



Challenges in Enggano Orthography Development

CHARLOTTE HEMMINGS 

I WAYAN ARKA 

ENGGA ZAKARIA SANGIAN 

DENDI WIJAYA 

MARY DALRYMPLE 

*Author affiliations can be found in the back matter of this article

ABSTRACT

This paper details three challenges encountered in developing an orthography for Enggano (ISO 639-3 Code eno), an endangered Austronesian language of Indonesia. Enggano, as spoken in the 1930s, was documented by Hans Kähler in a grammar, text collection, and dictionary (Kähler 1940, 1975, 1987). However, Kähler used different orthographic conventions in each publication. Moreover, the language has undergone significant sound changes since Kähler's time (see Yoder 2011). Hence, developing a standard orthography for contemporary Enggano is an important goal of our ongoing documentation project. Following Seifart (2006), we want the orthography to reflect the structure of the language, and also to be practical and easy to use for speakers who are familiar with Indonesian. We report on our progress in developing a phonemic orthography while also adopting the conventions of Indonesian as far as possible. In this paper, we discuss how Enggano central vowels, nasal vowels, and glides present a challenge to this endeavor, and outline the collaborative approach adopted to choose between different orthographic options.

ABSTRAK

Artikel ini menjelaskan tiga tantangan yang dihadapi dalam mengembangkan ortografi bahasa Enggano (ISO 639-3 Code eno), salah satu bahasa yang terancam punah di Indonesia. Enggano, seperti yang dibicarakan pada tahun 1930-an, telah didokumentasikan oleh Hans Kähler dalam tata bahasa, kumpulan teks, dan kamus (Kähler 1940, 1975, 1987). Namun, Kähler menggunakan konvensi ortografi yang berbeda di setiap publikasinya. Selain itu, bahasa Enggano telah mengalami perubahan bunyi yang signifikan sejak itu (lihat Yoder 2011). Oleh karenanya, mengembangkan ortografi standar untuk Enggano kontemporer merupakan tujuan penting dari proyek dokumentasi yang kami sedang lakukan. Merujuk

RESEARCH



CORRESPONDING AUTHOR:

Charlotte Hemmings

University of Oxford, UK
charlotte.hemmings@ling-phil.
ox.ac.uk

KEYWORDS:

Enggano; Indonesia;
Orthography Development;
Challenges; Language
Documentation

KATA KUNCI:

Enggano; Indonesia;
Pengembangan Ortografi;
Tantangan; Dokumentasi Bahasa

TO CITE THIS ARTICLE:

Hemmings, Charlotte, I Wayan Arka & Engga Zakaria Sangian & Dendi Wijaya & Mary Dalrymple. 2023. Challenges in Enggano Orthography Development. *Language Documentation and Description* 23(1): 4, 1-19. DOI: <https://doi.org/10.25894/ldd.329>

pada Seifart (2006), kami menginginkan sistem ortografi yang tidak hanya mencerminkan struktur bahasa, tetapi juga praktis dan mudah digunakan bagi penutur yang akrab dengan bahasa Indonesia. Oleh karena itu, kami menyampaikan kemajuan kami dalam mengembangkan ortografi fonemik sambil mengadopsi konvensi bahasa Indonesia sejauh mungkin. Dalam makalah ini, kami membahas bagaimana vokal lisan Enggano, vokal sengau, dan luncuran menghadirkan tantangan bagi upaya ini dan menguraikan pendekatan kolaboratif yang diadopsi untuk memilih di antara opsi ortografis yang berbeda.

1. INTRODUCTION

Orthography development is an important goal in language documentation (Seifart 2006): it is often the most tangible outcome of a language documentation project for speakers, and it facilitates the creation of other outputs, such as transcribed documentary corpora, teaching materials, dictionaries, grammars, and text collections. Nonetheless, the challenges faced in orthography development are not always explicitly discussed in the literature on language documentation (though see Cahill 2018; Jones & Mooney 2017; Lüpke 2011; Seifart 2006). This paper explores three challenges faced in developing an orthography for the Enggano language (IPA /ɛŋɡano/, ISO 639-3 Code eno) an endangered language of Indonesia spoken by approximately 1,500 speakers on Enggano Island, Sumatra. The primary aim of the paper is to reflect on the ongoing process of orthography development, providing a model for how to navigate such challenges in future language documentation projects. A secondary aim is to present the Enggano orthography, providing some motivation for the decisions that were reached.

The paper is structured as follows. Section 2 provides some background on the Enggano language and the history of its documentation. Section 3 reviews the existing literature on orthography development, and discusses major factors known to affect orthographic choices. Section 4 presents three case studies of challenges in Enggano orthography development, specifically the challenges of representing central vowel phonemes, nasal vowel phonemes, and allophonic glides. Finally, section 5 concludes with a presentation of the symbols and conventions of the Enggano orthography developed during the current documentation project, and reflections on the lessons learned for future orthography development projects.

2. BACKGROUND ON ENGGANO

The Enggano language is spoken on Enggano Island, situated off the southwest coast of Sumatra, Indonesia. Enggano Island is the southernmost of the Barrier Islands chain, which also includes Simeuluë, Nias, and the Mentawai Islands (Figure 1).



Figure 1 Map of Sumatra and the Barrier Islands. (© OpenStreetMap, Mapbox and Mapcarta).

Most scholars now agree that Enggano is an Austronesian language (Arka et al. 2022; Dyen 1965; Edwards 2015; Nothofer 1986; Smith 2017, 2020).¹ However, there is ongoing debate as to whether Enggano forms a subgroup with the Barrier Island languages and Batak languages of Sumatra, as claimed in Nothofer (1986), Smith (2017), and Billings & McDonnell (2022), or whether it is a primary branch of Malayo-Polynesian, as claimed in Edwards (2015). There are six main villages on Enggano, all located on the north coast of the island, as shown in Figure 2.



Figure 2 Map of Enggano Island (© OpenStreetMap, Mapbox and Mapcarta).

These include Enggano-speaking populations as well as non-Enggano speaker populations. As a whole, the language can be considered endangered, but the degree of endangerment varies (Arka et al. 2022). The language is most vital in the central villages: Meok, Apoho, and Malakoni. In northern and southern villages, non-Enggano populations are higher, which has resulted in increased influence of Indonesian, the national language of Indonesia, more code-mixing, and higher rates of language shift towards Indonesian in those villages.

Enggano has a relatively long history of documentation. The earliest records were collected during the Dutch colonial era in the mid-nineteenth century, and consist of word lists (Boewang 1854; Helfrich 1893, 1916; Helfrich & Pieters 1891; Oudemans 1879, 1889; van der Straaten & Severijn 1855; von Rosenberg 1855; Walland 1864; as well as the Holle list, collected in 1895 and published in van der Noord 1987) and some texts collected by Helfrich (1916). Based on a seven-month stay between 1937 and 1938, Hans Kähler produced the most comprehensive documentation to that point, consisting of a sketch grammar (Kähler 1940), a text collection (Kähler 1955, 1957, 1958, 1960a, 1960b, 1961, 1962, 1964, 1975), and a dictionary published posthumously as Kähler (1987). More recent linguistic work includes an unpublished word list collected by Bernd Nothofer in 1986 which formed the basis for historical work (Nothofer 1986); materials produced by government agencies in Indonesia, including the word list and grammar sketch produced by Pusat Pembinaan dan Pengembangan Bahasa [Center for Language Development and Cultivation] (Kasim et al. 1987; Nikelas, Rasyad & Semi 1994); the dictionary produced by Kantor Bahasa Provinsi Bengkulu [Regional Office for Language in Bengkulu Province] (Riswari et al. 2021); and word lists/example sentences in several linguistic theses (Butters 2021; Wijaya 2018; Yoder 2011).

¹ See Capell (1982) and Blench (2014) for earlier claims that Enggano is a non-Austronesian language with borrowed vocabulary from Austronesian neighbors. The debate rests on the fact that Enggano seems to have a surprisingly low cognate percentage with other Austronesian languages; in part this perception is the result of unusual sound changes in forms that are in fact Austronesian, coupled with lexical replacement of some forms for the purpose of taboo avoidance (see Nothofer 2021).

