ISO/IEC JTC1/SC2/WG2 N4655R
L2/15-004R
2015-12-03

Proposal to Encode the Soyombo Script in ISO/IEC 10646

Anshuman Pandey
Department of Linguistics
University of California, Berkeley
Berkeley, California, U.S.A.
anshuman.pandey@berkeley.edu

January 26, 2015
(Revised: December 3, 2015)

1 Introduction

A request to include Soyombo in the Unicode standard / Universal Character Set (ISO/IEC 10646), was made by the Mongolia and Japan national bodies in September 1998 (see N1855 L2/98-358). An update was provided in January 2000 by Takayuki K. Sato of Japan, which stated that the encoding project had stalled due to funding issues (see N2163 L2/00-055). The script was allocated to the Roadmap to the Supplementary Multilingual Plane (SMP) during WG2 Meeting 38 in March 2000 (see N2203 L2/00-234). There was no further action. This proposal seeks to fulfill the original request made 17 years ago.

This proposal supersedes the following documents:

- N3949 L2/10-399: Preliminary Proposal to Encode the Soyombo Script in ISO/IEC 10646
- N3986 L2/11-054: Determining the Encoding Model for Soyombo Vowels
- N4026 L2/11-125: Revised Preliminary Proposal to Encode Soyombo in the UCS
- N4142 L2/11-412: Proposal to Encode the Soyombo Script in ISO/IEC 10646
- N4414 L2/13-069: Revised Proposal to Encode the Soyombo Script in ISO/IEC 10646

The major departure from N4414 L2/13-069 is the proposed encoding model. Previously, the representation of consonant conjuncts was based upon the model for Tibetan. This model relied upon a set of subjoined forms for each consonant letter, which doubled the number of consonant characters in the repertoire. The new approach implements the *virama* model used for Brahmi-based scripts, in which a control character is placed between consonants to indicate that the adjacent letters are to be rendered as a conjunct stack. Characters for cluster-initial forms of four consonants used in language-specific orthographies have been retained. Changes have been made to character names and the ordering of the repertoire. Additional characters for head marks and terminal marks have been included. These topics are discussed throughout the proposal.

This revision contains editorial and informative changes suggested during a meeting with Mongolian experts in Tokyo, October 15–16, 2015. No substantive changes to the approved proposal or encoding are reflected here. Such changes are described in L2/15-247.

The proposed encoding for Soyombo in Unicode is an attempt to develop a character-encoding standard for the script. It is not an attempt to define an orthographic standard for the script. The proposed encoding is based upon an analysis of the script as it appears in various primary sources and in scholarly descriptions, and in conjunction with feedback from experts.

2 Background

Soyombo (, Mongolian: Соёмбо бичиг ᠰᠣᠶᠣᠮᠪᠣ ᠪᠢᠴᠢᠭ *soyombo bicig*) is a script used for writing Mongolian, Sanskrit, and Tibetan. Soyombo was designed in 1686 by Zanabazar (1635–1723), the first spiritual leader of Tibetan Buddhism in Mongolia, who also developed a Horizonal Square (*xebtee dörböljin*) script, also known as 'Zanabazar Square', which has been proposed for encoding in Unicode (see N4541 L2/14-024). The name *soyombo* is the Mongolian transcription of Sanskrit स्वयंभु *svayaṃbhu* 'self-existing'. It refers to the creation of the script by Zanabazar, who according to tradition saw the letterforms appear in the sky. The script is used mainly for producing Buddhist texts in Inner Asia. The majority of Soyombo records are manuscripts and inscriptions.

