TOTO ORTHOGRAPHY DEVELOPMENT

Toby Anderson[a] & Jonathan P Evans[b]

[a]SIL International and 3 Strands Consulting, [b]Institute of Linguistics, Academia Sinica,Taiwan
Corresponding author: jonathan@sinica.edu.tw

The Toto orthography, developed beginning in 2014, was created to address the inadequacies of using Bengali and Latin alphabets for writing Toto. These scripts failed to represent certain phonological features, such as front-rounded vowels and breathy vowels, posing challenges for accurate writing. The newly designed orthography, reflecting community input, unambiguously represents Toto phonology. Development of this writing system supports cultural identity, promotes literacy, and aids in educational efforts. Despite standardization, several challenges remain, such as transcribing elision, standardizing word breaks, long vowels, numbers and math symbols, punctuation, etc.

*Keywords:* Phonology, literacy, community, Dhimalish, Tibeto-Burman

1. Introduction

Toto (ISO 639-3: txo) is a Tibeto-Burman language spoken on the Indian side of the border with Bhutan. The entire language community of about 1700 speakers lives in one village, Totopara, in the Madarihat-Birpara CD block in the Alipurduar sub-division of the Alipurduar district of West Bengal (Figure 1). Toto is the smallest legally recognized language in India. Toto speakers live among many Nepali speakers, and most Toto speakers are at least bilingual. Public education is in Bengali, but school is about the only context in which Bengali is used. Toto speakers who have been to school are familiar with Bengali orthography but do not use it regularly, as there are few Bengali speakers in the environs.

Totopara is a highly isolated village. It can only be reached by one road, which is often impassable during the monsoon season, due to the flow of seasonal rivers. The high level of isolation may contribute to the vigorous status of the language, despite its small population and bilingualism in the socially dominant Nepali language. Ethnologue categorizes Toto at 6B "threatened" on the Expanded Graded Intergenerational Disruption Scale (EGIDS), but this might possibly be upgraded to 6A "vigorous". This corresponds to the fact that it is used by all generations in the community. Though the language itself is not in danger, it is undergoing rapid change as many native words are being replaced by Nepali equivalents.

Figure 1: *Totopara located within West Bengal*
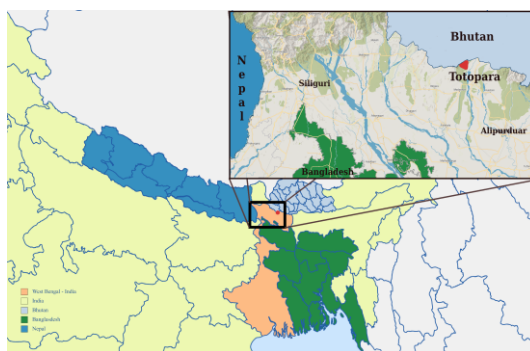
Toto is proposed to be a member of the Dhimalish group. Dhimalish is claimed to be either an independent group within Tibeto-Burman, or a sub-group within Central-Eastern Kiranti. Under either analysis, the Dhimalish group comprises three languages: Dhimal, Toto, and Lhokpu (Grollman & Gerber, 2018). There are two book-length treatments of Dhimal (King, 2009; Khatiwada, 2017). However, much less has been written about Toto (Perumalsamy, 2021) and Lhokpu (Gerber et al., 2016; Grollman & Gerber, 2018). Lhokpu is spoken nearby in Bhutan; Toto speakers are aware of it and say that Lhokpu is similar to their own language.

Members of the community desired a Toto orthography that could support mother-tongue education; e.g., at the Chitteranjan Toto Memorial Education Centre. It was also hoped that a writing system would facilitate communication in the digital age, as well as help to preserve Toto language and culture. The community expressed a preference for a unique set of letters, not based on any existing alphabet, reflecting their language's distinct identity. Orthographic development began in 2014 and the new alphabet was launched in the Toto community in May 2015. A font for the script was developed, and in July 2019 the new character set was accepted into the Unicode standard.

The following sections summarize relevant linguistic issues (sec. 2), deficiencies of pre-existing scripts (sec. 3), the method of script development (sec. 4), results thus far (sec. 5), and future prospects (sec. 6). 2.

2. Linguistic Issues

After conducting a standard phonemic analysis, Anderson (2014) found that Toto has the consonant and vowel phonemes presented in Tables 1 and 2.

