Representing Lota Ende in Unicode

Anshuman Pandey pandey@umich.edu

April 28, 2016

Introduction 1

This document offers an approach for representing the 'Lota Ende' or 'Ende' script in Unicode. This script is used for writing Ende (ISO 639: end), a Malayo-Polynesian language spoken on Flores, Indonesia. The language is known alternately as Li'o or Ende-Li'o. The script is an extension of Buginese with languagespecific letters and alternate forms. The script was described by Christopher Miller in Unicode Technical Note #35 "Indonesian and Philippine Scripts and Extensions" (2011: 39-41).

2 **Script Details**

Lota Ende contains the following letters:

11	/k/	5	$/\widehat{d_3}/$	
^♥, ∧	/g/	≈, ∞	/ɲ/	
Y	/n/	<u>~~. %</u>	/i/	

11	/k/	5	/d3/	\$, \$	/d/
∕∿, ∧∕	/g/	≈, ≈	/ɲ/	6	/f/
Х	/ŋ/	^^ , %	/j/	÷	/γ/
~,~	/p/	\$	/r/	X , Y	/mb/
х	/b/	\sim	/1/	v , X	/nd/
\checkmark	/m/	\sim	/w/	^V	/ŋg/
\wedge	/t/	/	/s/	\approx	$ \mathbf{I} $
\checkmark	/d/	ο, φ	/h/	Ø, X	/h/
\wedge	/n/	~~	/?/, /a/, 0	~~	/z/
\approx	/t͡ʃ/	S, 2	/6/		

and the following vowel signs:

Representing Lota Ende in Unicode

்	/i/	৻	/e/
਼	/u/	্ব	/0/

and the following vowel-silencing sign:

î -0

The structure of Ende is similar to that of Buginese. Each consonant letter possesses the inherent vowel /a/. This vowel is changed by applying dependent vowel signs, which attach to the left, right, above, and below the base consonant. A bare consonant is indicated by use of a sign, which indicates the silencing of the inherent vowel. Some prenasalized consonants are represented using distinctive letters. Several consonants have alternate forms that may co-occur with the regular forms of letters.

The letter ∞ is a vowel carrier and represents the independent form of the vowel /a/. Independent forms of vowels are represented by attaching vowels signs to ∞ , as shown below.

/a/	~~	\sim
/i/	$\dot{\sim}$	<u>≁+</u> ;
/i/	$\dot{\sim}$	∞ +∘
/e/	<~~	৵ + <
/0/	~~^	৵ + ৵

The vowel-silencing sign is used as follows:

/ka/	11	11
/k/	îi	//+ ℃
/ga/	\sim	\sim
/g/	ŵ	~v+ °

There are no script-specific digits or punctuation.

3 Comparison of Ende and Buginese repertoires

Several letters are shared between Ende and Buginese, but there are differences in the forms and values of letters, as well as letters used in Ende for sounds that are not represented in the standard Buginese script. A comparison is shown below, using Buginese as the basis for comparison as it is already encoded in Unicode.

3.1 Current repertoire for Buginese

3.1.1 Consonants

The Buginese block contains 23 letters. Of these, 11 are identical or nearly similar in Ende:

	Buginese	Ende			Buginese	Ende
KA	11	11		NA		\wedge
PA	シ	シ		YA	~~	~~
BA	ス	х		RA	*	\$
MA	\checkmark	\sim		VA	\sim	\sim
TA	^	\wedge		А	~~	\sim
DA	~	\checkmark				
			-			

Seven consonants have different forms:

	Buginese	Ende		Buginese	Ende
GA	<i>~</i>)	\sim	NYA	\sim	~
NGA	X	Х	LA	~	\sim
CA	~>	\approx	SA	\$	1
JA	~	2	HA	∞	٥

Note the differing value of \diamond in the two scripts: this letter represents /s/ in Buginese and is encoded with the name sA, while in Ende it represents /h/.