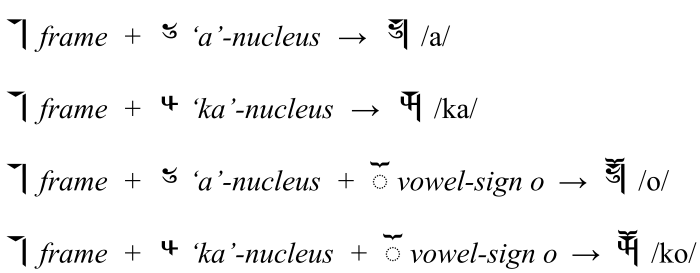
3 Script Details

3.1 Structure of the script

Soyombo is an alphasyllabic script that is written from left to right. Consonant letters possess the inherent vowel /a/, but in some contexts are purely alphabetic. Non-initial vowels are represented as signs. Initial and independent vowels are written using a vowel-carrier letter to which vowel signs are attached. The phonetic value of a consonant letter is altered by the addition of a vowel sign. Vowel length is indicated using a length mark that attaches to a letter or to a sequence of a letter and a vowel sign. Consonant clusters in Sanskrit and Tibetan are represented as conjuncts, which are produced either by attaching cluster-initial forms of letters to the left edge of a base consonant or by stacking non-initial letters beneath the base consonant. Syllable-final Mongolian consonants are indicated using signs.

3.2 Structure of characters

A Soyombo letter consists of a frame and a nucleus that represents a distinctive phoneme. The frame itself consists of two parts: a filled downwards-pointing triangle and a vertical bar positioned to the right of the triangle. The nucleus is placed beneath the triangle.



In the proposed encoding, an 'atomic Soyombo letter' is defined as a combination of the frame and a distinctive nucleus. For example, the letter '*ka*' is conceptualized as the singular unit 𑩜 and not a compound of the frame and the nucleus . This definition provides a practical means for uniquely identifying Soyombo letters.

There are two styles of frames: 'joined' and 'unjoined'. In the 'joined' frame the triangle and vertical bar connect. This frame is used with the majority of letters.

In the 'unjoined' frame the triangle and vertical bar do not connect and a gap is maintained between the two elements. The gap is preserved in conjuncts (see section 4.6.3). This frame occurs only with the following six letters: 𑩐 SOYOMBO LETTER KA, 𑩑 SOYOMBO LETTER KHA, 𑩒 SOYOMBO LETTER GA, 𑩓 SOYOMBO LETTER GHA, 𑩔 SOYOMBO LETTER NGA, 𑩕 SOYOMBO LETTER HA, 𑩖 SOYOMBO LETTER KSSA.

Vowels, final consonants, and other phonetic features appear as signs attached to various positions on a letter:

Several of these elements can occur with a single base letter:

3.3 Graphical structure of syllables

The structure of a graphical syllable in Soyombo is dependent upon the language being represented. Shown below are the structures for Mongolian, Tibetan, and Sanskrit. The notation uses the following abbreviations: V = vowel, C = consonant, M = mark, P = punctuation.

The structure of a Soyombo vowel syllable may be described as:

Mongolian:	$\mathbf{V_{carrier}}$ [V_{sign}] [M_{length}] [$V_{diphthong}$] [$C_{final\ sign}$]
Tibetan:	$\mathbf{V_{carrier}}$ [V_{sign}] [M_{length}]
Sanskrit:	$\mathbf{V_{carrier}}$ [V_{sign}] [M_{length}] [$M_{anusvara}$] [$M_{visarga}$]

The structure of a consonant syllable may be described as:

Mongolian:	**C** [V_{sign}] [M_{length}] [$V_{diphthong}$] [$C_{final\ sign}$]
Tibetan:	[C_{pre}] **C** [C*] [V_{sign}] [M_{length}] (P_{tsheg})
Sanskrit	[C_{pre}] **C** [M_{gem}] [C*] [V_{sign}] [M_{length}] [$M_{anusvara}$] [$M_{visarga}$]

3.4 Character repertoire and encoding order

The traditional Mongolian arrangement of the Soyombo script contains 90 characters:

The colorization is not part of the traditional presentation of the script, but has been added in order to illustrate the different classes of characters in the arrangement: head and terminal marks (red, section 4.9); vowels (blue, sections 4.1–4.2); vowel modifiers (cyan, section 4.3); consonant letters (black, section 4.4); final consonant signs (magenta, section 4.5); conjuncts (green, section 4.6).