In order to understand the structure of contemporary Enggano and address the debates surrounding its historical position in the Austronesian family, the authors of this paper are involved in an ongoing research project, “Enggano in the Austronesian family: Historical and typological perspectives”.² The project has two major strands, a documentation strand and a historical linguistics strand. The aim of the historical linguistics subproject is to assess the position of Enggano within the Austronesian family, while the documentation subproject’s goal is to produce a grammar of contemporary Enggano, an annotated documentary corpus, and teaching materials to aid in the process of language revitalization. In order to achieve these aims, it is necessary to have a standard orthography for writing the language.

Unfortunately, the previous literature does not provide a standard means of writing Enggano. Early records did not use a common orthography, as illustrated in Table 1.³ Moreover, there is little information on whether the orthography in these early word lists reflects a phonological analysis of the language or simply the compiler’s perception of words, which may have been influenced by their native language phonologies. Thus, it is unclear whether these are alternative spellings of the same form, or reflective of dialect differences and hence different pronunciations.

	‘two’	‘eye’	‘belly’
von Rosenberg (1855)	<i>adoea</i>	<i>bakka</i>	<i>koedei</i>
Waland (1864)	<i>adoeah</i>	<i>bah-kah</i>	<i>kidai</i>
Oudemans (1879)	<i>adoewa</i>	<i>ébāka</i>	<i>kiedāi</i>
Helfrich & Pieters (1891)	<i>adoewě</i>	<i>ěbakě</i>	<i>èkidai</i>

Table 1 Spelling variants in early records of Enggano.

Kähler’s work is based on a phonological analysis of the language as set out in the grammar (Kähler 1940: 83), the text collection (Kähler 1975: III–IV), and the dictionary (Kähler 1987: iii–iv). However, as discussed by Schmidt in the foreword to Kähler (1987) and summarised in Edwards (2015: 60), the orthography is not consistent across the different materials. This is illustrated in Table 2.

	‘finish’	‘like that’
Reconstructed pronunciation based on contemporary equivalents and discussion in Kähler grammar	/dihida/	/kinoʔoaha/
Kähler (1940)	<i>duhuda</i>	<i>kinō’ōāhā</i>
Kähler (1975)	<i>dāhāda</i>	<i>kino’oaha</i>
Kähler (1987)	<i>dāhāda</i>	<i>kinō’ōahā</i>

Table 2 Spelling variants in Kähler’s corpus.

The materials differ not only in the symbols they use—e.g. <u>, <ā>, and <ə> for the central vowel /i/—but also in the conventions adopted, e.g., whether to represent nasal vowels, as in the grammar and dictionary, or not, as in the 1975 texts.⁴

More recent works have tended to use either Indonesian orthography, which is practical but (as we will see) not necessarily suited to Enggano phonology, or the IPA transcription system, which is consistent and allows users to accurately represent the sounds of words, but is not particularly

² <https://enggano.ling-phil.ox.ac.uk/>.

³ The word lists are not consistent in including or excluding affixes like the noun marker *e-*, seen in several of the forms in Table 1.

⁴ Due to a phonological rule of nasal spreading, nasal vowels following nasal consonants are not contrastive. However, nasal vowels are contrastive in other contexts. See section 4.2 for further discussion.

practical as it includes symbols that are not familiar to speakers, such as ⟨i⟩ and ⟨ʔ⟩. Consequently, there is no orthography in the existing literature that accurately represents the sounds of Enggano in a way that is simple and easy to use.

Within the community, there is no standardized way of writing Enggano that is universally accepted and adopted. Instead, there are multiple competing orthographies in use. There is an ongoing project to translate the Bible into Enggano, and those closely involved in this project have suggested symbols such as ⟨è⟩ to represent the schwa vowel /ə/. Other speakers who are literate in Indonesian tend to adopt Indonesian conventions when writing Enggano, or combine these with regional practices, such as the use of ⟨eu⟩ for the central vowel, which is a digraph used in Sundanese (Wibawa et al. 2018) and Acehnese (Durie 1984).⁵ Adopting the conventions of the national language, Indonesian, is certainly the simplest option, and simplicity is desirable in an orthography (Seifart 2006). However, Enggano phonology differs from Indonesian phonology, and hence Indonesian orthography is not perfectly suited to the structure of Enggano, as we discuss in Section 4. Thus, the community does not have a single standard orthography that could be adopted in the documentation project either.

Consequently, as in many documentation projects (see Lüpke 2011), an important aim of the project is to develop a standard orthography for Enggano. Following Seifart (2006), we want the orthography not only to reflect the structure of Enggano, but also to be widely accepted and adopted—both in the community and by other researchers working with contemporary Enggano data. This will help to avoid the potentially divisive effects of multiple competing orthographies (cf. Lüpke 2011: 318). In order to do this, we need to take both linguistic and extra-linguistic factors into account and ensure that important stakeholders in the Enggano community are involved in the orthography development process. As an additional challenge, our work also took place during the COVID-19 pandemic, which imposed its own constraints on the process.

3. ORTHOGRAPHY DEVELOPMENT

Coulmas (2003: 35) defines an orthography as a set of graphic symbols (“graphemes”), the rules for how they are used, and the sounds they represent. Ideally, the orthography should not only reflect the structure of the language but also be practical and easy to use. Unfortunately, these desiderata are often in conflict, and consequently the orthography developer must balance the advantages and disadvantages of the different options. Following Seifart (2006: 287), decisions in two major areas must be made when developing an orthography in the context of minority/endangered languages: *orthographic depth* and *adopting existing conventions*.

Orthographic depth is the extent to which an orthography reflects pronunciation (Katz & Frost 1992). A shallow or phonemic orthography has one letter per sound, or rather one grapheme per phoneme, since allophonic variation and suprasegmental features are often not represented. In the context of endangered languages, where there may be many semi-speakers or learners, Seifart (2006: 283) suggests that a phonemic orthography may be particularly practical, since it clearly distinguishes important sound contrasts. This allows learners to understand how to write a word based on how it sounds and how to pronounce a word based on how it is written. A deep or morphemic orthography, however, preserves the graphic identity of words so that the same morpheme is written with the same letters even if it is pronounced differently in different contexts, as is the case for English *reduce* vs. *reduction* (Seifart 2006: 278). A “deeper” orthography can be easier for fluent speakers to read, and it is particularly useful in contexts where there is significant dialect variation and where languages undergo significant morphophonological alternations.

The second important decision is whether or not to adopt the existing conventions of a dominant language, which can make the orthography easier for speakers to learn. Endangered languages almost invariably exist in multilingual environments, where speakers are literate in the orthography of a dominant language (Lüpke 2011). Hence, adopting existing conventions may increase uptake

⁵ In Sundanese it represents an unrounded mid-back vowel /ɤ/ and in Acehnese an unrounded high-back vowel /ɯ/. Acehnese /ɤ/ and /ɯ/ are represented by ⟨è⟩ and ⟨ö⟩ respectively, at least in Durie (1984).

and usability. Following this principle, we adopted an alphabetic system, as is used in Indonesian, rather than a syllabary or logography (see e.g. Coulmas 2003 on different types of writing systems). However, the existing conventions of a dominant language may not fit the endangered language nor serve the emblematic function of expressing a distinct socio-cultural identity (see Cahill 2018; Sebba 2007). In the next section, we explore three challenges that arose in our attempt to develop a phonemic orthography for Enggano, while also adopting Indonesian conventions wherever possible. We show that adopting Indonesian orthography wholesale is disadvantageous in several respects, and we describe how we augmented and modified standard Indonesian orthography to make it suitable for Enggano. In doing so, we discuss both how to represent contrasts and how to decide which contrasts to represent (cf. Seifart 2006 on underrepresentation), weighing up the advantage of phonological accuracy against simplicity and learnability for users. Ultimately, decisions were guided by input from the Enggano community.

4. CHALLENGES IN ENGGANO

The documentation project began in 2018, and several field trips were conducted between 2018 and 2020, following standard methods for language documentation and description (see e.g. Austin 2016; Himmelmann 1998, 2006; Woodbury 2003, 2011). During the fieldtrips, co-authors I Wayan Arka and Mary Dalrymple established relationships with community elders in Enggano. They also built relationships with colleagues at Kantor Bahasa Provinsi Bengkulu [Regional Office for Language in Bengkulu Province], including co-author Dendi Wijaya, who were working on their own project to produce materials in the Enggano language.⁶ They provided training for native-speaker consultants, including co-author Engga Zakaria and primary-school teacher Fitri Yunita, to join the documentation project team and play a central role in data collection and analysis.⁷ During this time, a sizeable corpus was collected comprising word lists, elicitation, stimuli-based materials, and naturalistic texts. Many of the recordings were initially transcribed and translated into Indonesian by Engga Zakaria. Subsequently, the materials were translated from Indonesian into English by colleagues at Udayana University, Indonesia. In 2020, at the start of the COVID-19 pandemic, international travel ceased and all fieldwork moved online. Co-author Charlotte Hemmings joined the documentation project and began to work with the rest of the co-authors (henceforth the project team) to analyze recordings during regular meetings over Zoom.

