Endangered Languages in the Digital Public Sphere: A Case Study of the Writing Systems of Boro and Manipuri

Rachel Garton, Merrion Dale, L. Somi Roy & Prafulla Basumatary

Abstract. Languages around the world are rapidly disappearing, with thousands of minority languages facing potential extinction. A key aspect of language documentation and revitalization is having a viable writing system, as the ability to read and write one's native language allows for the particularly relevant in the digital age, with the Internet allowing communication on a global scale. However, despite the benefits of a writing system, the development and adoption of a writing system and orthographic standards are not easy and can become a lengthy process. Due to the rapid loss of languages, many communities do not have the luxury to allow orthographic norms to evolve naturally and must make rushed decisions on writing system use. Parallel to these initiatives, many language communities have taken to sharing their language and culture online via social media. Social media is much more informal than many other modes of writing and allows writers to explore alternative writing styles, spellings, and even scripts. To explore this further, this paper presents case studies of two vulnerable languages—Boro and Manipuri—through interviews with native speakers and the personal experiences of its co-authors.

1. Introduction

A critical piece of language revitalization and documentation is the development of an orthography. The ability to read and write one's native language opens avenues for using the more. It is also crucial for efficient language documentation, since an orthography brings written language standards and allows community members to be active participants in the documentation process, helping prevent what is known as the "transcription bottleneck" (Cahill and Rice, 2014; Council, 2020).

However, the development and adoption of orthographic standards is no small feat and has many potential hurdles, some relating to the intricacies of the language, and some relating to community identity (Cahill

Rachel Garton 0000-0002-7650-3478 E-mail: rachelgarton@my.unt.edu
Merrion Dale 0000-0003-0659-3743 E-mail: JessicaDale@my.unt.edu
L. Somi Roy E-mail: somiroythree@gmail.com
Prafulla Basumatary 0000-0003-3775-9240 E-mail: prafulla@gauhati.ac.in

Y. Haralambous (Ed.), *Grapholinguistics in the 21st Century 2022. Proceedings*
Grapholinguistics and Its Applications (ISSN: 2681-8566, e-ISSN: 2534-5192), Vol. 10.
Fluxus Editions, Brest, 2024, pp. 437–470. https://doi.org/10.36824/2022-graf-gart
ISBN: 978-2-487055-06-3, e-ISBN: 978-2-487055-07-0

and Rice, 2014; Devi and Choksi, 2022; Pappuswamy, 2017). Not only must communities decide on which writing system to adopt (e.g., Roman, Devanagari, a system with an indigenous script, etc.), but spelling conventions must also be agreed upon. Differing opinions on preferred scripts and orthographic norms can cause this process to be lengthy, which can then delay certain documentation and revitalization initiatives.

Due to the increasingly rapid global loss of many indigenous languages, many communities do not have the luxury of allowing orthographic norms to develop organically and must make rapid decisions on standards so that language documentation and revitalization can occur, some facing technological and infrastructural limitations along the way that can impede their progress. Additionally, speakers often find themselves operating in silos, each creating a variation of a potential orthography but unable to achieve a global consensus. Community input and collaboration is widely considered a critical facet of orthography development (Willis Oko, 2018) but it can be difficult to coordinate large workshops at a frequent cadence.

Social media has seen widespread adoption among indigenous communities for sharing their language and culture online (Buszard-Welcher, 2001; Carpenter et al., 2016; Cassels, 2019; Cru, 2018; Djomeni and Sadembouo, 2016; Emmanouilidou, 2014; B. D. Lillehaugen, 2016) and connecting individuals with other speakers of their language. This form of online written communication can be viewed as a way to gauge the state of a vulnerable stories and poems or religious texts, which may not accurately reflect spoken language and prove challenging for emerging readers to utilize.

Furthermore, standardized and regulated writing may have different lexical use and syntax patterns compared to spoken language (De Korne and Weinberg, 2021). Social media interactions are often much more informal and more accurately reflect daily conversion, meaning they may be a useful way to assess the organic use of a language's writing system and what patterns may emerge based on the types of written interactions that occur. This informal register can allow speakers to express themselves with a form of written language that best reflects their own linguistic realities, which can bolster language revival (ibid.).

To explore these ideas further, we present a case study of Boro (Bodo) (ISO 639-3: brx) and Manipuri (Meiteilon) (ISO 639-3: mni), two Sino-Tibetan languages spoken primarily in North-Eastern India. Both being scheduled languages per the Indian constitution and having available published literature (Pappuswamy, 2017). The writing systems of both languages have undergone notable transformations as speakers pursued scripts that were most representative of community identity and are still evolving today.

Section 2 will discuss the importance of written language for endangered orthographic standards. Section 3 will provide a cultural and sociolinguistic background of North Eastern India, the region in which both Boro and Manipuri are spoken. Section 4 and 5 will give in-depth information on the Boro and Manipuri writing systems, respectively. Section 6 will discuss the digital public sphere and the use of endangered and vulnerable languages online. Section 7 covers the current state of Boro and Manipuri use online, specifically on social media, and includes excerpts and observed trends from interviews conducted with native speakers of these languages as well as the third and fourth authors, who are native speakers of Manipuri and Boro, respectively.

As the authors discuss, the issue of orthography development for unwritten and endangered languages is often approached from a traditional linguistic perspective, though sociolinguists have recently applied critical theory to the topic (cf., De Korne and Weinberg 2021, Hernández, López-Gopar, and Sughrua 2017). Due to this perspective, written language is often viewed as secondary to spoken language and orthography development is seen as a "means to an end" for language documentation.

In an attempt to challenge some of these views and bring to light other facets of orthography development and use, the present paper seeks to include grapholinguistic theory (cf., Neef (2015) and Meletis (2020)) to the issue and connect the concerns of orthographic development to the larger grapholinguistic community. It is the authors' hope that this will bring greater awareness to the issue of orthography development for endangered the broader linguistics community how grapholinguistics can be applied to this critical topic.

2. Orthography Development

Having a corresponding writing system for one's language system provides opportunities to use the language in a wider variety of contexts. It allows not only for personal written communications (e.g., letters, dairies, etc.) but can also increase representation of the allows for the creation of educational materials and other publications. Also, if the script is available in Unicode, then it also allows for use of the language online (e.g., email, websites, social media), though there may still be hurdles to this, which is discussed in more detail in later sections.

Having a written language also helps to alleviate what is commonly referred to as the "transcription bottleneck" in language documentation. During documentation initiatives, there are typically hours upon hours of audio and video footage that is collected. The speech in these recordings must then be transcribed. Without a practical (i.e., non-academic) way to write the language, transcription falls on the shoulders of those

FIGURE 1. Photograph of sign in Meitei Mayek and Roman scripts, left (Bachaspatimayum, 2020)

FIGURE 2. Photograph of published Bodo Literature, right (Daimari, n.d.)

who are formally trained in linguistic transcription. A writing system allows community members to become active participants in the documentation process. However, development and adoption of a writing system and subsequent orthographic standards is a highly complex issue.

Before venturing into the larger discussion on orthography development, it is important to define the authors' use of terms such as "writing system," "orthography" and "script," as these terms are often used in a plethora of ways, often synonymously, among many linguists. We follow Neef (2015) and Meletis (2020), and others in our use of these terms.

A"writing system" is the combination of a script and how it is applied to a given language (i.e., graphematics). One may cite the "English writing system" which is a combination of the Roman script and the correspondences between the graphemes in the script and phonological units in the language. "Writing system" is very commonly conflated with "orthography," which will be addressed shortly.

A "script" is the physical implementation of a writing system, including its graphical features. For example, that the horizontal bar in <T> overhangs the vertical bar on both sides, as opposed to only one side as in <Γ>, is a feature of the Roman script and purely a graphetic concern, with no ties to a specific language.

An "orthography" is the prescriptive use of a writing system for a particular language, including spelling conventions and the correspondence of linguistic units to graphemes and grapheme clusters. Meletis (2020) describes three facets of an orthography—system, use, and norm. System is the actual writing system itself, as previously defined. This system is then put into use. Through more widespread and continued

use, certain norms begin to develop. This leads to the prescriptive use of the system, resulting in normative spellings. See Meletis (2022) for a more in-depth discussion on orthographic standardization.

Though many writing systems may have particular rules which allow for various spellings of a given word (e.g., "ryte" instead of "right," Meletis 2020, p. 156), orthographic norms dictate standardized spellings of written units. While there exist governing bodies to formally establish norms (cf., Académie française), historically many norms evolve organically over time. These norms are more enforced in certain domains, and variability is tolerated in some orthographic spaces better than others (ibid.). For example, published manuscripts are typically proofed for adherence to orthographic rules whereas written utterances shared on social media may be less standardized. This is of particular importance to the present study because the focus is on written language use on social media, where orthographic standards are typically relaxed and writers may feel more comfortable deviating from established norms or experimenting with alternative spelling conventions.