The proposed encoding for Soyombo contains 81 characters (see the code chart and names list). The proposed repertoire necessarily differs from the presentation of the script in traditional charts. Some forms shown as atomic characters in charts are logically treated in the encoding as composite forms consisting of multiple characters. For instance, vowel letters such as 𑩐 and Mongolian final-consonant syllables such as 𑩷 are decomposed into combinations of a base letter and combining sign, eg. 𑩅 + ◌͡◌ + ◌ॖ = 𑩐; 𑩅 + ◌ॖ = 𑩷; and forms such as 𑪊 are analyzed as conjunct stacks, not independent letters, 𑪉 + 𑪊. This analysis aligns with the underlying structure of the script and orthographic principles as preserved in traditional charts.

The encoding order of characters, namely consonant letters, follows a pattern based upon the order of the Tibetan script (see section 4.4.3 for details). The encoded order of Soyombo matches the order proposed for the encoding of the Zanabazar Square script (see N4541 L2/14-024).

3.5 Character names

An encoding for Soyombo requires assigning distinctive identifiers to characters. Mongolian charts of the script give identical values to multiple letters. For instance, 𑩊 is used for Mongolian /g/ and /γ/, and for Sanskrit and Tibetan /k/; however, 𑩋 is used for /g/ only in Sanskrit and Tibetan contexts. In Mongolian, letters used specifically for Tibetan and Sanskrit are called ‘гали’ *galig*, a term applied to characters used for the transcription of foreign sounds. Accordingly, in a Mongolian context 𑩊 and 𑩋 are referred to as *ga* and *galig ga*, respectively. The descriptor ‘*galig*’ was used in the names for several characters in the

original request from Japan and Mongolia to encode Soyombo (see N1855 L2/98-358). Along with the adoption of the Tibetan arrangement, the naming conventions of the Tibetan script provide an opportunity to assign distinctive names for Soyombo characters without the need for the '*galig*' descriptor. This approach aligns with transliterations given in scholarly studies, such as Shagdarsürüng (2001), as well as the proposed encoding for the related Zanabazar Square script (see N4541 L2/14-024).

Character names appear throughout this document in small capitals, eg. SOYOMBO LETTER KA, SOYOMBO LETTER KHA, etc. The full name is given at the first mention of a character, while subsequent references contain shortened names, eg. KA, KHA, etc. Transliterations are in italics, eg. *ka*, *kha*, etc.

3.6 Glyphic representation of characters

The representative glyphs for Soyombo are illustrative. Although they resemble forms found in various sources, they are intended as serving as the standard or normative form. Users may design Soyombo fonts in accordance with their preferred styles.

The Soyombo font used in this document is based upon the 'JG Soyombo' font designed by Jason Glavy, with some glyphs sourced from the font designed by Oliver Corff for the "Soyombo for LaTeX" package, with several modifications and additions made by the proposal author.

4 Proposed Encoding

4.1 Vowel letter

There is 1 vowel-carrier letter:


SOYOMBO LETTER A

This letter represents either the vowel *a* or a zero vowel, depending upon phonotactical conditions. When combined with a vowel sign it represents an independent or initial vowel and assumes the value of the sign.