Toto obstruents exhibit a two-way laryngeal distinction, described here as voiced vs voiceless. This structure contrasts with Nepali and Bengali which both contrast two levels of aspiration and two levels of voicing; e.g., /t, tʰ, d, dʰ/.

The Toto vowel system is marked by phonemic murmur or breathy distinction on all five front vowels. Origin of breathiness in Toto is still under investigation. The breathy voice contrast yields four front rounded vowel phonemes (Table 2). Front rounded vowels often surface as glide-vowel sequences in rapid speech, but are produced as front rounded monophthongs in careful speech: /jægø/ -> [jægue] ~ [jægø] 'hill'. Crosslinguistically, front rounded vowels are much less common than front unrounded vowels and back rounded vowels. The allophonic separation into the glide [w] and an unrounded front vowel conforms to typological expectations.

Breathy phonemes occur in geographic pockets across the Tibeto-Burman family. Breathy consonants occur in a swath across Bhutan and Nepal: Dzongkha, Dhimal, Kiranti, and Chepang. Breathy vowels are documented in the same area. In some cases, they occur as a breathy/modal tone contrast (Tamangic, Bodish, Magaric). Eastward of this area, breathy vowel phonemes occur in Northern Naga, Paang (Kuki-Chin), Lolo-Burmese, Karenic and Bai. Before the present analysis, Toto speakers perceived breathiness as a consonantal feature, due to their education in Bengali. Table 3 demonstrates the breathy/modal contrast in Toto vowels; the lack of initial consonant in /ø̤wɑ/ 'beating' confirms that breathiness is a vowel feature.

Table 1: *Toto consonant phonemes*

	bilabial	alveolar	palatal	(labio-) velar	glottal
nasal		m	n		ŋ
plosive	p b	t d		k g	
affricate		tʃ dʒ			
fricative		s			h
approximant		ɹ	j	w	
lateral approximant		l			

Table 2: *Toto vowel phonemes*

i	y	u		i̤	y̤
e	ø	o		e̤	ø̤
æ		ɑ		æ̤	

Table 3: *Toto breathy vowels*

æ vs æ̤	dæsiwɑ/	black	/dæ̤se/	jackfruit
e vs e̤	/tewɑ/	walking	/te̤wɑ	salty
i vs i̤	/piwɑ/	plucking	/pi̤wɑ/	weaving
ø vs ø̤	/øwɑ/	vomiting	/ø̤wɑ/	beating a person
y vs y̤	/ky̤t/	egg	/ky̤/	hand

Table 4: *Long vs short vowels*

/gɑwɑ/	sitting	/gɑːwɑ/	scattering
/owɑ/	peeling	/oːwɑ/	announcing

Table 5: *Contrast of toneless and tonal words*

/nowɑ/ thinking	/nŏwɑ/ coming out
/no:wɑ/ carrying on the head	/nŏ:wɑ/ measuring
/leŋwɑ/ braiding	/lĕŋwɑ/ laughing

Toto also contrasts vowel length (Table 4). Vowel length distinctions are often lost in connected speech but show up in careful speech. It is not yet clear if all vowels contrast in length.

Toto contrasts rising tone /ă/ with level (toneless) syllables, on both long and short vowels (Table 5). In connected speech, the tonal rise surfaces on the last syllable of its phrase, rather than where it is lexically specified.

A less common phenomenon is word-level nasalization, primarily on animal names: /kũwɑ̃/ 'tiger'. A similar phenomenon has been observed in Dhimal (King 2009).

Breathiness, vowel length, tone, and nasalization entail decisions that remain to be finalized in the standardization of Toto orthography.

In addition to Totos, the village of Totopara is home to about 2000 Nepali speakers. Thus, Nepali is the main L2 in the community, although most speakers are not literate in Nepali. However, members of the Baptist church, which uses the Nepali Bible, are often literate in Nepali. The prevalence of Nepali is reflected in the large number of Nepali loanwords within Toto.

Bengali is the language of education in government schools, resulting in a degree of fluency and literacy in Bengali. In fact, literacy is historically associated with the Bengali script; early attempts at writing Toto used Bengali script. However, as Toto children are exposed to Bengali solely within the formal educational context, most do not attain substantial fluency or literacy in Bengali. Despite this limited proficiency, familiarity with the Bengali script exerts a notable influence on the Totos' metalinguistic awareness, as noted above for breathiness.

