. The project outcomes involving ethnographic insight include Hisa, Mahuze and Arka (2017; 2018), Arka et al (forthcoming), and Rajeg et al (2025), among others.

Importantly, my positionality as an Indonesian linguist of Balinese background has conferred certain advantages. Although I am not a native speaker of Enggano and Marori, my fluency in Indonesian, which functions as the lingua franca across the archipelago, facilitates smooth communication and collaboration. Moreover, my familiarity with Indonesia's ethnolinguistic diversity has helped me build rapport and navigate local protocols with cultural sensitivity. Unlike foreign researchers, I do not face the same degree of cultural distance or bureaucratic restriction, allowing for sustained and reciprocal engagement with the community throughout the project lifecycle.[6]

In sum, the Enggano and Marori case studies demonstrate the value of co-creative, hybrid teams**,** where outsider and insider perspectives are not only combined but continually negotiated in the research process. The integration of linguistic training, corpus development, ethnographic insight, and mutual respect has enabled our projects to meet—and in many respects exceed—their original goals. This experience affirms the argument made throughout this chapter: that the most productive and ethically sound linguistic work emerges when native and non-native linguists collaborate as equal partners, drawing on their respective strengths to produce scholarship that is both academically rigorous and locally meaningful and useful.

7. Recommendations for Practice, From Extraction to Co-Creation: Strategies for Ethical and Effective Collaboration

The foregoing sections clearly underline the point that to foster truly inclusive linguistic science, impactful language documentation research depends not just on scholarly rigour, but on cultivating co-creative, ethically grounded relationships with local communities. This means moving beyond short-term, outsider-driven extraction toward models of shared responsibility, local leadership, and strategic capacity building. In contexts like Indonesia—marked by linguistic, ethnic, and political complexity—these collaborative strategies must be finely attuned to local ecologies, which must also respect national language ideologies (Arka 2013:74–76).

[6] A reviewer raised the point that being Indonesian can be a disadvantage in certain parts of Indonesia, particularly in Tanah Papua. I appreciate and fully acknowledge this important observation. In some areas—especially the central highlands—where Indonesian state presence is often perceived as colonial occupation, being Indonesian can indeed be a liability. In extreme cases, this has resulted in harassment and even lethal attacks on Indonesian civilians, including teachers, health workers, and, in other sectors, policemen and soldiers. While I have not personally experienced such hostility in Tanah Papua, nor am I aware of fellow Indonesian linguists encountering it, my knowledge is limited to colleagues—both Papuan and non-Papuan—who are typically affiliated with government institutions (e.g., universities, BRIN, or local branches of the National Language Board/Balai Bahasa). To my knowledge, none of these linguists have faced the kinds of threats or violence noted by the reviewer. Nonetheless, I recognise that such risks are real and must be taken into account when considering the positionality and safety of Indonesian researchers working in politically sensitive areas of Tanah Papua.

One foundational strategy is transparent expectation management from the outset. As noted in Sawaki & Arka (2018:263–264*)*, avoiding terms like *proyek*, which local communities often associate with state-driven aid programs, helps minimise misunderstandings. Instead, we use terms like *aktivitas* or *program* and host community meetings to explain project goals, timelines, and mutual benefits. These meetings are opportunities to build trust and include elders and younger, educated locals who may later become co-researchers or documentation leaders.

Crucially, recruiting and mentoring a capable local CRC has proven key to success. In my Marori work, Mr. Agus Mahuze became more than a consultant—he was a cultural liaison, logistics coordinator, and linguistic co-analyst (Arka 2018; Hisa, Mahuze, and Arka 2017; Hisa, Mahuze, and Arka 2018). His long-term involvement ensured continuity and legitimacy within the community. As I argue in Arka (2025) and Arka & Mullan (to appear), empowering such local figures supports not only more nuanced and accurate analyses, but also community investment in the outcomes of the research.