4.2 Vowel signs and length mark

There are 10 dependent vowel signs:


SOYOMBO VOWEL SIGN I


SOYOMBO VOWEL SIGN O


SOYOMBO VOWEL SIGN UE


SOYOMBO VOWEL SIGN AI


SOYOMBO VOWEL SIGN U


SOYOMBO VOWEL SIGN AU


SOYOMBO VOWEL SIGN E


SOYOMBO VOWEL SIGN VOCALIC R


SOYOMBO VOWEL SIGN OE


SOYOMBO VOWEL SIGN VOCALIC L

and 1 mark for indicating vowel length:


SOYOMBO VOWEL LENGTH MARK

The vowel signs and length mark attach to letters in the following positions:

The vowel signs and the length mark can combine with LETTER A and with consonant letters. Initial and independent forms of vowels are represented by attaching vowel signs to 𑩐 LETTER A. Long vowels are represented by attaching the ◌𑩇 VOWEL LENGTH MARK. The ten vowel syllables shown in traditional charts are represented using the 10 vowel signs and the length mark as follows:

a 𑩐 <𑩐 LETTER A>

ā 𑩐𑩇 <𑩐 LETTER A, ◌𑩇 VOWEL LENGTH MARK>

i 𑩑 <𑩐 LETTER A, ◌𑩑 VOWEL SIGN I>

ī 𑩑𑩇 <𑩐 LETTER A, ◌𑩑 VOWEL SIGN I, ◌𑩇 VOWEL LENGTH MARK>

ü 𑩒 <𑩐 LETTER A, ◌𑩒 VOWEL SIGN UE>

ǖ 𑩒𑩇 <𑩐 LETTER A, ◌𑩒 VOWEL SIGN UE, ◌𑩇 VOWEL LENGTH MARK>

u 𑩓 <𑩐 LETTER A, ◌𑩓 VOWEL SIGN U>

ū 𑩓𑩇 <𑩐 LETTER A, ◌𑩓 VOWEL SIGN U, ◌𑩇 VOWEL LENGTH MARK>

e 𑩔 <𑩐 LETTER A, ◌𑩔 VOWEL SIGN E>

ē 𑩔𑩇 <𑩐 LETTER A, ◌𑩔 VOWEL SIGN E, ◌𑩇 VOWEL LENGTH MARK>

ö 𑩕 <𑩐 LETTER A, ◌𑩕 VOWEL SIGN OE>

ō̈ 𑩕𑩇 <𑩐 LETTER A, ◌𑩕 VOWEL SIGN OE, ◌𑩇 VOWEL LENGTH MARK>

o 𑩖 <𑩐 LETTER A, ◌𑩖 VOWEL SIGN O>

ō 𑩖𑩇 <𑩐 LETTER A, ◌𑩖 VOWEL SIGN O, ◌𑩇 VOWEL LENGTH MARK>

ai 𑩗 <𑩐 LETTER A, ◌𑩗 VOWEL SIGN AI>

au 𑩘 <𑩐 LETTER A, ◌𑩘 VOWEL SIGN AU>

r̥ 𑩙 <𑩐 LETTER A, ◌𑩙 VOWEL SIGN VOCALIC R>

r̥̄ 𑩙𑩇 <𑩐 LETTER A, ◌𑩙 VOWEL SIGN VOCALIC R, ◌𑩇 VOWEL LENGTH MARK>

l̥ 𑩚 <𑩐 LETTER A, ◌𑩚 VOWEL SIGN VOCALIC L>

l̥̄ 𑩚𑩇 <𑩐 LETTER A, ◌𑩚 VOWEL SIGN VOCALIC L, ◌𑩇 VOWEL LENGTH MARK>

4.2.1 Sanskrit vocalic sounds

The ◌𑩗 VOWEL SIGN VOCALIC R and ◌𑩘 VOWEL SIGN VOCALIC L, and the corresponding long forms, are used for representing Sanskrit vocalic sounds. In Sanskrit contexts, the ◌𑩙 VOWEL SIGN UE is used for representing *u* and in such cases it is transliterated as *u* instead of *ü*.