The block contains four letters that are not used in Ende. These represent prenasalized consonants of the Bugis language:

	Buginese	Ende
NGKA	~	
MPA	22	
NRA	<u>م</u> ک	_
NYCA	~	—

3.1.2 Vowel signs

Buginese contains 5 vowel signs, of which 4 are identical in Ende:

	Buginese	Ende
VOWEL SIGN I	ċ	்
VOWEL SIGN U	਼	਼
VOWEL SIGN E	<	<
VOWEL SIGN O	্ব	্ব

The following sign is not used in Ende:

	Buginese	Ende
VOWEL SIGN AE	े	

3.1.3 Punctuation

Both Buginese marks of punctuation are used in Ende:

	Ende	Buginese
PALLAWA	***	***
END OF SECTION	¢	¢

3.2 Missing characters

The 'Buginese' block does not have characters that correspond to the following 10 letters required for representing Ende:

S	/6/	*	/nd/
S	/ď/	~v	/ŋg/
6	/f/	~	$ \mathbf{I} $
-3	/ɣ/	Ø	/h/
×	/mb/	~~	/z/

Additionally, there are alternate forms of 10 letters that have the potential of being treated as distinctive characters rather than as glyphic variants:

<i>α</i> , φ	/h/	Ø, ×	/h/
^^ , %	/j/	v , X	/nd/
≈, ∞	/ɲ/	X , Y	/mb/
∼>, ~∨	/p/	<i>د</i> م, ۶	/d/
№, №	/g/	W, N	/6/

Buginese does not have a vowel-silencing sign, but such a character is used in Ende:

	Ende	Buginese	
े	-0	_	

4 Approach for encoding

The 'Buginese' block contains 30 characters: 23 consonant letters, 5 vowel signs, and 2 punctuation signs. Representing Ende in Unicode requires 34 characters: 29 letters, 4 combining vowel signs, and 1 VIRAMA. Of these letters, 17 are distinctive, while 12 can be represented using existing Buginese characters. Of the vowel signs, 4 are identical and 1 does not occur in Ende. In total, a minimum of 18 new characters is required for Ende. There is a potential to encode an additional 10 characters for alternate forms of letters.

Ende should be represented in Unicode as an extension of Buginese, rather than as a separate script. This involves the following actions:

- 1. As the 'Buginese' block in the BMP has only two spaces remaining, and as there is no free space in the BMP, a new block should be created in the SMP with the name 'Buginese Extensions'. The block should encompass at least 5 columns to accommodate characters from other Buginese orthographies.
- 2. Encode the following 17 letters in 'Buginese Extensions'. As some letters may also used in Buginese orthographies for other languages, character names should be generic and not specific to Ende:

\sim	ga
٢	nga
\approx	са
2	ja
~	nya
\sim	la

1 sa bba 3 S dda 6 fa -3 gga mba $\boldsymbol{\infty}$ ≈ nda ngga ^.∕ \approx rra hha 0 ~~ za

3. Add the following annotation to the names list for the 'Buginese' block:

- 4. Encode the following combining sign in the existing 'Buginese' block (see L2/16-075):
 - **ô** BUGINESE SIGN VIRAMA-2
- 5. Identify the status of the following alternate forms as distinctive letters or glyphic variants:
 - *∧* alternate ga
 - \sim alternate pa
 - \Leftrightarrow alternate nya
 - × alternate ya
 - φ alternate ha
 - ∼ alternate bba
 - S alternate dda
 - $\boldsymbol{\gamma}$ alternate mba
 - **X** alternate nda
 - × alternate hha

A formal proposal for encoding letters of Ende and other Buginese-based scripts is forthcoming.

