

Supporting Tulu language written in the Kannada script

Shriramana Sharma, jamadagni-at-gmail-dot-com, India

2012-Jun-07

§1. Introduction

In addition to the regular vowel repertoire of other South Indian languages – a, ā, i, ī, u, ū, e, ē, ai, o, ō, au (and the Sanskrit-based vocalic ɾ etc) – the Tulu language has four additional vowels. In the Tulu Lexicon, these are represented in romanization as ʊ, ʊʊ, ɛ, ɛɛ. In this document, however, we shall designate the long vowels as ū and ē after the ISO pattern.

In the Kannada orthography for Tulu adopted by the Tulu Lexicon, these additional vowels are written as follows:

	ʊ	ū	ɛ	ē
Independent vowels :	ಆ್	ಆ್ಲ	ವ್	ವ್ಲ
Dependent vowels :	್	್ಲ	್	್ಲ

Scans from the Tulu Lexicon (courtesy Vaishnavi Murthy) follow as attestation.

List of Tulu Independent Vowels:

ಆ್	ಆ್ಲ	ಅ	ಆ	ಇ	ಈ	ಉ	ಊ	ಋ	ೠ
ʊ	ʊʊ	a	aa	i	ii	u	uu	ɾ	ɽ
ವ್	ವ್ಲ	ಎ	ಏ	ಐ	ಒ	ಓ	ಔ		
ɛ	ɛɛ	e	ee	ai	o	oo	ov		

Special vowels in Tulu:

		Word initial position	After a consonant
Centralised back vowel }	ʊ	ಆ್	್
	ʊʊ	ಆ್ಲ	್ಲ
Lower mid front vowel }	ɛ	ವ್	್
	ɛɛ	ವ್ಲ	್ಲ

್
್ಲ

stands for ʊ after consonants
stands for ɛ after consonants

Given this existing attested practice, this document will discuss how it may be implemented in Unicode and what problems are associated therewith.

§2. Independent Vowels

As far as the independent vowels are concerned, there is not much problem. The existing independent vowels 0C85 KANNADA LETTER A ಅ and 0C8E KANNADA LETTER E ಎ may be combined with 0CCD KANNADA SIGN VIRAMA ಿ and 0CD5 KANNADA LENGTH MARK ಿಲ appropriately to achieve the desired effect as follows:

0C85 ಅ	+	0CCD ಿ	=	ಅಿ		
0C85 ಅ	+	0CCD ಿ	+	0CD5 ಿಲ	=	ಅಿಲ
0C8E ಎ	+	0CCD ಿ	=	ಎಿ		
0C8E ಎ	+	0CCD ಿ	+	0CD5 ಿಲ	=	ಎಿಲ

The only issues involved are that rendering engines (especially OpenType-based ones) should recognize these sequences as valid and not spew out dotted circles, and that fonts (i.e. Kannada fonts desiring to cater to Tulu language writing) should provide the appropriate glyphs and glyph substitutions.

§3. Dependent Vowels

As for the dependent vowels, supporting ಂ and ಃ is also not a serious issue. The same characters 0CCD KANNADA SIGN VIRAMA ಿ and 0CD5 KANNADA LENGTH MARK ಿಲ can be appropriately combined with the use of 0CC6 KANNADA VOWEL SIGN E ಿ as follows:

0C95 ಳ	+	0CC6 ಿ	+	0CCD ಿ	=	ಳಿ		
0C95 ಳ	+	0CC6 ಿ	+	0CCD ಿ	+	0CD5 ಿಲ	=	ಳಿಲ

Again, it is merely a matter of the rendering engine accepting the sequences and the font providing appropriate glyph substitutions.