4.2.2 Mongolian diphthongs

In Mongolian contexts the signs ◌𑩛 VOWEL SIGN AI and ◌𑩜 VOWEL SIGN AU represent the secondary vowels *i* and *u*, and are used with other vowel signs for representing diphthongs. The encoded representations of diphthongs identified in Soyombo records analyzed for this proposal are as follows:

ii 𑩈𑩛 <𑩅 LETTER A, ◌𑩍 VOWEL SIGN I, ◌𑩛 VOWEL SIGN AI>

iu 𑩈𑩜 <𑩅 LETTER A, ◌𑩍 VOWEL SIGN I, ◌𑩜 VOWEL SIGN AU>

īi 𑩈𑩎𑩛 <𑩅 LETTER A, ◌𑩍 VOWEL SIGN I, ◌𑩎 VOWEL LENGTH MARK, ◌𑩛 VOWEL SIGN AI>

īu 𑩈𑩎𑩜 <𑩅 LETTER A, ◌𑩍 VOWEL SIGN I, ◌𑩎 VOWEL LENGTH MARK, ◌𑩜 VOWEL SIGN AU>

üi 𑩈𑩙𑩛 <𑩅 LETTER A, ◌𑩙 VOWEL SIGN UE, ◌𑩛 VOWEL SIGN AI>

ui 𑩈𑩚𑩛 <𑩅 LETTER A, ◌𑩚 VOWEL SIGN U, ◌𑩛 VOWEL SIGN AI>

ei 𑩈𑩏𑩛 <𑩅 LETTER A, ◌𑩏 VOWEL SIGN E, ◌𑩛 VOWEL SIGN AI>

oi 𑩈𑩐𑩛 <𑩅 LETTER A, ◌𑩐 VOWEL SIGN O, ◌𑩛 VOWEL SIGN AI>

4.3 Vowel modifiers

There are two vowel modifiers used for transliterating words of Sanskrit origin:

◌𑩀 SOYOMBO SIGN ANUSVARA

◌𑩁 SOYOMBO SIGN VISARGA

The sign ANUSVARA is used for indicating nasalization. The sign VISARGA represents post-vocalic aspiration. Independent forms of these modifiers, as shown in traditional script charts, are represented by combining them with 𑩅 LETTER A:

aṃ 𑩅𑩀 <𑩅 LETTER A, ◌𑩀 SIGN ANUSVARA>

aḥ 𑩅𑩁 <𑩅 LETTER A, ◌𑩁 SIGN VISARGA>

They attach to letters in the following positions:

In encoded text they always occur after any accompanying vowel sign; and if ANUSVARA and VISARGA occur together, then the ANUSVARA is placed before the VISARGA in the encoded ensequence:

oṃ	𑩐𑩕𑪖	<𑩐 A, ◌𑩕 VOWEL SIGN O, ◌𑪖 SIGN ANUSVARA>
āḥ	𑩐𑩛𑪗	<𑩐 A, ◌𑩛 VOWEL LENGTH MARK, ◌𑪗 SIGN VISARGA>
aṃḥ	𑩐𑪖𑪗	<𑩐 A, ◌𑪖 SIGN ANUSVARA, ◌𑪗 SIGN VISARGA>

4.4 Consonant letters

There are 40 consonant letters:

