

14: Language Documentation in South Asia

1. INTRODUCTION

Our chapter surveys the history and some noteworthy current examples of language documentation in selected areas of the larger South Asian geographic region. We consider both the history and these examples within the broader question of how language documentation may be defined and what its goals should be. A complete analysis of language documentation in the greater South (and Southeast) Asian context is beyond the limited scope of this overview. Instead, we briefly contextualize concepts and practices of language documentation in samples of activity from three representative countries in which our own work and collaborations are primarily situated: Nepal, India, and Pakistan.

1. What is “Language Documentation” in the South Asian Context?

We begin with foundational definitions provided by others and then consider these definitions as they may apply in South Asia. Woodbury (2011:159) defines language documentation as “the creation, annotation, preservation, and dissemination of transparent records of a language.” According to Himmelmann (2006: 1), language documentation is “a field of linguistic inquiry and practice in its own right which is primarily concerned with the compilation and preservation of linguistic primary data and interfaces between primary data and various types of analyses based on these data.” For Himmelmann, language documentation methods and outputs should also include stable and accessible archives, metadata, standard conventions of representation and annotation, and a responsible sense of accountability to the needs of the community.

Despite Himmelmann’s comments about community accountability and needs, by and large, the output of documentation in the conceptualization of Himmelmann and Woodbury remains largely institutional in focus: language documentation is done *for* and *by* communities of *scholars*, ideally but *not necessarily* involving community co-creators of the materials per se, and their indexing for such future work of this nature. The applicability of this model in the South Asian context deserves an examination. In this area, historically, language documentation initiatives have been done by and for academic communities, although there is a recent shift towards more community-centred and community-originating work. At the same time, long-standing institutional and social barriers remain very complex and difficult to negotiate for many low-status communities. We return to these challenges in Sections 2 and 3 and in the conclusion (Section 5). Section 4 describes some developments and newer documentation initiatives that illustrate shifts towards expanded definitions of documentation that centre this work around communities and expand the value of the outputs to additional stakeholders.

Reflections on the history and goals of documentation are ongoing, as they should be, as in McDonnell et al. (2018), including important discussions surrounding alternatives to extractive, colonial methods, the importance of public awareness about languages, and the

importance of local cultural meanings and considerations. Before moving into our case studies and their implications, we also summarize Chelliah (2018), which, even eight years on, still accurately assesses the language documentation situation in India: What she observes for India may historically (and currently) still be the norm for the areas that we consider, although there have been some positive trends, examples that we will highlight. One major takeaway from her report is the persistence of conflicting ideologies about what constitutes a sufficient or even comprehensive record of a language along with ongoing power dynamics in the academy and in society that privilege larger languages over smaller, under-resourced ones. Additionally, language documentation in India has (and continues to be) done in very hierarchical, caste-based structures. This often has the result of entrenching community member passivity and engagement/involvement in the goals and outputs of this work. Furthermore, ongoing political and economic shifts can limit infrastructure (in particular, digital infrastructure) and the sustainability of resources that allow documentation outputs to translate into maintenance or revitalization.

At the same time, Chelliah notes some positive trends in the Indian context, trends that have largely maintained their momentum. These include new institutes and institutional incubators which provide new computational tools and infrastructure and include community collaboration (e.g., The Centre for Endangered Languages & Mother Tongue Studies at the University of Hyderabad in Telangana and The People's Linguistic Survey of India), newer and revitalized academic societies and conferences (e.g., Northeast Indian Linguistic Society), and the emergence of some sustainable grassroots community movements, notably efforts in communities to document verbal arts, folktales, and oral histories, as well as organisations that fund and train community documenters (e.g., the [Living Tongues Institute for Endangered Languages](#)).

While some advances have been made in recent years, especially with the interface between tech-oriented enterprises and language communities, both large and mid-sized in India, and in a handful of cases small languages as well, that have largely focused on creation of tools relevant to modern information technology, language documentation efforts in India remain clearly rooted in older standards and practices. Some models are explicitly extractive in orientation, such as recent work on Nicobarese and Andamanese languages coming out of the Central Institute for Indian Languages [CIIL], e.g., Abbi and Vysakh (2020a-c), Abbi and Pachori (2020a-b), Rajasingh (2019). At many universities in India, the main access to primary language data is either more or less random due to the presence of a speaker of a minority language amongst the student body or in the contexts of fieldwork trips for the training of linguistics students who are speakers of primarily Scheduled (official) languages of the country, which number twenty-two. In both institutional contexts, whether government-directed (CIIL) or university-based (e.g., University of Delhi, JNU, University of Hyderabad), the focus is on elicitation of targeted data based on questionnaires and sociolinguistic surveys, and the goals of such processes and the end products or deliverables of such are rather narrowly focused studies by students to achieve a degree (MA, MPhil, PhD typically) or to produce some government

sanctioned output. Another primary or secondary concern adopted in the Indian institutional contexts is applied outputs, especially orthographies and/or basic literacy materials for beginning students to meet benchmarks in the Mother Language Education initiatives. Some specific exceptions to this model are detailed below. Evidence of this process is demonstrated in the range of papers presented at any linguistics conference in India.

That said, in many contexts in India, there is a pronounced difference between the methodologies and associated products pursued from the top down in ‘mainland’ North, Central, East and South India, which are perhaps subject to more careful scrutiny by government and institutional hierarchies and local politics, and the more grassroots-level, community-based initiatives that are somewhat more commonly found in the northeastern part of the country. Note that Indian censuses exclude ‘inconveniently’ small language communities from official reckoning—usually ones with fewer than 10,000 speakers are ignored, but on the local level this is often not so, such that the most recent official census of India reckoned around 120 languages, but that number itself is thought to be found in the northeast of the country alone. Language ideologies and the oppressive creep of the ever more aggressive Hindu nationalist politics that dominate modern Indian society and demand more space for Hindi, further complicate the matter. So, while the difference between considering a specific lect as a dialect or a language is always grounded in socio-political concepts, perhaps nowhere is this clearer than in India, where there is, on the one hand, a generalized disdain for non-Hindu minority communities, and a Sanskritic understanding of language as requiring a dedicated writing system, literacy and script. Such an ecology affords movements towards ethnolinguistic erasure where such conditions are insufficiently met.

On the other hand, there is an overt fetishization and sexualization of so-called ‘tribal’ communities, whose existence is tolerated and celebrated as long as it is in the service of framing the State as a benevolent and enlightened benefactor and bringer of ‘civilization’ (Pfeffer 2014)¹. The vast majority of Indian scholars in mainstream India are Hindu, with much smaller numbers being Muslim or in some areas Christian. Hindu scholars often internalize the biases against minority communities held by the majority of their fellow citizens. The consequence of socio-political-economic-religious ideologies for mainstream Indian linguistics with respect to language documentation is that extractive methodologies are not critically assessed for their appropriateness, and new generations of scholars are created holding such beliefs anew.

