1 Introduction

This document is a revision of “Preliminary Proposal to Encode the Soyombo Script in ISO/IEC 10646” (N3949 L2/10-399). The major change is the encoding model for vowels. Vowel letters are now to be written using a vowel-carrier letter and a set of combining signs (see N3986 L2/11-054). Additional changes include minor revisions to character names and properties. Some issues identified during additional research are enumerated in Section 5. Resolution of these issues is necessary before a final encoding for Soyombo may be established.

The Soyombo font used here is based upon the font developed by Oliver Corff in August 1996 for his “Soyombo for LATEX” package. Some modifications have been made to Corff’s font and new glyphs added by the present author. A new font is being developed and will accompany the formal proposal.

2 Background

Soyombo (Mongolian: Соёмбо бичиг soyombo bicig) is a script used for writing Mongolian, Sanskrit, and Tibetan. It was used mainly for producing ornamental Buddhist texts.1 The script was designed in 1686 by Zanabazar (1635–1723), the first spiritual leader of Tibetan Buddhism in Mongolia, who also developed the Xäwtää Dörboljin (Horizontal Square) script.2 The name soyombo (𑡒𑡍𑡺𑡶𑡖) is derived from Sanskrit स्वयंभु ‘self-existing’.

There are several records in Soyombo, most of which are manuscripts, such as that shown in Figure 9. Soyombo is described in many secondary sources, such as studies of Mongolian scripts, eg. Boldsaikhan, et al. (2005), Shagdarsüreng (2001), etc. Soyombo is currently in use, as attested by a poem that was recently typeset using a digitized font (see Figure 5).

3 Writing System

3.1 Structure

Soyombo is an alphasyllabic script based upon the Brahmi model and patterned after Tibetan. Consonant letters have the inherent vowel a. Vowels are represented using a vowel-carrier letter and dependent vowel signs. There is no virama-like character. Final consonants are represented using contextual forms of regular letters. Consonant conjuncts are written using subjoined forms.

---

1 Atwood 2004: 518.  
2 See N3956 L2/10-411.
The graphical structure for Soyombo characters is ‘two-dimensional, not linear’ (Corff 1996; see Figure 3). Consonant and vowel letters are designed upon the basic frame 𑁄. Signs are written above the frame, eg. 𑁄; inside the frame, below the base element of the letter, eg. 𑁃; attached to the frame, eg. 𑁄; or to the right, eg. 𑁅. Final consonants are written as combining signs inside the frame of the letter, eg. 𑁆. Consonant conjuncts are represented similarly, with reduced forms of 𑁄 written inside the frame of 𑁅.

3.2 Character Names

Names for characters are based upon those given in secondary sources, such as Shagdatsürgün (2001). Names for certain consonants follow those given in N1855 L2/98-358.

3.3 Vowels

The following vowel letters are attested:

a ā i ī u ū ư ũ ę ē o ō ư ũ ă ā o ō ư ũ u ū ă ā o ō ư ũ

The graphical decomposition of these vowel letters is:

a ā i ī u ū ư ũ ę ē o ō ư ũ u ū ă ā o ō ư ũ

Soyombo vowel letters and signs follow the Tibetan system of representing independent vowels by writing dependent signs with a ‘vowel-carrier’, ie. 𑁄 + 0F68 TIBETAN LETTER A. As illustrated, the letter 𑁄 a functions as a vowel-carrier letter in Soyombo. The above decomposition indicates that dependent signs exist for basic vowels; long vowels are indicated by writing a basic vowel sign and a 𑁄 length mark.

All independent and dependent vowels may be written using the following set of characters (the ‘-’ is an abbreviation for ‘VOWEL SIGN’):


The 𑁄 LENGTH MARK is written after the accompanying vowel sign:

A + -I + LENGTH → ī
There is an additional vowel letter used for representing Tibetan:

\[
\text{ŋa}
\]

**3.4 Consonant Letters**

The basic set of consonant letters is:

\[
\begin{array}{ccccccccccccc}
\text{Ga} & \text{ka} & \text{nga} & \text{ja} & \text{ca} & \text{nya} & \text{da} & \text{ta} & \text{na} & \text{ba} & \text{pa} & \text{ma} & \text{ya} & \text{ra} & \text{va} \\
\text{la} & \text{sha} & \text{sa} & \text{ha} & \text{kssa}
\end{array}
\]

Although it is a conjunct, the letter ꐇ KSSA appears in Soyombo charts at the end of the basic consonants.