5 References

Banda, Maria Matildis. 2014. "*Lota* Characters in Ende, Flores". Presented at the International Workshop on Endangered Scripts of Island Southeast Asia, Tokyo University of Foreign Studies, February – March 2014.

http://lingdy.aacore.jp/doc/endangered-scripts-issea/maria_matildis_banda_paper.pdf

- Miller, Christopher. 2010. "Unicode Technical Note #35: Indonesian and Philippine Scripts and Extensions". http://www.unicode.org/notes/tn35/
- Pandey, Anshuman. 2016. "Proposal to encode VIRAMA signs for Buginese". L2/16-075. http://www.unicode.org/L2/L2016/16075-buginese-virama-signs.pdf
- van Suchtelen, B. C. C. M. M. 1921. *Endeh (Flores)*. Mededeelingen van het Bureau voor de Bestuurszaken der Buitengewensten, bewerkt door het Encyclopædisch Bureau. Aflevering XXVI. Weltevreden: "Papyrus".

De taal. *)

De Endeneesche taal, in verschillende dialecten gesproken door ongeveer 120.000 zielen, vertoont vrij-veel overeenkomst met de Maleische, niet zoozeer door het aantal overeenkomende woorden – dit is betrekkelijk gering en bepaalt zich tot eenige tientallen hoogstens – als wel door den opbouw in het algemeen. Ze is echter veel minder ontwikkeld.

De in het Maleisch veel gebruikte woorden van vreemden oorsprong vindt men in het Endeneesch gedeeltelijk terug. Waar ze zich burgerrecht zijn gaan veroveren, is dit echter van zeer recenten datum, ze kwamen er waarschijnlijk via het Maleisch.

Wij noemen een paar: sabo-zeep, karata-papier, ada-gewoonte, habatijding, hoekoe-vonnis, roepa of hoepa-soort, basa-taal, parakara-zaak, pëtikist, kapa-schip, nona-gehuwde vrouw, teh-thee, biola-viool, tjepatoe-schoen, gëredja-kerk, kereta-rijtuig, boti-flesch, sorodadoe-soldaat, seresa-sergeant, letěna-luitenant, sakèbě-gezaghebber, běsětěri-bestuurs-assistent, doi-geld, gëla-glas. Vroeger waren dit, voor den bergbewoner althans, meest geheel vreemde begrippen. Ze deden hun intrede met kapala-hoofd, mandomandoer, goeroe-onderwijzer, djoeroebasa-tolk, soera-brief, gěra of gělakoper (van mal. gělang-armband)

De Endehsche taal is kort en bondig; zij gebruikt weinig woorden om veel te zeggen.

Het Endeneesch heeft dezelfde consonanten en vocalen als het Maleisch, vermeerderd met een palatale d, een gutturaal palatale ghr, de Nederlandsche f, eenige voorslagklanken bp, mb, ngg, nd, en de g (van ons goed) doch verminderd met de slot kâf σ of i en tj en nj, welke laatste oorspronkelijk niet in de taal voorkwamen.

Een Maleisch sprekende Endesche bergbewoner zegt "Nona soena soeroe soesi go", waarmee hij in het dagelijksch Maleisch bedoelt: "Njonja tjoema soeroeh tjoetji got".

Per onderscheiding van de g (als in garçon) hebben wij de weinige woorden, die de Hollandsche g-klank hebben aangeduid met ch (bijv. bacha, berispen). Ze zijn trouwens alleen in gebruik bij die stammen, die het langst geisoleerd zijn gebleven (Nageh, Roja), sterk komt dit uit bij de z.g. Rokkaneezen (Ngada), de tot voor kort wildste en meest ongenaakbare stam op Flores.

Figure 1: Description of Ende (from van Suchtelen 1921: 221).

^{•)} Voor de samenstelling van deze uiteraard zeer beknopte Endeneesche grammatica werd, ook om vergelijking te vergemakkelijken "het Overzicht der grammatica van de Maleische taal" van C. Spat op den voet gevolgd.

De klemtoon valt immer op de penalt, ook als deze de stomme \check{e} vraagt. De meeste woorden hebben maar twee lettergrepen.