Some scholars stand apart from this such as the above-mentioned Professor Shobhana Chelliah of the South Asian archive CoRSAL, which is notably based in the United States and different intellectual-academic expectations are in play. She and her colleagues have been devoting considerable time to training scholars and activists in modern documentation and archiving techniques as a counterbalance to many top-down approaches still in force in Indian academia. Similar efforts have also been undertaken in Pakistan by CoRSAL co-director

¹ The term ‘tribal’ is highly problematic in the Indian context (see Anderson and Jora 2017, 2022 for discussion).

Professor Sadaf Munshi, who has organised several documentation and archiving workshops for Pakistani scholars and language community members.

2. CHRONOLOGICAL CONSIDERATIONS

As our case studies in Sections 3 and 4 elaborate, the histories of language documentation, including approaches, methods, and outcomes, show considerable variation in the greater South Asian Region. We divide our discussion into focused consideration of India, Nepal, and Pakistan in this and other sections and return to their implications for the field of language documentation in Section 5.

2.1 India

Indigenous grammatical traditions for Indic or Dravidian have been active for more than a millennium in India, but these do not constitute language documentation in a modern sense of course. Moreover, the religious contextualization of such work is central, not linguistic or ethnocultural. Before Independence, most description of non-major languages in India was carried out by European missionaries or British civil servants: For example, the massive *Linguistic Survey of India* edited by Grierson consists largely of brief descriptions of the languages based on translations of the prodigal son parable submitted by missionaries. Collaborative work on minor Indian languages involving Indian and non-Indian scholars began in the mid 20th century. Norman Zide's Munda Languages Project based at University of Chicago trained both American and Indian scholars in then standard practice in linguistic analysis. Notable among these is Ram Dayal Munda who was the first Mundari speaker to receive a PhD in Linguistics. However, most energy in the linguistics space in India continued to either be traditional/Paninian or historical in orientation or to primarily focus on issues in the formal grammatical analysis and Chomskyan syntax that simply does not afford any attention to the core aspects of language documentation of minority languages of India. In addition, occasional questionnaire-based descriptions of languages have periodically come out from CIIL with partial descriptions of various minor languages. But until recently, and despite various government initiatives at documenting endangered languages through the creation of various university- and regionally-based administrative documentation programs, this model of investigation has endured in Indian language documentation, with only a few notable exceptions.

2.2 Nepal

Language documentation in Nepal has only in the past 50 or so years gained momentum in the sense of shifting methods and tools, community collaboration and community-utilised outputs, and approaches to language description that emphasize daily language practices. Until the mid-20th century, Nepal was essentially closed to outsiders, although trade relations between Nepal and India and Tibet were still active (see van Spengen 1987; Rogers 2004 for histories of

economic opportunities, in particular permission to engage in international trade provided to residents of the upper Manang and Mustang regions by varying royal dynasties).

An etymological dictionary of Nepali was compiled by Sir Ralph Lilley Turner while he served with Gurkha regiments in India in the 1920s (Collections & Library, 2015). In the late 1940s and 1950s, some researchers were permitted into Nepal, for example by an Agreement of Commerce and Friendship signed in 1947 between the U.S. and Nepal (*Historical Documents - Office of the Historian*, n.d.). Despite these early initiatives, and as covered in Eagle (1999), Awasthi (2004), and Turin (2006), a “one king, one country, one culture” policy stigmatized and minimized recognition and opportunities for indigenous linguistic and cultural expression until the late 1980s. (Hough et al. 2009: 160). Shifting official policies on linguistic pluralism over time have more recently culminated in Nepalese government recognition of the right of individuals to receive primary education in first (indigenous) languages, as enshrined in the 1990 Constitution, although there has been little support for implementation of this right beyond community-level initiatives.

With anthropology being linked closely to language documentation, the first anthropologist of Nepal, Dor Bahadur Bista, received his training as an assistant to Christoph Fuhrer-Haimendorf in the early 1950s, publishing “People of Nepal” (Bista 1967). The Centre for Nepal and Asian Studies (CNAS) and the Department of Sociology and Anthropology at Tribhuvan University were founded in 1977 and 1981, respectively. The Linguistic Society of Nepal was founded at TU in 1979, which now publishes the academic journal *Nepalese Linguistics* and hosts an annual conference.

Another entity deserving of commentary is the Summer Institute for Linguistics (SIL), an evangelical Christian nonprofit organisation whose main goal is to document and describe lesser-known languages. SIL played a foundational role in early language documentation work in Nepal, entering into an agreement with Tribhuvan University in the 1960s and producing the first descriptions of previously undocumented Nepalese languages in the 1960s and 1970s, including Nepali and Maithili (Indo-European) and also Gurung, Bantawa, Newar, Chepang, Sunwar, and Sherpa (Tibeto-Burman) (Hale 2019). In a paper that was published, and which then apparently disappeared from an Oxford publication (Calvet 2007), the SIL was expelled from Nepal due at least in part to strict anti-proselytization laws, although SIL linguists have maintained informal and productive relationships with academics and communities through the years.

Despite border controls and stifling language policies, there have been several efforts since the 1800s to survey the 100 + languages of Nepal, which span four language families (Indo-European, Sino-Tibetan, Dravidian, Austro-Asiatic), plus an isolate (Kusunda). The National Statistics Office of Nepal conducts a linguistic census every ten years (National Population and Housing Census 2021/2025). Perhaps the most comprehensive effort, the Linguistic Survey of Nepal (LinSuN), was launched in 2008, directed by Tribhuvan University and with authorization from the Nepalese government. Gautam (2019) provides a historical summary and critique of the LinSuN, including a more comprehensive bibliography of sociolinguistic descriptions of specific languages that emerged from this work. Although now

very likely outdated due to the constant ebb and flow of language communities and larger socio-political and migratory forces, the LinSuN is beneficial in that linguists worked with speech communities in-situ to gather data. Furthermore, it provides a snapshot of language practices across many under-described and under-resourced Nepalese languages. It also provides comparative vocabulary sets, which are useful for establishing etymological cognates or getting a glimpse into the effects of language contact (lexical borrowing). Additionally, there are sentence sets, which provide some basic syntactic information. But inevitably, a large-scale effort like the LinSuN is not intended for, and therefore does not capture, the more intricate and nuanced phonetic, phonological, and lexico-grammatical details found in more focused and longer-term documentation efforts of a single or set of languages.

Starting in the 1980s and 1990s, and often with financial support from federal agencies in North American and Europe, Australia and parts of Southeast Asia, and private foundations and sponsors including the Volkswagen Stiftung, The Endangered Languages Programme, The Firebird Foundation, and the Endangered Language Fund, language documentation in Nepal has pivoted towards longer-term relationships built through time between the linguist/anthropologist and often one or a small handful of speech community representatives. These approaches have yielded more expansive and representative language corpora and grammatical descriptions that have often fed into rich and detailed reference grammars that include text collections, as well as some (early) digital archives.