As previously mentioned, "orthography" is commonly conflated with "writing system," but the two are not synonymous. "Orthography" is strictly the prescriptive aspects of a writing system. This is an important distinction because if descriptive linguistics focuses on orthographies rather than writing systems then the descriptions become inherently prescriptive in nature (ibid.).

That is not to say that development/adoption of orthographic norms should not be prioritized, but they should be prioritized for the right reasons. If the end goal of developing orthographic standards is only so that it can be used to document spoken language, does that actually add any value to members of the language community beyond language preservation? Furthermore, critical theories have stressed the potential dangers of forcing a prescriptive orthography in a language community too early, as it can lead to literacy elitism and harmful social distinctions of "literate" and "illiterate" (De Korne and Weinberg, 2021; Hernández, López-Gopar, and Sughrua, 2017).

Grapholinguistic theory also stresses the fact that the orthographic module of writing systems (i.e., the prescriptive rules of writing system use) are not obligatory and written language can exist without strict conventions, as it has done so historically (Meletis, 2020). Rather than force prescriptive norms and rush decisions, the development of a written culture within a language community should reflect traditional communicative practices and evolve within the cultural context of the community as much as possible.

Cultural context is important. There are many non-linguistic factors that influence orthography development (Cahill, 2014), some of which will be explored in greater detail in later sections. These fall into what

is called "sociocultural fit"[1]. Writing and written language are deeply rooted in cultural identity and influenced by a myriad of sociopolitical factors, including religion, national identity, politics, etc. (Cahill, 2014; Meletis, 2020). Script, and even the typeface of said script, is strongly associated with particular groups and, thus, the choice to adapt an existing writing system can be seen as aligning culturally with those associated groups. This is seen to be the case in the North East Indian region, as will be discussed in the next section. No matter how good a linguistic fit a particular writing system may be, it will never be adopted by the writing community if it is a poor sociocultural fit (Cahill, 2014).

Technology is also a key factor in the modern, digital age. While a writing system is necessary for communicating online (Pappuswamy, 2017), the nature of the script and orthography also dictate how easy it is for individuals to engage in writing practices. Not all scripts are supported by Unicode and typing certain graphemes, including those with diacritics, on conventional keyboards can be cumbersome, leading to some writers using shorthand digraphs as an alternative to diacritic use (Chelliah and Garton, 2023).

Furthermore, prior exposure to majority language scripts and orthographic standards can influence the preferences of the language community. It is possible that an individual's native language (L1) is not the language they use in school, where they learn to read in a majority language and in another script. This native script (S1) (cf., Gnanadesikan 2020) may differ from the script used for the L1 writing system (L1WS) and features of the S1 may influence an individual's acquisition of their L1WS. This influence, or transference, may be positive or negative depending on the nature of the L1WS and the L2WS. If the L1WS, which would be S2, is a different script than S1 then the acquisition of S2 may be tedious and frustrating (ibid.). Adoption of a new writing system may go poorly for a language community if too many members have an S1 that differs from the script of the new system.

There are also linguistic factors of orthography development (Cahill and Rice, 2014), which determine "linguistic fit"[2]. In other words, how well does a writing system fit with the language system that it represents? When adopting a writing system, decisions must be made about how to represent certain linguistic features. These decisions are typically focused around how to represent the results of morphophonological processes (Chelliah and Garton, 2023). Should surface or underlying forms be represented? Or should a mix be proposed, as Snider (2014) recommends? Questions such as what constitutes an orthographic word and punctuation conventions must also be addressed. For an in-depth

1. See Meletis (2020), chapter 8 for an in-depth review of sociocultural fit.
2. See Meletis (ibid.), chapter 6 for an in-depth review of linguistic fit.

discussion and examples of these in a Tibeto-Burman context, see Chelliah and Garton (2023).

This plethora of factors means that the questions of writing system and orthography are different for every language and community of writers, though there are common threads and lessons to be learned from all stories. With that, the present paper focuses on the stories of two – focusing on the use of these languages' orthographies in digital public spheres.

3. Linguistic Landscape of North Eastern India

India is home to over 170 languages and 540 dialects, according to the Linguistic Survey of India (Grierson, 1921), and is a linguistically rich country. The North Eastern region of India alone is home to over 400 languages and dialects, 80 of which are considered endangered (Pappuswamy, 2017). These languages and dialects are spread across four language families—Tibeto-Burman, Indo-Aryan, Austria-Asiatic, and Tai-Kadai—the majority belonging to the Tibeto-Burman grouping (Haokip, 2021).

Language status in India is no less complex. While many countries merely classify languages as "official," the multilingual environment of India necessitates a more layered system. Special language status in India is determined by the Eighth Schedule of the Indian Constitution, resulting in some languages being referred to as "scheduled Hindi, though business and government work may also be conducted in English.

There are 22 scheduled languages[3], and these languages have special status in India. Scheduled languages are guaranteed cultural representation and the Indian government has a responsibility to promote them (Pappuswamy, 2017). Scheduled languages are also often the third language taught in schools per India's Three-Language Formula, which stipulates that schools teach Hindi, English, and a third Indian language (Eucation, 2020). The general purpose of scheduled languages is to preserve and promote linguistic diversity in India and protect indigenous languages of the region. However, this in no way guarantees the vitality of these considered vulnerable by UNESCO (Moseley and Nicholas, 2010; Pappuswamy, 2017). UNESCO cites the lack of a viable writing system and consistent script as a factor of this vulnerability (Moseley and Nicholas, 2010).

3. All 22 languages are: (1) Assamese, (2) Bengali, (3) Gujarati, (4) Hindi, (5) Kannada, (6) Kashmiri, (7) Konkani, (8) Malayalam, (9) Manipuri, (10) Marathi, (11) Nepali, (12) Oriya, (13) Punjabi, (14) Sanskrit, (15) Sindhi, (16) Tamil, (17) Telugu, (18) Urdu (19) Bodo, (20) Santhali, (21) Maithili and (22) Dogri (Governent of India, 2017).

Language and writing are deeply rooted in culture and, as such, are strongly tied to cultural identity. As previously discussed, the choice of writing system and script is often heavily influenced by socio-political and religious factors. This is absolutely the case in North Eastern India. The religious landscape of India is complex, with a spectrum of Buddhism, Hinduism, and Christianity. Writing systems are often heavily associated with religion (Lillis and McKinney, 2013; Pappuswamy, 2017), with Devanagari seen as Hindu, Latin seen as Christian, and Tibetan seen as Buddhist (Pappuswamy, 2017). Furthermore, deciding power over script choice is seen as a matter of local autonomy and can result in socio-political tension, as was the case with the Boro script shift in the 1970s (covered more in section 5).

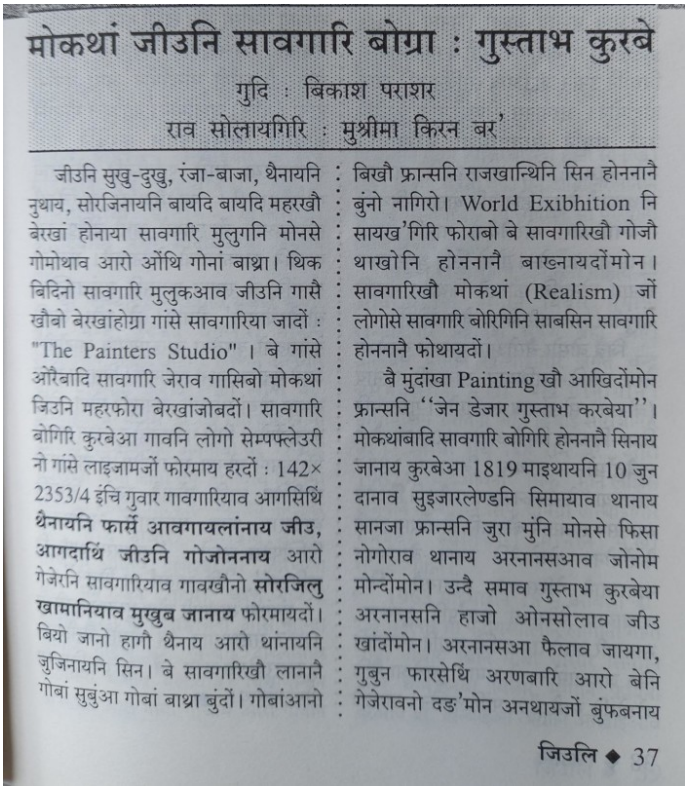
The North Eastern region of India has dozens of tribal bodies, many of which have historically felt culturally repressed and economically exploited (Prabhakar, 1974). While great efforts have been made by the Indian government to weave these tribes into the larger tapestry of Indian culture, many individuals still feel closely tied to their tribal identities rather than a homogenous national identity. Script use and orthographic norms are a strong part of this identity and the use of a particular orthography can be seen as a way of expressing this identity (ibid.).