Training, however, must be contextualised within the broader ecology of language endangerment, ethics and curriculum reform. As I note in Arka (2013:81-82), minority language speakers often do not perceive language loss as an immediate threat. Instead, Indonesian is embraced as an "opportunity"—a tool for education, mobility, and survival. Language maintenance can seem counterintuitive, especially to youth, unless it is explicitly linked to broader aspirations for wellbeing and identity continuity. In this light, capacity-building must address not only linguistic skills but also strategic motivation, helping communities understand how local language knowledge can coexist with, and even enhance, broader life goals. Furthermore, recent work highlights the importance of embedding decolonial and Indigenous research methodologies within linguistics curricula (Tsikewa 2021; Hudley et al. 2024), and of reframing fieldwork as a collaborative, socially accountable practice, with institutional recognition of community-driven scholarship as central to the discipline. Building on these insights, I argue that incorporating community voices is essential at every stage—from project conception to implementation, including capacity-building training—to ensure sustainable local collaborations.

In line with this emphasis on sustainability and community collaboration, capacity-building must be paired with practices that ensure research is non-extractive and gives back to the community—particularly in relation to the materials collected. This includes clear communication from the outset about data ownership, long-term accessibility, and how materials will be used after the project's completion. In both the Marori and Enggano projects, local communities were informed early on that all recordings and other data would be collected under strict ethical guidelines and archived in secure online repositories such as PARADISEC[7] and ELAR[8], with the community retaining the right to decide levels of public access. In practice, only a few individuals expressed interest in receiving copies of recordings, often for personal reasons (e.g., as mementos of family members who had participated as language consultants). Nonetheless, both projects have made it a priority to leave copies of recordings and related data with local communities, where they are now managed by CRCs, reflecting our commitment to socially responsible, non-extractive research, in line with Hudley et al.'s (2024) call for community members to act as long-term stewards and decision-makers in linguistic projects. In the Marori context, materials relating

[7] https://www.paradisec.org.au/
[8] https://www.elararchive.org/

to ethnobotany and ritual sites have also proven valuable for non-linguistic purposes, particularly in supporting eco-tourism and Marori Cultural Centre initiatives led by Agus Mahuze. For Enggano, while we were unable to obtain the early 20th-century legacy recordings collected by German linguist Kähler (Kähler 1975), we were able to return copies of his legacy texts to the Enggano community.

A critical insight from Arka (2013) is that strategic language work in Indonesia is filtered through national ideologies like *Bhinneka Tunggal Ika* 'Unity in Diversity' and political constructs such as "threat" vs. "opportunity." Activities that appear benign or beneficial to linguists may be misinterpreted—especially in sensitive regions like *Tanah* Papua, where any assertion of cultural identity risks being classified as *ancaman* 'threat' by the state (Arka 2013:85–86). While in principle political neutrality is desirable, in practice this is not always possible or even safe, particularly where research permits and long-term access depend on maintaining good relations with multiple stakeholders, including government actors. My own extended fieldwork experience in Merauke—a region with a strong government presence—has shown that early engagement with authorities (e.g., *Kesatuan Bangsa* (*kesbang) or* 'National Unity*'* offices) and symbolic gestures—like displaying presidential photos in workshops— can be crucial for securing institutional support and minimising political friction. At the same time, I recognise that in other areas, particularly in parts of the central Papuan highlands where armed conflict persists, such displays could have the opposite effect and place researchers at risk. In these contexts, the safest and most productive strategy is to follow local advice on the best approach for maintaining long-term engagement not only with the speech community but also with local authorities, with the aim of gaining their support for work that ultimately benefits the community.

Finally, the choice of theoretical framework also plays a crucial role. As noted earlier (cf. Arka & Mullan, to appear), modular architectures such as LFG offer the flexibility to incorporate culturally grounded concepts—such as Balinese speech levels and deference or politeness markers—within a formal analytical model. However, one of the strengths of LFG is that, due to its modular design, it can be adopted in a simplified form, allowing linguists to capture rich, language-specific insights without explicating or relying on the full formal apparatus. This adaptability makes it especially valuable in collaborative settings: it enables native-speaker collaborators—even those without formal linguistic training—to recognise their intuitions and cultural knowledge in the resulting grammatical description. In turn, this strengthens their sense of ownership and fosters deeper engagement with the research process.

In sum, successful collaboration in language documentation—especially for non-local linguists—requires a dynamic interplay of cultural sensitivity, institutional negotiation, theory-practice integration, and capacity building. These strategies help transform documentation research from a one-time research intervention into a sustainable, community-driven process of knowledge preservation and revitalisation.