Additional consonant letters used for writing Sanskrit are shown below. The names for these characters begin with GALIG, eg. SOYOMBO LETTER GALIG GA:

\[
\begin{array}{ccccccccccccc}
\text{Ga} & \text{gha} & \text{ja} & \text{jha} & \text{tta} & \text{ttha} & \text{dda} & \text{ddha} & \text{nna} & \text{da} & \text{dha} & \text{ba} & \text{bha} & \text{ssa}
\end{array}
\]

Other consonant letters for writing Tibetan:

\[
\text{tsa} \quad \text{tsha} \quad \text{dza} \quad \text{zha} \quad \text{za}
\]

**3.5 Final Consonant Signs**

The following signs are used for writing final consonants:

\[
\begin{array}{ccccccccc}
\text{-g} & \text{-k} & \text{-ng} & \text{-d} & \text{-n} & \text{-b} & \text{-m} & \text{-r} & \text{-sh} & \text{-s} & \text{-ang}
\end{array}
\]

**3.6 Consonant Conjuncts**

The following consonant conjuncts are attested:

\[
\begin{array}{ccccccccccccccc}
\text{ky}a & \text{kr}a & \text{kl}a & \text{km}a & \text{kk}a & \text{ng}ka & \text{ny}ca & \text{nnt}a & \text{n}ta & \text{mp}a & \text{l}ka & \text{sh}ka & \text{ska} & \text{r}ka
\end{array}
\]

Some of these conjuncts are atomic ligatures, eg. ꐇ NNTTA, while others may be analyzed as a combination of a base letter (regular or half) and a subjoined form. Conjuncts are produced by stacking character elements within the letter frame.

Conjuncts with ꐇ KA as \(C_1\) are written using ꐇ (the half-form of ꐇ GA) and a subjoined form of \(C_2\):
The following conjuncts may be analyzed as consisting of a base letter + subjoined form:

Given the above, the following subjoined and reduced forms may be identified:

At present, the attested conjuncts are proposed for encoding as independent characters. If additional research uncovers the use of subjoined or reduced forms, then the encoding model will be revised (see Section 5).

3.7 Various Signs

- ANUSVARA: ॐ; eg. ओम

- VISARGA: 

3.8 Punctuation

- DANDA I and DOUBLE DANDA II

- WORD SEPARATOR:

3.9 Head Marks

There are two ‘head marks’: HEAD MARK and TERMINAL MARK. These characters are used for marking the beginning and end of a text section.

3.10 Symbols

The SYMBOL MONGOLIA was developed by Zanabazar, the creator of the script. It is an officially-recognized symbol in Mongolia and appears on the flag of Mongolia (see Figure 10) and coat of arms (see Figure 11). It is not part of the script, but it should be included in the Soyombo block.
4 Implementation Details

4.1 Collation

To be determined.