During the online meetings, the project team worked with the corpus to better understand the structure of Enggano. Specifically, we listened back to recordings and asked Engga Zakaria to repeat slowly in order to clarify the sounds of words and their internal structure. In particular, we looked for minimal pairs in order to establish a list of contemporary Enggano phonemes, and we identified related forms in our word list in order to understand the morphophonological processes in Enggano. Building on Yoder’s (2011) analysis and the findings of these regular meetings, we established the phonemes in Tables 3 and 4 for contemporary Enggano.⁸

	BILABIAL	ALVEOLAR	PALATAL	VELAR	GLOTTAL
Stop	p b	t d		k	ʔ
Nasal	m	n			
Trill		r			
Fricative			ç		h

Table 3 Consonants.

⁶ Dendi Wijaya is now affiliated with Badan Riset dan Inovasi Nasional (BRIN) [National Research and Innovation Agency].

⁷ At the start of the project, the team also included another native-speaker consultant, Afrizon Rafles.

⁸ Note that Enggano does not have either /ŋ/ or /g/ as a phoneme. The language name /ɛngano/ is the Malay/Indonesian term. It is thought that the name may have originated from Portuguese. Today no other name is known for the language or the island. Enggano elders may also say /ekano/.

	FRONT		CENTRAL		BACK	
High	i	ĩ	ɨ	ĩ	u	ũ
Mid	e	ẽ	ə	ã	o	õ
Low			a	ã		

Table 4 Vowels.

The voiceless consonants /p/, /k/ and /ʔ/ are represented using the symbols <p>, <k> and <ʔ>. Although Standard Indonesian orthography uses the letter <k> for both [k] and [ʔ], we represent this phonemic distinction orthographically, since it has a relatively high functional load: e.g., it allows users to distinguish between common words like /məʔ/ the relativizer (cf. Indonesian *yang*) and /mæk/ ‘many’, and morphemes like the 1SG possessive suffix /-ʔ/ and the 1PL.INCL possessive suffix /-k/ (see section 4.1 for further discussion on the importance of representing phonemic contrasts). The phonemic status of [t] and [r] are not entirely clear. Kähler (1940) treats both as allophones of /d/. However, [t] and [r] both occur frequently in our corpus, and there are some near minimal pairs suggesting they may have phonemic status in contemporary Enggano: e.g. [i top] ‘below’ vs. [dop] ‘land/earth’; [di] ‘tongue’ vs. [ari] 2PL, [kitahaʔ] ‘stab’ vs. [karahaʔ] ‘my body’. Consequently, we write these as <t>, <r>, and <d> respectively.⁹ The voiced stop /b/ is written as and the nasals /m/ and /n/ are written as <m> and <n>. Note that the voiced stops /b/ and /d/ are sometimes realised as [m] and [r] word-finally, which results in variant pronunciations, e.g. [jub] vs. [jum] ‘house’ and [hiid] vs. [hiir] ‘woman’. The variants with [m] and [r] are particularly associated with younger speakers.

The phoneme /h/ has several allophones: [ç] following the oral high front vowel /i/, [x] following non-front high oral vowels and [h] in all other contexts. In Kähler’s work, these variants are represented orthographically ([ç] and [x] as <x>, and [h] as <h>), but we treat this as allophonic and hence write all as <h>. The phoneme /ç/ is variously realised as [s], [t], [ç] and [tʃ] depending on the speaker. It is not attested word-initially and normally occurs word-finally after /i/. Given that [ç] is also an allophone of /h/, Yoder (2011) treats this sound as underlyingly /h/. However, historically it comes from a different source and hence we represent it orthographically as <c>. Note that /ç/ appears to have [t] as a variant realisation in non-final positions, e.g. *yic* [jiç] ‘word/sound’ → *yitik* [jitik] ‘our words/sounds’. However, it is not to be considered an allophone of /t/ in contemporary Enggano since /t/ can also occur word-finally following /i/, e.g. *i dit* [idit] ‘there’. Finally, the vowels have nasal allophones following nasal consonants (see section 4.2) and /i/ and /e/ have glide allophones in onset position (see section 4.3 and Yoder 2011: 22).

Once we had established the phonemic inventory, we were able to identify several orthographic issues that arose in the tension between adhering to the phonemic principle of one symbol per sound and following Indonesian orthographic conventions as closely as possible. We first discussed each of these issues within the project team and developed a list of options, weighing up the advantages and disadvantages. We then consulted with other researchers familiar with Enggano, including Bernd Nothofer, Erik Zobel, and Daniel Krausse, who are working within the historical linguistics strand of the Enggano project, and Yanti Riswara Idris of Kantor Bahasa Provinsi Bengkulu, who was co-ordinating the production of an Enggano-Indonesian dictionary (Riswari et al. 2021). Finally, Engga Zakaria and Dendi Wijaya (henceforth referred to as “the local team”), who were based in Bengkulu and able to travel to Enggano Island, took the options to the Enggano community and explained the motivations for our choices to important stakeholders. The local team sought feedback from community elders via focus group interviews, and from schoolchildren and learners of Enggano via sample tasks from the teaching materials. This allowed us to validate the choice of symbols and begin to socialize the proposed orthography among learners.

⁹ The lateral approximant [l] is also treated as a dialect variant of /d/ in Kähler (1940) and has a marginal status in contemporary Enggano. As discussed in Yoder (2011: 24), it mainly occurs in borrowings, e.g. [kaʔblau] ‘blue’ or as a variant realization of /d/ or /r/, e.g. [kalel] vs. [karer] ‘soft’. Like all non-adapted borrowings, it is written as in the source language.

4.1 CENTRAL VOWELS

The first challenge was that Enggano distinguishes seven values for vowels, while Indonesian, which adopted the Roman alphabet, has only five vowel letters. This creates a challenge when representing the two central vowels /ə/ and /i/ which can be shown to be separate phonemes in light of the minimal pairs in Table 5 below.

/e/ vs. /ə/	/ə/ vs. /i/	/i/ vs. /u/	/ə/ vs. /a/	/ə/ vs. /o/	/i/ vs. /i/
[be] ‘dog’	[kəɾ] ‘swallow’	[pi] ‘see’	[kikə] ‘open’	[kəɾ] ‘swallow’	[pi] ‘see’
[bə] ‘water’	[kiɾ] ‘live’	[pu] ‘run’	[kika] ‘close’	[koɾ] ‘lift’	[pi] ‘garden’
[keɾ] ‘cry’	[hiəɾ] ‘worm’	[hiir] ‘woman’	[dəɾ] ‘current’	[dəxə] ‘hear’	[kiɾ] ‘turtle’
[kəɾ] ‘swallow’	[hiir] ‘woman’	[hiur] ‘dust’	[daɾ] ‘husband’	[doɦa] ‘boat’	[kiç] ‘ant’
[eʔ] ‘put’	[kəx] ‘mountain’	[jiɾ] ‘life’	[kəx] ‘mountain’	[ʔudəp] ‘taro’	[hi] ‘stop’
[əʔ] ‘you (sg)’	[kiɾ] ‘turtle’	[juɾ] ‘head’	[kaɦ] ‘go’	[dop] ‘land’	[hi] ‘with’

Table 5 Minimal pairs identifying oral vowel phonemes.

The three options we considered for representing central vowels were the following:

1. Use the same symbol for multiple phonemes: /e/, /ə/, and /i/.
2. Use diacritics for /ə/ and /i/.
3. Use digraphs for /ə/ and /i/.

Option 1 is familiar to speakers and writers of Indonesian, since both /e/ and /ə/ are represented by the letter <e> in Standard Indonesian orthography. This is the option adopted by Kantor Bahasa in their 2021 Enggano-Indonesian dictionary (Riswari et al. 2021). Since it adopts existing conventions, this option would be easy to learn and would make the orthography consistent with other resources. However, it would mean abandoning the phonemic principle and would result in words that sound very different being written in the same way. For example, a word written as <anek> could potentially be pronounced in three ways: /anek/ (a grass species), /anək/ (friend-1PL INCL ‘my friend’) or /aniɿk/ (a verb root meaning ‘pull’). A variant of this option would be to use <a> for /a/ and /ə/. This option was discarded at an early stage since it is neither phonemic, nor familiar from the conventions of writing Indonesian.