8. Conclusion

This chapter has explored how collaborations between native- and non-native-speaker linguists can diversify and deepen linguistic science. Drawing on personal experience and

institutional reflections from over two decades of research in Indonesia and echoing ideas first by Ameka (2006) and also later by others (Florey and Himmelmann 2010:122; Dobrin and Berson 2011; Czaykowska-Higgins 2009; Bischoff and Jany 2018, among others), I have argued that inclusive, co-creative models of linguistic description are not only possible but essential—particularly in linguistically rich, socially complex, and historically under-resourced contexts like Indonesia.

Several key lessons have emerged. First, the success of linguistic documentation efforts depends fundamentally on rethinking who is positioned as a knowledge producer. Native speakers are not merely data sources; when equipped with the right tools, training, and support, they become co-analysts, co-authors, and stewards of their linguistic heritage. Their insights not only enhance the descriptive depth of linguistic work but also ground it in culturally embedded, socially meaningful knowledge systems.

Second, rigorous linguistic training—whether field-based or academic—plays a transformative role in enabling local collaborators to move from participation to co-creation and leadership. This training includes not just linguistic instruction (from descriptive to typological and theoretical) but also hands-on experience with analytic tools, scholarly writing, and interdisciplinary perspectives. As demonstrated through the Balinese, Marori, and Enggano case studies, such capacity-building efforts yield outcomes that are richer in both scientific and community impact.

Third, the role of non-native linguists remains significant—but not as extractive outsiders. Rather, they function best as facilitators, collaborators, and bridge-builders—particularly where local capacity is still emerging. When practiced ethically and reflexively, non-native participation can help catalyse long-term, locally anchored research ecologies. The experiences shared here also affirm that field-based training and mentorship models—especially those fostering continuity through local research assistants/collaborators or graduate students—are vital for embedding documentation within communities over time.

Finally, while acknowledging that research is inherently non-neutral (cf., Czaykowska-Higgins 2009), this chapter has underscored that linguistic research today must be reimagined as a co-constructed, socially accountable practice. This view, increasingly shared across the discipline (cf. Tsikewa 2021, Hudley et al 2024), aligns with broader shifts in modern documentary and field-based linguistics: away from extractive paradigms and toward sustained, reciprocal partnerships that reflect a plurality of epistemologies. Such a shift requires structural changes—not only in how projects are designed and funded, but also in how institutions recognise and support the long-term engagement needed to produce collaborative, ethical, and impactful research. This movement toward decolonizing research praxis ensures that community voices are central to linguistic knowledge production, echoing Hudley et al.'s (2024) call for fundamental transformations to the structures, incentives, and epistemologies that shape the discipline.

In short, the lessons learned from the Indonesian context offer a compelling model for the emerging paradigm in linguistic science: one that is pluralistic, interdisciplinary, co-creative, and grounded in mutual respect. As we move forward, diversifying science must mean more than increasing representation; it must entail transforming the very practices through which knowledge is made—ensuring that the voices and values of speaker communities are central to how we document, analyse, and understand the world's languages.

References

Ameka, Felix K. 2006. "Real descriptions: Reflections on native speaker and non-native speaker descriptions of a language." In *Catching language : the standing challenge of grammar writing*, edited by Felix K Ameka, Alan Dench and Nicholas Evans. Berlin: Walter de Gruyter.

Arka, I Wayan. 2003. *Balinese morphosyntax: a lexical-functional approach*. Canberra: Pacific Linguistics.

Arka, I Wayan. 2008a. "Local autonomy, local capacity building and support for minority languages: field experiences from Indonesian." *Language Documentation & Conservation Special Publication* 1 66-92.

Arka, I Wayan. 2008b. "Voice and the syntax of =a/-a verbs in Balinese." In *Voice and grammatical relations in Austronesian languages*, edited by P.K. Austin and S. Musgrave, 70-89. Stanford: CSLI.

Arka, I Wayan. 2010. "Ritual dance and song in language documentation: vera in Rongga and the struggle over culture and tradition in modern Manggarai-Indonesia." In *Language endangerment in the Austronesian world: challenges and responses*, edited by M. Florey, 90-109. Oxford: Oxford University Press.