While historically the use of a preferred script was highly localized, the advent of the Internet, the increased accessibility of multilingual keyboards (e.g., Keyman, see Computational Resource for South Asian Languages (CoRSAL) (n.d.[c])) and the expansion of Unicode have given speakers of minority languages new, globally visible platforms upon which they can engage and express their linguistic identities.

4. Boro Writing System Shifts and Current Orthography

The Boro language, also known as Bodo, is spoken in Assam, particularly the autonomous Bodoland Territorial Region in North Eastern India (CoRSAL, n.d.). It is a member of the Sino-Tibetan language family and spoken by over 1.4 million people (India, 2011). Despite the large number of speakers, UNESCO still considers Boro a vulnerable language, with the writing system being a contributing factor (Moseley and Nicholas, 2010).

Originally, the language was written with a mixture of Roman, Assamese, and Bengali writing systems, with no orthographic conventions established (Sarmah, 2014). In the 1950s, the Assamese script was settled on, but this was short-lived. Since the 1950s, there have been continued discussions over which script would best represent the identity of the Bodo people. Assamese was not the preferred script for the Bodo



मोकथां जीउनि सावगारि बोग्रा : गुस्ताभ कुरबे

गुदि : बिकाश पराशर

राव सोलायगिरि : मुश्रीमा किरन बर'

जीउनि सुखु-दुखु, रंजा-बाजा, थैनायनि नुथाय, सोरजिनायनि बायदि बायदि महरखौ बेरखां होनाया सावगारि मुलुगनि मोनसे गोमोथाव आरो ओंथि गोनां बाथ्रा। थिक बिदिनो सावगारि मुलुकआव जीउनि गासै खौबो बेरखांहोग्रा गांसे सावगारिया जादों : "The Painters Studio" । बे गांसे ओरैबादि सावगारि जेराव गासिबो मोकथां जिउनि महरफोरा बेरखांजोबदों। सावगारि बोगिरि कुरबेआ गावनि लोगो सेम्पफ्लेउरी नो गांसे लाइजामजों फोरमाय हरदों : 142× 2353/4 इंचि गुवार गावगारियाव आगसिथि **थैनायनि फार्से आवगायलांनाय जीउ, आगदाथि जीउनि गोजोननाय** आरो गेजेरनि सावगारियाव गावखौनो **सोरजिलु खामानियाव मुखुब जानाय** फोरमायदों। बियो जानो हागौ थैनाय आरो थांनायनि जुजिनायनि सिन। बे सावगारिखौ लानानै गोबां सुबुंआ गोबां बाथ्रा बुंदों। गोबांआनो बिखौ फ्रान्सनि राजखान्थिनि सिन होननानै बुंनो नागिरो। World Exibhition नि सायख'गिरि फोराबो बे सावगारिखौ गोजौ थाखोनि होननानै बाख्नायदोंमोन। सावगारिखौ मोकथां (Realism) जों लोगोसे सावगारि बोरिगिनि साबसिन सावगारि होननानै फोथायदों।

बै मुंदांखा Painting खौ आखिदोंमोन फ्रान्सनि "जेन डेजार गुस्ताभ करबेया"। मोकथांबादि सावगारि बोगिरि होननानै सिनाय जानाय कुरबेआ 1819 माइथायनि 10 जुन दानाव सुइजारलेण्डनि सिमायाव थानाय सानजा फ्रान्सनि जुरा मुंनि मोनसे फिसा नोगोराव थानाय अरनानसआव जोनोम मोन्दोंमोन। उन्दै समाव गुस्ताभ कुरबेया अरनानसनि हाजो ओनसोलाव जीउ खांदोंमोन। अरनानसआ फैलाव जायगा, गुबुन फारसेथि अरणबारि आरो बेनि गेजेरावनो दङ'मोन अनथायजों बुंफबनाय

जिउलि ◆ 37

FIGURE 3. Photo of an article about French painter Gustave Courbet (Boro, 2005)

people, who wanted autonomy and a way to express their cultural identity (George, 1994; Prabhakar, 1974).

In the 1960's the first Boro medium schools were introduced, where, as the name suggests, Boro was the medium for all classroom instruction. This prompted accelerated conversations about the need for a standardized Boro writing system to be used in textbooks (C. Basumatary, 2014; P. Basumatary, 2017; BC, 2019; Wary, 2017). Throughout the 1970s, the Bodo Sahitya Sabha (Bodo Literacy Society) became a major participant in the push for a new script for the Boro writing system. Originally, the Roman script was proposed and school materials began to be published in Roman script following proposed orthographic rules (Prabhakar, 1974). However, other parties pushed for Devanagari and eventually the Bodo Literacy Society, Bodo Sahitya Sabha, shifted their stance, formally agreeing to adopt Devanagari for Boro in 1975 (Sarmah, 2014). Despite this formal adoption, discussion about which writing sys-

FIGURE 4. Photograph from a Boro orthography workshop on December 12, 2010 (photo taken by co-author Prafulla Basumatary)

tem and script to adopt are ongoing. Co-author Prafulla Basumatary made an observation regarding personal preference for one script over another on Facebook. He noted that Bodos historically have used both scripts but that a section of the population still has a preference for using exclusively Roman script and he observes that this preference carries over into their script choice on Facebook.

Today, the Bodo Sahitya Sabha remains a critical organization which continues to work with the Boro community on orthographic standards for the language. Workshops are frequently held and the literacy society has been known to leverage social media to announce workshop decisions. This will be discussed in more detail in section 7. Boro is largely written in Devanagari (see Fig. 3), but some writers will use the Roman script in certain contexts. As the authors explore in section 7, there are numerous motivations for a Boro speaker to opt to write in either Roman or Devanagari script.

5. Manipuri Writing System History

The Manipuri language, also known as Meiteilon (also, Meithei, Meitei, Meitheiron, Meetei), is widely spoken in the Indian state of Manipur (CoRSAL, Manipur) as well as in Assam, Tripura, Bangladesh and Myanmar. The official Census of India 2011 recognized 1.7 million speakers, but other estimates are closer to 3 million due to the distribution of speakers beyond Manipur (Roy, 2017).

FIGURE 5. Digitized image of a Manipuri manuscript (Shobhana Lakshmi Cheliah, n.d.)

FIGURE 6. Photo of Manipuri speaker Chanam Hemchandra displaying a Manipuri manuscript (Roy, 2008)

FIGURE 7. Photograph of Manipuri manuscript collection (Molinaro, 2008b)

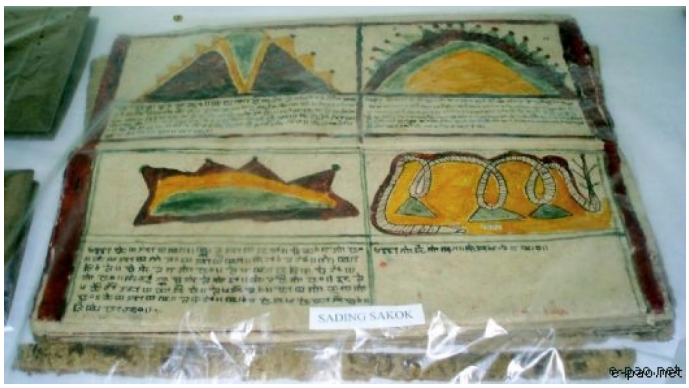


FIGURE 8. Photograph of a Manipuri manuscript: Sading Sakok (Molinaro, 2008a)

Despite these large numbers, UNESCO reports still consider the language vulnerable due to a number of factors, including lack of a unified writing system (Moseley and Nicholas, 2010). Like Boro, Manipuri has been written with more than one script and this flux has prevented a unified writing system to be accessible for all generations.

Prior to the 18th century, the Manipuri language was written using an abugida called Meitei Mayek (Laithangbam, 2017). The script is from a group of Tibetan scripts originating from the Gupta Brahmi script (Shobhana Lakshmi Chelliah, 2011). In the early 18th century, Hindu missionaries spread Hinduism amongst the region, prompting the converted King of Manipur (and, later, the British) to decree the use of the Bengali script (Bangla) instead (Laithangbam, 2017; Roy, 2017). Bangla initially coexisted with Meitei Mayek throughout the 18th, 19th, and

20th centuries but then gradually gained ascendance after the Anglo-Manipuri War of 1891. Today, roughly 40,000 manuscripts in Meitei Mayek remain, in part because Hindu missionaries torched many writings in Meitei Mayek (Roy, 2017). See Figs. 4 and 5 for a scan of a surviving manuscript and photo of a bound book.

While Meitei Mayek was originally replaced with Bangla, the script was officially reinstated as official for the Manipuri writing system in 2006. Now shop signs, newspapers, and other public written materials in Meitei Mayek can be found (Laithangbam, 2017). It was added to the Unicode Standard in 2009 (Roy, 2017) which has further spurred its adoption. The Supplementary Document by co-author L. Somi Roy details this history in his account of his attempt to bridge the gap in the usage of the two scripts.

