Challenges in Enggano Orthography Development

Charlotte Hemmings, I Wayan Arka, Mary Dalrymple, Engga Zakaria Sangian and Dendi Wijaya

Introduction

- Background on Enggano
- Orthography Development
- Challenge 1: oral vowels
- Challenge 2: nasal vowels
- Challenge 3: glides
- Conclusion

Background on Enggano

- Enggano is spoken by approx. 1,500 speakers on Enggano Island, Sumatra, Indonesia

- There is some debate around sub-grouping but most people now agree that Enggano is Austronesian (Nothofer 1992, Edwards 2015, Smith 2020)

- Our documentation project (2018-present) has largely focused on central dialects (e.g. Meok)

- The aim of the project is to produce a documentary corpus, a grammar of Enggano and a set of teaching materials
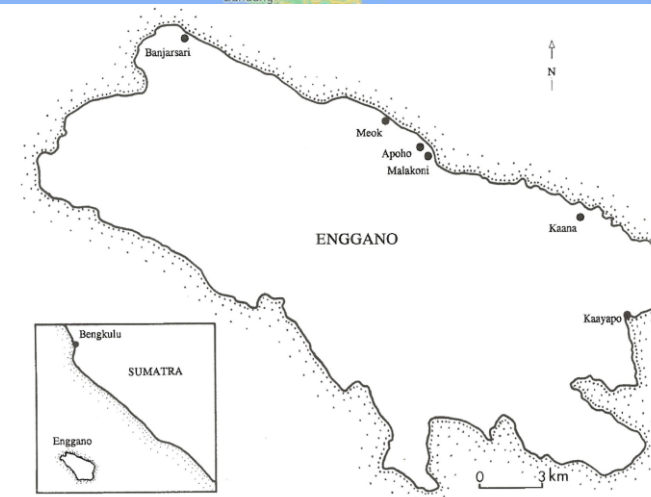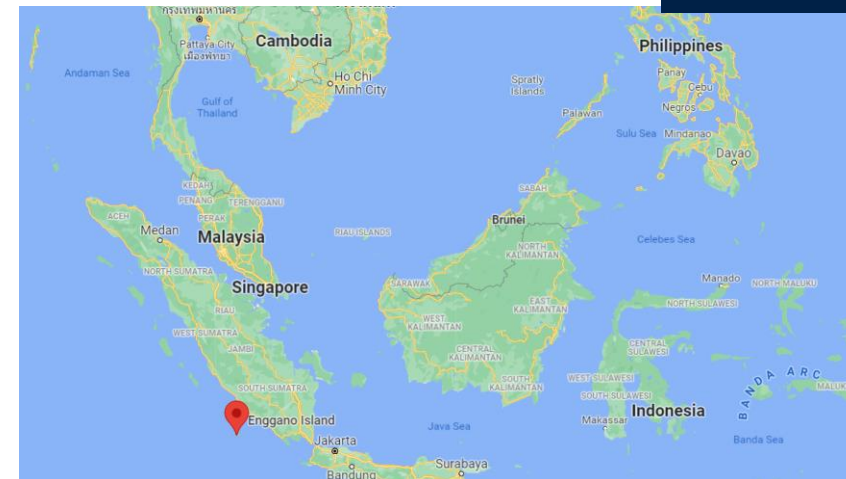

Fig. 15: Map of Enggano Island (Indonesia). (Drawn by Frans Stelling)

map from ter Keurs (2006: 134)

History of Documentation

1850-1900	Early Wordlists	Von Rosenberg 1855, Van der Straaten & Severijn 1855, Walland 1864, Oudemans 1879 Helfrich & Pieters 1891, Helfrich 1893, 1916 Holle List 1895
1930s	Hans Kähler	Grammar Sketch (Kähler 1940) Text Collection (Kähler 1955, 1957, 1958, 1960, 1961, 1962, 1964, 1975) Dictionary (published posthumously, Schmidt 1987)
1980s	Bernd Nothofer	Wordlist & Historical Work
1980s-2020s	Kantor Bahasa	Nikelas et al (1994), Wijaya (2018), ongoing projects
2011	Brendan Yoder	Masters Thesis on Phonology