However, the matter of the vowels ಣ and ಣ is quite different. What we see in the attestation is that they have re-used the Kannada script virama sign as a vowel marker to denote this sound. While this requires almost no additional glyph/engine design, and the appropriate sequences might seem obvious:

0C95 ಕ + 0CCD ಿ = ಕಿ
 0C95 ಕ + 0CCD ಿ + 0CD5 ಿ = ಕೀ

... the implementation is not straightforward in the case of the short vowel. While the long vowel can be represented with no issues, representing the short vowel by the 0CCD KANNADA SIGN VIRAMA ಿ alone will cause problems that are clear when one examines entire words (rather than isolate sequences). Observe some examples from the Lexicon:

Example:

ಅ	ಅಂದ	ಅಜ್ಜ	ಅ	ಅಂದ	ಅಃ
u	unda	ujju	a	and <u>u</u>	aḥ
ಅಕಡ್	ಅನ್ನ	ಅಜ	ಅಜಂ	ಅಜಕಲ್	
ak <u>u</u> ḍu	a <u>n</u> ṇa	aja	ajam	ajak <u>u</u> ḷu	
ಅಜಕಾಯಿ	ಅನ್ನ	ಅಳ್	ಅಲ್		
ajakaayi	a <u>n</u> ṇa	aḷu	aḷu		

The examples circled in red have the special vowel u in word medial position: akuḍu and ajakuḷu. How are these Tulu words written in the Kannada script to be encoded? It is not possible to simply use just 0CCD KANNADA SIGN VIRAMA ಿ as follows:

0C85 ಅ + 0C95 ಕ + 0CCD ಿ + 0CA1 ಡ + 0CCD ಿ ?= ಅಕಡ್

... because as is the standard with Indic scripts, the sequence 0C95 0CCD 0CA1 being of the pattern CONSONANT + VIRAMA + CONSONANT will produce combining behaviour:

0C85 ಅ + 0C95 ಕ + 0CCD ಿ + 0CA1 ಡ + 0CCD ಿ = ಅಕಡ್

... which is obviously *not* what is desired. It is also not possible to suggest that a simplified Kannada font *without* the required combining forms should do for Tulu, because language content should be represented by encoded text and not by font changes and because Tulu itself uses consonant clusters which are represented by the regular stacking behaviour of the Kannada script, as seen in the examples above circled in blue.

Therefore one is faced with a dilemma as to what to do.

§4. Possible solutions to the problem with dependent vowel ʉ

§4.1. Invisible spacing characters

While inserting 200C ZERO WIDTH NON-JOINER to prevent the combining behaviour would achieve the desired *appearance*:

0C85 ಅ + 0C95 ಕ + 0CCD ಿ̣ + 200C | + 0CA1 ಡ + 0CCD ಿ̣ = ಅಕ್ಡ

... this is an insufficient solution because this character is default-ignorable in text processes such as searching and collation, resulting in minimal pairs (such as if there existed a word akḍ* which differs from akḍ only by a medial ʉ) being mutually indistinguishable. This would go against established collation orders for Tulu orthography in Kannada script and is hence unacceptable.

2060 WORD JOINER (which was a choice as it does not cause a line-break opportunity unlike the other invisible space characters) in lieu of 200C ZWNJ was considered but is inappropriate because, as per TUS 6.1 p 546 (p 576 of PDF):

... inserting a word joiner between two characters has no effect on their ligating and cursive joining behavior. The word joiner should be ignored in contexts other than word or line breaking.

... it seems it cannot prevent the combining behaviour of the CONSONANT + VIRAMA + CONSONANT combination (although it is unclear whether the above statement of TUS applies to Indic contexts too), and furthermore it also appears to be default ignorable (as it has GC=Cf; vide http://www.unicode.org/reports/tr44/#Default_Ignorable_Code_Point) which again makes it no different than ZWNJ.

§4.2. Alternative written representation for ʉ based on u

It is clear* that the ʉ sound of Tulu is morphologically related to the regular u and is in fact cognate to the “kurriyalukaram” of Tamil or the “saṃvṛtōkāram” of Malayalam (except for the fact that in Tulu it is phonemic rather than just an allophone of /u/).

* “Old Kannaḍa has the inherited ten vowels /i e a o u ī ē ā ō ū/. ... Tuḷu has the same core system but it has added ε (front low unrounded vowel, historically from -ay word-finally) and ̄i (high central unrounded) (Bhat 1998), which

Given this, in retrospect it might have been better if the Tulu Lexicon editors (who developed the Kannada script orthography for Tulu) had paralleled the writing of the same sound in old Malayalam (and in pedagogical old Tamil) and defined ು to be written in the Kannada script as VOWEL SIGN U ು + VIRAMA ು just like in Malayalam the “saṃvṛtōkāram” is written as VOWEL SIGN U ು + VIRAMA ು. [See TUS 6.1 p 319.]