6. Endangered Languages in the Digital Age

The 21st century has brought a new wave of technologies, some referring to society's current digital transformation as the "Fourth Industrial Revolution" (Schwab, 2018). The rapid advancement of information technologies, which allow users to easily engage with one another on a global scale, has given rise to new public spheres which exist only digitally.

The concept of "public sphere" (German: *Öffentlichkeit*) was originally coined by Jürgen Habermas, a German philosopher. Habermas defined a public sphere as a community "made up of private people gathered together as a public and articulating the needs of society with the state" (Soules, 2007). While Habermas' definition does include virtual communities which do not "necessarily exist in any identifiable space" (ibid.), the authors specify public spheres which exist online as "digital public spheres" to better differentiate the phenomenon from more traditional and analog public spheres.

Among these digital public spheres exist those comprised of speakers of endangered languages who are communicating and connecting online to share their language and culture. Before the digital age, speakers of endangered languages had much more siloed communities in which language use, especially written language use, might take place. The language might be written locally within a tight geographic space, but writings rarely spanned beyond that and, if they did, it was likely in private communications or formal (often religious) writings, not in a broader social capacity. As a result, languages used in such a confined environment were at a stronger risk of dying out, as speakers defaulted to the more wide-spread and accepted majority languages of their area. For details on the many factors which contribute to language endangerment, see Shobhana L. Chelliah (2021).

Enter the age of the Internet, digital communication, and social networking platforms. Suddenly, the reach of small language communities is expanded globally so not only can individuals communicate across the world in real time, but their writings can be easily diffused among others. Buszard-Welcher's 2001 publication *Can The Web Help Save My Language* explored the then-emerging phenomenon of websites dedicated to the use, preservation, and promotion of endangered languages, and the potential role these sites could play in the future of language revitalization. She analyzed these sites in terms of their creators, users, content, and appeal to young speakers. Since 2001, the arrival of social media has provided many significant platforms for endangered

Social media can serve as an access point for many individuals to other people who speak their language, or who are affiliated with the language through their heritage (Cassels, 2019; Scannell, 2012). Many of these platforms allow the creation of folksonomies, which ensure that the tagged content is searchable by others and can boost the content's visibility. This property alone has valuable implications for the promotion of minority and endangered languages as evidenced in examples of videos on TikTok using the hashtag #cajunfrench which, at the time of writing, have garnered a collective 30.7 million views. Selecting this tag brings up every video that has used it (often an individual sharing a recording of an older family member telling a story in Cajun French or teaching songs, stories, and new vocabulary). Similarly, separate digital language activism campaigns and case studies on the Gwitch'in, Sami, and Zapotec languages have resulted in the creation of Twitter hashtags to encourage use of these indigenous languages, and in the case of Zapotec, helped to establish a more active online "writing culture" (Billock, 2015; B. D. Lillehaugen, 2016).

The creation and utilization of Facebook groups for language learners/communities has served in some respects as a modern interpretation of endangered language websites. Group members are able to engage in discourse with one another and share questions, media from other platforms, and even archived materials related to the community's shared language and culture (Dale et al., n.d.; De Falco and Cesarano, 2016). For example, the University of North Texas' Computational Resource for South Asian Languages (CoRSAL) has worked with linguists and language collection depositors for the Boro, Lamkang, Azamgarhi, and Burushaski language collections to create such groups on Facebook as a means to connect members from each speaker community to their archived easily accessible for promotion and revitalization purposes for speaker communities who might otherwise face internet connectivity struggles, making it more feasible to access this data through Facebook on their mobile devices.

Community-based language and policy planning prioritizes the agency of representation and use (McCarty, 2018). The participatory

nature of the digital public sphere has allowed for certain areas to serve as grounds for grassroots language reclamation that can manifest through vernacular literacies, as well as language ideology and image planning discourse (Androutsopoulos, 2013; Barton and Lee, 2012; Cru, 2018). A 2018 case study examined the comments section for two videos of Yucatec Mayan and Mapuche rap music and found that the positive feedback left by members of each respective language community were expressed through multiple orthographic systems and representative of sentiments of prestige and image planning among the speakers of these indigenous languages (Cru, 2018). Cru noted that, in contrast to regulated environments like schools, "user-generated digital spaces are opening up opportunities for the production of meaningful writing practices in indigenous further noted that these spaces are more free from expectations of standardization. This is a sentiment that B. Lillehaugen (2019) and B. D. Lillehaugen (2016) supports in her work with the use of Zapotec on Twitter. She notes that it is not crucial for a language to have a standardized writing system for them to participate in writing their language on social media, and that such standardization can often arise out of the ongoing writing practices of the speaker population (p. 365).

6.1. Representation and Negotiation in the Digital Public Sphere

While the technology exists to facilitate online use, there are still developmental hurdles for many languages. If a language's orthography utilizes a script that is not yet available in Unicode, there is still work to be done. Harkening back to the early exploration of dedicated endangered language websites, Buszard-Welcher identified the issue of font (un)availability, noting that there was considerable discussion of this issue across these websites she studied:

> Of the sites in the sample, 28% had content on writing systems. This includes discussions of fonts (7 sites), offering fonts for free download (7 sites), and discussions of writing systems (7 sites). The latter are usually short descriptions of the development of particular writing systems accompanied by a chart of symbols. One site, Cherokee, has online lessons for learning to use the writing system (Buszard-Welcher, 2001, p. 334).

The need for bridging the digital language divide, and ensuring greater diversity in web/social media compatible scripts persists, but there have been notable efforts put forth by the speakers of affected Osage language (ISO-639-3: osa) is a Native American language spoken in modern day Oklahoma, United States and is written with a script that represents the movements in traditional dances. The project to incorporate the graphemes into the Unicode Standard was long and expensive,

requiring coordination with community leaders and Unicode developers (Martucci, 2014). Thankfully, this hard work paid off and today Osage speakers can write their language digitally (Nation, 2014).

Despite the inclusion of many scripts into the Unicode Standard, usability is still a major concern. Multilingual keyboard software and mobile applications have become fairly advanced, but the usability gap between more widespread scripts (e.g., Roman, Cyrillic) and specialized scripts (e.g., Meitei Mayek) is wide. These usability concerns add a layer of complexity to the already-complex issue of orthography development and script choice. In fact, as explored in Section 5, there have been proposals to "Romanize" Manipuri in order to facilitate easier digital use, particularly on mobile devices (Roy, 2017). Additionally, writers of some languages, such as Hakha Lai (ISO 639-3 code: cnh) and Mizo (ISO 639-3 code: lus), may also resort to shorthand representations of graphemes which are cumbersome to type. Both the Hakha Lai and Mizo orthographies utilize the grapheme <ṭ> but writers will often substitute this with <tt> and <tr>, respectively (Chelliah and Garton, 2023).

These examples raise an interesting question with respect to the case studies at hand: How are Boro and Manipuri orthographic standards negotiated in a digital space? What factors may dictate or motivate individual script choice/spelling conventions? As discussed, the use of endangered languages online has surged in recent years, resulting in far more written material than ever before. Could digital platforms, like social media, expedite more natural development of community orthographic norms? The following sections will explore written language use by Boro and Manipuri speakers online, specifically as it pertains to their use of social media, and will discuss potential implications for how social media might be leveraged by language communities continuing to develop orthographic rules for their chosen writing system.

7. Boro and Manipuri Online

In order to understand the lived experiences of Boro and Manipuri writers, the authors interviewed Boro and Manipuri speakers who use their respective language to engage online, either actively or passively. The focus of the interviews was on Facebook, as both aforementioned Bodo Literacy Society, have a Facebook presence, but interviews also touched on other social media and digital platforms, such as WhatsApp, YouTube, Instagram, and email. Themes from the interviews were then validated against the observations and experiences of the third and fourth authors as users of social media platforms such as Facebook and WhatsApp and native Manipuri and Boro speakers, respectively. The

following sections discuss the individual results for Boro and Manipuri before summarizing the findings and identifying key themes.

7.1. Boro Use Online

Although, unsurprisingly, every internet/social media user will have behaviors and tendencies that shape the varying observations of their user unique experiences, among the Boro speakers interviewed there were trends in their observations of the factors influencing digital Boro writing. Interviewees reported varying writing domains of original posts on Instagram and WhatsApp. They also reported active commenting on YouTube videos as well as both individual and group Facebook posts.

There was a unanimous consensus among interviewees that a primary motivating factor for the choice between using Roman or Devanagari script online was that of convenience. Perception of convenience was motivated in part by the assessment that Roman script could be employed with a greater speed. Additionally, co-author Prafulla Basumatary notes that the Roman keyboard may be regarded as easier to use and present fewer spelling challenges for users. Individual technological proficiency was another factor that one interviewee cited as contributing to the convenience of Roman script, particularly with respect to an individual's ability to incorporate another keyboard like Devanagari onto their (primarily mobile) devices. This was a notable factor in one interviewee's observations of Roman script usage on WhatsApp.