- 𑩜 SOYOMBO LETTER KA
- 𑩝 SOYOMBO LETTER KHA
- 𑩞 SOYOMBO LETTER GA
- 𑩟 SOYOMBO LETTER GHA
- 𑩠 SOYOMBO LETTER NGA
- 𑩡 SOYOMBO LETTER CA
- 𑩢 SOYOMBO LETTER CHA
- 𑩣 SOYOMBO LETTER JA
- 𑩤 SOYOMBO LETTER JHA
- 𑩥 SOYOMBO LETTER NYA
- 𑩦 SOYOMBO LETTER TTA
- 𑩧 SOYOMBO LETTER TTHA
- 𑩨 SOYOMBO LETTER DDA
- 𑩩 SOYOMBO LETTER DDHA
- 𑩪 SOYOMBO LETTER NNA
- 𑩫 SOYOMBO LETTER TA
- 𑩬 SOYOMBO LETTER THA
- 𑩭 SOYOMBO LETTER DA
- 𑩮 SOYOMBO LETTER DHA
- 𑩯 SOYOMBO LETTER NA
- 𑩰 SOYOMBO LETTER PA
- 𑩱 SOYOMBO LETTER PHA
- 𑩲 SOYOMBO LETTER BA
- 𑩳 SOYOMBO LETTER BHA
- 𑩴 SOYOMBO LETTER MA
- 𑩵 SOYOMBO LETTER TSA
- 𑩶 SOYOMBO LETTER TSHA
- 𑩷 SOYOMBO LETTER DZA
- 𑩸 SOYOMBO LETTER ZHA
- 𑩹 SOYOMBO LETTER ZA
- 𑩺 SOYOMBO LETTER -A
- 𑩻 SOYOMBO LETTER YA
- 𑩼 SOYOMBO LETTER RA
- 𑩽 SOYOMBO LETTER VA
- 𑩾 SOYOMBO LETTER LA
- 𑩿 SOYOMBO LETTER SHA
- 𑪀 SOYOMBO LETTER SSA
- 𑪁 SOYOMBO LETTER SA
- 𑪂 SOYOMBO LETTER HA
- 𑪃 SOYOMBO LETTER KSSA

4.4.1 Notes on consonants

The 𑩐 SOYOMBO LETTER KSSA represents the Sanskrit cluster *kṣa* (/kʂa/). In Soyombo, this letter represents a phoneme that is phonetically a consonant cluster, but it has the structure of an atomic letter. It is encoded as a letter because in all cases consonant clusters are written as conjunct stacks in Soyombo, not as ligatures. While in some scripts the written form for /kʂa/ has an encoded representation as a character sequence, such an approach would not be consistent with this script.

The 𑩇 SOYOMBO LETTER -A is used for representing ཀ U+0F60 TIBETAN LETTER -A.

4.4.2 Phonetic values of consonant letters

The phonetic values for consonant letters in Mongolian ('M'), Sanskrit ('S'), and Tibetan ('T') contexts are given below. The primary difference between the three occurs in Mongolian contexts where letters for Sanskrit voiceless sounds (ie. 𑩋 KA, 𑩎 CA, 𑩑 TA, 𑩔 PA) are used for voiced stops, while the letters for voiceless aspirated sounds (ie. 𑩌 KHA, 𑩏 CHA, 𑩒 THA, 𑩕 PHA) are used for voiceless stops.

		M	S	T			M	S	T
𑩋	KA	*g*, *γ*	*ka*	*ka*	𑩔	PA	*b*	*pa*	*pa*
𑩌	KHA	*k*, *q* (*x*)	*kha*	*kha*	𑩕	PHA	*p*	*pha*	*pha*
𑩍	GA		*ga*	*ga*	𑩖	BA		*ba*	*ba*
𑩎	GHA		*gha*	*gha*	𑩗	BHA		*bha*	*bha*
𑩏	NGA	*ng*	*ṅa*	*nga* (*ṅa*)	𑩘	MA	*m*	*ma*	*ma*
𑩐	CA	*j*	*ca*	*ca*	𑩙	TSA			*tsa*
𑩑	CHA	*c*	*cha*	*cha*	𑩚	TSHA			*tsha*
𑩒	JA		*ja*	*ja*	𑩛	DZA			*dza*
𑩓	JHA		*jha*	*dzha*	𑩜	ZHA			*zha*
𑩔	NYA	*ny*	*ña*	*ña*	𑩝	ZA			*za*
𑩕	TTA		*ṭa*	*ṭa*	𑩇	-A			*'a*
𑩖	TTHA		*ṭha*	*ṭha*	𑩞	YA	*y*	*ya*	*ya*
𑩗	DDA		*ḍa*	*ḍa*	𑩟	RA	*r*	*ra*	*ra*
𑩘	DDHA		*ḍha*	*ḍha*	𑩠	LA	*l*	*la*	*la*
𑩙	NNA		*ṇa*	*ṇa*	𑩡	VA		*va*	*wa*
𑩚	TA	*d*	*ta*	*ta*	𑩢	SHA	*š*	*śa*	*śa*
𑩛	THA	*t*	*tha*	*tha*	𑩣	SSA		*ṣa*	*ṣa*
𑩜	DA		*da*	*da*	𑩤	SA	*s*	*sa*	*sa*
𑩝	DHA		*dha*	*dha*	𑩥	HA	*h*	*ha*	*ha*
𑩞	NA	*n*	*na*	*na*	𑩐	KSSA		*kṣa*	*kṣa*