Besides Nepali and Bengali, some Toto community members are familiar with other languages as well. English is mainly known by prominent Toto families and teachers in private schools. Some community members understand Dzongkha, the national language of Bhutan. Before Covid-related policies closed the international border in 2020, many Totos worked in Bhutan as day laborers; hence their familiarity with Dzongkha. Additionally, exposure to television and social media has given some Totos a basic understanding of Hindi.

Despite this level of multilingualism, many Toto people, especially women and children, are monolingual in Toto. On the other hand, there are some Totos who do not speak their language at all, which is an issue of concern within the community.

3. Deficiencies of pre-existing scripts

Before the development of the autochthonous Toto script, the writing of Toto was performed using non-standardized Bengali or Latin orthography. Currently, Bengali letters are still used most often, and Latin letters are often used in digital communication, such as texting.

Toto speakers encounter several challenges when using exogenous orthographies to transcribe their language. As shown in Table 1, Toto has an alveolar series of consonants /n, t, d, s, ɹ, l/. Bengali has dental and retroflex places of articulation but does not mark an alveolar series. This difference is salient to the Totos, who find it unclear which Bengali series to use when transcribing alveolars. When using the Latin alphabet, they write {t, d, s, n, r, l}.

Toto has the front rounded vowels /y, ø, ǿ/. Neither Bengali nor English have front-rounded vowels. When using Bengali orthography, Toto speakers represent front rounded vowels with a symbol for /u, o/, followed by a symbol for the front unrounded vowel of the same height, either /i, e/. When using Latin orthography, front rounded vowels are written {ui}, {oe}: /mømbø *moemboeh* 'millet'.

The Bengali alphabet does not distinguish between breathy and modal vowels. However, it does differentiate between aspirated and

unaspirated consonants. Consequently, when writing syllables with breathy vowels, a Toto writer often defaults to using the letter for an aspirated consonant from Bengali. This approach introduces complexities when the breathy vowel syllable either lacks an onset consonant (/ɵwɑ/ 'beating') or has an onset that cannot be aspirated in Bengali (/nɛwɑ/ 'licking clean'). When writing in Latin script, {h} can be written after a breathy vowel, which causes fewer ambiguities than writing it before the vowel, where it could be perceived as part of an onset: /jɵ yeoh/ 'grass'.

Vowel length (Table 4) and tone are not standardized in either Bengali or Latin orthographies. Nasalized vowels are marked in Bengali script: {এ, অ্যা} /e, æ/ vs. {এঁ, অ্যাঁ} /ẽ, æ̃/, but not in Latin orthography. The deficiencies in marking vowel length, tone, and nasalization are not as serious as those discussed above, because these distinctions are marginal. In the case of tone, speakers are often unaware of the distinction. Table 6 summarizes the ways that Toto sounds are written in these two orthographies, as well as in the new Toto script. Many sounds have multiple ways they can be transcribed in Bengali script.

Table 6 shows that Toto orthography distinguishes two laryngeal settings among obstruents, phonologized as voiced and voiceless. Each front vowel has its symbol, and a consistent diacritic mark under the vowel indicates breathy voicing. Thus, the system addresses the main problems with the Bengali and Latin systems. At present, as in the exogenous orthographies, there are no markings for tone, vowel length, and nasality.

Table 6: *Methods of transcribing Toto sounds*

IPA	Toto	Latin	Bengali
m	Π	m	ম
n	ᴔ	n	ন ণ
ŋ	ᴓ	ng	ঙ ০ং
p	Π	p	প
b	ᴩ	b	ব
t	ᴕ	t	ত ট

d	ᴚ	d	দ ড
k	ᴘ	k	ক
g	ᴝ	g	গ
tʃ	◁	ch	চ
dʒ	ᴲ	j	জ
s	ᴧ	s	শ ষ স
h	ᴒ	h	হ
ɹ	<	r	র ড়
j	▽	y	য়
w	ᴕ	w	ওয়
l	ᴖ	l	ল
i	ᴣ	i	ই ঈ
e	ᴆ	e	এ
æ	ᴴ	ae	এ "আা
y	⊙	ui	ঐ উই
ø	ᴐ	oe	ঔ ঐ ওই
u	ᴗ	u	★ উ
o	ᴪ	o	ও
ɑ	ᴧ	a	আ
i̥	ᴣ	ih	ইহ ঈহ
e̥	ᴒ	eh	এহ
æ̥	ᴴ	aeh	এহ "যাহ
y̥	⊙	uih	ঐহ উইহ
ø̥	ᴐ	oeh	ঔহ ঐহ ওইহ