The interviewees indicated that educational medium background could be a similarly influential factor in a Boro speaker's script choice, word choice, and spelling conventions. The difference between a Boro medium educational background and another language medium, such as English, Assamese, or Bengali, impacts a young Boro speaker's exposure to their medium don't have the same access to Boro language classes where Devanagari spelling conventions are taught.

One of the primary aims of the Bodo Sahitya Sabha, according to their official website[4] is to impart a sense of urgency regarding the need for more mother-tongue education and to "expand the Bodo language by producing standard literature and Text Books in [Bodo]" (Sabha, n.d.). These efforts are part of co-author Prafulla Basumatary's current postdoctoral research on the Initial Literacy Measures of Boro Learners. Of the three Boro speakers interviewed, only one of them had received Boro medium educational training. As a result, they expressed a feeling that it was their responsibility to share Boro content on social media using Devanagari, and to also provide transliteration and translation as they

4. The official statement from the Bodo Sahitya Sabha (Bodo Literacy Society) can be found on their website at `https://bodosahityasabha.org/aims_objects.html`.

saw fit. Alternatively, the following quote is from an interviewee who received an English medium education, and details the way this background impacted their understanding of Standard Boro in Devanagari script:

> [...] most of us didn't really have Boro as a subject in our school since we were... from the English medium school, we would say that. So, we don't really, to be honest, even I'm just learning the Boro script. We didn't really have any idea about how it's written in Boro actually. So, since we just speak Boro... we just write it in Roman alphabets. Because that's easier for us rather than writing in Devanagari script since we don't really have the actual idea of how it's written in Devanagari...So, I can read it, but the problem was we didn't really understand a lot of what was written because the written words are quite different from the everyday words that we use when we're speaking. So yeah, we could read, most of us could read the Boro, but... it was like we didn't really understand. But yeah, now I think a lot of us are trying to understand what it is, by maybe taking the help of dictionary and stuff. Yeah. That's how it is.

This interviewee reported seeing Standard Boro on Instagram captions and shared posts in WhatsApp and giving up trying to read them because they were too hard for eir to understand. Another Boro speaker with an English medium background reported a higher comfort level using Standard Boro, the formal written register taught in Boro medium schools, and reported switching between writing in English or in Boro with Roman and Devanagari script depending on the nature of the post. Formality of topic was a common theme among interviewees contributing to an individual's script choice across digital platforms. Interviewees noted a difference between Standard Boro and a more informal and conversational Boro, where the primary difference between the two is word choice. They described Standard Boro as being common to domains of online posts including discussions or posts about Bodo culture, literature, and other topics of academic or intellectual significance. Additionally, groups on Facebook, particularly those disseminating information to a large audience, were reported to commonly write in Standard Boro. Posts created by Bodo news media accounts on Facebook exclusively write using Standard Boro and commonly use Devanagari. Interviewees noted that Roman script is not limited to conversational Boro and can be seen used for formal topics, however one interviewee observed that, in eir experience, those who use Devanagari on social media would be strictly using Standard Boro.

In addition to topic formality, the script choice of one's online conversation partner may also have some bearing on one's personal script choice, particularly in the comment sections of platforms like Facebook or Youtube. An interviewee noted that in their personal experience many individuals in their Facebook network (including themselves) use English as their prominent language for captioning their posts, but that

even when e creates a Facebook post in English e will engage in Boro comments with script and formality variance depending on the topic and conversation partner.

Another interviewee, who reported no longer being an active Facebook user, recalled their memories from the many years they were active on Facebook in addition to the exchanges they witnessed on other digital platforms. They reported seeing debates taking place in Facebook groups and noted that the formality of word choice might change the more heated an argument got. They occasionally observed script switching depending on the users' educational backgrounds. For example, if the digital interlocutors shared a Boro medium background the exchange might stay in Devanagari, and if not the exchange might switch from Devanagari to Roman the longer they continued to converse. They reported that online arguments might switch from English/Assamese to Boro in public groups or shared posts. Additionally, the interviewee observed that YouTube comment sections differ in formality and script choice by video domain. Often more informal content (such as music videos or funny videos) will yield a more informal form of writing in both Roman and Devanagari scripts, but in their experience, formal content like Boro political/news videos have more formal commenting and individuals more heavily utilize Devanagari. A different interviewee reported that in their experience Roman script is what they see utilized the most on YouTube, but that tone and word choice would vary depending on the content of the video in question.

That being said, the informal written register of Boro is reported to be seen across digital platforms and is commonly seen in the comment sections on Facebook and YouTube, where individuals are writing in the same way they would be likely to speak to one another in person. In one interviewee's experience Facebook comment sections tended to frequently display both Roman and Devanagari scripts, particularly when the original post was from an account or group with a large following. When prompted as to whether individuals might be motivated to use one script or another the interviewee noted the following:

> [As] far as I have noticed, they may have their own writing but all of them know the standard one. So if the other person is writing in standard form of Boro, then the other, like for me, then I will reply in there in standard form. So, like that, if I want to comment on some one friend who I have known now like from earlier, then I will write how I speak.

With respect to any type of metalinguistic discourse, all respondents agreed that this was not common practice among Boro online conversation. The Boro speaker who indicated e had observed metalinguistic discourse noted that, with respect to addressing or announcing the spelling conventions at large, they have only witnessed two individuals post anything of that nature on Facebook. E did mention, however,

that e had more frequently seen people involved in Bodo news media pointing out spelling errors to other Boro-speaking Facebook users if those individuals were sharing intellectual content, particularly relating to Bodo identity and culture.

The aforementioned interviewee's remark that for most users, Boro writing on social media is "more about the message" than the spelling, was echoed by another interviewee who felt these conversations seldom take place. Eir impression was that spelling discourse and graphol-inguistic correction was rare because many people would not see the value or impact that altering their spelling choices would have on their everyday life. This interviewee shared memories of their own personal transliteration work converting Boro from Devanagari to Roman script. E recalled consulting eir friends on WhatsApp for their guidance on the spellings of certain words and encountered two different responses:

> I didn't really get an answer from them, because they... all they would tell me is like, "This is what we have been writing, since a very long time..." Yes. Or maybe sometimes someone would say, like, when I tell them "Okay, but this person told me that this is written right," and then they'll ask me, "Okay. Where... are they from? No no, they are wrong. This is how you should write." And then again, there's another reaction to it somewhere else: "Oh, maybe you can also spell it like that." Yeah, there are like two reactions to it. Someone sometimes tells me that "No, this is wrong, you should read what I've written, I've told you," and when someone tells me "No, you can also write this one...." And then I think they accepted all at times, sometimes they even... they are like you know, "Just accept it all because, like it also... it like, you know, it's because of the dialect probably. That's why it's different so since we have to include all the dialects of the Boro, so why don't... why not include all of it all the spelling?"

Dialectal variation in spelling was reported by more than one interviewee and points to the notion that this is a common justification for said variance. Similarly, co-author Prafulla Basumatary has observed that individuals who do not know correct spelling may shorten the spelling of words in Roman or Devanagari script either by choice or without knowing their spelling is inaccurate. After years of work, Basumatary believes that the Boro Sahitya Sabha is close to the goal of having a finalized orthography. He posits that including more linguists into the workflow in addition to those who are from literature and other academic backgrounds, may also be beneficial to completing the process.

7.2. Manipuri Use Online

Similar to the frequent use of Roman script for Boro, there was consensus by Manipuri respondents that Roman script is the predominant script used online, despite Roman script having no history of an official

script. There was a general consensus among respondents that Roman is easier to use on conventional keyboards compared to Bangla or Meitei Mayek so it is often preferred. Additionally, while Meitei Mayek was added to Unicode in 2009, it is not always supported on mobile devices and there is minimal awareness of Meitei Mayek digital keyboards. One respondent shared that despite having Unicode, people do not realize and thus default to Roman script:

> I also use in Roman script, because... in the mobile phone, Android phones, or computer, ... Bengali is already there but Meitei script is something difficult to use ... we have the Meitei script ... I think we have Unicode also but people [are] not aware about that. That's why most of the people use Roman script.

As this respondent also reported, even if an individual is aware that Meitei Mayek can be used on their personal devices, it is difficult to use and Roman is more convenient for rapid online input. However, while convenience was frequently reported as a contributing factor, perhaps the most notable reason for the use of Roman script being preferred online is the sentiment that Roman script is the common script to bridge the gap between those literate in Bangla and those taught in Meitei Mayek. When Meitei Mayek became the official script of Manipuri in 2006, it resulted in a schism between those who were educated in Bangla and the younger generation who became fluent in Meitei Mayek. Younger Manipuri speakers are unlikely to know Bangla and many older speakers have not fully mastered or become comfortable with Meitei Mayek. Roman script, however, is mostly known by all speakers, despite never being an official script used by the language. Therefore, to ensure one's message can be read by all, Roman script may be preferred. As one respondent reported:

> Older people [use] Bengali but the new generation does not know about the use of Bengali script so they use Meitei script or Roman script. Because, since 2006, [we] started to use the Meitei script instead of Bengali script. That's why the new generations know only the Meitei script.