Option 2 involves combining existing letters with diacritics to create separate graphemes for each phoneme. Building on the community proposal to use <è> for /ə/, we proposed <ù> for /i/, which distinguishes the words above as *anek*, *anèk*, and *anùk*. This option has the advantage of making a phonemic orthography possible. However, it has the disadvantage that the symbols may not be universally familiar. More importantly, diacritics are notoriously cumbersome to type.¹⁰

Option 3 involves digraphs or symbols where two letters together represent a single sound, much like <sh> for /ʃ/ in English. This would also allow for a phonemic orthography, but one where some sounds are represented by two letters. This has some advantages over the use of diacritics. Firstly, the use of digraphs to represent central vowels is common among languages of the region, as discussed above, and <eu> is already used informally among some speakers of Enggano for /i/. Secondly, digraphs are easier to type than diacritics. However, the major disadvantage is that it is

¹⁰ At an early stage in the discussion, the project team also considered using <é> (e with acute accent) as this may be easier to adapt for mobile phones and keyboards. We discussed the following symbols: <é> for /e/, <è> for /ə/ and <ù> for /i/, since <é> was previously used to represent /e/ in the orthography used to write Malay/Indonesian prior to spelling reforms (see e.g. Vikør 1988) and is relatively common practice among Indonesian languages that distinguish /e/ and /ə/. This option was ultimately rejected on the basis that representing nasal phonemes using the tilde <̃> requires combining diacritics. /è/ occurs much more frequently in Enggano words than /ə/ and /i/, and in some really commonly used words like the demonstratives [ʔèʔ] ‘this’, [ʔèðʔ] ‘that’. Hence, it was felt that it would be unnecessarily cumbersome to use stacked diacritics, as in <é̃> for /è/, in these common words. Another option we discussed was using <é> to represent /ə/ rather than /e/. This avoids the pitfall of stacking diacritics in common words, but it runs counter to expectations that <é> is to be read [e].

difficult to distinguish digraphs representing a single sound from the representation for diphthongs and vowel sequences, which are also common in Enggano, and appear in words like /kahi/ ‘one’ and /panau/ ‘speech’ (see Yoder 2011). Although the sequence /eu/ itself is not attested in Enggano, other sequences such as /eo/, /ea/, and /oe/ are attested, and are typically pronounced as separate syllable peaks: e.g. [he.o] ‘in’, [he.o] ‘bone’, [ki.pa.ko.ko.eh] ‘squat’. Hence, the use of digraphs could lead to mispronunciation of the central vowels as diphthongs or vowel sequences.

Weighing up the relative advantages and disadvantages of the options, the project team agreed that the preferred option was to use characters with diacritics (the second option), since this allows for a phonemic orthography and introduces symbols that are uniquely associated with the Enggano language, symbolizing a distinct Enggano identity. Hence, this option has the ability to increase both the usability and acceptability of the orthography (cf. Cahill 2018). The local team presented this option to the community elders and explained the advantages in terms of distinguishing between minimal pairs. This, in particular, was felt to be important to the elders, who volunteered the examples discussed above during an interview. Their clear message was that an orthography that distinguishes different pronunciations was better, both from a pedagogical perspective but also in terms of highlighting the uniqueness of the Enggano language. Consequently, the current orthography, as summarized in section 5, uses <ê> for /ə/ and <û> for /i/. To counteract the difficulty of typing diacritics, the team developed a custom Keyman keyboard (https://keyman.com/_legacy/keyboards/enggano) which allows users to type characters with diacritics in the Enggano orthography on laptops and mobile phones.

4.2 NASAL VOWELS

The second challenge is representing nasal vowels. In many languages, nasal vowels are simply allophones of oral vowels that occur in the context of nasal consonants [m], [n], and [ŋ]. In Enggano, however, (near) minimal pairs suggest that there are seven nasal vowel phonemes in addition to the oral vowel phonemes (see Table 6).

/i/ vs. /ĩ/	/ĩ/ vs. /ĩ̃/	/u/ vs. /ũ/	/o/ vs. /õ/	/ə/ vs. /ẽ/	/e/ vs. /ē/	/a/ vs. /ã/
[pi] ‘garden’	[pi] ‘see’	[ku] ‘tree’	[jok] ‘vein’	[kə] ‘tuber’	[kep] ‘not yet’	[jah] ‘go’
[pi] ‘jungle snail’	[pĩ] ‘fireplace’	[kũ] ‘count’	[jõkõ?] ‘headscarf’	[kẽ] ‘try’	[kēp] ‘island’	[jãh] ‘place’

Table 6 Minimal pairs identifying nasal vowel phonemes.

The easiest option for representing nasalization—particularly since diacritics are already part of the orthography for representing oral vowels—is to use the tilde diacritic <̃> that represents nasalisation in the IPA and which was also used in the Kähler corpus. Hence, we adopted the following symbols for the nasal vowels: <ĩ>, <ũ>, <õ>, <ẽ>, <ē>, <ã>.¹¹ These are included in the custom Enggano Keyman keyboard.

However, there is an additional complexity in Enggano phonology. In words containing voiceless oral consonants, nasalization is phonemic, as seen in Table 6. However, when words contain the nasal consonants, /m/ and /n/, all the vowels are nasalised. In contrast, when words contain the voiced oral consonants, /b/ and /d/, all the vowels are oral vowels, and nasal vowels do not occur, as seen in Table 7.

/b/	/m/	/d/	/n/
[be] ‘dog’	[mã] ‘why’	[di] ‘tongue’	[ni] ‘name’

Table 7 Nasal harmony in Enggano.

¹¹ Note that Standard Indonesian does not have symbols for nasal vowels. Hence, another option would be to use the same symbol for oral and nasal variants, as in Kähler’s (1975) texts and in the recent dictionary produced by Kantor Bahasa (Riswari et al. 2021). This option has the advantage that it is simpler to type; however, as with central vowels, it was felt that a phonemic orthography that distinguished minimal pairs was preferable. As discussed above, the symbols <û> and <ê> for /i/ and /ẽ/ involve diacritic stacking. As pointed out by a reviewer, this is difficult to type and more visually complex, which would make them harder to process. However, these nasal phonemes are not particularly frequent. Out of approximately 700 words in the Yoder (2011) word list, there are only nine containing /ẽ/ and three containing /ĩ/.

Thus, in these contexts, oral and nasal vowels are in complementary distribution and could be considered allophones. Yoder (2011: 34) supports this analysis, treating [mē] ‘why’ and [nī] ‘name’ as underlyingly /me/ and /ni/. The nasal realisation of oral vowels is triggered by a predictable phonological process of nasal spreading, which also happens across morpheme boundaries. Consider the different allomorphs of verbal prefixes like *bu-*, a prefix marking verbs in basic realis clauses and in dependent clauses, as seen in Table 8 below.

ALLOMORPH	ROOT	DERIVED FORM
/b-/ ~ /bu-/	[ʔu] ‘say’	[buʔu]
	[pi] ‘see’	[bupi]
/m-/ ~ /mū-/	[ūē] ‘cry’	[mūē]
	[nō] ‘eat’	[mūnō]

Table 8 Allomorphs of the *bu-* prefix.

The prefix is realised as [bu-] when it attaches to stems like [ʔu] ‘say’ and [pi] ‘see’. However, when the same prefix attaches to a stem containing either nasal vowels, like [ūē] ‘cry’, or nasal consonants and vowels, like [nō] ‘eat’, the prefix is realised as [mū-]. Thus, as Yoder (2011: 34) states, vowels in Enggano are underlyingly oral in roots containing nasal consonants and are allophonically nasalised via nasal spreading.

This extra complexity leaves us with another orthographic decision: should nasal vowels be represented with the tilde wherever nasalisation occurs, regardless of the phonological status of the sounds, or should the tilde mark only vowels that are underlyingly nasal, and not the result of predictable phonological processes?

Options for nasal vowels

1. Use tilde whenever nasalisation occurs.
2. Only write nasals when underlying.

The two options would result in the spellings in Table 9:

	[kū] ‘count’	[nō] ‘eat’	[mū-nō] ‘BU-eat’	[m-ūē] ‘BU-cry’
Option 1	kū	nō	mūnō	mūē
Option 2	kū	no	muno	mūē

Table 9 Representing nasalisation.