4. Origin and present state of the orthography

Ideally, an orthographic system should reflect language structure while being user-friendly. However, balancing these goals can be tricky. For example, one key decision is how closely the writing system should mirror Toto pronunciation (orthographic depth). A simple option is a "phonemic" orthography, where each sound has one written symbol. This can be helpful for learners of the writing system because it indicates salient sound differences (Seifart 2006: 283). Alternatively, a "morphemic" orthography prioritizes consistent spelling for morphemes, even if pronunciation varies, as in the orthographic stability of English signs and signatures. Morphemic consistency can be easier for fluent Toto readers, and also provides transcriptional stability across dialect differences.

Another weighty decision is the extent to which a

new orthography resembles writing systems in the larger social context. In some cases, speakers want their writing system to serve as a bridge for their children to learn English or other languages of wider communication; e.g., Tawrã (Evans & Manyu 2021) In other instances, speakers want their orthography to reinforce their distinct social and linguistic identities (Cahill 2018).

The UK-based non-profit organization, Bright Eyes Global Action Project (BEGAP), identified a significant need for mother-tongue education within the Toto community. In response, they established the Chitteranjan Toto Memorial Education Centre in 2014. Seeking to integrate a mother-tongue curriculum into the school, BEGAP director Lissa Davies requested assistance from SIL International to develop a Toto orthography. SIL members Toby and Mary-Grace Anderson then came to reside in the community and they noted a strong desire among the Toto people for a standardized writing system. They recognized that such an orthography could benefit the community in terms of language preservation, cultural continuity, and as an on-ramp to learning. Developing an official script would elevate the language's status on the Expanded Graded Intergenerational Disruption Scale (EGIDS), and contribute to the long-term vitality of the Toto language and culture. For children in school, learning literacy in a language they already know facilitates comprehension and engagement with the educational material.

To evaluate the most suitable script for the new orthography, the community was consulted about the two main existing options: the Latin and Bengali scripts. It was found that the Toto people preferred an entirely novel set of letters rather than adopting an existing alphabet. After the Toto phonemic categories were established, Dhaniram Toto was approached to create a unique letter for each phoneme. To achieve this, Dhaniram based the designs on everyday objects and concepts that start with the corresponding phonemes, making the letters resemble these items. For instance, the Toto character for /d/ (ठ) is designed to evoke a jackfruit ठप्ऱठ /dæse/ hanging from a branch. The new alphabet debuted in May 2015, and since then, there have been initiatives to promote

literacy in the Toto language using this orthography. A font (fonts.google.com) and keyboarding software (keyman) were created in 2015, and by July 2019, the characters were incorporated into the Unicode 14.0 standard. Presently about 100 children are enrolled in mother-tongue education at the Chitteranjan Toto Memorial Education Centre. This is a multilingual school, in which instruction in most classes occurs in English or Nepali. There is one class per week for Toto literacy. Despite the script being in existence less than ten years, about 10% of the children now have some exposure to Toto script literacy lessons (Village Square, 2024).

The Baptist church in the community uses a Nepali version of the Bible, but in recent years has been translating it into Toto. The New Testament is almost ready for publication, and it has been decided to publish it as a di-script text, with the Toto words written in both the Bengali script and the newly developed Toto script, side-by-side on each page. It is hoped that this will facilitate church members' learning of the new script.

Appendix A demonstrates the orthography and the phonemic analysis of Toto. For the linguist, the chart demonstrates the phonemic analysis behind the new orthography. For Toto language educators and learners, it exemplifies the range of combinations of consonants and vowels.

The chart demonstrates the challenge of writing long vowels. This is clear in the last row, where two adjacent identical vowels are written in the same way as one long vowel: {ΛΛ\'oQ\'ə<QꝒQΛ} /ɑɑŋpetʃpuwɑ/ 'getting a shock', vs. {ꝒꝒQΛ} /ø̞:wɑ/ 'beating someone'. Presently there is no proposed strategy for distinguishing repeated vs long vowels. Further community testing is needed to determine whether resolving this ambiguity forms a perceived need by the users.

We exemplify the orthography with a text collected from Mercy Toto, as she describes her feelings about the new Toto orthography (1).