This schism in script was often mentioned by respondents and has had significant impact on the writing culture of Manipur, particularly in more informal registers such as social media. More background the schism is discussed in the Supplementary Document appended to the end of the present manuscript.

Since Roman is not, and never had been, an official script, there are no spelling conventions for its use. Writers will often use invented spellings that they feel most accurately represent the phonology of the utterance. While this does result in variation across texts, respondents shared that spellings are close enough that, with the help of context, communication is not hindered in any way. Due to lack of codified

spelling conventions, there was also a sentiment among respondents that standardized spelling was not a major concern when using Roman script on social media: "in social media it doesn't matter about the spelling or whatever, any word they like they can use and they can use whatever spelling." Roman is viewed more as a convenient way to transcript Manipuri in a script that all can understand rather than a script being preferred for official use. Some of the spelling variations reported by respondents is inconsistent use of double vowels to indicate phonemic vowel length (e.g., <aa> vs. <a> for [a:]), which is a common spelling variation seen in other languages of the region (see Chelliah and Garton (2023)). Other observed spelling variation centers around representation of aspirated consonants, with one respondent reporting that he encounters <f> and <ph> both used for [p^h].

These reports highlight the persistent issues faced by writers of uncommon and minority scripts globally, not just in North Eastern India. Accessibility of a script may not be optimal, leading many writers to opt for a majority script in certain situations. This inaccessibility can take many forms. Firstly, Unicode representation can be difficult to obtain. The Unicode consortium will only incorporate new characters once they have been vetted and approved and typefaces must be designed so that they conform to Unicode standards, which can be a tedious process requiring support from Unicode developers (Martucci, 2014).

Once Unicode support is accomplished, there are still issues with the distribution of the fonts and input methods for devices. As mentioned by Manipuri respondents, many people are unaware of the availability of digital Manipuri keyboard options. Even if awareness is achieved, there is then the issue of adoption by users. Some input methods can have issues of usability and ergonomics. Many conventional input methods, such as digital keyboard layouts and predictive text, are highly alphabet-centric and not designed with other writing system typologies in mind (Rowe, 2022). The user experience for writers of other writing system typologies can be lacking. These issues can compound and lead to cases like Manipuri, where writers default to a majority script like Roman when the register of writing is informal and ease-of-typing is a priority.

The informal register of social media could be a factor of script choice as well. Respondents did not report many original writings in Roman script on social media despite the Manipuri community having a very active writing culture, such as the publication of *Asangba Nongjabi*, "Crimson Rainclouds," a famous play written by MK Binodini Devi, who signed her works as Binodini. For more background on this publication, please see the Supplementary Document appended to the end of the present manuscript. Formal writings such as Crimson Rainclouds are published in Meitei Mayek or, if published prior to 2006, in the Bangla script. Other official publications such as learning materials, signage,

newspapers, etc. are all also in Meitei Mayek. Roman script is strictly used in informal contexts, such as social media discussions. due to convenience.

The topics of social media discussions, though informal, can be broad. Respondents reported that topics discussed online typically varied by age group and also ethnic groups and tribal affiliations. Political discussions and current events are often discussed, as seen with other languages on social media. Metalinguistic, or metagrapholinguistic, discussion was not commonly reported. However, a notable occurrence was mentioned by a Manipuri interviewee who said that when Meitei Mayek was added to Unicode there was some discussion online about this advancement. Interestingly, though many Manipuri speakers are undoubtedly aware of the inclusion of Meitei Mayek in Unicode, awareness of keyboards to input the script appears to remain lacking.

7.3. Summary of Findings

When comparing interviews across both languages, themes began to emerge. These themes can be grouped into four general categories—the writers' script choice, the topics being discussed online, spelling conventions and word choice, and metalinguistic/metagrapholinguistic discussion.

Script Choice

By far the most prevalent factor influencing script choice was ease of use and convenience. For both Boro and Manipuri writers, the use of the Roman script is by far the easiest in a digital format and, thus, is the most commonly used. Toggling back and forth between the keyboards is clunky, and in the case of Manipuri speakers not all users have access to a Meitei Mayek keyboard.

Education was also a notable factor. Not education level, but rather educational background and the year(s) in which the individual was educated. In the case of Boro, it is not the case that all Boro speakers receive a Boro medium education, and as a result this impacts non-Boro medium students' proficiency in reading and writing their language using Devanagari script and contributes to their preference for Roman script. With respect to Manipuri, as noted above, the official shift from Bangla to Meitei Mayek caused a schism, so those educated before vs. after 2006 were educated in different scripts.

Lastly, the nature of the topic being discussed was also mentioned as an influencing factor for some interviewees. For example, if a post is longer or more formal then a writer may opt to use Devanagari for Boro, as it is the official script. One interviewee also mentioned that

if they are in a conversation with someone they will often adapt to the script used by their partner so the conversation flows in one script rather than a back-and-forth. This can be considered a kind of written code-switching.

Topics Discussed

For topics discussed, speakers of both languages reported a wide range of concepts across various digital platforms. Everything from news and politics to culture, such as poetry, art or musical performances. This is a positive indication of a vibrant and healthy online writing culture, which is very promising for the long-term vitality of these

Spelling/Word Choice

Interviewees frequently reported that spelling conventions were not often regarded on social media, likely due to the informal register. One Manipuri interviewee even said "no one cares about spelling online," though some respondents said they observed social media users correcting one another's spelling. Formality of certain topics in Boro were one factor that contributed to the formality of the register. Correction of spelling and grammar seemed to be more common as a response to more formal and official posts, such as posts from official pages. One Boro respondent explained that this is because official pages are seen as representatives of the language and have higher visibility. Therefore, proper language use is perceived as more important.

Dialect was also cited as a reason for some of the spelling variation. If someone is unsure of how to spell a word, they will make a guess based on how it sounds. The phonology may differ from dialect to dialect, making these invented spellings differ in tandem.

One final theme for spelling is the issue of loan words. Respondents from both languages mentioned loanwords being a point of uncertainty with spelling because some writers wish to preserve the original spelling of the borrowed word and others may wish to nativize it and adapt the spelling to native conventions. This is particularly common when the loan word is from a language with a shared script (e.g., English, Hindi, etc.)

Meta(grapho)linguistic Discussion

One final theme is the occurrence of metalinguistic discussion, or more specifically metagrapholinguistic discussion (i.e., discussing written language in written posts). While such discussion was not as common as originally expected, Bodo respondents did mention some notable occurrences. As already mentioned, some users would correct one another's

spelling, which could be seen as a form of metagrapholinguistic interaction. Additionally, one Manipuri respondent reported discussion about the inclusion of Meitei Mayek into Unicode, which can also be considered a form of metagrapholinguistic discussion. More notable, however, was the occurrence of announcements about changes to or decisions made about Boro orthographic standards, often due to an orthography workshop. These announcements might also yield isolated comments and discussion.

8. Conclusion and Looking to the Future

In the digital age, it is propitious for minority languages to expand online, as this provides more channels for communication in the increase exposure to the international community. However, writing one's keyboards for one's native script. In these situations, writers may opt to use an already-available script as a stop-gap until Unicode expands or input devices improve. The issue of Unicode availability and the usability of digital keyboards remain serious issues for minority languages and underrepresented scripts. It would behoove the linguistic community to focus on these issues as a serious concern for language revitalization and representation.

Through the interviews conducted with native speakers and from the third and fourth authors' experiences, this paper examined the online writing culture of Boro and Manipuri speakers and explored the history and use of each language's writing system on social media.

Despite script changes and revisions of standards, both languages have active online writing cultures. A major commonality between these two writing cultures is that Roman is often the preferred script for use online despite not being official for either language. With Roman script use as well as use of other scripts, orthographic conventions are not always adhered to, but this appears to have little or no impact on communication and online assertions.

Boro and Manipuri represent two of many vulnerable languages across the globe, many of which are still growing their online presence. As seen with the cases of both Boro and Manipuri, and as urged by B. D. Lillehaugen (2016) and others, vulnerable language communities need not wait for official orthographic standards to begin writing their language. An online writing culture can flourish without codified standards of standards may not even be adhered to online even when they do exist.

This paper was written as an effort to assess the current state of Boro and Manipuri digital writing. While the objective was not to predict the shape these languages' orthographic standardization will ultimately take, we expect that it will be of continuing significance to observe how

the digital writing of both languages evolves alongside updates made to their writing systems.

Acknowledgements

The authors would like to express immense thanks and gratitude for the invaluable insight and support from Dr. Shobhana Chelliah. Her assistance with connecting us to Boro and Manipuri informants as well as her comments and feedback helped make the present paper a possibility. The authors would also like to thank the comments from the Boro and Manipuri speakers who were interviewed as a part of this project. Their shared experiences provided the foundation of the present research. Finally, the authors wish to acknowledge the immense effort by Dr. Yannis Haralambous and his colleagues for organizing and hosting this conference.