2.3 Pakistan

The history of documentation in what is now Pakistan consists of contributions from foreign scholars, colonial administrative officials and military officers, and, more recently, community-based efforts to record and maintain their own languages. Although early linguistic documentation in the region is often associated with 19th-century colonial scholarship, a small number of indigenous traditions of language description also existed. One notable example is Muhabbat Khan's *Riyaz al-mohabbat*, an early grammatical description of Pashto written in 1806 (Blumhardt & Mackenzie, 1965). Nevertheless, the majority of early documentation was carried out by British administrators and scholars whose work sought to classify, compare, and describe the languages of the Indian subcontinent. Major contributions from this period include Trumpp's (1872) work on Sindhi, Biddulph's (1880) descriptions of languages of the Hindukush region, and Leitner's (1880, 1893) collections of linguistic and ethnographic works documenting numerous languages and communities of the Hindukush region. The most ambitious undertaking of the colonial era was Grierson's *Linguistic Survey of India* (1903–1928), which, along with descriptions of languages across the Indian subcontinent, included comparative data and descriptions for numerous languages spoken in present-day Pakistan. Coverage, however, was uneven across languages. While Grierson dedicated substantial attention to major regional languages such as Western Punjabi and Sindhi, many lesser-spoken languages, particularly in northern Pakistan, were represented only by brief grammatical sketches, texts, or wordlists, often based on data from a small number of speakers.

Following the partition of British India and the establishment of Pakistan in 1947, language policy became closely intertwined with nation-building. The promotion of Urdu as the national language was intensified in an effort by the government to foster a shared national identity across a linguistically diverse population (Rahman 2006). One consequence of this governmental emphasis was that relatively little sustained attention was devoted to documenting the country's many minority and endangered languages. Although language remained an important marker of local and regional identity, linguistic research during the early decades of Pakistan's existence as a nation focused largely on issues of language policy, standardization, and the major regional languages, while documentation of smaller speech communities remained sparse.

Beginning in the mid-20th century and accelerating from the 1970s onward, documentary work in Pakistan increasingly shifted from broad classificatory and comparative surveys towards more detailed descriptive works on individual languages. A central figure in this period was Georg Morgenstierne, whose extensive fieldwork across the northwest regions of South Asia produced foundational descriptions of numerous Indo-Aryan, Iranian, and Nuristani languages (e.g., Kalasha, by Morgenstierne 1973). Subsequent decades saw the publication of increasingly comprehensive grammars, dictionaries, and text collections for previously under-described languages. Important examples include Elfenbein's work on Balochi (1966), Bashir's (1988) research on Kalasha, Baart's (1999) description of Kalam Kohistani, Munshi's (2006) work on Burushaski, and Liljegren's (2016) grammar of Palula. During this period, large-scale language survey efforts also emerged, most notably the 5-volume *Sociolinguistic Survey of Northern Pakistan*, conducted by SIL International (Rensch et al. eds. 1992) in collaboration with the National Institute of Pakistan Studies beginning in the 1980s. These projects greatly expanded the availability of documentation materials on Pakistani languages, with a particular focus on the languages of northern Pakistan, a region with a particularly rich linguistic diversity due to the natural barriers between communities created by rugged mountainous terrain and river systems.

More recently, documentation in Pakistan has increasingly reflected broader developments in documentary linguistics, including long-term fieldwork projects, digital archiving efforts, native speaker participation, and language revitalization initiatives. While a small number of foreign researchers continue to participate in documentation projects on Pakistani languages, there has also been a marked growth in documentation led by Pakistani community members themselves. These efforts have largely been spurred on by local non-profit organisations such as the Forum for Language Initiatives (FLI) and Idara Baraye Taleem wa Tarqi (IBT), which provide training in language documentation skills, promote the collection and dissemination of literature and poetry, and publish grammars, dictionaries, and surveys of endangered Pakistani languages. While academic institutions in Pakistan remain largely unengaged with sustained documentation efforts, grassroots documentation developments have created increased excitement and investment in language documentation within the language communities themselves, offering promising directions for the future of documentation in the country.

3. SHIFTING METHODOLOGIES

We provide examples and discussion in this section about how the methods of language documentation in India, Nepal, and Pakistan have changed through time. Using Leonard & Haynes (2011) as a lens through which to discuss these changes, we observe movements from what they call the “traditional view”, where the project goals are at the centre and researcher needs and expertise are privileged over community needs and expertise, to methods that are more “true collaboration”, where researcher needs and expertise are on equal footing with community needs and expertise.

The languages considered in Sections 3 and 4 are displayed on Map 1, recognizing that in some cases, language communities are not adequately or accurately represented by a single point.

INSERT MAP 1 HERE

3.1 India

Rather than academic, institution-specific initiatives such as those by either CIIL or universities like Delhi University or JNU, some of the most encouraging action being seen are coming out of grassroots initiatives. This is particularly clearly seen among communities located in the northeastern States of India like Arunachal Pradesh, Nagaland, etc., rather than ‘mainland’ majority India. One noteworthy example of this can be seen with respect to achievements made by the Wancho community of Arunachal Pradesh. They developed an orthography and have begun using this to represent the language. They have been very active on the decolonial digital lexicographic platform known as the Living Dictionaries where no fewer than twenty-two speakers have collaborated with each other as Dictionary Managers that cites Wancho headwords in this new orthography. This is entirely independent of universities or other non-autochthonous organisations, and they have used the tool to create a locally derived and shared multimedia resource.

3.2 Nepal

As noted in Section 2, the past twenty years have seen more the rise of focused and long-term documentation projects to provide rich and detailed descriptions more resembling the definition of language documentations given by Woodbury and Himmelmann. In such scenarios, documentarians spend months and even years working with one or more speakers, combining more controlled or even stimulus-based methods of lexical and morpho-syntactic elicitation complemented by the recording, detailed annotation and transcription of extended stretches of discourse, sometimes involving multiple participants. These discourse stretches may also range across many different genres, from casual conversation, to verbal arts (stories, legends, myths, songs, poems), to formal orations, to liturgical/ritual language use. In some cases, the stimulus-based elicitations are part of other initiatives that aim at typological analysis of linguistic phenomena. An example of these approaches is work done by Carol Genetti, who first began

language documentation work on the Dolakha variety of Newar (Newaric) in 1987 as a graduate student and has maintained relationships with the speech community ever since, with many academic publications including an [award-winning](#) reference grammar (Genetti 2007), and culminating in a community-collaborative production of a traditional Dolakha Newari play, [Misāmi](#), in the same language in Spring 2026.

One example of an interdisciplinary, longitudinal team-based project with a central goal of documenting language and cultural practices with large amounts of naturalistic data is the Chintang and Puma (Kiranti) Documentation project, with a project team consisting of linguists, psycholinguists, and anthropologists (summarized in Gaenszle et al. 2005) and with data housed at DOBES (<https://dobes.mpi.nl/projects/chintang/links/>). CPDP researchers included European and Nepalese scholars and graduate students and included ethnographic methods involving both adult speakers and children. Of special note is the Child Language Acquisition study, research which is otherwise quite rare and sheds light on language socialization in an area of intense cross-family language contact, as well as another study of social organisation and settlement structure, which sheds light on clan affiliations and hierarchies.