Het Endeneesch heeft geen oorspronkelijk letterschrift. De strandbewoners van de hoofdplaats, het eiland Endeh, Baraai, Noemba en Nanga-Panda kennen, met eenige wijzigingen (zie tabel), het Boegineesche schrift. Dit vermag echter, evenmin als het Latijnsche alle Endesche klanken weer te geven.

In verband met hetgeen wij boven omtrent de Holl. g mededeelden, is het begrijpelijk dat het Boegineesche schrift dien klank niet kan weergeven. De Endeher zal daarom "bacha" schrijven Σ

Figure 2: Description of Ende (from van Suchtelen 1921: 222).

De toonlooze	schrijft men uitgesproken met de vocaien:					ODMEDVINCEN	
lettergrepen	a	e	i	0	oe	OPMERKINGEN.	
ě	~	~~~	~~	7.07	<i>∧</i> ָ́∧		
b	Y	M.	ż	자	Ķ		
t	^	~	~	~7	Ŷ		
dj	7	10	->	-57	Ŀ.		
tj	≍	~~~	××	₩1	××	vreemde oorsprong	
h	6	×6	6	67	Ģ	schijnt in het Boeg. in één	
d	Ÿ	^V	ŵ	rv	Ÿ	vorm voor te komen	
d	S	100	Ś	57	Ś	als onder <i>h</i> .	
r	*	* *	\$	ר א	Ŷ		
r	*	<i>~≈</i>	*	ר≈	<i>≈</i> ਼		
S	1	~	1	-1			
ghr	rg v	13	-3	-31	-is-	schijnt in het Boeg. niet voor	
ng	٢	え.	1	ᇧ	Ý	te komen	
f	و	2	ف	-91	و.	als onder ghr	
p	~	~~	~	~1	~	•	
k	1	14	1	1 ء	!		
g	ン	m	~	~~7	<u>ب</u> ب		
l ;	15	200	ふ	~~1	۲. ۲.		
m	v	~~	Ý	רי	Ŷ		
n	•	~~	ķ	١	Ŷ		
W	~	m	ŵ	~1	\sim		
h	. ^^	100	, , ,	۲.	ŝ		
j	%	1%	· ;;	r	×.		
nj	\approx	~~		\mathfrak{S}_1	\approx	vreemde oorsprong	
bp	Ś	2	3	י∿	Ś	als onder ghr	
mb	$\boldsymbol{\uparrow}$	~~	\uparrow	רי	Ϋ́		
ngg	~~	100	Ň	ריא	Ň		
nd	y	12	2	۲	. .	als onder ghr	

Figure 3: Description of Ende (from van Suchtelen 1921: 223).

Aangezien de Endehsche taal geen gesloten lettergrepen-systeem kent, kan het Boegineesche schrift, alleen uit open lettergrepen bertaande, uitstekend voor weergave van het Endehsch gebruikt worden.

Het Endehsche alphabet kent feitelijk geen letters, maar open lettergrepen, uitgesproken met de verschillende vocalen.

De stamvormen zijn die uitgesproken met de a.

Door toevoeging van diverse teekens verandert deze vocaal in de andere. De teekens voor de lettergrepen heeten *iroe nia ko lota* lett. neus gezicht van den lontar (op welk blad meestal het schrift gegrift wordt). De vocaal wijzigers hebben geen verzamelnaam, maar heeten voor de

> e — boko ∧ i — rateh • o — rolo 7 oe — rawa •

Een a lettergreep verandert achtereenvolgens in een e, i, o, oe lettergreep respectievelijk door

vóðrplaatsing	van	de	boko	~
bovenplaatsing	"	"	rateh	•
achterplaatsing	"	"	rolo	7
onderplaatsing	"	"	rawe	٠

Het schrift loopt van links naar rechts. Het kent geen afzonderlijke teekens voor de getallen.

