Title: Determining the Encoding Model for Soyombo Vowels
Author: Anshuman Pandey (pandey@umich.edu)
Status: Individual Contribution
Action: For consideration by UTC and WG2
Date: 2011-02-05

1 Introduction

This document presents some questions about the encoding model for Soyombo vowel letters and signs, which were originally discussed in N3949 L2/10-399.

2 Background

The set of Soyombo vowel letters is:

A	AA	I	II	UE	UUE	U	UU	E	EE	O	OO	OE	OOE	AU	AI

R	RR	L	LL

These are formed by writing the letter A with the dependent sign for each vowel:

-A	-AA	-I	-II	-UE	-UUE	-U	-UU	-E	-EE	-O	-OO	-OE	-OOE	-AU	-AI

-R	-RR	-L	-LL

Additionally, the long vowels are produced by writing the basic vowel sign with the sign ◌ (-AA).

The above vowel letters and signs are proposed for independent encoding in N3949.

3 Discussion

Soyombo is modeled upon the Tibetan script and the graphical structure of the vowel letters and signs follow the Tibetan system of representing independent vowels by writing dependent signs with a 'vowel-carrier', the ཨ U+0F68 TIBETAN LETTER A. Given this, it may be more appropriate to adopt a similar model for Soyombo vowel letters and sign instead of encoding them all as independent characters. Two such possibilities are discussed below:

1. **Encode a 'vowel-carrier letter' and all vowel signs**

   The graphical structure of Soyombo vowels suggests that a 'vowel-carrier letter' approach may be practical. If ᠌ A is treated as a vowel-carrier letter, then independent forms of vowels may be produced by writing a dependent sign with A. This model would require only the following unique characters for representing all vowel letters and signs:

A	-AA	-I	-II	-UE	-UUE	-U	-UU	-E	-EE	-O	-OO	-OE	-OOE	-AU	-AI

-R	-RR	-L	-LL
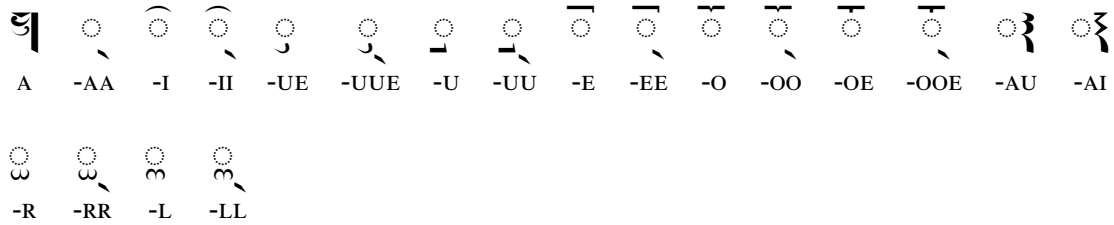
   Independent vowels would be produced by combining a vowel sign and A:

 $$\text{A} \; + \; \text{-II} \; \rightarrow \; \bar{\imath}$$
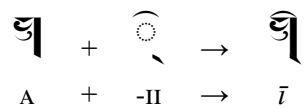
2. **Encode a 'vowel-carrier letter' and only basic vowel signs**

   As the signs for long vowels may be decomposed into basic vowel sign + a length mark, the model suggested above may be further extended. This second approach requires the encoding of only the 'vowel-carrier' ᠌ A, the basic vowel signs, and a ᠌ *LENGTH MARK, which is the more semantically appropriate name for the character currently named VOWEL SIGN AA. With this model all vowel letters and signs may be written using the following set:

A	-I	-UE	-U	-E	-O	-OE	-AU	-AI	-R	-L	LENGTH

   The long forms of each vowel would be produced by combining the basic vowel sign and the ᠌ *LENGTH MARK:

   $$\text{A} \; + \; \text{-I} \; + \; \text{LENGTH} \; \rightarrow \; \bar{\imath}$$
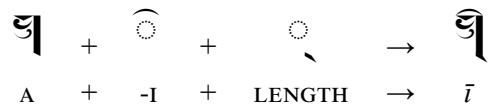
4   Recommendation

It is recommended that instead of independently encoding the Soyombo vowel letters and signs, one of the two encoding models described above be selected. Of the two, the second approach offers a model comparative to the Tibetan system, upon which Soyombo is based.