The earthquakes that significantly impacted both Kathmandu and Himalayan highland communities in 2015 led to the convening and training of young adults originating from the Gorkha, Manang, and Mustang Districts in language documentation methods (Hildebrandt et al. 2019). These residents, who were not professionally trained documentary linguists, learned from foreign and Nepalese linguists and anthropologists about audio and video recording techniques, best practices in documentation metadata, and basic annotation practices. They returned to their impacted communities and conducted semi-structured interviews and recorded stories from friends and relatives about their lived experiences and cultural understanding of natural disasters, in their own mother tongues. Languages represented in the corpora housed at the University of Virginia's [Tibetan and Himalayan Digital Library](#) include Gurung, Manange and Nar-Phu (Tamangic) as well as Nubri, Kutang, Tsum, Lowa, several varieties of Tibetan (Bodish), as well as Nepali (Indo-European). Note that in addition to being an example of a more collaborative approach as advocated by Leonard & Haynes (2010), this project in some ways resembles the methods advocated by Everett (2001) in monolingual field research, in which native speakers conducted recordings amongst groups of their own kind and friends, where the languages live and are practiced continuously. They then worked with linguists who did not speak their languages to provide greater nuanced details about phonological, grammatical, and lexical patterns.

Most team-structure documentation is done with funding external to Nepal, and these projects are often led by foreign researchers. However, one notable exception is the [Baram language documentation project](#), which, while funded by the Endangered Language Programme, was entirely organised and conducted by Nepalese linguists from Tribhuvan University.

3.3 Pakistan

As noted in section 2, documentation efforts in Pakistan during the 20th century largely followed the “traditional” documentation model, in which foreign researchers initiated documentation projects with their academic and professional goals largely shaping the intended deliverables, with community needs incorporated to varying degrees. More recently, however, Pakistan has witnessed the emergence of a documentation ecosystem that increasingly enables community members to participate as equal collaborators rather than solely as language consultants.

A major catalyst for this shift has been the emergence of local community-based organisations dedicated to language documentation and revitalization. The recent surge in documentation efforts initiated and carried out primarily – and in some cases entirely – by community members themselves has resulted in new opportunities to pursue what Leonard & Haynes (2010) refer to as “truly collaborative” documentation projects. More specifically, the formation of NGOs such as the Forum for Language Initiatives (FLI) has stimulated community interest in the documentation of Indigenous Pakistani languages. Founded in 2003 by the Swedish linguist Henrik Liljegren, the FLI aims to train and equip native speakers of Pakistani languages to lead documentation and revitalization efforts within their own communities. While foreign researchers were heavily involved in the early years of the FLI, trained community members have increasingly assumed responsibility for its leadership (Liljegren & Akhunzada, 2017). Today, the executive board of the FLI is constructed entirely of native speakers of Pakistani languages.

Organisations like the FLI have provided crucial training in documentation skills, as well as local resource sharing across endangered language communities. Not only have these efforts enabled community members to gain fluency in documentation practices and to cultivate networks of invested speakers capable of promoting sustained language development over many years, but they have also empowered community members to identify and articulate local priorities from the outset of documentation projects. By equipping community members with documentation expertise before foreign researchers become involved, organisations such as the FLI fundamentally alter the traditional researcher-consultant relationship. Rather than serving primarily as language consultants, community members enter collaborations as well-trained partners capable of shaping project goals, methodologies, and outputs.

A good illustration of this truly collaborative model is the documentation of Gwarbati, spoken by around 15,000 people in northern Pakistan. In 2016, the FLI initiated a language documentation project on the language, which produced a wide variety of academic publications, high-quality Gwarbati recordings from a variety of domains and genres, and essential training in language documentation for community members. In conjunction with work initiated by FLI-supported Gwarbati speakers, Henrik Liljegren at Stockholm University secured funding to document the language. In collaboration with Gwarbati speakers trained through the FLI, the joint project produced dozens of hours of high-quality recordings, data from more than a hundred speakers, lexicons, hours of word-level speech transcriptions, and numerous research

publications, including works co-authored by both foreign researchers and native Gawarbatī speakers.

The establishment of community-led language organisations, coupled with continued participation from foreign researchers, has created an environment for effective and sustained collaboration between community members and foreign scholars. Grassroots organisations like the FLI and IBT have provided community members in Pakistan with the necessary resources and training to develop clear ideas of what the community hopes to get out of a documentation project (literacy materials, dictionaries, orthography development, multilingual education infrastructure, etc.). At the same time, foreign researchers contribute access to funding opportunities, archival infrastructure, specialised equipment, and broader international research networks. Although the priorities of foreign researchers and language communities are not always identical, there is often substantial overlap in the foundational work required to achieve both sets of goals. The collection and long-term preservation of accessible, high-quality recordings across diverse communicative domains, together with careful phonological, morphological, syntactic, and semantic analysis, not only support theoretical linguistic research but also provide the resources necessary for orthography development, pedagogical materials, dictionaries, and other revitalization outputs.

It is also important to recognise that the needs, desired outcomes, and levels of desired participation in the documentation project by the language community are not monolithic. In practice, the desires of one community member often differ substantially from those of another. While some speakers seek to take active leadership roles in directing documentation projects and determining their outcomes, others are content with participating as paid language consultants or are simply pleased that their language is receiving increased attention – and thereby an increased level of prestige – from documentation projects conducted by fellow community members and/or foreign researchers. Recognizing this diversity of perspectives helps avoid treating “the community” as a single voice regarding documentation goals.

The picture that emerges from the ground in Pakistan, then, is somewhat distinct from the struggles related to documentation efforts in other parts of the world. As Holton (2009) notes, what is considered highly unethical and “extractive” in some contexts may be perfectly fine, if not commended, in another. In Pakistan, for instance, it is certainly considered unethical to reject collaboration with interested community members and not to consider the needs and desired outcomes of the community. However, academic and theoretical goals pursued by researchers are not necessarily viewed as orthogonal to community members' needs. Instead, such work is typically treated as another avenue for promoting, documenting, and revitalizing the language.

Despite these encouraging grassroots developments, sustained institutional investment from Pakistani universities and government agencies remains limited. While a small number of institutions have invested in language documentation initiatives, most universities offer little formal training in documentary linguistics for interested community members, and government

funding opportunities for documentation projects remain scarce. Consequently, much of the recent progress has depended on community organisations and international collaborations rather than long-term institutional support within Pakistan. Ideally, the growing enthusiasm for language documentation at the grassroots level will increasingly be matched by sustained investment from Pakistani universities and governmental agencies, ensuring that community-led initiatives continue to expand and flourish.

4. SHIFTING OUTPUTS

In this section we explore some more recent outputs and results of these shifting methods in this region of South Asia, with a particular eye towards community collaboration and outputs that cross disciplinary boundaries.