(1) ȣΛ<Λóôȝ ꝒΛ<9 ꝒΛꝒꝒ9 ⊲ꝺꝒð ȝ QΛQΛ ꝒΛꝒꝒ9 ⊽ꝶȝȝ �Λ QΛꝒ9 ⊲ꝺQΛ ꝒðõΛ ꝒΛꝒꝒ9⅄99Λ 99Ꝓȝ ȣꞪꝒ9 ȝ⊲9 ⊲ꝺꝒð ठ9⅄<Λ ठȝóΛꝒȝꝒ9 QꞪꝒ99Λ

ᴣʌᓬᎧᕒᎰᐊᓴᎧᕽʌ ᕒᐁᎧᕽʌ ᕒʌᕒᴣᎥᒡᕽᎧᕽ ᐁᕒᴣᕃᎧᕽʌ ᒐᕒᐸᴣᎧᕽ
ᕒʌᎧᕽᎰ ᎰᎰᕒᴣᕽᕒᴴᎰʌ

/gaɹani taɹo kabko tʃube i lawa kabko jøði ha
lako tʃuwabeɹa kabkosota totbi dʒæko itʃo tʃube
dosɹa diŋabiko pækota dʒaŋkotʃuwa bepa
kabihiŋ ebdʒuwa skɹipt kapko totbikæta/

Hello, let me say that this new way of writing our
language enables the Toto tribe to express
ourselves to the people around us with the Toto
language in this script.

ᕒʌᎧᕽʌᎱᕽᴣᴣʌᕒᎰᒐᕽᎰᎰʌ ᒍʌᎧᕽʌᎰʌᎧᕽʌ ᒍʌ ᕒʌ
ᕒʌᕒᴣᒡʌ ᴣᎧᕒᎰ ᐊᎱᎧᕽʌ ᒭᎧᕒᐊᕒᐸᐸᎧᕽᴣ ʌᴣᎰᐸᕒ ᕒᎧᕒᎧʌ
ᕒʌᎧᕽᎰ ᴣᎧᕽᴣᎰʌ Ꮁʌᒭᒐʌᐸᐸ ᎰᴣᎧᕽʌᎰ ᐊᎱᐸᎰ ᕒʌᕒᕒᎰʌ
ᎰᕒᎰ Ꮁʌᴣᴣ ᕒʌ ᴣʌᎧᕒʌᎧᕽʌ

/banajpijakosota haŋpatana ha ka kabiha jøko
tʃuwa meŋtʃeɹaŋga adoɹe neha kapko jøðiha
lamsaɹe tiŋpako tʃɹɹo kabeja oko lagi ka
jaŋtana/

Only now that this has been made, we are able to
present our language in writing. For this I am
glad.

ʌᒐʌ ᎰᎧʌᒐʌᐸᐸᎰ ᎧʌᎰ ᴣ ᴣᒭʌʌ ᕒʌᕒᕒᎰ ᴣʌᎱʌ ᎰᎧᎰᎧᎱ
ᴣᎱᎧᕽʌᎰ ᐊᎱᕒᎰᕒᎰᎧʌ ᐁᕒᴣᎧᕽʌ ᕒʌᕒᕒᎰ ᴣᎧᒡᴣᎧʌ
ᎰᎧᎰᎧᎱ ᴣᎱᎧʌ ᎱᎱᐸᕒᕒᎰᒐᔣᎰᎰʌ

/asa nasaɹaŋ yse dʒimsa kabko dʒaja naŋaj
gøpako tʃukobewa ebdʒuwa kabko jøðiha naŋaj
gøpa puɹakosota/

Furthermore, after much learning of this
[alphabet], we will be able to advance our culture
and language and move them forward.

5. Next steps

Up to this point, the focus of orthographic
development has been on isolated lexical items.
However, some decisions remain to be made, for
the orthography to become fully functional, and
for speakers to have a shared perspective on how
to use it.

Phonemically, linguistic stakeholders should come
to an agreement about the representation (or not)
of vowel length, tone, and nasality. At this point,
these features seem to be sufficiently marginal
that there is no need to devise orthographic
representations for them.

During connected speech, there is substantial
phonetic reduction; e.g., of inflectional
morphology. However, under careful speech, the
component morphemes are recoverable. As part of
literacy training, speakers will need to become
aware of the reduction process and how to undo it
for spelling consistency.