De onvolkomenheid van het schrift komt het duidelijkste uit in de weergave der vocalen. De korte a (van jas) en de toonlooze \check{e} brengen geen veranderlng in de stamlettergreep. Wel neemt men om een lange aweer te geven het hulpmiddel van de h te pas.

a en h zijn trouwens, blijkens hun schrijfwijze resp. \mathcal{M} en \mathcal{M} , nauw verwant.

rechts,	wordt	geschreven	N MAA
raken,	"	"	NA
mand,	"	"	= M A
koud,	"	"	$= \Lambda$
zitten,	,,	"	1 11/1
uitslag,	"	19	102 .
kind,			MAMA
zand,	"	n	MA
	raken, mand, koud, zitten, uitslag, kind,	raken, " mand, " koud, " zitten, " uitslag, " kind, "	mand, "" koud, "" zitten, "" uitslag, "" kind, ""

Het verschil tusschen de i (van ik) en *ee* (van eer) kan niet aangegeven worden. In onze schrijfwijzen verwaarloosden wij daarom ook het verschil.

kěteh, uitgesproken met de i (van ik,) guur, wordt geschreven = ^^ kateh, jeuk, ", " = ^^^ Evenmin wordt verschil gemaakt tusschen de ee (van heet) en e (van met) mbéna, plat, wordt geschreven ^ ^ ^ ^<</p>

mbèteh (een eigennaam), wordt geschreven $/\gamma/\Lambda$

Figure 4: Description of Ende (from van Suchtelen 1921: 224).

Geen verschil bestaat in weergave van korte en lange oe, noch in die van de verschillende o-klanken.

Een zekeren graad van volmaaktheid lijkt ons daarentegen het Endehsche schrift te bezitten voor de weergave van de consonanten.

Een Europeesch oor ontgaat geheel of grootendeels het verschil tusschen

d,	v	en	d,	S
r,	\approx	"	r ,	R
b,	2	"	bp,	6

Het verschil tusschen h, φ en h \mathcal{M} laat zich nog het beste opmerken: h, φ is de duidelijk uitgesproken h bijv. in den uitroep Heh, $\wedge \varphi$, terwijl h \mathcal{M} meer den overgang tusschen twee lettergrepen vormt, de aangeademde h.

Maar ons ontgaat ten eenen male het in de h 's van

hi, φ alang en hia \dot{N} juichen hoekoe, $\varphi =$ treffen "hoe'oe \dot{N} \dot{N} poeder holo, φ \dot{N} hoofd "hongga \dot{N} \dot{N} scheren, evenzeer het verschil in de d's van

dadi, $\sim M \sim$, bevallen en daga $\vee M N$, pagaai dědi, $\sim \sim$, druppelen " děga $\vee N$, tikken, ook tusschen de r 's van

ratoe, 2 zijn en rawoe 2 MM inhoud

rach, \bigwedge nog niet " rangga $\gg M$ \bigwedge overtreden, of tusschen b en bp van

baba, vader en ba, 🗢 bord.

Zelfs is het verschil tusschen de ghr -3 en de g $\cancel{}$ dikwijls lastig te bepalen.

moga $\sqrt{7}$ \sim \sim \sim $\sqrt{7}$ $\sqrt{7}$

Het feit dat het End. schrift alleen open lettergrepen vermag weer te geven levert een groot bezwaar op tegen het gebruik er van voor andere talen bijv. het Maleisch.

De ontwikkelde Endeher heeft er echter wat op gevonden.

Hij trekt er dus nog een open lettergreep bij, bijv. toea wordt toeana, maar breekt hem met een "tanda mata", lett. teeken dood, af in den vorm van een kleine \wedge rechts boven.

Voor de \check{e} lettergreep \bigwedge zagen wij ook wel χ gebruiken. Dat schijnt echter ouderwetsch te zijn.

Figure 5: Description of Ende (from van Suchtelen 1921: 225).

.

.