THE MANIPURI LEARNING MODULE
Using a Modern Manipuri Play as a Learning Tool

L. Somi Roy

"Binodini's Crimson Rainclouds: A Learning Module" is a set of digital documents in English and Manipuri based on a modern Manipuri play. It is designed as a learning program for people in the arts and humanities anywhere in the world who wish to learn Manipuri.

It grows out of "Crimson Rainclouds," my English translation of a Manipuri play called *Asangba Nongjabi*. The play is a work for radio and stage by the Manipuri writer, my mother, MK Binodini Devi, who signed her works as Binodini. It is about an artist caught between his art, and two women in his life. The play is celebrated in Manipur for the expressiveness and power of its realistic dialog.

I used the play "Crimson Rainclouds" to develop a learning tool for Manipuri for anyone anywhere. From this developed the Manipuri Learning Module on Zenodo which was later migrated it to the University of North Texas in Denton. Free universal access to the Module documents are available for scholars and researchers on both Zenodo and UNT Libraries portals.

The Module is made up of five parts: "Crimson Rainclouds," the English translation of the play; the Manipuri original *Asangba Nongjabi* in the Bangla script that Binodini wrote in; its transliteration into Meitei Mayek; a Roman transliteration; and, finally, an audio-recording of the play in Manipuri. One can therefore read the play in English, or the Manipuri original in Bangla, Meitei Mayek or Roman scripts. They can also listen to the play. There is also a note on the Romanization system I

FIGURE 9. MK Binodini Devi (1922-2011)

have employed, which will be discussed further below. I added a primer of the Meitei Mayek alphabet too, since the script is not well-known like Bangla.

FIGURE 10. Asangba Nongjabi by Binodini, 1967

"Crimson Rainclouds" was published in 2012 by Thema Books of Kolkata as a two-language, three-script volume. The English translation is followed by the Manipuri original in Bangla and Meitei Mayek. The two Manipuri versions were published on facing pages. The text of each page in one script corresponds to the text on the facing page in the other script.

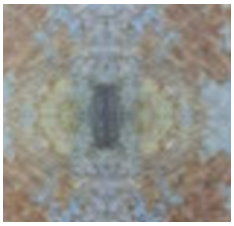


FIGURE 11. Crimson Rainclouds, 2012

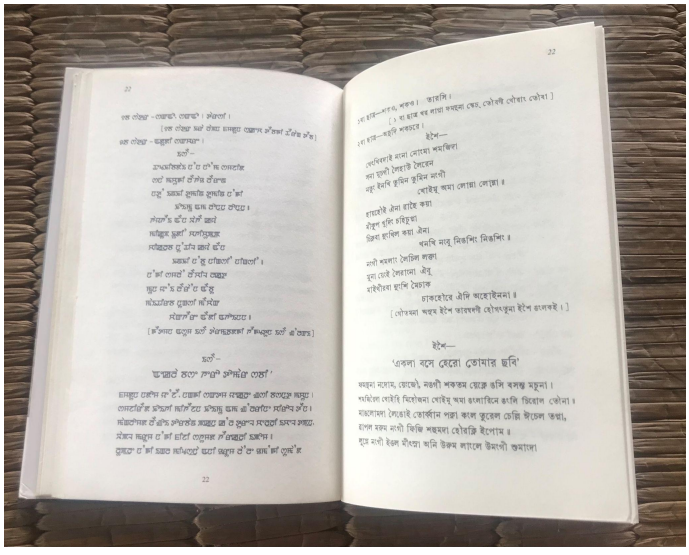


FIGURE 12. Interfacing Bangla and Meitei Mayek scripts in "Crimson Rainslouds"

The reason for the inclusion of the Meitei Mayek transliteration was because the general reader in Manipur cannot read this script. It was true in 2012 when the book came out. It is true even today.

The backstory is that language activists burned down the Central Library in Manipur in 2005. The appalling incident was the result of growing frustration at the state government's failure—over the preceding two decades—to implement its own law to replace the Bangla script with the indigenous Meitei Mayek. Their hand forced, the government hurriedly enacted the replacement. It immediately resulted in an abrupt schism in Manipuri orthography. And today, 19 years on, the below-22 generation can read Manipuri only in Meitei Mayek, while the general readership above age 22 can read it only in the Bangla script.

FIGURE 13. Manipuri primer in Meitei Mayek, archived in CoRSAL (https://digital.library.unt.edu/ark:/67531/metadc1213732/)

The dual-script approach of "Crimson Rainclouds" from Thema Books was therefore to build a bridge across this orthographic chasm. However, the non-Manipuri would still not be able to read the Manipuri texts if they do not know or first learn either of these two scripts.

To date, there is no official Romanization system developed by linguists for Manipuri in the way they have for most of the world's major text or post in Manipuri, with no uniformity and consistency, and in every and any way they choose. So in 2015, I created a Romanization system for the purpose of creating the Manipuri learning tool using the Manipuri play by my mother. I call it Mobile Manipuri because it does not use diacritical marks so as to keep it user-friendly in the manner of informal romanization currently used on mobile phones in Manipur today. Therefore, it is not IPA based. It updates the 1964 romanization system of the great scholar N. Khelchandra with elements I saw emerging with Facebook, WhatsApp and other social media. The intent was to make it simple and usable for people who wished to read *Asangba Nongjabi* or to learn Manipuri. The target users include the increasing number of Manipuris who, having been educated or grown up outside the state as a result of increased geographical mobility in India, cannot read or write either Bangla or Meitei Mayek. I used this system to transliterate *Asangba Nongjabi* into Roman to create the Manipuri Learning Module for non-Manipuri researchers and students of Manipur and Manipuri culture all over the world.

A new publication called "Binodini's Crimson Rainclouds: A Learning Module" of an expanded Manipuri Learning Module with the addition of the Roman version of the play is planned for this year.

FIGURE 14. Pundit N. Khelchandra with Manipuri Manuscripts

References

Androutsopoulos, Jannis (2013). "Performing participatory culture and metalinguistic discourse: Performing and negotiating german dialects on YouTube." In: *Discourse 2.0: Language and new media*. Baltimore: Georgetown University Press, pp. 47–71.

Bachaspatimayum, Sunzu (2020). "Manipur is now on the war tourism map. But preserving Koirengei airfield is key." In: *ThePrint*. https://theprint.in/opinion/manipur-is-now-on-war-tourism-map-preserving-koirengei-airfield-key/575627/.

Barton, D and C. K. M. Lee (2012). "Refining vernacular literacies in the age of Web 2.0." In: *Applied Linguistics* 33.3, pp. 282–298.

Basanta, Kumar (n.d.). *Meitei Mayek Primer*. Manipur: Prime.

Basumatary, Chithum (2014). "Marginalised Bodo people in Education by the Government-A Sciological Study." In: *IOSR Journal Of Humanities And Social Science* 19.1, pp. 64–66.

Basumatary, Prafulla (2017). *Verbal Semantics in a Tibeto-Burman Language: The Bodo verb*. Vol. 42. Contemporary Studies in Descriptive Linguistics. Oxford: Peter Lang.

BC, Preetam (2019). "Bodo-medium day observed in Assam." In: *APN News*.

Billock, Jennifer (2015). "How Social Media Is Helping to Save Endangered Languages." https://www.mentalfloss.com/article/68609/how-social-media-helping-save-endangered-languages.

Boro, Kiran (2005). "Mwkhthang jiuni saogari bwgra Gustave Courbet." Archive Location: India—Assam—Udalguri District—Udalguri Publisher: Boro Women Literacy Academy.

Buszard-Welcher, Laura (2001). "Can the Web Help Save My Language?" In: *The Green Book of Language Revitalization in Practice*. https://brill.com/view/book/edcoll/9789004261723/B9789004261723-s027.xml. Leiden, Netherlands: Brill, pp. 331–345.

Cahill, Michael (2014). "Non-Linguistic Factors in Orthographies." In: *Developing Orthographies for Unwritten Languages*. SIL International, pp. 9–25.

Cahill, Michael and Keren Rice (2014). *Developing orthographies for unwritten languages*. Vol. 6. Publications in language use and education. SIL International.

Carpenter, Jennifer et al. (2016). "Digital access for language and culture in First Nations communities." In: *Knowledge Synthesis Report for Social Sciences and Humanities Research Council of Canada*.

Cassels, Morgan (2019). "Indigenous languages in new media: Opportunities and challenges for language revitalization | Working Papers of the Linguistics Circle." In: *Working Papers of the Linguistics Circle of the University of Victoria* 29. https://journals.uvic.ca/index.php/WPLC/article/view/18767, pp. 25–43.

Chelliah, Shobhana and Rachel Garton (2023). "Orthography Development for Languages of the South Central Branch of Tibeto-Burman: Lessons from Lamkang." In: *Himalayan Linguistics* 22.

Chelliah, Shobhana L. (2021). *Why Language Documentation Matters*. London: Springer.