Note that roots containing nasal vowels like [ūē] can also trigger nasal spreading into prefixes like /bu-/. Hence, option 2 is not quite as simple as suppressing the tilde if the word contains an <m> or <n>, but requires an understanding of the morphosyntactic structure of words and the morphophonological processes that occur. Luckily, there are not many verbal roots that contain underlying nasal vowels, and hence exceptions to this rule (nasal consonant → no need for tilde) could simply be learned.¹²

Option 1 (use tilde whenever nasalisation occurs) reflects pronunciation and is easy to learn and use: whenever you hear a nasal, you write it. It is also the convention adopted in Kähler’s early work (Kähler 1940, 1955, 1957, 1958, 1960a, 1960b, 1961, 1962, 1964) and the dictionary (Kähler 1987), though not the 1975 text collection, which doesn’t represent nasal vowels at all. However, it does not accurately reflect the structure of Enggano, and it would require using lots of diacritics, which can be troublesome to type, as discussed above. Option 2 (only write nasals when underlying) reflects the underlying phonemic structure more accurately and may therefore be simpler for fluent speakers, particularly since it reduces the number of diacritics needed. However, learners would have to learn the relatively complex conventions for when they write a tilde and when they don’t. As we have seen, this requires an understanding of morphology and the difference between stems and affixes.

¹² We have only identified 23 roots containing nasal vowels without nasal consonants to date.

There is psycholinguistic evidence that the best orthography for learners may not be the same as for a fluent reader or speaker (see Seifart 2006: 282). Ultimately, the project team concluded that only writing nasals when underlying (option 2) represented a simpler and more linguistically motivated orthography, and hence was preferable given the vitality of the language, even if it presented a steeper initial learning curve for learners. To help with this process, we intend to provide speakers with lists of roots and word forms containing underlying nasal vowels as a part of written community guidelines for using the orthography.

The local team presented these conventions to Enggano stakeholders, and again received positive feedback. Using the tilde diacritic to distinguish between minimal pairs, such as those in Table 6, was felt to be useful for the same reasons as above. However, avoiding unnecessary diacritics in cases where nasalisation is predictable was also felt to be simpler and more intuitive. Whether the community will adopt the convention of consistently representing roots containing nasal vowels with the tilde, even in morphologically complex forms with /mu-/, remains to be seen. Nonetheless, the current orthography follows the convention that nasal phones are written with a tilde when underlying, but written without diacritics when nasalisation is the result of a predictable phonological process in roots containing nasal consonants.

4.3 GLIDES

The final challenge is representing glides, particularly the palatal glide [j̥] which can also be realised as [j] and is henceforth transcribed as such. Standard Indonesian has a palatal approximant as part of its consonant inventory that is regularly represented in the orthography using the symbol <y> (Soderberg & Olson 2008). However, following Yoder (2011), we treat glides in Enggano as allophones of vowels that occur as syllable onsets in high-low vowel sequences word-initially, and as offglides in low-high diphthongs. For example, Table 10 shows the palatal glide allomorph of the 3SG agreement prefix *i-*, which occurs on verbs following negation and in some other dependent contexts:

ALLOMORPH	ROOT/STEM	DERIVED FORM
[i]-	[puak] ‘leave’	[i.pu.ak] ‘3SG.leave’
	[kəkə] ‘walk’	[i.kə.kə] ‘3SG.walk’
	[dər] ‘hold’	[i.dər] ‘3SG.hold’
[j]-	[ah] ‘go’	[jah] ‘3SG.leave’
	[ãʔkĩh] ‘dry’	[jãʔkĩh] ‘3SG.dry’
	[er] ‘cry’	[jer] ‘3SG.cry’
	[ēʔēh] ‘cough’	[jēʔēh] ‘3SG.cough’

Table 10 Allomorphs of the 3SG agreement prefix *i-*.

When the root or stem is vowel-initial, the agreeing form begins with a glide. However, when the root or stem begins with a consonant, the prefix is realised as a high front vowel in its own syllable peak. Since the sounds are in complementary distribution, we treat them as allophones. Hence, this is a case where Standard Indonesian orthography has an extra symbol for a contrast that is not phonemic in Enggano—in other words, the opposite issue to central vowels, where Enggano makes more distinctions in central vowels than can be represented in Indonesian orthography.

Options for glides (initial version)

1. Use the existing symbols <i> and <y> for [i] and [j].
2. Use the symbol <i> to represent both allophones of /i/.

Option 1 reflects the practice in Kähler’s work, where <y> is frequently used for the palatal glide. It also represents the general practice among the community, as shown by the written answers of schoolchildren who took part in a session to test our Enggano teaching materials, and who

often write words like [jah] as <yah>. Hence, it would be easy to use for speakers, which is a key factor in successful adoption. However, it would not be consistent with the principle of a shallow phonemic orthography, where each letter represents a unique phoneme, since, as argued above, the glides are really allophones of vowel phonemes.

For this reason, the project team initially proposed option 2, using the symbol <i> to represent both allophones of /i/, on the basis that it more accurately reflects the structure of contemporary Enggano and is in keeping with the principle of a phonemic orthography. However, when collecting feedback on the proposal, the local team faced increasing resistance from the community, who were keen to continue the established practice of using <y> for the glide (option 1). Subsequently, the project team worked intensively with Milson Kaitora, a speaker of Enggano who is closely involved in the Bible translation project, to explore this issue. Milson Kaitora was invited to spend two weeks at the University of Oxford with the project team to discuss, among other things, his views on Enggano orthography, and at this meeting it became apparent that there is an important contrast for speakers between vowel sequences word-initially and those following a word-initial glottal stop. Specifically, /i/ is pronounced [i] after a glottal stop and [j] in vowel-initial forms, creating some near minimal pairs such as [ʔiok] ‘beach’ vs. [jok] ‘vein’ as seen in Table 11.

GLOTTAL-STOP INITIAL WORDS	VOWEL-INITIAL WORDS
[ʔiok] ‘beach’	[jok] ‘vein’
[ʔiu] ‘sea’	[jub] ‘house’
[ʔie] ‘stone’	[jiç] ‘word/sound’

Table 11 Vowel-initial words vs. glottal-stop initial words.

The nominal forms in Table 11 are derived from glottal-initial and vowel-initial roots in the Kähler corpus respectively, as shown in Table 12, but seem to have undergone a process of vowel-copying (see Edwards 2015 for discussion in Old Enggano), perhaps triggered by the *e-* direct case marker that was obligatory with non-human nouns in Kähler but seems to be optional in contemporary Enggano. Before /a/, the copied vowel is realised as [e] following a glottal stop, or [e̞] as an initial glide.

	KÄHLER FORM	CONTEMPORARY FORM
‘beach’	e-’oki	[ʔiok]
‘vein’	e-oko	[jok]
‘sea’	e-’ue	[ʔiu]
‘house’	e-uba	[jub]
‘stone’	e-’ea	[ʔie]
‘word/sound’	e-ici	[jiç]
‘bone’	e-’aa	[ʔea]
‘foot’	e-ae	[e̞a]

Table 12 Old and contemporary forms.

The natural inclination for speakers of Contemporary Enggano is to insert a non-phonemic glottal stop before vowel-initial roots. Consequently, vowel-initial human nouns that never occurred with *e-* in Kähler are pronounced with an initial glottal, as seen in Table 13.

	KÄHLER FORM	CONTEMPORARY FORM
‘beach’	ūpū	[ʔūp]
‘vein’	āmā	[ʔām]

Table 13 Old and contemporary forms.

Similarly, the distinction between the 1SG agreement prefix ‘*u-*’ and the 2SG prefix *u-* that was documented in Kähler (1940) is not found in contemporary Enggano, with both realised as [ʔu]. For this reason, when speakers see a word written as *iok* they are naturally inclined to pronounce it with an initial glottal, as opposed to an initial glide. Since this can make a meaningful difference, we agreed on a new proposal:¹³

Proposal for glides, final version:

- Use <y> for [j] in onset position of vowel sequences and intervocalically.
- Write <i> elsewhere, including in vowel sequences after glottal stops, and in diphthongs.
- Do not write glottal stop initially, but assume that all vowel-initial words are pronounced with a glottal.