4.1 India

A small number of high-quality reference grammars of minority languages of India have appeared in well regarded series, e.g. on Karbi by Konnerth (2020), Mongsen Ao by Coupe (2007) or Meithei by Chelliah (1997). Looking at the catalogues of modern language documentation archives reveal that there has been significant output in contemporary formats using modern documentation methodologies that have been produced in recent years on a wide range of underdocumented or endangered languages of India by Indian and non-Indian scholars independently and collaboratively. These include annotated audio and video files, photos and text data with accompanying metadata. ELAR has at least one file in over sixty languages of India (although some 30-plus languages not from India come up in a search using the country filter ‘India’, so one must approach use of this data with some caution). Included among these are a number of lesser-known Indo-Aryan languages in addition to minor languages belonging to the Trans-Himalayan, Austroasiatic and Dravidian families as well as small groups and isolates like Nihali and Andamanese. PARADISEC has annotated deposits from more than forty languages of India with a heavy skewing towards non-Indic varieties including several belonging to the Tai family and at least ten varieties subsumed under the Tangsa macro-ethnicity. Unsurprisingly, CoRSAL has significant deposits of materials from languages of India numbering over forty in total. These primarily skew to languages in the northeastern part of India and involve digitized print materials as well as annotated materials that were born digital. Several of these deposits were made by indigenous researchers, e.g., Yankee Modi’s work on Milang. Included in both PARADISEC and CoRSAL, materials deposited by the research team at Living Tongues Institute can be found in large numbers. For example, nearly three quarters of the files associated with India in the PARADISEC archive come from three projects run by the collaboration between western and tribal Indian scholars at this organisation from three languages belonging to the Munda family. These deposits include annotated videos, audio files, photos with metadata, etc. They also include community-based output like a basic children’s reader and ethnobotanical survey found in the Birhor deposit at CoRSAL.

An example of grassroots citizen science in India in collaboration with an extra-academic NGO can be seen in the Munda Languages Initiative carried out by Living Tongues Institute for Endangered Languages with members of the Mundari, Santali, Ho, Birhor, Sora, Gta', Remo, Gutob, Juang, Gorum, Kharia and Korku-speaking communities of Jharkhand, Odisha, Chhattisgarh, Madhya Pradesh, Maharashtra and Assam. Several PhD scholars who are tribal have participated in this project for many years (more than a decade each) and even documented languages other than their own mother or heritage tongues even, which is not overly common anywhere globally really for indigenous PhD scholars of Linguistics. Many different kinds of outputs have been achieved and being created now for different Munda languages. This includes mainstream linguistic output (Horo et al. 2020, Gogoi et al. 2024, Anderson 2023) other scientific domains such as ethnobotany (Lakra et al. 2021) and community-based output (Jora & Birhor 2022), the first ever children's reader book in the newly adopted orthography for the Birhor language, annotated film clips on Sora traditional cultural practices and music available on YouTube in Sora, English and Hindi for widespread and local consumption.

4.2 Nepal

Outputs of language documentation in Nepal have increasingly included high-quality reference grammars published by presses including De Gruyter Brill (including the Languages of the Greater Himalayan Region series, edited by George van Driem), Language Science Press, and Lingcom, among others. An increasing number of documentation efforts are now hosted on digital archives. One example is the Kusunda, a linguistic isolate that was thought to be extinct since the early 2000s. Bodt & Aaley (to appear) summarize the documentation and revitalization efforts of Kusunda. Although the language is technically moribund in the traditional sense, with only a single living speaker remaining, their work (and work by others) has resulted in several publications and a rich audio-video archive deposited in Zenodo (Aaley & Bodt 2019 <https://zenodo.org/records/14249883>).

The Manang Languages Project was a multidisciplinary documentation initiative between 2012-2019, including linguistic description, geo-tagging to study dialect variation, and educational research to study the function of local languages in relation to Nepali and English in early childhood education in the Manang District of Nepal. While this project may be considered as more "traditional" in the sense of Leonard & Haynes (2010), the decisions about which community members participated were made by local leaders. Outputs from this project range from traditional publications accessible primarily to other academics, to a community dictionary (Gyalsumdo-Nepali-English) co-constructed with Gyalsumdo speakers, to projects located more in the digital humanities, namely an [online, interactive atlas](#) of language attitudes and practices.

Pangloss (originally named LACITO, <https://pangloss.cnrs.fr/?lang=en>) was pioneered by a collaboration between linguist Boyd Michailovsky, who has conducted extensive documentation in Nepal, and engineer John B. Lowe (Michailovsky 2008; Michailovsky et al. 2014; Adamou 2025). It now hosts audio, primarily running discourse, from over 250 endangered or understudied languages from around the world, including examples from Nepali

dialects, Tamang, Newar, and some Kiranti languages. It also hosts digital dictionaries, including some Nepalese languages.

In addition, there has been an increase in linguistic analyses that go “beyond the grammar”, namely done by Lauren Gawne and colleagues on gesture, onomatopoeia, reported speech, and evidentiality (Gawne 2025; Gawne et al. 2025; Gawne 2021; Gawne & Hildebrandt 2020; Gawne & Hill (eds.) 2017), as well as work done collaboratively on cultural impacts of big events/migration (Childs et al. forthcoming/2026).

4.3 Pakistan

The evolution of documentation outputs in Pakistan has largely mirrored the broader developments in the field of language documentation. Early work focused on comparative and historical surveys detailing the number, diversity, and classification of languages. Toward the end of the 20th century, documentation outputs increasingly consisted of grammatical descriptions and wordlists. More recently, however, documentation has expanded beyond these traditional products to include durable digital archives alongside resources designed to support the long-term needs of language communities.

While grammatical descriptions and other similar work have continued, there has been a marked shift in emphasis by foreign researchers toward the creation of durable digital archives that preserve a wide variety of linguistic materials, including poetry, personal narratives, wordlists, folktales, phonetic experiments, and linguistic analyses and commentary. Digital archives are particularly valuable because they allow a single documentation effort to serve multiple audiences simultaneously, preserving primary data for unforeseen future linguistic research while also making culturally and pedagogically significant materials accessible to language communities themselves.

One such example is the Mankiyali Language Resource, a collection of linguistic materials for Mankiyali, a language spoken by less than 500 people in the northern part of the country. Developed by Prof. Sadaf Munshi and her research team, funded by the NSF’s Documentation of Endangered Languages program, and housed within the Computational Resource for South Asian Languages (CoRSAL), the Mankiyali collection includes more than 350 documentary items. These include photographs depicting community life and material culture, hundreds of annotated recordings, and experimental data examining the unique properties of the language’s stress system, among many other items. In addition to the archive, the project resulted in various phonological (Paramore 2021) and morphosyntactic (Englert 2022) descriptions of the language, as well as a formal orthography developed jointly by native speakers in conjunction with Professor Munshi’s research team. The project, therefore, illustrates how a single documentation effort can simultaneously generate archival, descriptive, theoretical, pedagogical, and community-oriented outputs.