This would mirror the other special vowel ε which is written as VOWEL SIGN E ು + VIRAMA ು. Further, the long vowel ು could also be written simply by adding the length mark to follow the existing pattern. Observe:

$$\begin{array}{rclclcl}
 0C95 \text{ ು} & + & 0CC1 \text{ ು} & + & 0CCD \text{ ು} & = & \text{ಠ} \\
 0C95 \text{ ು} & + & 0CC1 \text{ ು} & + & 0CCD \text{ ು} & + & 0CD5 \text{ ು} & = & \text{ಠ}
 \end{array}$$

(On a typographic note, it seems that here the vowel sign U ು should probably take its allograph ು [which it takes with PA and VA] to avoid overlapping with the virama sign.)

It is obvious that if the dependent ು were to thus be written using VOWEL SIGN U ು + VIRAMA ು, the problem caused by using the virama alone for ು would be avoided, as now the sequence CONSONANT + VIRAMA + CONSONANT will not be used except when there is actually a consonant cluster in the underlying language content.

Of course, in this case, the independent vowel also would be changed accordingly:

$$\begin{array}{rclclcl}
 0C89 \text{ ು} & + & 0CCD \text{ ು} & = & \text{ಠ} \\
 0C89 \text{ ು} & + & 0CCD \text{ ು} & + & 0CD5 \text{ ು} & = & \text{ಠ}
 \end{array}$$

Now while this is an entirely viable option technically and linguistically, it requires buy-in from the Tulu scholars and user community, which would of course be based on the degree of their need of representing their language in Kannada Unicode.

mainly occur finally. i and u result from a split of older /u/ and i corresponds to the enunciative vowel of the other Southern languages. The long counterparts of i and ε are extremely restricted.” – The Dravidian Languages, Bhadriraju Krishnamurthy, Cambridge University Press, Cambridge, 1st South Asian Edition, 2003, ISBN: 978-0-521-77111-5, p 52. On the “enunciative vowel”, ibid p 49: “A short /u/ following a word-final stop (in Modern Tamil any consonant) is phonetically a back unrounded vowel [i] which was called the ‘enunciative vowel’”.

§4.3. Alternative written representation for all special vowels

All the above methods involve devising new Kannada vowel signs for Tulu. This imposes much work on designers of fonts and rendering software since requisite glyphs need to be constructed and the above recommended sequences need to be programmed into the script grammar. It is unlikely that this would happen soon at a large scale and would prove a serious stumbling block for those desiring to write Tulu immediately in Kannada Unicode.

A simple solution would be to use an existing simple character as a vowel modifier. I considered 02BC MODIFIER LETTER APOSTROPHE as it is already used in Devanagari for Bodo etc, but the Tulu native users objected to this as it would be confusable with quotation marks. A viable alternate is using 02C2 MODIFIER LETTER LEFT ARROWHEAD and doubling it for length:

Independent Vowels:

u	0C85 ಉ	+	02C2	◀	=	ಉ◀
ū	0C85 ಉ	+	02C2 02C2	◀◀	=	ಉ◀◀
ε	0C8E ಎ	+	02C2	◀	=	ಎ◀
ē	0C8E ಎ	+	02C2 02C2	◀◀	=	ಎ◀◀

Dependent Vowels:

ku	0C95 ಕ	+	02C2	◀	=	ಕ◀			
kū	0C95 ಕ	+	02C2 02C2	◀◀	=	ಕ◀◀			
kε	0C95 ಕ	+	0CC6 ಿ	◌	+	02C2	◀	=	ಕ◌◀
kē	0C95 ಕ	+	0CC6 ಿ	◌	+	02C2 02C2	◀◀	=	ಕ◌◀◀

Again, despite the simplicity of this model (all that is required would be a font on one's system providing this character and a simple method to input it), it is unknown as to what extent the Tulu scholars and user community will be ready to adopt this new orthography.

§5. Conclusion

This document has discussed how Tulu written in the Kannada script may be implemented in Unicode. Initiative from the native scholars will be required for further action.

-o-o-o-