A final complication that then became apparent is that the historical process of vowel-copying in nouns has resulted in some minimal pairs with morphologically complex verbal/nominal forms where the glide corresponds to an underlying /i/ as in Table 14:

[ʔea] ‘bone’	[ɛar] ‘child’	[ɛāh] ‘younger sibling’
[ɛa] ‘foot’	[jar] ‘don’t’	[jāh] ‘place’ ¹⁴
[ja] ‘3SG-come’ (v)		

Table 14 Minimal pairs.

We use <ye> to represent [ɛ] so that these contrasts are captured, as this was important to the community. This results in the spellings in Table 15.¹⁵

Thus, our orthography uses the symbols <i> for [i], <y> for [j], <e> for [e] and <ye> for [ɛ]. Although this new proposal moves away from the phonemic principle proposed elsewhere in this paper, this is a good example of how involving the community in an ongoing discussion about orthography can not only reveal important contrasts that may have been overlooked, but also ensures that the resulting orthography matches with speakers’ expectations, resulting in a system that is maximally practical and therefore more likely to be adopted and used.

13 Another option would be to write initial glottal stops with the symbol <ʔ>, either all the time or, in line with the discussion concerning nasals, only in those instances where it is underlying, e.g. [ʔu] ‘1SG’. However, we decided against this for two reasons: firstly, it is not necessarily easy for speakers to distinguish between underlying glottal stop and non-phonemic inserted glottal, especially in nouns and pronouns that—unlike verb stems—do not provide any morphological clues as to their underlying form. Secondly, and perhaps more importantly, when speakers see an initial vowel they are likely to pronounce it with a glottal stop anyway. Hence, the spellings ‘*iok*’ ‘beach’ and *iok* ‘vein’ would most likely be pronounced the same. The preference of the community was therefore to distinguish *iok* [ʔiok] ‘beach’ and *yok* [jok] ‘vein’.

14 This is related to the verb *i* ‘exist’ + a locative nominalization suffix. Suffixation in contemporary Enggano triggers some complicated morphophonological processes, such as the resurfacing of the final vowel of the Kähler root *ia*. Though this makes the derivational process less transparent, the initial vowel is underlying /i/ as opposed to the copy vowel found in other nouns.

15 It remains a matter of ongoing research whether vowels apart from /i/ and /e/ have glide allophones. Yoder (2011: 48) analyzes the high and mid-vowels as having glide allophones after non-initial glottal consonants. However, this position does not appear to be as prominent for speakers as glides in initial position. Initial /u/ tends to be realised with a preceding glottal rather than as [w], at least in agreeing forms. However, [w] does occur intervocalically, particularly in borrowed words, e.g. *kawer* ‘marry’ (< kawin), and is written as such. In words containing nasals, it is possible for the palatal glide to be realised as [ɲ] or [j̠] rather than [j]. For now, we use the symbol <y> for all of these glide realisations. The decision to use <ye> needs further testing to see what support it receives in the wider community.

MEANING	PRONUNCIATION	SPELLING
'beach'	[ʔiok]	iok
'vein'	[jok]	yok
'sea'	[ʔiu]	iu
'house'	[jub]	yub
'stone'	[ʔie]	ie
'word/sound'	[jiç]	yic
'bone'	[ʔea]	ea
'foot'	[ɛa]	yea
'come'	[ja]	ya
'child'	[ɛar]	year
'don't'	[jar]	yar
'younger sibling'	[ɛāh]	yēāh
'place'	[jāh]	yāh

Table 15 Orthography for glides.

5. CONCLUSION

In conclusion, this paper has outlined three patterns of Enggano phonology that present challenges when trying to develop an orthography with reasonably shallow orthographic depth (following the principle that phonemic orthographies are most transparent for learners) while also adopting many of the existing conventions of Standard Indonesian (following the principle that staying close to an already familiar orthography is the simplest option and hence most likely to be readily adopted). In each case, it was shown that the symbols and conventions familiar to speakers from Indonesian were at odds with the desire for an orthography that accurately and consistently represented the morphophonological structure of Enggano. This necessitated a multi-stage approach to developing orthographic proposals that involved the following steps:

1. weighing up the advantages and disadvantages of different options within the project team
2. seeking feedback from linguists via Zoom meetings, conference presentations, informal discussions
3. seeking feedback from the community via focus group interviews, and testing out the proposals in schools
4. (where necessary) revising proposals on the basis of community feedback and repeating the cycle

The outcome of this process is a working orthography that is now being implemented in the documentary corpus, grammar and teaching materials. This will hopefully make the project outputs maximally accessible to speakers of Enggano and aid in the process of language revitalization (see Cahill 2018). The current orthography adopts the symbols in Table 16. It also adopts the convention that predictable and regular phonological processes, such as nasal spreading and glide-insertion between vowels in a vowel sequence, are not represented for the sake of simplicity. However, where variant pronunciations are optional or reflect dialectal/sociolinguistic variation, the convention is to write the word as it was pronounced by the speaker.

PHONEME	GRAPHEME	EXAMPLE
/p/	p	<i>pi</i> ‘garden’
/b/	b	<i>be</i> ‘dog’
/t/	t	<i>it</i> ‘banana’
/d/	d	<i>dar</i> ‘husband’
/k/	k	<i>kak</i> ‘people’
/ʔ/	’	<i>’u</i> ‘speak’
/m/	m	<i>mék</i> ‘many’
/n/	n	<i>no</i> ‘eat’
/r/	r	<i>rer</i> ‘soft’
/ç/	c	<i>yic</i> ‘word/sound’
/h/	h	<i>hũũ</i> ‘fruit’
/i/	i for [i] y for [j]	<i>pi</i> ‘garden’ <i>yub</i> ‘house’
/ĩ/	ĩ	<i>pĩh</i> ‘squeeze’
/i/	ù	<i>pù</i> ‘see’
/ĩ/	ũ	<i>pũũ</i> ‘fireplace’
/u/	u	<i>pu</i> ‘run’
/ũ/	ũ	<i>kũ</i> ‘count’
/e/	e for [e] ye for [ɛ]	<i>be</i> ‘dog’ <i>yea</i> ‘foot’
/ē/	ē	<i>kēp</i> ‘island’
/ə/	è	<i>bè</i> ‘water’
/ə/	ě	<i>kě</i> ‘try’
/o/	o	<i>po</i> ‘coconut’
/ō/	ō	<i>kōp</i> ‘grave’
/a/	a	<i>pa</i> ‘child’
/ā/	ā	<i>kāp</i> ‘tribal leader’

Table 16 Enggano orthography.

The process of orthography development and corpus building is still ongoing, and we may encounter further issues as our project progresses. Once the orthography design and the spellings of particular words have been further tested and approved by community leaders, educators, and other important stakeholders, the next step will be to implement the orthography and promote literacy in the wider community. This may not be straightforward, given that there are already competing orthographies. However, to maximize acceptability among speakers, we will explain the linguistic motivations behind orthographic decisions. We will also highlight the potential of the orthography not only to fulfill practical, pedagogical functions in aiding learners to read and write, but also to serve as an identity-marking function that represents the distinctiveness of Enggano culture (see Cahill 2018). During the implementation, we will remain open to any changes that

make the orthography more acceptable to the Enggano community, such as including <eu> as a variant grapheme for <ü>, and we will continue to view this as a collaborative and ongoing process. One factor that may increase the acceptability of the orthography will be ease of use with modern technology including smartphones. Our Keyman keyboard allows for the use of diacritics and combined diacritics using any font.¹⁶ As part of a future project, we hope to develop an app that will enable predictive text and auto-correction when typing on smart phones in order to make the orthography easy to use and promote uptake among the Enggano community.

The main lesson that we have learned from this project is how important it is to be open to continually developing and re-negotiating orthographic decisions, either when motivated by linguistic factors or when motivated by extra-linguistic concerns within the community. In the context of language documentation, it is to be expected that the linguist's understanding of the structure of the language will change over the course of the project. Nonetheless, it is helpful to think about orthographic representation from the beginning of a project, since, as shown in relation to glides, community feedback on orthographic proposals can provide direct insight into language structure. In our project, we were forced by the COVID-19 pandemic to do much of the analysis and discussion online. However, this has turned out to have several advantages that could be borne in mind for future documentation and orthography development projects. Firstly, having regular meetings via Zoom allowed us to connect scholars and language activists in different countries and time zones who could bring their own expertise and experience to the challenge of orthography development, particularly in the initial stages of weighing up the advantages and disadvantages of the different options. Secondly, and more importantly, it meant that we had no choice but to invest time in capacity building and training for the local project team in Indonesia, who played an important role in collecting feedback from the community and facilitating the re-development of orthography proposals. This, we believe, led to a more collaborative approach, where important stakeholders were directly involved in the process and empowered to make informed orthographic choices and communicate these to the Enggano community at large (cf. Czaykowska-Higgins 2009: 24). Working collaboratively and collecting feedback as we go ensures greater community engagement. Ultimately, this is the only way to really achieve the goal of an orthography that is linguistically motivated but also practical, easy to use, and widely accepted.