Much of the recent work produced by local linguistic non-profits has focused on language revitalization objectives, such as mother-tongue education programs, literacy and

orthography initiatives, and poetic anthologies. All of these outputs reflect the desire of language communities to transmit the use of and pride in their indigenous languages to the next generation, to improve early education opportunities for children who exclusively (or primarily) speak a lesser-known language, and to combat utilitarian shifts to more dominant languages, the national language, or the de facto language of the Pakistani elite, English.

Alongside these community-oriented products, the documentation of interesting typological features of Pakistani languages as well as unique features of individual languages has resulted in a small but growing body of theoretically-oriented research, often conducted jointly by foreign scholars and Pakistani scholars (Paramore & Aurangzeb 2025; Paramore 2026; Kula & Syed 2020; Zubair 2016; Hussain 2020; 2021). Together, these scholarly outputs complement the community-focused products described above, illustrating that contemporary documentation projects increasingly serve both academic and revitalization goals. Nevertheless, continued growth in theoretically oriented documentation research will likely depend upon greater investment from Pakistani universities and research institutions. Without stronger institutional support, community organisations will understandably continue to prioritize documentation outputs that most directly address the immediate needs of language communities.

5. CLOSING COMMENTS

We close this chapter with a return to the Woodbury and Himmelmann definitions of language documentation, and ask: Do these definitions apply to work now being done in India, Nepal, and Pakistan? In some senses, yes; language documentation (in the sense of Woodbury) in these regions have involved the creation, annotation, presentation, and dissemination of (increasingly) transparent records of many historically under-described and under-resourced languages. In the sense of Himmelmann, we also have seen the rise of stable and accessible archives (in particular, via CoRSAL, PARADISEC, and ELAR), standardized metadata and annotation/transcription conventions, and an increasing focus on community needs.

We also consider McDonnell et al. (eds.) (2018) and the contributions in that edited volume. In their introduction, they advocate for documentation resulting in greater public awareness about language diversity and vitality. They call also for decolonial methods. While there have been several highly recognised and award-winning outputs, many of these are still primarily accessible only to academic audiences. Genetti's collaboration with the Dolakha Newar community in the production of *Misāmi* stands as an exception to this. At the same time the increase in funded documentation work done by citizens, and better-yet, community members, is a sign that language documentation work in these regions is decreasingly extractive. Additionally, McDonnell et al. and Riesburg (2018, same volume) call for descriptions of languages not as objects of study enshrouded in opaque theories, but that rather are transparent and recognise languages as a living part of the community. This is also happening to a certain

extent, via ethnobotany and music projects in Birhor and Sora in India and via studies of population churn and linguistic impacts in Himalayan highlands Nepal.

At the same time, much funded and university-supported documentation work is still done in a hierarchical (often caste-based) arrangement, which hampers attempts to achieve truly collaborative methods as advocated for by Leonard & Haynes (2010). This persistence spills into the limited types of outputs, ones that ultimately lack truly useful pedagogical materials and educational support for teachers, or else impose outside orthographies and linguistic descriptions that are simplified, or worse-yet, that are typologically skewed to portray languages from other families as more similar to dominant (often Indo-European) ones. In the Nepal context, the challenges to language promotion and viability are surveyed by Regmi (2017), where he identifies problems ranging from a lack of a reliable socio-linguistic understanding of language practices, to conflicts between communities for official recognition, to outward migration trends, to the lack of a formal Academy of Nepalese languages. There is also a need for more mass-consumption press and information portals to convey the value of languages and their properties in ways that can be ethically valued by wider populations.

These advances and persistent challenges ultimately point to a discipline very much in flux in this region. Despite persistent political and economic pressures and the challenges of nationalistic movements throughout greater South Asia, language communities (both in-situ and émigré) are increasingly recognizing the importance of documentation initiatives for their own means. Academics from fields including linguistics, geography and environmental sciences, education/pedagogy, and the social sciences are increasingly partnering to reveal rich and nuanced descriptions and resources. It is our hope that this nascent raising of awareness and intellectual and cultural cross-fertilization will generate a positive feedback loop that will increase capacity for new initiatives that not only reflect the definitions of documentation defined and advocated for by the references that we have cited but will also feed into continuously expanded and refined definitions that resemble the changing needs and identities of both language and academic communities.

(7229 words, excluding references)

References

- Aaley, U.R. and Bodt, T.A., 2019. *Kusunda Audio and Video Files Day 6 04082019 – Part 1*. Zenodo. Available at: <https://doi.org/10.5281/zenodo.3380206>.
- Abbi, A. and Pachori, R., 2020a. *A grammatical sketch of Sanenyo*. Mysore: Central Institute of Indian Languages (SPPEL).
- Abbi, A. and Pachori, R., 2020b. *Dictionary of Sanenyo: Nicobarese–Sanenyo–English–Hindi: Images and sounds*. Mysore: Central Institute of Indian Languages (SPPEL).
- Abbi, A. and Vysakh, R., 2020a. 'Aspects of word formation in Luro: The endangered language of the Nicobar Islands', *Asian Language and Linguistics*, 1(1), pp. 9–33.
- Abbi, A. and Vysakh, R., 2020b. *A grammatical sketch of Luro*. Mysore: Central Institute of Indian Languages (SPPEL).
- Abbi, A. and Vysakh, R., 2020c. *Dictionary of Luro: Nicobarese–Luro–English–Hindi: Images and sounds*. Mysore: Central Institute of Indian Languages (SPPEL).
- Adamou, E., Guillaume, S. and Michaud, A., 2025. 'The Pangloss Collection: Opening up research data on endangered and underdocumented languages', *Language*, 101(1), pp. e38–e59. Available at: <https://doi.org/10.1353/lan.2025.a954237>.
- Anderson, G.D.S. and Jora, B., 2017. 'The linguistic ecology of "particularly vulnerable tribal" groups in Middle India: State-mediated ethnolinguistic erasure and neocolonialist "development"', *Texas Linguistics Forum*, 60, pp. 1–14.
- Anderson, G.D.S. and Jora, B., 2022. "If you eat beef, you cannot learn": An exploration of tribal identity, internal neocolonialism, "development" and contact effects on the languages of two "particularly vulnerable tribal groups" who are speakers of Munda languages in Jharkhand and Odisha, India', in Anderson, G.D.S. (ed.) *Munda Linguistics: Typological, Descriptive and Diachronic Perspectives*. Pune: Deccan College, pp. 59–126.
- Anderson, G.D.S., 2023. 'The many ways to be invisible in Sora', in Sidwell, P. and Alves, M. (eds.) *Papers from the Ninth and Tenth International Conferences on Austroasiatic Linguistics. Journal of the Southeast Asian Linguistics Society*, Special Publication 12. Honolulu: University of Hawai'i Press, pp. 19–33.
- Awasthi, L.D., 2004. *Exploring monolingual school practices in multilingual Nepal*. PhD thesis. Copenhagen: Danish University of Education (Danmarks Pædagogiske Universitet).
- Baart, J., 1999. *A sketch of Kalam Kohistani grammar*. Studies in Languages of Northern Pakistan, Vol. 5. Islamabad: National Institute of Pakistan Studies, Quaid-i-Azam University, and Summer Institute of Linguistics.