An issue that is more systematically challenging is
that word breaks remain to be standardized. The
Toto language is morphologically rich, and
deciding what gets written together as one word,
where the boundaries are, etc. will require shared
writing experience, as well as time and discussion.
As the usage of the Toto script expands, speakers
will also need to resolve whether to use Arabic or
Bengali numerals, punctuation, etc., or whether
they will create their own.

6. Conclusion

The development of a community-centered
orthography for the Toto language, initiated in
2014, represents a significant step toward
preserving the language, supporting mother-
tongue based education, and enhancing cultural
identity among its speakers. Unlike previous
orthographic strategies (Bengali, Latin), the new
script accurately represents important
phonological features, such as front-rounded
vowels and breathy vowel phonation. Planned
further developments include standardizing word
breaks and addressing phonetic reductions. The
successful implementation of this writing system
is expected to have a lasting impact on the
preservation of the Toto language, fostering a
sense of pride and continuity among its speakers,
and ultimately contributing to the language's
vitality in the face of external pressures.

7. Supplementary materials

Font resource: https://fonts.google.com/noto/
specimen/Noto+Serif+Toto
Keyboarding resource: https://keyman.com/
keyboards/txo_toto

References

Anderson, T. (2014). *Toto language fieldnotes* [Unpublished manuscript].

Cahill, M. (2018). Orthography design and implementation for endangered languages. In K. L. Rehg & L. Campbell (Eds.), *The Oxford handbook of endangered languages* (pp. 327–346). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780190610029.013.17

Evans, J. P., & Manyu, J. (2021). The sounds of Tawrã, a Tibeto-Burman language. *Linguistics of the Tibeto-Burman Area, 44*(1), 1–26. https://doi.org/10.1146/noto_serif_toto

Gerber, P., Gerber, T., & Grollmann, S. (2016). Links between Lhokpu and Kiranti: Some observations. Academia.edu. https://www.academia.edu/30259338/

Grollmann, S., & Gerber, P. (2018). Linguistic evidence for a closer relationship between Lhokpu and Dhimal: Including some remarks on the Dhimalish subgroup. *Cahiers de Linguistique Asie Orientale, 47*(1), 1–96. https://doi.org/10.1163/19606028-04701004

Khatiwada, K. (2017). *A reference grammar of Dhimal.* Lincom Europa.

King, J.T. (2009). *A grammar of Dhimal.* Brill

Perumalsamy, P. (2021). *A linguistic description of Toto language spoken in West Bengal (Vol. 1, Pt. 1).* Office of the Registrar General, India, Language Division.

Seifart, F. (2006). Orthography development. In J. Gippert, N. P. Himmelmann, & U. Mosel (Eds.), *Essentials of language documentation* (pp. 275–300). De Gruyter Mouton. https://doi.org/10.1515/9783110197730.275

Village Square. (2024). Learning the ABC of Toto language. Village Square. https://www.villagesquare.in/learning-the-abc-of-toto-language/

Appendix A: *Toto initial consonant and vowel combinations*

Modal vowels:

	ᶾ i	☉ y	℧ u	ð e	Ɔ ø	9 o	H æ
p	pika 'cow'	pyty 'honey'	pusæ 'dandruff'	peta 'insect'	pøŋ 'helper'	po:wa 'beat stick'	pæto 'tadpole'
b	biŋa 'wind'	byta 'travel bundle'	buŋba 'termite'	bekoŋ 'skin'	bødu 'navel'		bæro 'friend'
t	ti 'water'	tyja 'house fly'	tuŋsiŋ 'heart; chest'	tewa 'walking'	tøti 'ground container'	toto 'Toto person'	tætʃo 'ten'
d	ditʃo 'four'		duwa 'elder brother'	deŋti 'cold water'	dø 'stick'	dowa 'finding'	dæsiwa 'black'
k	kibi 'us'	kyty 'egg'	kupu 'thumb'	keka 'chicken'	kø:wa 'scraping'	kokø 'leg'	kæ 'language'
g	gibe 'deer'	gywa 'hide under blanket'	guwaj 'areca nut'	gewa 'knowing'	gø 'tent'	goŋto 'back of neck'	gædoŋ 'pass entrance'
m	miŋwa 'ripening'	myry 'firefly'	musa 'father'	meŋta 'spot'	mø:jwa 'inviting'	moŋmoŋwa 'warm'	mæŋwa 'searching by touch'
n	ni 'salt'	ny 'lap'	nuwa 'bow down'	netoŋ 'neck'	nøbø 'nose'	noko '2.sg.poss'	nætæ 'road'
ŋ				siŋe 'tree'		ŋoka 'monkey'	
s	sikæ 'wild chicken'	syma 'bee'	suŋtʃo 'three'	setaŋ 'tooth'	søja 'wasp'	sowa 'exchange'	sæti 'back basket'
tʃ	tʃiduŋ 'bag'	tʃyti 'oil'	tʃube 'new'	tʃepa 'sweat'	tʃøwa 'buying'	tʃopoj 'bowl'	tʃæbe 'she animal'
j	jiwa 'shaking out water'	jywa 'speaking'	ju 'millet beer'	jeme 'younger sister'	jødi 'word'	jo:wa 'picking up'	jæwa 'destroying'
w	wiwa 'asking'			wejwa 'opening (v)'		woŋta 'stone hand mill'	wæwa 'breaking food with hands'
dʒ	dʒi 'leopard'	dʒywa 'pushing'	dʒupa 'body'	dʒebe 'old (thing)'	dʒøwa 'dropping from	dʒowa 'doing'	dʒæ 'clan/tribe'