Arka, I Wayan. 2013. "Language management and minority language maintenance in (eastern) Indonesia: strategic issues." *Language Documentation and Conservation* 7:74-105.

Arka, I Wayan. 2015. "On the Dynamics of Glocalisation and Minority Language Conservation in Contemporary Indonesia." In *International Conference on Language, Culture and Society*, edited by Obing Katubi, 1-10. Jakarta: LIPI.

Arka, I Wayan. 2018. "Reflections on the diversity of participation in language documentation." *Language Documentation & Conservation* Language Documentation & Conservation Special Publication No. 15, edited by Bradley McDonnell, Andrea L. Berez-Kroeker, and Gary Holton:132-139.

Arka, I Wayan. 2025. "Language documentation and the multi-dimensionality of capacity building: framing research diversity in an Indonesian ethno-ecological context." In *Projecting Voices: Essays in language and linguistics in honour of Jane Simpson*, edited by C O'Shannessy, J Gray and D Angelo. Canberra: Asia-Pacific Linguistics.

Arka, I Wayan, Arono, Dendi Wijaya, and Engga Zakaria Sangian. forthcoming. "The Socio-ethnolinguistic Landscape of Contemporary Enggano." In *Enggano: Historical and Contemporary Perspectives*, edited by I Wayan Arka, Mary Dalrymple and Charlotte Hemmings. Canberra: EL Publishing.

Arka, I Wayan, and Keira Mullan. to appear. "Social Cognition, Typology, and Modularity in Linguistic Theory: An Indonesian Perspective." In *Linguistic worlds and social minds: what language documentation and linguistic diversity can tell us about the nature and origin of language*, edited by I Wayan Arka, Barth Danielle and Henrik Bergqvist.

Bischoff, Shannon T. , and Carmen Jany, eds. 2018. *Insights from Practices in Community-Based Research, From Theory To Practice Around The Globe*. Berlin, Boston: De Gruyter Mouton.

Briggs, Charles L. 1986. *Learning how to ask: a sociolinguistic appraisal of the role of the interview in social science research*. Cambridge: Cambridge University Press.

Curry, Mary Jane, and Theresa M. Lillis. 2018. "An Academic Literacies Approach to English for Research Publication Purposes: Issues, Practices and Perspectives." In

Publishing Research in English as an Additional Language: Practices, Pathways and Potentials, edited by Margaret Cargill and Sally Burgess, 41–60. Adelaide: University of Adelaide Press.

Czaykowska-Higgins, Ewa. 2009. "Research models, community engagement, and linguistic fieldwork: Reflections on working within Canadian indigenous communities " *Language Documentation & Conservation* 3:15–50.

Dobrin, Lise M. , and Josh Berson. 2011. "Speakers and Language Documentation." In *The Cambridge Handbook of Endangered Languages*, edited by Peter K. Austin and Julia Sallabank, 187-211. Cambridge: Cambridge University Press.

Dobrin, Lise M., and Saul Schwartz. 2016. "Collaboration or Participant Observation? Rethinking Models of 'Linguistic Social Work." *Language Documentation & Conservation* 10:253–257.

Enfield, N. J. 2002. *Ethnosyntax: Explorations in Grammar and Culture.* Oxford: Oxford University Press.

Evans, Nicholas, and Alan Dench. 2006. "Introduction: Catching language." In *Catching language : the standing challenge of grammar writing*, edited by Felix K Ameka, Alan Dench and Nicholas Evans, 1-39. Berlin: Walter de Gruyter.

Florey, Margaret, and Nikolaus Himmelmann. 2010. "New Directions in Field Linguistics: Training Strategies for Language Documentation in Indonesia." In *Endangered Languages of Austronesia*, edited by Margaret Florey, 121-140. Oxford: Oxford Univesity Press.

Himmelmann, Nikolaus. 1998. "Documentary and descriptive linguistics." *Linguistics,* 36:161-195.

Hisa, La , Agustinus Mahuze, and I Wayan Arka. 2017. "The ethnobotanical-linguistic documentation of Sago: a preliminary report from Merauke." *Linguistik Indonesia* 35 (2):187-200.

Hisa, La, Agus Mahuze, and I Wayan Arka. 2018. *Etnobotani: Pengetahuan lokal suku Marori di Taman Nasional Wasur Merauke*. Merauke: Balai Taman Nasional Wasur Merauke.