Met lette op het verschil in schrijfwijze van de volgende veel op elkaar lijkende woorden:

maEh,	ziel, VMM niet, (mal. djangan) VMM
maweh,	wachten, of langzaam 🗸 🗥
nggaEh	zoeken, AV X edele, AV AAA geest, AV AAA
kai,	hij of zij, 🛲 😕
kahi,	openen, // M
	oudere, 🖉 🦯 🔨
kari,	hij, = 🛪
	, , , , , , , , , , , , , , , , , , ,
aeh,	water, 🔨 🎊
aj,	weet ik het, kan wel, min/".
hai,	voet, M M
haki,	echtgenoot, 🔊 🚽
toewa,	heer, $?$
toea,	schoonvader, 🔨 👭
	(Allah), $\uparrow \land \land$

Uit deze voorbeelden blijkt hoe 'n machtige hulp de kennis van het schrift kan zijn voor den Europeeschen beoefenaar van de Endehsche taal-

Als een aardige demonstratie, hoe het schrift zich wijzigt, voegen wij hierbij de copie van een tabel, een halve eeuw geleden opgemaakt door den Controleur S. Roos. (Men zie van zijn hand: lets over Endeh: T.B.G. XXIV.)

Figure 6: Description of Ende (from van Suchtelen 1921: 226).

13

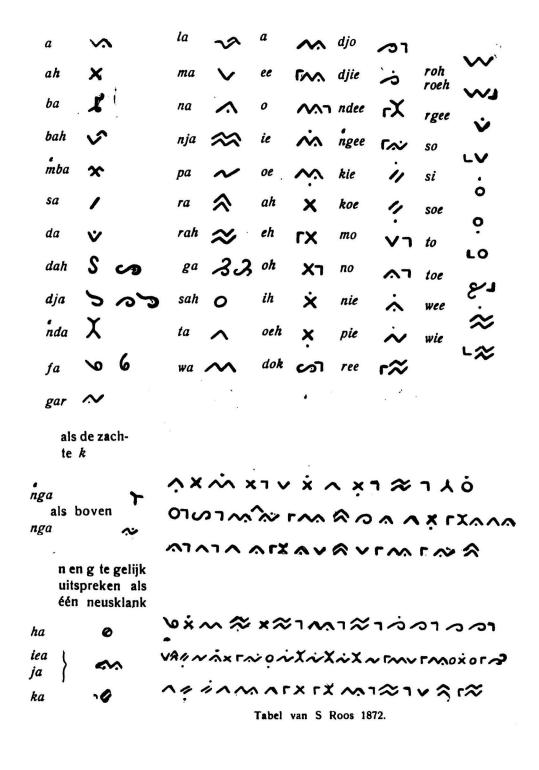


Figure 7: Description of Ende (from van Suchtelen 1921: 227).

No.	Latin Characters	Ende Characters	Bugis Characters	Remarks	No.	Latin Characters	Ende Characters	Bugis Characters	Remarks
1.	a	M	\sim	the same	16	mpa		ú	
2.	ba	7	5		17	na	4		
3.	bha	$\hat{\mathbf{v}}$			18	nda	≈		
4	ca		~5		19	nga	X	\succ	
5	da	Ý	U		20	ngga	Ň		
6	dha	S			21	ngka		5	
7	fa	V			22	nra		\approx	
8	ga	Ň	N		23	nya		\sim	
9	gha	3			24	nyca		$\tilde{\mathcal{X}}$	
10	ha	0	∞		25	pa	N	N	the same
11	ja	С	0		26	ra	~	~	the same
12	ka	11, =	11		27	rha	~		
13	la	N	N		28	sa	1,-	0	
14	ma	V	U	the same	29	ta	1		the same
15	mba	~~			30	wa	M	m	the same
-					31	ya		00	

The Comparison of the Ende Characters and Bugis Characters

Figure 8: Comparison of Ende and Buginese (from Banda 2014: 12).