- Bashir, E., 1988. *Topics in Kalasha syntax: An areal and typological perspective*. PhD dissertation. University of Michigan.
- Biddulph, J., 1880. *Tribes of the Hindoo Koosh*. Calcutta: Office of the Superintendent of Government Printing.
- Bista, D.B., 1967. *People of Nepal*. Kathmandu: Department of Publicity, Ministry of Information and Broadcasting, His Majesty's Government of Nepal.
- Blumhardt, J.F. and Mackenzie, D.N., 1965. *Catalogue of Pashto Manuscripts in the Libraries of the British Isles*. London: The Trustees of the British Museum and Commonwealth Relations Office.
- Bodt, T. and Aaley, U.R., forthcoming. 'Revitalising a language isolate with a single speaker: The case of Kusunda', in Hildebrandt, K., Modi, Y., Peterson, D. and Suzuki, H. (eds.) *The Oxford Handbook of Tibeto-Burman Languages*. Oxford: Oxford University Press.
- Calvet, L.-J., 2007. *Language policy and imperialism: The Summer Institute of Linguistics MS*. Available at: <https://people.bu.edu/manfredi/CalvetCh14anglaisSIL.pdf>.
- Chelliah, S., 2018. *Reflections on language documentation in India*. Honolulu: University of Hawai'i Press. Available at: <http://hdl.handle.net/10125/24826>.
- Chelliah, S., 1997. *A grammar of Meithei*. Berlin: Mouton de Gruyter.
- Childs, G., Craig, S., Ghale, T., Gurung, T.W., Gurung, Y.T., Hildebrandt, K., Lama, R.L. and Lama, S., forthcoming/2026. 'Population churn in Highland Nepal: Demographic, social, and linguistic implications', *Himalaya*.
- Collections, School of Oriental and African Studies Library, 2015. *SOAS history: Ralph Lilley Turner – Special Collections, SOAS Library*. Available at: <https://blogs.soas.ac.uk/archives/2015/08/04/soas-history-ralph-lilley-turner/> (Accessed: 28 May 2026).
- Coupe, A., 2007. *A grammar of Mongsen Ao*. Berlin: Mouton de Gruyter.
- Eagle, S., 1999. 'The language situation in Nepal', *Journal of Multilingual and Multicultural Development*, 20(4–5), pp. 272–327. Available at: <https://doi.org/10.1080/01434639908666382>.
- Elfenbein, J., 1966. *The Baluchi language: A dialectology with texts*. Royal Asiatic Society Monographs, Vol. XXVII. London: The Royal Asiatic Society of Great Britain and Ireland.
- Englert, E., 2022. *A first look at Mankiyali morphology*. Master's thesis. University of North Texas.

Everett, D.L., 2001. 'Monolingual field research', in Newman, P. and Ratliff, M. (eds.) *Linguistic Fieldwork*. Cambridge: Cambridge University Press, pp. 166–188.

Gaenszle, M., Bickel, B., Banjade, G., Lieven, E., Paudyal, N.P., Rai, A., Rai, I.P., Rai, M., Rai, N.K., Rai, V.S., Sharma, N.P. and Stoll, S., 2005. 'Research report: The Chintang and Puma Documentation Project (CPDP)', *European Bulletin of Himalayan Research*, 28, pp. 95–103.

Gautam, B.L., 2019. 'Sociolinguistic survey of Nepalese languages: A critical evaluation', *Language Ecology*, 3(2), pp. 189–208. Available at: <https://doi.org/10.1075/le.19004.gau>.

Gawne, L., 2025. *Gesture: A slim guide*. Oxford: Oxford University Press.

Gawne, L., 2021. "“Away” gestures associated with negative expressions in narrative discourse in Syuba (Kagate, Nepal) speakers', *Semiotica*, 239, pp. 37–59.

Gawne, L., Hildebrandt, K.A. and Styles, S., 2025. 'Natural disasters elicit spontaneous multimodal iconicity in onomatopoeia and gesture: Earthquake narratives from Nepal and New Zealand', *Australian Journal of Linguistics*, 45(3), pp. 448–465. Available at: <https://doi.org/10.1080/07268602.2025.2506628>.

Gawne, L. and Hildebrandt, K.A., 2020. 'Reported speech in earthquake narratives from six Tibeto-Burman languages', *Studies in Language*, 44(2), pp. 461–499.

Gawne, L. and Hill, N.W. (eds.), 2017. *Evidential systems of Tibetan languages*. Berlin: De Gruyter Mouton.

Genetti, C., 2007. *A grammar of Dolakha Newar*. Berlin: De Gruyter Mouton.

Gogoi, P., Horo, L. and Anderson, G.D.S., 2024. 'Phonetic and phonological analysis of the Mundari vowel system', in Sidwell, P. (ed.) *Austroasiatic Linguistics: In Honour of Gérard Diffloth (1939–2023)*. Chiang Mai: Chiang Mai University Press, pp. 75–90.

Grierson, G.A., 1903–1928. *Linguistic Survey of India*. 11 vols. Calcutta: Office of the Superintendent of Government Printing.

Hale, A., 2019. 'Reflections on the beginnings of linguistics at Tribhuvan University', *Nepalese Linguistics*, 34(1), pp. 87–89. Available at: <https://doi.org/10.3126/nl.v34i1.41171>.

Hildebrandt, K.A., Burge-Beckley, T. and Sebok, J., 2019. 'Language documentation in the aftermath of the 2015 Nepal earthquakes: A guide to two archives and a web exhibit', *Language Documentation & Conservation*, 13, pp. 618–651. Available at: <http://hdl.handle.net/10125/24914>.

Himmelmann, N.P., 2006. 'Language documentation: What is it and what is it good for?', in Gippert, J., Himmelmann, N.P. and Mosel, U. (eds.) *Essentials of Language Documentation*.

Berlin and New York: De Gruyter Mouton, pp. 1–30. Available at:
<https://doi.org/10.1515/9783110197730.1>.

Office of the Historian, n.d. *Historical documents*. Available at:
<https://history.state.gov/historicaldocuments/frus1950v05/d836> (Accessed: 28 May 2026).

Holton, G., 2009. 'Relatively ethical: A comparison of linguistic research paradigms in Alaska and Indonesia', *Language Documentation & Conservation*, 4(2), pp. 161–175.

Horo, L.P.S. and Anderson, G.D.S., 2020. 'Acoustic phonetic study of the Sora vowel system', *Journal of the Acoustical Society of America*, 147(4), pp. 3000–3011. Available at:
<https://doi.org/10.1121/10.0001011>.

Hough, D., Thapa Magar, R.B. and Yonjan-Tamang, A., 2009. 'Privileging Indigenous knowledges: Empowering multilingual education in Nepal', in Skutnabb-Kangas, T., Phillipson, R., Mohanty, A.K. and Panda, M. (eds.) *Social Justice through Multilingual Education*. Bristol: Multilingual Matters, pp. 159–176.

Hussain, Q., 2020. 'Dawoodi (Pakistan) – Language snapshot', *Language Documentation & Description*, 19, pp. 130–137.