Hudley, Anne H. Charity , Christine Mallinson, and Mary Bucholtz, eds. 2024. *Decolonizing Linguistics*. New York: Oxford University Press.

Hudley, Anne H. Charity , Ignacio L. Montoya, Christine Mallinson, and Mary Bucholtz. 2024. "Introduction: Decolonizing Linguistics." In *Decolonizing Linguistics*, edited by Anne H. Charity Hudley, Christine Mallinson and Mary Bucholtz, 1-45. New York: Oxford University Press.

Kähler, Hans. 1975. *Texte von der Insel Enggano (Berichte über eine untergehende Kultur). .* Berlin: Dietrich Reimer.

Kroeger, Paul. 2007. "LFG as a Framework for Descriptive Grammar." In *Proceedings of LFG07*, edited by M Butt and T. H King. Stanford: CSLI, http://cslipublications.stanford.edu/LFG/12/lfg07.html.

Montrul, Silvina. 2016. *The Acquisition of Heritage Languages*. Cambridge: Cambridge University Press.

Musgrave, Simon, and Nick Thieberger. 2021. "The language documentation quartet." *Proceedings of the 4th Workshop on the Use of Computational Methods in the Study of Endangered Languages* Volume 1 (Papers): 6–12. doi: Online: Association for Computational Linguistics.

Oka Granoka, Ida Wayan., Ida Bagus Udara Naryana, I Wayan Jendra, I Nyoman. Medra, I G.N. Putra Yasa, and I Wayan Bawa. 1985. Tata Bahasa Bali. Denpasar: Proyek Pengembangan Bahasa dan Sastra Indonesia dan Daerah Departemen Pendidikan dan Kebudayaan.

Polinsky, Maria, and Olga Kagan. 2007. "Heritage languages: In the “wild” and in the classroom. ." *Language and Linguistics Compass* 1 (5):368–395.

Rajeg, Gede P. , C. Hemmings, I Wayan Arka, and Engga Z. Sangian. 2024. "Enggano Middle Voice and Valence Over Time." 15th International Conference of the Association for Linguistic Typology (ALT XV) NTU, Singapore, 4-6 December 2024.

Rajeg, Gede Primahadi Wijaya , Charlotte Hemmings, Engga Zakaria Sangian, Dendi Wijaya, and I Wayan (with Milson Kaitora Arka, Harun Kaharubi, M. Raflizen Kaitora (alm.), Aron Kaitora (alm.), Johansen Kaharubi, Ishar Timius Kaitora, Marlansius Kaharubi, Adam Kurniawan Kauno, & Resiawati Kaitora). 2025. *Kamus Bahasa Enggano (1st ed)*: University of Oxford & Zara Abadi.

Sangian, Engga Zakaria, Charlotte Hemmings, Arono Arono, Gede Primahadi Wijaya Rajeg, and I Wayan Arka. 2024. *Pihia' Yic Enggano VII*: University of Oxford & Zara Abadi. .

Sawaki, Yusuf, and I Wayan Arka. 2018. "Reflections on linguistic fieldwork and language documentation in eastern Indonesia." *Language Documentation & Conservation* Special Publication No. 15, edited by Bradley McDonnell, Andrea L. Berez-Kroeker, and Gary Holton:256-266.

Tsikewa, Adrienne. 2021. "Reimagining the current praxis of field linguistics training: Decolonial considerations." *Language* 97 (4):e293–e319. doi: DOI: https://doi.org/10.1353/lan.2021.0072.

Udayana, I Nyoman, G. P. W. Rajeg, and Ida A Made Puspani. 2025. "The (non)canonical status of the ka- passive in Balinese." *Linguistics Vanguard*:1-15. doi: doi.org/10.1515/lingvan-2023-0122.

Woodbury, Anthony C. 2003. "Defining documentary linguistics. ." In *Language Documentation and Description*, edited by Peter K. Austin, 35-51. London: SOAS.

Woodbury, Anthony C. 2011. "Language documentation." In *The Cambridge Handbook of Endangered Languages*, edited by Peter K. Austin and Julia Sallabank, 159-185. Cambridge: Cambridge University Press.