FUNDING INFORMATION

Our work is funded by the Arts & Humanities Research Council, UK (Grant AH/S011064/1) and runs from 2019 to 2024. Besides the authors, project partners are Prof. Bernd Nothofer and collaborators at Goethe-Universität Frankfurt am Main, Universitas Udayana in Bali, Universitas Bengkulu, Kantor Bahasa Provinsi Bengkulu and BRIN (Badan Riset dan Inovasi Nasional). The project builds upon earlier research supported by grants from the Endangered Language Fund and the University of Oxford's John Fell Fund.

COMPETING INTERESTS

The authors have no competing interests to declare.

AUTHOR AFFILIATIONS

Charlotte Hemmings  orcid.org/0000-0002-3076-5544
University of Oxford, UK

I Wayan Arka  orcid.org/0000-0002-2819-6186
Australian National University/Universitas Udayana, AU

Engga Zakaria Sangian  orcid.org/0009-0000-8802-6819
Universitas Dehasen Bengkulu, ID

¹⁶ Instructions on how to use the Keyman keyboard are available here: <https://enggano.ling-phil.ox.ac.uk/static/keyman.html>.

REFERENCES

- Arka, I Wayan, Arono, Dendi Wijaya & Engga Zakaria. 2022. Critical ecological factors in ethnolinguistic vitality: evidence from Enggano. *Paper presented at the Austronesian and Papuan Languages and Linguistics Conference*, Berlin, Germany.
- Austin, Peter K. 2016. Language documentation 20 years on. In Luna Filipović & Martin Pütz (eds.), *Endangered languages and languages in danger: Issues of documentation, policy, and language rights*, 147–170. Amsterdam/Philadelphia: John Benjamins. DOI: <https://doi.org/10.1075/impact.42.07aus>
- Billings, Blaine & Bradley McDonnell. 2022. Subgrouping Malayo-Polynesian languages of Sumatra and the Barrier Islands. *Paper presented at the 30th Annual Meeting of the Southeast Asian Linguistics Society*, University of Hawai'i at Mānoa.
- Blench, Roger. 2014. The Enggano: Archaic foragers and their interactions with the Austronesian world. Unpublished draft. Online at <http://www.rogerblench.info/Language/Austronesian/Enggano/Enggano%20and%20its%20history.pdf>. Accessed 2023-05-12.
- Boewang, Djoeragan. 1854. Verslag omtrent het eiland Engano [Report on the island of Engano]. *Tijdschrift voor Indische Taal-, Land- en Volkenkunde* 2: 379–93.
- Butters, Marielle. 2021. *Negation in four languages of Indonesia*. Boulder, Colorado: University of Colorado at Boulder dissertation.
- Cahill, Michael. 2018. Orthography design and implementation for endangered languages. In Kenneth L. Rehg & Lyle Campbell (eds.), *The Oxford handbook of endangered languages*, 327–346. Oxford: Oxford University Press. DOI: <https://doi.org/10.1093/oxfordhb/9780190610029.013.17>
- Capell, Arthur. 1982. Bezirkssprachen im UAN-Gebiet [District languages in the UAN area]. In Rainer Carle, Martina Heinscke, Peter Pink, Christel Rost & Karen Stadlander (eds.), *Gava': Studies in Austronesian languages and cultures dedicated to Hans Kähler*, 1–14. Berlin: Dietrich Reimer.
- Coulmas, Florian. 2003. *Writing systems: An introduction to their linguistic analysis*. Cambridge: Cambridge University Press. DOI: <https://doi.org/10.1017/CBO9781139164597>
- Czaykowska-Higgins, Ewa. 2009. Research models, community engagement, and linguistic fieldwork: Reflections on working within Canadian Indigenous communities. *Language Documentation & Conservation* 3(1): 15–50.
- Durie, Mark. 1984. *A grammar of Acehnese on the basis of a dialect of north Aceh* (Verhandelingen van Het Koninklijk Instituut Voor Taal-, Land- En Volkenkunde 112). Dordrecht/Cinnaminson, NJ: Foris Publications.
- Dyen, Isidore. 1965. *A lexicostatistical classification of the Austronesian languages* (Indiana University Publications in Anthropology and Linguistics, Memoir 19 of the *International Journal of American Linguistics*). Baltimore: Waverly Press.
- Edwards, Owen. 2015. The position of Enggano within Austronesian. *Oceanic Linguistics* 54(1): 54–109. DOI: <https://doi.org/10.1353/ol.2015.0001>
- Helfrich, Oscar Louis. 1893. Aanvullingen en verbeteringen op de Maleisch–Nederlandsch–Enganeesch woordenlijst [Additions and corrections to the Malay-Dutch-Enggano vocabulary list]. *Tijdschrift voor Indische Taal-, Land-en Volkenkunde* 35: 228–33.
- Helfrich, Oscar Louis. 1916. Nadere bijdrage tot de kennis van het Enganeesch [Further contribution to the knowledge of Enggano]. *Bijdragen tot de Taal-, Land-en Volkenkunde van Nederlandsch-Indië* 71(3/4): 472–555. DOI: <https://doi.org/10.1163/22134379-90001719>
- Helfrich, Oscar Louis & J. A. J. C. Pieters. 1891. Proeve van eene Maleisch-Nederlandsch-Enganeesch woordenlijst [Sample of a Malay-Dutch-Enggano vocabulary list]. *Tijdschrift voor Indische Taal-, Land- en Volkenkunde* 34: 539–623.
- Himmelman, Nikolaus P. 1998. Documentary and descriptive linguistics. *Linguistics* 36: 161–195. DOI: <https://doi.org/10.1515/ling.1998.36.1.161>
- Himmelman, Nikolaus P. 2006. Language documentation: What is it and what is it good for. In Jost Gippert, Nikolaus P. Himmelman & Ulrike Mosel (eds.), *Essentials of language documentation*, 1–30. Berlin/New York: De Gruyter Mouton. DOI: <https://doi.org/10.1515/9783110197730.1>
- Jones, Mari C. & Damien Mooney. 2017. *Creating orthographies for endangered languages*. Cambridge: Cambridge University Press. DOI: <https://doi.org/10.1017/9781316562949>