4.2 Character Properties

\[
\begin{align*}
11840; & \text{SOYOMBO LETTER A;Lo;0;L;;;;;N;;;;; } \\
11850; & \text{SOYOMBO LETTER GA;Lo;0;L;;;;;N;;;;; } \\
11851; & \text{SOYOMBO LETTER KA;Lo;0;L;;;;;N;;;;; } \\
11852; & \text{SOYOMBO LETTER NGA;Lo;0;L;;;;;N;;;;; } \\
11853; & \text{SOYOMBO LETTER JA;Lo;0;L;;;;;N;;;;; } \\
11854; & \text{SOYOMBO LETTER CA;Lo;0;L;;;;;N;;;;; } \\
11855; & \text{SOYOMBO LETTER NYA;Lo;0;L;;;;;N;;;;; } \\
11856; & \text{SOYOMBO LETTER DA;Lo;0;L;;;;;N;;;;; } \\
11857; & \text{SOYOMBO LETTER TA;Lo;0;L;;;;;N;;;;; } \\
11858; & \text{SOYOMBO LETTER NA;Lo;0;L;;;;;N;;;;; } \\
11859; & \text{SOYOMBO LETTER BA;Lo;0;L;;;;;N;;;;; } \\
1185A; & \text{SOYOMBO LETTER PA;Lo;0;L;;;;;N;;;;; } \\
1185B; & \text{SOYOMBO LETTER MA;Lo;0;L;;;;;N;;;;; } \\
1185C; & \text{SOYOMBO LETTER YA;Lo;0;L;;;;;N;;;;; } \\
1185D; & \text{SOYOMBO LETTER RA;Lo;0;L;;;;;N;;;;; } \\
1185E; & \text{SOYOMBO LETTER VA;Lo;0;L;;;;;N;;;;; } \\
1185F; & \text{SOYOMBO LETTER LA;Lo;0;L;;;;;N;;;;; } \\
11860; & \text{SOYOMBO LETTER SHA;Lo;0;L;;;;;N;;;;; } \\
11861; & \text{SOYOMBO LETTER SA;Lo;0;L;;;;;N;;;;; } \\
11862; & \text{SOYOMBO LETTER HA;Lo;0;L;;;;;N;;;;; } \\
11863; & \text{SOYOMBO LETTER KSA;Lo;0;L;;;;;N;;;;; } \\
11865; & \text{SOYOMBO VOWEL SIGN I;Mn;0;NSM;;;;;N;;;;; } \\
11866; & \text{SOYOMBO VOWEL SIGN U;Mn;0;NSM;;;;;N;;;;; } \\
11867; & \text{SOYOMBO VOWEL SIGN UE;Mn;0;NSM;;;;;N;;;;; } \\
11868; & \text{SOYOMBO VOWEL SIGN E;Mn;0;NSM;;;;;N;;;;; } \\
11869; & \text{SOYOMBO VOWEL SIGN O;Mn;0;NSM;;;;;N;;;;; } \\
1186A; & \text{SOYOMBO VOWEL SIGN OE;Mn;0;NSM;;;;;N;;;;; } \\
1186B; & \text{SOYOMBO VOWEL SIGN AU;Mn;0;NSM;;;;;N;;;;; } \\
1186C; & \text{SOYOMBO VOWEL SIGN AI;Mn;0;NSM;;;;;N;;;;; } \\
1186D; & \text{SOYOMBO VOWEL LENGTH MARK;Mn;0;NSM;;;;;N;;;;; } \\
1186E; & \text{SOYOMBO CONSONANT SIGN G;Mn;0;NSM;;;;;N;;;;; } \\
1186F; & \text{SOYOMBO CONSONANT SIGN K;Mn;0;NSM;;;;;N;;;;; } \\
11870; & \text{SOYOMBO CONSONANT SIGN NG;Mn;0;NSM;;;;;N;;;;; } \\
11871; & \text{SOYOMBO CONSONANT SIGN D;Mn;0;NSM;;;;;N;;;;; } \\
11872; & \text{SOYOMBO CONSONANT SIGN N;Mn;0;NSM;;;;;N;;;;; } \\
11873; & \text{SOYOMBO CONSONANT SIGN ANG;Mn;0;NSM;;;;;N;;;;; } \\
11874; & \text{SOYOMBO CONSONANT SIGN GHA;Lo;0;L;;;;;N;;;;; } \\
11875; & \text{SOYOMBO CONSONANT SIGN JHA;Lo;0;L;;;;;N;;;;; } \\
11876; & \text{SOYOMBO CONSONANT SIGN TTA;Lo;0;L;;;;;N;;;;; } \\
11877; & \text{SOYOMBO CONSONANT SIGN TTHA;Lo;0;L;;;;;N;;;;; } \\
11878; & \text{SOYOMBO CONSONANT SIGN DDA;Lo;0;L;;;;;N;;;;; } \\
11879; & \text{SOYOMBO CONSONANT SIGN DDHA;Lo;0;L;;;;;N;;;;; } \\
1187A; & \text{SOYOMBO CONSONANT SIGN NNA;Lo;0;L;;;;;N;;;;; } \\
1187B; & \text{SOYOMBO CONSONANT SIGN JSSA;Lo;0;L;;;;;N;;;;; } \\
1187C; & \text{SOYOMBO CONSONANT SIGN SH;Mn;0;NSM;;;;;N;;;;; } \\
1187D; & \text{SOYOMBO CONSONANT SIGN S;Mn;0;NSM;;;;;N;;;;; } \\
1187E; & \text{SOYOMBO CONSONANT SIGN ANG;Mn;0;NSM;;;;;N;;;;; } \\
1187F; & \text{SOYOMBO LETTER GALIG GA;Lo;0;L;;;;;N;;;;; } \\
11880; & \text{SOYOMBO LETTER GALIG GHA;Lo;0;L;;;;;N;;;;; } \\
11881; & \text{SOYOMBO LETTER GALIG JA;Lo;0;L;;;;;N;;;;; } \\
11882; & \text{SOYOMBO LETTER GALIG JHA;Lo;0;L;;;;;N;;;;; } \\
11883; & \text{SOYOMBO LETTER GALIG TTA;Lo;0;L;;;;;N;;;;; } \\
11884; & \text{SOYOMBO LETTER GALIG TTHA;Lo;0;L;;;;;N;;;;; } \\
11885; & \text{SOYOMBO LETTER GALIG DDA;Lo;0;L;;;;;N;;;;; } \\
11886; & \text{SOYOMBO LETTER GALIG DDHA;Lo;0;L;;;;;N;;;;; } \\
11887; & \text{SOYOMBO LETTER GALIG NNA;Lo;0;L;;;;;N;;;;; }
\end{align*}
\]
5 Issues

1. Are subjoined or reduced forms of consonants used for writing conjuncts involving other consonants besides those shown in script charts?

2. Is the ◌\(\text{v}\) ‘gemination’ mark used with other consonants besides \(\text{k}\)\(\text{a}\)?

3. Is there a need to encode subjoined or reduced forms separately? Or is encoding the attested conjuncts as independent characters sufficient?