The letters NYA, NNA, CA, CHA, JA, and ZA may occur in Mongolian contexts.

4.4.3 Encoded order of consonants

The primary difference between the traditional arrangement and the proposed repertoire is the ordering of consonants. The traditional arrangement shows three sets of consonants, shown below separated by '…':

𑩐	𑩑	𑩒	𑩓	𑩔	𑩕	𑩖	𑩗	𑩘	𑩙	𑩚	𑩛	𑩜	𑩝	𑩞
ga	*ka*	*ṅa*	*ja*	*ca*	*ña*	*da*	*ta*	*na*	*ba*	*pa*	*ma*	*ya*	*ra*	*va*
𑩟	𑩠	𑩡	𑩢	𑩣	…	𑩑	𑩒	𑩔	𑩕	𑩗	𑩘	𑩙	𑩚	𑩛
la	*śa*	*sa*	*ha*	*kṣa*		*ga*	*gha*	*ja*	*jha*	*ṭa*	*ṭha*	*ḍa*	*ḍha*	*ṇa*
𑩜	𑩝	𑩞	𑩤	𑩥	…	𑩦	𑩧	𑩨	𑩩	𑩪	𑩫			
da	*dha*	*ba*	*bha*	*ṣa*		*tsa*	*tsha*	*dza*	*zha*	*za*	*'a*			

The first set (*ga .. kṣa*) contains the 20 main consonants required for writing Mongolian texts including loanwords and sounds foreign to Mongolian. These 20 letters in addition to the 14 consonants of the second set (*ga .. ṣa*) are required for representing Sanskrit; and these 34 letters along with the 6 letters of the third set (*tsa .. 'a*) provide all the 40 letters needed for Tibetan, including loan words. While this arrangement reflects a Mongolian analysis of the script, it obscures the phonological pattern upon which the script is based, as is evidenced by the similar structure of letters belonging to particular classes of articulation:

velar	𑩐	𑩑	𑩒	𑩓	𑩔	𑩥	𑩤	
	ka	*kha*	*ga*	*gha*	*ṅa*	*kṣa*	*ha*	
palatal	𑩕	𑩖	𑩗	𑩘	𑩙	𑩧	𑩨	𑩩
	ca	*cha*	*ja*	*jha*	*ña*	*tsa*	*tsha*	*dza*
retroflex	𑩚	𑩛	𑩜	𑩝	𑩞			
	ṭa	*ṭha*	*ḍa*	*ḍha*	*ṇa*			
dental	𑩖	𑩗	𑩦	𑩧	𑩨			
	ta	*tha*	*da*	*dha*	*na*			
labial	𑩙	𑩚	𑩤	𑩥	𑩛			
	pa	*pha*	*ba*	*bha*	*ma*			
liquid	𑩜	𑩝	𑩟	𑩞				
	ya	*ra*	*la*	*va*				
fricative	𑩠	𑩡						
	za	*zha*						
sibilant	𑩢	𑩣	𑩡					
	śa	*ṣa*	*sa*					
glottal	𑩫							
	'a							