History of Orthography

- Early works did not use a common orthography, e.g. *ietebe, itèbè, itcébe* 'on top'

- Kähler used different symbols and conventions in the grammar, texts and dictionary

1940 Grammar	1975 Texts	1987 Dictionary	
du̱hu̱da	*dā̱hāda*	*dəhəda*	'finish'
kinõ̱'õ̱ãhã	*kino'oaha*	*kinõˀõahã*	'like that'

- Yoder (2011) uses IPA transcriptions: *dɨhɨr* 'finish', *kinaʔah* 'like that'

Writing Enggano in the community

- There is no formalised orthography

- However, there is an ongoing project to produce a bible translation in Enggano

- Some symbols have been proposed for this (e.g. **ė** for [ə])

- Speakers are literate in Indonesian and hence most tend to adopt Indonesian conventions (e.g. use **e** for both [ə] and [e])

- However, Indonesian conventions may not reflect the structure of Enggano...

Our Challenge

- In order to document contemporary Enggano, we need a standard orthography

- There is no existing orthography – so we need to develop one! (cf. Lüpke 2011)

- We want it to reflect the **structure of Enggano** but we also want it to be widely accepted and adopted (by the community and other researchers)

- We need to take both linguistic and extra-linguistic factors into account and ensure that important stakeholders are involved in the orthography development process (Seifart 2006).

- (added challenge in COVID 19 pandemic: how can we achieve this online?)

Orthography Development

- An orthography is a set of graphic symbols ("graphemes") and rules for how they are used (Coulmas 2003)

- Ideally an orthography should reflect the structure of the language, but also be practical and easy to use.

- **These issues may be in conflict!**

- Two important decisions (Seifart 2006):

    (1) Orthographic Depth

    (2) Adopting existing conventions

- Ultimately, the orthography developer has to balance the advantages and disadvantages of different options

How did we approach the challenge?

- Before the pandemic, the team had already begun to collect a corpus

- Since the pandemic, we have been having regular meetings via Zoom.
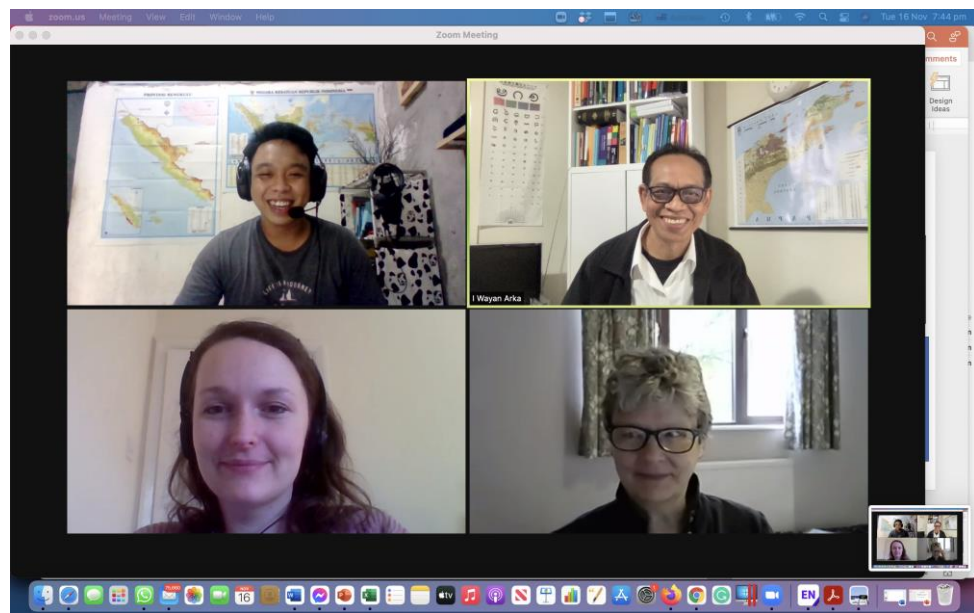
- The aim is to better understand the structure of Enggano

- In particular, we looked for minimal pairs in order to establish a list of **phonemes**

- We could then propose a phonemic orthography and begin to compile a list of orthographic issues.