Chelliah, Shobhana Lakshmi (2011). *A Grammar of Meithei*. Vol. 17. Mouton Grammar Library (MGL). Berlin: De Gruyter.

———— (n.d.). "Photograph of a Manipuri manuscript: Court Chronicle of the Kings of Manipur." Archive Location: India—Manipur.

Computational Resource for South Asian Languages (CoRSAL) (n.d.[a]). "Boro Language Resource." Digital Library. https://digital.library.unt.edu/explore/collections/BLR/.

———— (n.d.[b]). "Manipur Language Resource." https://digital.library.unt.edu/explore/collections/MLR/.

———— (n.d.[c]). "Using KeyMan." https://corsal.unt.edu/curriculum.

Cook, Vivian and Bene Bassetti (2005). "An Introduction to Researching Second Language Writing Systems." In: *Second Language Writing Systems*. Second Language Acquisition. Multilingual Matters, pp. 1–67.

Council, First People's Cultural (2020). "Orthographies." https://fpcc.ca/resource/orthographies/.

Cru, Josep (2018). "Micro-level language planning and YouTube comments: destigmatising indigenous languages through rap music." In: *Current Issues in Language Planning* 19.4, pp. 434–452.

Daimari, Rejoice (n.d.). "Our maternal language: BODO language." https://thebodotribe.blogspot.com/2015/06/our-maternal-language-bodo-language.html.

Dale, Merrion et al. (n.d.). "Case study of using Facebook groups to connect community users to archived CoRSAL content." In: *Language Documentation and Conservation* ().

De Falco, Dario and Alfonso Cesarano (2016). "Endangered languages in the era of social media: the case of Kenyah Lebu' Kulit language." In: *Linguapax Review 2016 Digital Media and Language Revitalisation*. Linguapax Review 4. Linguapax International.

De Korne, Haley and Miranda Weinberg (2021). ""I Learned That My Name Is Spelled Wrong": Lessons from Mexico and Nepal on Teaching Literacy for Indigenous Language Reclamation." In: *Comparative Education Review* 65.2, pp. 288–309.

Devi, Heisnam Olivia and Nishaant Choksi (2022). "Graphic diversity and ethnic identification in contemporary Manipur." In: *Contemporary South Asia* 30.3.

Djomeni, Gabriel D. and Etienne Sadembouo (2016). "African languages and digital media: Practice, challenges and perspectives in Cameroon." In: *Linguapax Review 2016 Digital Media and Language Revitalization*. Linguapax Review 4, pp. 39–54.

Emmanouilidou, Lydia (2014). "For Rare Languages, Social Media Provide New Hope." In: *NPR*.

Eucation, Shreya Arya (2020). "What is 3 Language Formula?" In: *Law Times Journal*.

George, Sudhir Jacob (1994). "The Bodo Movement in Assam: Unrest to Accord." In: *Asian Survey* 34.10, pp. 878–892.

Gnanadesikan, Amalia E. (2020). "S_1: The Native Script Effect." In: *Proceedings of Graphonlinguistics in the 21st Century, 2020*. Ed. by Yannis Haralambous. Vol. 4. Grapholinguistics and Its Applications. Brest: Fluxus Editions, pp. 103–123.

"Constitutional Revisions Relating to Eighth Schedule" (2017). https://www.mha.gov.in/sites/default/files/Eighth_Schedule.pdf.

Grierson, George Abraham (1921). *Linguistic survey of India*. Vol. 10. Calcutta: Office of the Superintendent of Government Printing, India.

Haokip, Pauthang (2021). "Orthography of Tribal languages of NE India." https://www.academia.edu/45153983/Orthography_of_Tribal_languages_of_NE_India. Virtual Session.

Hernández, Lorena Córdova, Mario E. López-Gopar, and William M. Sughrua (2017). "From Linguistic to Semiotic Landscape: Indigenous Language Revitalization and Literacy." In: *Studie z Aplikovane Lingvistiky* 2, pp. 7–21.

India, Census of (2011). "Statement 4—Scheduled Languages in Descending Order of Speakers' Strength—2011." https://censusindia.gov.in/nada/index.php/catalog/42458/download/46089/C-16_25062018.pdf.

Laithangbam, Iboyaima (2017). "Banished Manipuri script stages a comeback." In: *The Hindu*.

Lillehaugen, Brook (2019). "Tweeting in Zapotec: Social Media as a Tool for Language Activists." In: *Indigenous Interfaces: Spaces, Technology, and Social Networks in Mexico and Central America*. Tucson: University of Arizona Press.

Lillehaugen, Brook Danielle (2016). "Why write in a language that (almost) no one can read? Twitter and the development of written literature." In: *Language Documentation and Conservation*, pp. 356–393.

Lillis, Theresa and Carolyn McKinney (2013). "The sociolinguistics of writing in a global context: Objects, lenses, consequences." In: *Journal of Sociolinguistics* 17.4, pp. 415–439.

Martucci, Brian (2014). "The Endangered Osage Language Gets a Unicode-Friendly Alphabet." In: *TheLineMedia*.

McCarty, Teresa L. (2018). "Community-Based Language Planning." In: *The Routledge Handbook of Language Revitalization*. Routledge Handbooks Online.

Meletis, Dimitrios (2020). *The Nature of Writing. A Theory of Graphолinguistics*. Vol. 3. Grapholinguistics and Its Applications. Brest: Fluxus Editions.

———— (2022). "Types of orthographic standardization: A sociolinguistic approach." Presentation at the *Grapholinguistics in the 21st Century* Conference, https://www.youtube.com/watch?v=nvd3srmNqTk. Brest.

Molinaro, Mary (2008a). "Photograph of a Manipuri manuscript collection." https://digital.library.unt.edu/ark:/67531/metadc1631615/. Archive Location: India—Manipur.

———— (2008b). "Photograph of a Manipuri manuscript: Sading Sakok." https://digital.library.unt.edu/ark:/67531/metadc1631574/m1/1/. Archive Location: India—Manipur.

Moseley, Christopher and Alexandre Nicholas (2010). *Atlas of the world's languages in danger—UNESCO Digital Library*. Paris: UNESCO Publishing.

Nation, The Osage (2014). "Osage Font & Keyboard Installation." https://www.osagenation-nsn.gov/who-we-are/language-department/osage-font.

Neef, Martin (2015). "Writing systems as modular objects: proposals for theory design in grapholinguistics." In: *Open Linguistics* 1.1.

Pappuswamy, Uma (2017). "Issues and Challenges in search of Effective Orthography for Unwritten Languages of North-East India." In: *Unwritten Languages of India*. Delhi: Sahitya Akademie, pp. 117–159.

Prabhakar, M. S. (1974). "The Politics of a Script: Demand for Acceptance of Roman Script for Bodo Language." In: *Economic and Political Weekly* 9.51, pp. 2097–2102.

Rowe, Jacqueline (2022). "A language of no importance: the consequences of neglecting marginalized languages in the digital world." In: *OpenGlobalRights*.

Roy, L. Somi (2008). "Photograph of Khaba with a Manipuri manuscript." https://digital.library.unt.edu/ark:/67531/metadc1631594/. Archive Location: India—Manipur.

———— (2017). "Mobile Manipuri: Transliteration and the Making of a Romanization System." https://www.media-studies.ca/articles/habermas.htm. Archive Location: India—Manipur.

Sabha, Bodo Sahitya (n.d.). "Aims and Objectives." https://bodosahityasabha.org/index.html.

Sarmah, Satyendra Kumar (2014). "Script Movement Among the Bodo of Assam." In: *Proceedings of the Indian History Congress* 75, pp. 1335–1340.

Scannell, Kevin (2012). "Translating Facebook into Endangered Languages." In: *Proceedings of the 16th Foundation for Endangered Languages Conference*, pp. 106–110.

Schwab, Klaus (2018). *The Fourth Industrial Revolution*. In: *Encyclopædia Britannica Anniversary Edition: 250 Years of Excellence (1768–2018)*. Chicago, Ill.: Encyclopædia Britannica, Inc.

Snider, K. (2014). "Orthography and Phonological Depth." In: *Developing Orthographies for Unwritten Languages*. Ed. by M. Cahill and K. Rice. Dallas, TX: SIL International, pp. 27–48.

Soria, Claudia (2016). "What is digital language diversity and why should we care?" In: *Linguapax Review 2016 Digital Media and Language Revitalisation*. Linguapax Review 4. Linguapax International, pp. 13–28.

Soules, Marshall (2007). "Jürgen Habermas and the Public Sphere." https://www.media-studies.ca/articles/habermas.htm.

Wary, Ajoy (2017). "Revisiting Bodo Medium Secondary School Education in Udalguri." In: *Journal of Tribal Intellectual Collective India,* 4.2, pp. 1–28.

Willis Oko, Christina M (2018). "Orthography Development for Darma (The Case That Wasn't)." In: *Language Documentation & Conservation* 12, pp. 15–46.