4. The Tibetan model is appropriate for encoding Soyombo vowels. If subjoined forms of consonants are to be encoded, which model would facilitate the easiest implementation?

6 References


Acknowledgments

This project was made possible in part by a grant from the United States National Endowment for the Humanities, which funded the Universal Scripts Project (part of the Script Encoding Initiative at the University of California, Berkeley). Any views, findings, conclusions or recommendations expressed in this publication do not necessarily reflect those of the National Endowment of the Humanities.
Figure 1: Proposed code chart for Soyombo.
Consonants
11840 ḷ SOYOMBO LETTER GA
• Also represents Sanskrit ka
11841 ḷ SOYOMBO LETTER KA
• Also represents Sanskrit kha
11843 ḷ SOYOMBO LETTER NGA
11844 ḷ SOYOMBO LETTER JA
• Also represents Sanskrit ca
11845 ḷ SOYOMBO LETTER CA
• Also represents Sanskrit cha
11846 ḷ SOYOMBO LETTER NYA
11847 ḷ SOYOMBO LETTER DA
• Also represents Sanskrit ta
11848 ḷ SOYOMBO LETTER TA
• Also represents Sanskrit tha
11849 ḷ SOYOMBO LETTER NA
1184A ḷ SOYOMBO LETTER BA
• Also represents Sanskrit pa
1184B ḷ SOYOMBO LETTER PA
• Also represents Sanskrit pha
1184C ḷ SOYOMBO LETTER MA
1184D ḷ SOYOMBO LETTER YA
1184E ḷ SOYOMBO LETTER RA
1184F ḷ SOYOMBO LETTER VA
11850 ḷ SOYOMBO LETTER LA
11851 ḷ SOYOMBO LETTER SHA
11852 ḷ SOYOMBO LETTER SA
11853 ḷ SOYOMBO LETTER HA
11854 ḷ SOYOMBO LETTER KSSA

Vowel additions for Sanskrit
11878 𑢔 SOYOMBO SIGN VOCALIC R
11879 𑢓 SOYOMBO SIGN VOCALIC L

Various signs for Sanskrit
1187A 𑢎 SOYOMBO SIGN ANUSVARA
1187B 𑢔 SOYOMBO SIGN VISARGA

Ligatures
1187C 𑢒 SOYOMBO LIGATURE KYA
1187D 𑢓 SOYOMBO LIGATURE KRA
1187E 𑢔 SOYOMBO LIGATURE KLA
1187F 𑢕 SOYOMBO LIGATURE KMA
11880 𑢖 SOYOMBO LIGATURE KKA
11881 𑢗 SOYOMBO LIGATURE NGKA
11882 𑢘 SOYOMBO LIGATURE NNYA
11883 𑢙 SOYOMBO LIGATURE NNTA
11884 𑢚 SOYOMBO LIGATURE NTA
11885 𑢛 SOYOMBO LIGATURE MPA
11886 𑢜 SOYOMBO LIGATURE LKA
11887 𑢝 SOYOMBO LIGATURE SHKA
11888 𑢞 SOYOMBO LIGATURE SKA
11889 𑢟 SOYOMBO LIGATURE RKA

Additions for Tibetan
1188A 𑢢 SOYOMBO LETTER TSA
1188B 𑢣 SOYOMBO LETTER TSHA
1188C 𑢤 SOYOMBO LETTER DZA
1188D 𑢥 SOYOMBO LETTER ZHA
1188E 𑢦 SOYOMBO LETTER YA
1188F 𑢧 SOYOMBO LETTER SMALL A

Punctuation
11890 𑢠 SOYOMBO DANDA
11891 𑢡 SOYOMBO DOUBLE DANDA
11892 𑢢 SOYOMBO WORD SEPARATOR

Symbols
11893 𑢣 SOYOMBO HEAD MARK
11894 𑢤 SOYOMBO TERMINAL MARK
11895 𑢥 SOYOMBO SYMBOL MONGOLIA
Figure 3: Basic graphical structure of Soyombo characters (adapted from Corff (1996).
Figure 4: Photograph of a chart of the Soyombo script (from “Histoire du livre” 2010).
Figure 5: Poem for the Green Tara typeset in a digitized Soyombo font (from Wikimedia 2009b).
2.2. Монгол хэлэнд хэрэглэгдэх цагаан толгой.

Figure 6: Characters of the Soyombo script (from Boldsaikhan, et al. 2005: 367).
Figure 7: Characters of the Soyombo script (from Kapaj 2002).
Figure 8: Characters of the Soyombo script (from Kapaj 2002).
Figure 9: A folio from a Soyombo manuscript (from Mongolwiki 2008).
Figure 10: The flag of Mongolia showing the Soyombo symbol (from Wikimedia 2008).

Figure 11: The coat of arms of Mongolia showing the Soyombo symbol (from Wikimedia 2009a)