How did we approach this challenge?

- We then worked with our research assistant, Engga, to develop proposals for how to deal with these issues

- The local research team took the proposals to Enggano to **gather feedback** from community elders and test out the orthography in schools

- We also discussed this with other researchers working on Enggano (including members of our project team and Kantor Bahasa Bengkulu)

Phonemes of Enggano (see Yoder 2011)

	Bilabial		Alveolar		Palatal	Velar	Glottal
Stop	p	b	t	d		k	ʔ
Nasal		m		n			
Rhotic			r				
Fricative					ç		h

	Front		Central		Back	
High	i	ĩ	ɨ	ɨ̃	u	ũ
Mid	e	ẽ	ə	ə̃	o	õ
Low			a	ã		

Challenge 1: seven oral vowels

- There are seven vowel phonemes but only five vowel letters in the Indonesian alphabet

[be] 🔊	'dog'	[bə] 🔊	'water'
[kər] 🔊	'swallow'	[kɨr] 🔊	'live'

- Option 1: use the same symbol for multiple sounds (e for [e] and [ə])

- Option 2: use diacritics (ė for [ə] and u̇ for [ɨ])

- Option 3: use digraphs (eu for [ɨ])

Challenge 1: seven oral vowels

	Advantages	**Disadvantages**
Option 1	• adopts conventions • easy to learn & simple to write	• letters ≠ phonemes • hard to pronounce (anek)
Option 2	• letters = phonemes • ė proposed by community	• symbols not familiar (kėr vs kůr) • hard to type (especially on phones)
Option 3	• symbols = phonemes • eu used informally by speakers	• difficult to distinguish digraphs from diphthongs (kahai') • Could lead to mispronunciation (keur)

*with option to use digraphs informally

Challenge 2: nasal vowels

- Minimal pairs in words containing voiceless oral consonants /p/, /t/ and /k/ suggest that nasal vowels are phonemes

[ku]	🔊	'tree'	[kũ]	🔊	'count'
[kə]	🔊	'tuber'	[kə̃]	🔊	'try'

- To achieve a phonemic orthography we use ~ diacritic to represent nasals: ĩ, ɨ̃, ũ, ẽ, ə̃, õ, ã

Challenge 2: nasal vowels

- Roots with nasal consonants, /m/ and /n/, only contain nasal vowels, and roots with voiced oral consonants, /b/ and /d/, only contain oral vowels
- **Nasal spreading**: when a stem contains nasal consonants or nasal vowels, nasalisation spreads to the affixes

Affix	Root	Derived Form
bu-	*'u* 'say'	bu-'u
	pù 'see'	bu-pù
	ũẽ 'cry'	m-ũẽ
	no 'eat'	mu-no

- Consequently, Yoder (2011: 34) argues that vowels are underlying oral in roots containing nasal consonants

Challenge 2: nasal vowels

- Option 1: use the tilde (~) whenever nasalisation occurs
- Option 2: only write nasals when they are underlying

Option 1	Option 2	
ũẽ	ũẽ	cry
m-ũẽ	m-ũẽ	bu-cry
nõ	no	eat
mũ-nõ	mu-no	bu-eat

	Advantages	Disadvantages
Option 1	- reflects pronunciation - easy to use	- lots of diacritics - doesn't reflect phonemes
Option 2	- reflects the phonemes - Simpler for fluent speakers	learners will have to learn the conventions

Challenge 3: glide allophones of vowels

- Glides ([j] and [w]) occur as allophones of vowels when they occur syllable initially (including at the start of words):

[jis]	🔊	'word/sound'
[jub]	🔊	'house'

- Glides are also (optionally) inserted by a process of palatalization that occurs when a high front vowel precedes a glottal consonant (/h/ and /ʔ/)

Affix	Root	Derived Form
ki-	'u hẽk	ki'iu 'say' kihiẽk 'sit'

Challenge 3: glide allophones of vowels

- The palatal glide [j] is very frequent word-initially in contemporary Enggano.