Hussain, Q., 2021. 'Fundamental frequency and phonation differences in the production of stop laryngeal contrasts of endangered Shina', *Languages*, 6, p. 139.

Jora, B., Birhor, B. *et al.*, 2022. *Abun Ari Re: In our surroundings*. Ranchi: Living Tongues Institute for Endangered Languages.

Konnerth, L., 2020. *Grammar of Karbi*. Berlin: Mouton de Gruyter.

Kula, N.C. and Syed, N.A., 2020. 'Non-myopic nasal spreading in Saraiki', *Radical: A Journal of Phonology*, 1, pp. 126–182.

Lakra, S.K., Jora, B. and Anderson, G.D.S., 2021. *A survey of the ethnobotanical knowledge of the Birhor Tribe of India*. Ranchi: Living Tongues Institute for Endangered Languages.

Leitner, G.W., 1880. *The languages and races of Dardistan*. Calcutta: Office of the Superintendent of Government Printing.

Leitner, G.W., 1893. *Dardistan in 1866, 1886, and 1893: Being an account of the history, religions, customs, legends, fables, and songs of Gilgit, Chilas, Kandia (Gabrial), Yasin, Chitral, Hunza, Nagyr, and other parts of the Hindukush, as also an appendix on the origin, nomenclature, pronunciation, grammar, and vocabulary of thirty-three languages*. London: W.H. Allen & Co.

Leonard, W.Y. and Haynes, E., 2010. 'Making "collaboration" collaborative: An examination of perspectives that frame linguistic field research', *Language Documentation & Conservation*, 4, pp. 268–293.

Liljegren, H., 2016. *A grammar of Palula*. Studies in Diversity Linguistics, Vol. 8. Berlin: Language Science Press.

Liljegren, H. and Akhunzada, F., 2017. 'Linguistic diversity, vitality and maintenance: A case study on the language situation in northern Pakistan', *Multiethnica*, 37, pp. 60–79.

McDonnell, B., Berez-Kroeker, A.L. and Holton, G., 2018. 'Reflections on documentary linguistics', in McDonnell, B., Berez-Kroeker, A.L. and Holton, G. (eds.) *Reflections on Language Documentation 20 Years after Himmelmann 1998*. Language Documentation & Conservation Special Publication No. 15. Honolulu: University of Hawai'i Press, pp. 1–11.

Michailovsky, B., 2008. 'Digitized resources for languages of Nepal', in Saxena, A. and Borin, L. (eds.) *Lesser-known Languages of South Asia: Status and Policies, Case Studies and Applications of Information Technology*. Berlin: Mouton de Gruyter, pp. 243–262.

Michailovsky, B., Mazaudon, M., Michaud, A., Guillaume, S., François, A. and Adamou, E., 2014. 'Documenting and researching endangered languages: The Pangloss Collection', *Language Documentation & Conservation*, 8, pp. 119–135. Available at: <http://hdl.handle.net/10125/4621>.

Morgenstierne, G., 1973. *Indo-Iranian Frontier Languages. Volume 4: Kalasha*. Oslo: Universitetsforlaget.

Munshi, S., 2006. *Jammu and Kashmir Burushaski: Language, language contact, and change*. PhD dissertation. University of Texas at Austin.

National Statistics Office, 2025. *National Population and Housing Census 2021: Languages in Nepal*. Thapathali, Kathmandu: National Statistics Office.

Paramore, J.C., 2021. *Mankiyali phonology: Description and analysis*. Master's thesis. University of North Texas. Available at: <https://doi.org/10.12794/metadc1833465>.

Paramore, J.C., 2026. *From abstract to episodic: Representations in speech production*. PhD dissertation. University of California, Santa Cruz.

Paramore, J.C. and Mankiyali, A., 2025. 'The acoustic correlates of word-level stress and focus-related prominence in Mankiyali', *Journal of the International Phonetic Association*, 55(3), pp. 289–330. Available at: <https://doi.org/10.1017/S0025100325100650>.

Pfeffer, G., 2014. 'Ethnographies of states and tribes in Highland Odisha', *Asian Ethnology*, 73(1–2), Special Issue: *The Bison and the Horn: Indigeneity, Performance and the State in India*, edited by R.K. Wolf and F. Heidemann, pp. 259–280.

Rajasingsh, V.R., 2019. *Nicobarese–Hindi–English Dictionary: Pâu–Sanènyo–Lurèòò–Mâuèot–Lamãongsæe–Tèokahãaçnilãahngão–Hindi–English*. Mysore: Central Institute of Indian Languages.

Rahman, T., 2006. 'Language policy, multilingualism and language vitality in Pakistan', in Saxena, A. and Borin, L. (eds.) *Lesser-known Languages of South Asia: Status and Policies, Case Studies and Applications of Information Technology*. Berlin: Mouton de Gruyter, pp. 73–106.

Regmi, D.R., 2017. 'Convalescing the endangered languages in Nepal: Policy, practice and challenges', *Gipan*, 3(1), pp. 139–149.

Rensch, C.R., Decker, S.J., Hallberg, D.G., Backstrom, P.C., Radloff, C.F., Hallberg, C.E., O'Leary, C.F. and Decker, K.D. (eds.), 1992. *Sociolinguistic Survey of Northern Pakistan*. Vols. 1–5. Islamabad: National Institute of Pakistan Studies, Quaid-i-Azam University, and Summer Institute of Linguistics.

Riesberg, S., 2018. 'Reflections on descriptive and documentary adequacy', in McDonnell, B., Berez-Kroeker, A.L. and Holton, G. (eds.) *Reflections on Language Documentation 20 Years after Himmelmann 1998*. Language Documentation & Conservation Special Publication No. 15. Honolulu: University of Hawai'i Press, pp. 151–156.

Rogers, C., 2004. *Secrets of Manang: The story behind the phenomenal rise of Nepal's famed business community*. Kathmandu: Mandala Publications.

van Spengen, W., 1987. 'The Nyishangba of Manang: Geographical perspectives on the rise of a Nepalese trading community', *Kailash*, 13(3–4), pp. 131–277. Available at: <http://www.dspace.cam.ac.uk/handle/1810/227353>.

Trumpp, E., 1872. *Grammar of the Sindhi language*. London and Leipzig: Trübner & Co. and F.A. Brockhaus.

Turin, M., 2006. 'Minority language policies and politics in Nepal', in Saxena, A. and Borin, L. (eds.) *Lesser-known Languages of South Asia: Status and Policies, Case Studies and Applications of Information Technology*. Berlin: De Gruyter, pp. 61–72.

Woodbury, A.C., 2011. 'Language documentation', in Austin, P.K. and Sallabank, J. (eds.) *The Cambridge Handbook of Endangered Languages*. Cambridge Handbooks in Language and Linguistics. Cambridge: Cambridge University Press, pp. 159–186.

Zubair, S., 2016. *A phonological description of Wadiyara, a language spoken in Pakistan*. Master's thesis. Payap University.