- Kähler, Hans. 1940. Grammatischer Abriss des Enggano [Grammatical outline of Enggano]. *Zeitschrift für Eingeborenen Sprachen* 30: 81–117, 182–210, 296–320.
- Kähler, Hans. 1955. Ein Text von der Insel Enggano (Westküste von Sumatra) [A text from the island of Enggano (west coast of Sumatra)]. *Afrika und Übersee* 39: 89–94.
- Kähler, Hans. 1957. Ein Text von der Insel Enggano (Westküste von Sumatra) [A text from the island of Enggano (west coast of Sumatra)]. *Afrika und Übersee* 41: 153–56.
- Kähler, Hans. 1958. Texte von der Insel Enggano (Westküste von Sumatra) [Texts from the island of Enggano (west coast of Sumatra)]. *Afrika und Übersee* 42: 179–89.
- Kähler, Hans. 1960a. Texte von der Insel Enggano (Westküste von Sumatra) [Texts from the island of Enggano (west coast of Sumatra)]. *Afrika und Übersee* 43: 299–312.
- Kähler, Hans. 1960b. Texte von der Insel Enggano (Westküste von Sumatra) [Texts from the island of Enggano (west coast of Sumatra)]. *Afrika und Übersee* 44: 41–71.
- Kähler, Hans. 1961. Texte von der Insel Enggano (Westküste von Sumatra) [Texts from the island of Enggano (west coast of Sumatra)]. *Afrika und Übersee* 45: 205–24.
- Kähler, Hans. 1962. Texte von der Insel Enggano (Westküste von Sumatra) [Texts from the island of Enggano (west coast of Sumatra)]. *Afrika und Übersee* 46: 127–45.
- Kähler, Hans. 1964. Texte von der Insel Enggano (Westküste von Sumatra) [Texts from the island of Enggano (west coast of Sumatra)]. *Afrika und Übersee* 48: 46–57.
- Kähler, Hans. 1975. *Texte von der Insel Enggano (Berichte über eine untergehende Kultur)* [Texts from the island of Enggano (Reports about a declining culture)]. Berlin: Dietrich Reimer Verlag.
- Kähler, Hans. 1987. *Enggano-Deutsches Wörterbuch [Enggano-German Dictionary]*. Edited by Hans Schmidt. Vol. 14. Berlin: Dietrich Reimer Verlag.
- Kasim, Yuslina, Marjusman Maksan, Syamsir Arifin, Zailoet & Muhamad Yamin. 1987. *Pemetaan bahasa daerah di Sumatra Barat dan Bengkulu [Mapping of local languages in West Sumatra and Bengkulu]*. Jakarta: Pusat Pembinaan dan Pengembangan Bahasa, Departemen Pendidikan dan Kebudayaan.
- Katz, Leonard & Ram Frost. 1992. The reading process is different for different orthographies: The orthographic depth hypothesis. In Ram Frost & Leonard Katz (eds.), *Orthography, phonology, morphology, and meaning*, 67–84. Amsterdam: Elsevier North Holland Press. DOI: [https://doi.org/10.1016/S0166-4115\(08\)62789-2](https://doi.org/10.1016/S0166-4115(08)62789-2)
- Lüpkke, Friederike. 2011. Orthography development. In Peter K. Austin & Julia Sallabank (eds.), *The Cambridge handbook of endangered languages*, 312–336. Cambridge: Cambridge University Press. DOI: <https://doi.org/10.1017/CBO9780511975981.016>
- Nikelas, Syahwin, Halipami Rasyad & M. Atar Semi. 1994. *Morfologi dan sintaksis bahasa Enggano [Morphology and syntax of the Enggano language]*. Jakarta: Pusat Pembinaan dan Pengembangan Bahasa, Departemen Pendidikan dan Kebudayaan.
- Nothofer, Bernd. 1986. The Barrier Island languages in the Austronesian language family. In Paul Geraghty, Lois Carrington & S. A. Wurm (eds.), *FOCAL II: Papers from the Fourth International Conference on Austronesian Linguistics*, 87–109. Canberra: Pacific Linguistics.
- Nothofer, Bernd. 2021. Enggano revisited: the word for ‘window’. In Irene Schneider & Holder Warnk (eds.), *Knowledge, science, and local tradition*, 215–222. Wiesbaden: Harrassowitz Verlag.
- Oudemans, Anthonie Cornelis. 1879. Woordenlijst van de talen van Enggano, Mentawai en Nias [Glossary of the languages of Enggano, Mentawai, and Nias]. *Tijdschrift voor Indische Taal-, Land-en Volkenkunde* 25: 484–88.
- Oudemans, Anthonie Cornelis. 1889. Engano (bewesten Sumatra) zijne geschiedenis, bewoners en voortbrengselen [Enggano (west of Sumatra): Its history, inhabitants, and produce]. *Tijdschrift van het Koninklijk Nederlandsch Aardrijkskundig Genootschap* 6: 109–64.
- Riswari, Yanti, Azmi Ridwan Fauzi, M. Yusuf, Resy Novalia, Zainal Arifin Nugraha & Olga Chaesa Novianti. 2021. *Kamus Dwibahasa: Enggano-Indonesia, Indonesia-Enggano [Bilingual Dictionary: Enggano-Indonesian, Indonesian-Enggano]*. Bengkulu, Indonesia: Kantor Bahasa Provinsi Bengkulu, Badan Pengembangan dan Pembinaan Bahasa, Kementerian Pendidikan, Kebudayaan, Riset dan Teknologi.
- Sebba, Mark. 2007. *Spelling and society: The culture and politics of orthography around the world*. Cambridge: Cambridge University Press. DOI: <https://doi.org/10.1017/CBO9780511486739>
- Seifart, Frank. 2006. Orthography development. In Jost Gippert, Nikolaus P. Himmelmann & Ulrike Mosel (eds.), *Essentials of language documentation*, 275–300. Berlin, New York: De Gruyter Mouton. DOI: <https://doi.org/10.1515/9783110197730.275>
- Smith, Alexander D. 2017. The Western Malayo-Polynesian problem. *Oceanic Linguistics* 56(2): 435–490. DOI: <https://doi.org/10.1353/ol.2017.0021>
- Smith, Alexander D. 2020. Nasalization in Enggano historical phonology. *Oceanic Linguistics* 59(1/2): 347–365. DOI: <https://doi.org/10.1353/ol.2020.0015>

- Soderberg, Craig D. & Kenneth S. Olson. 2008. Indonesian. *Journal of the International Phonetic Association* 38(2): 209–213. DOI: <https://doi.org/10.1017/S0025100308003320>
- van der Noord, Abs. 1987. Enggano. In W. A. L. Stokhof (ed.), *Holle lists: Vocabularies in languages of Indonesia*, 189–205. Canberra: Pacific Linguistics.
- van der Straaten, J. & P. Severijn. 1855. Verslag van een in 1854 bewerkstelligd onderzoek op het eiland Engano [Report of an investigation conducted on the island of Engano in 1854]. *Tijdschrift voor Indische Taal-, Land- en Volkenkunde* 3: 338–69.
- Vikør, Lars S. 1988. *Perfecting spelling: spelling discussions and reforms in Indonesia and Malaysia, 1900–1972, with an appendix on Old Malay spelling and phonology*, (Verhandelingen van het Koninklijk Instituut voor Taal-, Land- en Volkenkunde). Dordrecht: Foris Publications.
- von Rosenberg, H. 1855. Beschrijving van Engano en van deszelfs bewoners [Description of Engano and its inhabitants]. *Tijdschrift voor Indische Taal-, Land- en Volkenkunde* 3: 370–386.
- Walland, J. 1864. Het eiland Engano [The island of Engano]. *Tijdschrift voor Indische Taal-, Land- en Volkenkunde* 14: 93–124, 329–339.
- Wibawa, Jaka Aris Eko, Supheakmungkol Sarin, Chen Fang Li, Knot Pipatsrisawat, Keshan Sodimana, Oddur Kjartansson, Alexander Gutkin, Martin Jansche & Linne Ha. 2018. Building open Javanese and Sundanese corpora for multilingual text-to-speech. *Proceedings of the eleventh international Conference on Language Resources and Evaluation (LREC 2018)*. Online at <https://aclanthology.org/L18-1255.pdf>. Accessed 2023-05-12.
- Wijaya, Dendi. 2018. *Pemarkah Gramatikal Verba Bahasa Enggano Dalam Hubungan Keaspekan (Kajian Morfosintaksis) [The grammatical markers of verbs in the Enggano language in relation to aspectuality (a study of morphosyntax)]*. Jakarta, Indonesia: Universitas Indonesia master's thesis.
- Woodbury, Anthony C. 2003. Defining documentary linguistics. *Language Documentation and Description* 1: 35–51.
- Woodbury, Anthony C. 2011. Language documentation. In Peter K Austin & Julia Sallabank (eds.), *The Cambridge handbook of endangered languages*, 159–186. Cambridge: Cambridge University Press. DOI: <https://doi.org/10.1017/CBO9780511975981.009>
- Yoder, Brendon. 2011. *Phonological and phonetic aspects of Enggano vowels*. Grand Forks, ND: University of North Dakota master's thesis.

TO CITE THIS ARTICLE:

Hemmings, Charlotte, I Wayan Arka & Engga Zakaria Sangian & Dendi Wijaya & Mary Dalrymple. 2023. Challenges in Enggano Orthography Development. *Language Documentation and Description* 23(1): 4, 1–19. DOI: <https://doi.org/10.25894/ldd.329>

Submitted: 18 January 2023

Accepted: 04 May 2023

Published: 07 July 2023

COPYRIGHT:

© 2023 The Author(s). This is an open-access article distributed under the terms of the Creative Commons AttributionNonCommercial 4.0 International License (CC-BY-NC 4.0), which permits unrestricted distribution, reproduction and adaptation in any medium, provided the original author and source are credited, and that the material is not used for commercial purposes. See <https://creativecommons.org/licenses/by-nc/4.0/>.

Language Documentation and Description is a peer-reviewed open access journal published by Aperio.