- We hypothesize that the *e-* nominal prefix could trigger palatalization in vowel/glottal/h-initial words but the glide was reinterpreted as part of the root when the prefix became optional

Kahler Form	Contemporary Form	Meaning
e-uba	[jub]	'house'
e-ada	[jar]	'child'
e-ici	[jiç]	'word/sound'
e-huda	[hiɨr]	'woman'

reasonably high functional load

no minimal pairs, hence they are allophones

Challenge 3: glide allophones of vowels

- Option 1: use the symbol 'y' to represent the glide as in Indonesian, but 'i' elsewhere for the high, front vowel (yic, yub)

- Option 2: use 'i' to represent both the glide and the vowel (iic, iub)

	Advantages	Disadvantages
Option 1	• adopts conventions • easy to use	• the orthography is not consistently phonemic
Option 2	• reflects phonemes • more consistent	• learners will have to adapt the conventions that they know from Indonesian

Enggano Orthography

phoneme	grapheme	example
/p/	p	*pi* 'garden'
/b/	b	*be* 'dog'
/t/	t	*it* 'banana'
/d/	d	*dar* 'husband'
/k/	k	*kak* 'people'
/ʔ/	'	*i'ĩẽ'* 'here'
/m/	m	*mėk* 'many'
/n/	n	*no* 'eat'
/r/	r	*hiủr* 'woman'
/ç/	c	*iic* 'word/sound'
/h/	h	*hĩũ* 'fruit'

phoneme	grapheme	example
/i/	i	*pi* 'garden'
/ĩ/	ĩ	*pĩh* 'squeeze'
/ɨ/	ủ	*pủ* 'see'
/ɨ̃/	ũ	*pũ'* 'fireplace'
/u/	u	*pu* 'run'
/ũ/	ũ	*kũ* 'count'
/e/	e	*be* 'dog'
/ẽ/	ẽ	*kẽp* 'island'
/ə/	ė	*bė* 'water'
/ə̃/	ẽ	*kẽ* 'try'
/o/	o	*po* 'coconut'
/õ/	õ	*kõp* 'grave'
/a/	a	*pa* 'child'
/ã/	ã	*kãp* 'tribal leader'

Conclusion

- We now have a working orthography which we are implementing in the corpus and teaching materials

- This may not be the final orthography and we will likely encounter further issues as our linguistic analysis progresses.

- However, working collaboratively, and collecting feedback as we go, ensures **community engagement and empowerment** (Czaykowska-Higgins 2009: 24).

- 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.

Summary

- The COVID pandemic forced much of the linguistic analysis to take place online

- This has the potential disadvantage of the linguists not observing the language in its natural environment

- However, it has allowed regular meetings and connect people in different time zones/locations

- Most importantly, we have had no choice but to invest time in capacity building/training of our local team – leading to a more collaborative approach and greater **community engagement**.

With thanks to…

I Wayan Arka, Australian National University
Mary Dalrymple, University of Oxford
Dendi Wijaya, Kantor Bahasa Bengkulu
Engga Zakaria Sangian, Universitas Dehasen Bengkulu

With thanks to…

Pak Harun, Pak Raflizen, Pak Johansen and the Enggano Community

With thanks to…

Bernd Nothofer, Goethe-Universität Frankfurt am Main
Erik Zobel, independent researcher
Daniel Krausse, Goethe-Universität Frankfurt am Main
Yanti Riswara Idris, Kantor Bahasa Bengkulu

Audience at LDLT6!