	꒳ i	⊙ y	Ʊ u	ð e	Ɔ ø	9 o	Ħ æ
					trees'		
ꗷ **h**	hini 'today'	hyti 'milk'	huwa 'following'	hepa ga 'how?'	hø:wa 'untying'	howa 'going (NW)'	hæpkuŋ 'morning'
< **ɹ**	muɹi 'chilli'		-ɹuŋ 'clf:rock'	-ɹeŋ 'clf:fruit.'	køɹø 'finger'	-ɹo '-fut'	
Ꝗ **l**	limu 'grave'	lyty 'bat'	lu:wa 'undressing'	lebe 'tongue'	løwa 'climbing'	lo 'get up'	læwa 'narrating'
ø **ø**	i 'this; one'	yse 'many'	udaŋ 'tall; long'	eŋ 'ginger'	ø:wa 'vomiting'	o:wa 'digging'	ælywa 'red'

Breathy vowels (and /ɑ/):

	ʌ a	꒳ i	⊙ y	ð ẹ	Ɔ ø	Ħ æ
p	pataŋ, machete	piwa, weaving	pyia, snake	peṭa, pot stand	pøpø̥wa, ready for seed	
b	bakuŋ, drum	biwa, blowing	bywa, gradually grow	høŋbe, small hole	mømø, millet	
t	tatawa, good tasting	tise, necklace	tỹwa, laying egg	tewa, strong tasting	tøwa, hanging	tæwa, slippery
d	daŋ, horn	dika, buffalo	dy, bamboo verandah	depawa, coming		dæse, jackfruit
k	kaɹø, finger	kija, dog	ky, hand	kewa, scratching the ground	køwa, scraping skin clean	kæwa, sound
g	gawa, sitting	gimwa, cooling down	gywa, work (v)	gebe, branch	gøwa, wanting	
m	mata, leader	miṭʃo, eye	mywa, behave as if dying		incur small injury	mæwa, grinding; smashing
óo	óoʌ93	óo꒳…		óo…		óoħ…

	ʌ a	ʒ i	⊙ y	ə e̞	ɔ ø	ꭢ æ
n	nɑti you	n̩iso two		licking bowl clean		næɡu bed
ó ŋ	dʒaŋa mosquito					
ʎ s	sa house	si̠ke branching	sysywa whistling	se̞wa bearing fruit		sæwa killing
⊲ tʃ	tʃawa eating	tʃi̠wa tearing	tʃyti dew	tʃe̞wa one swing cut	tʃøwa plowing	
▽ j	japuwa tired			mixing	jø grass	jæwa sour
 w	waŋwa coming	wi̠wa milling				wæti rain
 dʒ	dʒaŋsaj red weaver ant	dʒi̠ja bird	dʒyj a mouse	dʒe̞be̞ŋ carp	combing (hair)	dʒæwa animal screaming
∩ h	hahaŋwa yawning		falling down	dragging	unscrewing	hæwa spitting
< ɹ	-ɹa clf:leaf; clf:money					
 l	lapa leaf	li̠wa pull down	fighting	singing	being well fed	læ needle
ø	aaŋpetʃpuwa Getting a shock	i̠wa pushing over	ya getting down	e̞w a putting	øw a beating someone	æwa decaying; spoiling