The proposed ordering of consonant letters in accordance with the pattern of the Tibetan script is shown below. It offers a more natural arrangement as it orders related letters contiguously, which in turn highlights the underlying congruence between letter-forms and the phonological system of Sanskrit, as adapted and expanded by Tibetans:

ka	*kha*	*ga*	*gha*	*ṅa*	*ca*	*cha*	*ja*	*jha*	*ña*	*ṭa*	*ṭha*	*ḍa*	*ḍha*	*ṇa*
ta	*tha*	*da*	*dha*	*na*	*pa*	*pha*	*ba*	*bha*	*ma*	*tsa*	*tsha*	*dza*	*zha*	*za*
'a	*ya*	*ra*	*la*	*va*	*śa*	*ṣa*	*sa*	*ha*	*kṣa*					

4.4.4 Consonant half-forms

In most cases the half-form of a letter consists of the nucleus. For letters whose glyphs contain a free-standing ◌ element, the half-form may be produced by both removing or retaining the element.

KA			TTA			PA			-A		
KHA			TTHA			PHA			YA		
GA			DDA			BA			RA		
GHA			DDHA			BHA			LA		
NGA			NNA			MA			VA		
CA			TA			TSA			SHA		
CHA			THA			TSHA			SSA		
JA			DA			DZA			SA		
JHA			DHA			ZHA			HA		
NYA			NA			ZA			KSSA		

4.4.5 Glyphic variants of consonant letters

The glyphic representations of consonant letters are quite uniform across the available sources. There are stylistic differences in the shape of the triangle and the 'swoosh' (see section 4.14.1), but the forms of the consonant nucleii are regular. Nevertheless, glyphic variants have been observed for the following letters:

	regular	variant
KSSA	𑩜	𑩝
ZA	𑩞	𑩟
LA	𑩠	𑩡

The letters 𑩠 LA and 𑩞 ZA have similar appearances. In some sources ZA is written as 𑩟 in order to distinguish it from 𑩠 LA.

4.5 Final consonant signs

Mongolian syllable-final consonants are written using the following 12 combining signs:

◌𑩇 SOYOMBO FINAL CONSONANT SIGN G
◌𑩈 SOYOMBO FINAL CONSONANT SIGN K
◌𑩉 SOYOMBO FINAL CONSONANT SIGN NG
◌𑩊 SOYOMBO FINAL CONSONANT SIGN D
◌𑩋 SOYOMBO FINAL CONSONANT SIGN N
◌𑩌 SOYOMBO FINAL CONSONANT SIGN B
◌𑩍 SOYOMBO FINAL CONSONANT SIGN M
◌𑩎 SOYOMBO FINAL CONSONANT SIGN R
◌𑩏 SOYOMBO FINAL CONSONANT SIGN L
◌𑩐 SOYOMBO FINAL CONSONANT SIGN SH
◌𑩑 SOYOMBO FINAL CONSONANT SIGN S
◌𑩒 SOYOMBO FINAL CONSONANT SIGN -A

The names for these signs are based upon the phonetic values of the Mongolian codas. The signs attach to the frame below the nucleus of a letter. The exception is FINAL CONSONANT SIGN -A, which attaches to the right of the frame, raised slightly above the baseline in order to distinguish it from VOWEL LENGTH MARK.

The signs are condensed forms of the nucleus of consonant letters or derived from a fragment of the nucleus, as shown below. The exception is ◌𑩒 FINAL CONSONANT SIGN -A, which is not derived from 𑩅 LETTER -A as would be expected; rather it is based upon ◌𑩗 VOWEL LENGTH MARK.

The final-consonants shown in traditional charts of Soyombo are represented as follows:

ag, *aγ* 𑩄𑩇 <𑩄 LETTER A, ◌𑩇 FINAL CONSONANT SIGN G>

ak, *aq* 𑩄𑩈 <𑩄 LETTER A, ◌𑩈 FINAL CONSONANT SIGN K>

ang 𑩄𑩉 <𑩄 LETTER A, ◌𑩉 FINAL CONSONANT SIGN NG>

ad, *at* 𑩄𑩊 <𑩄 LETTER A, ◌𑩊 FINAL CONSONANT SIGN D>

an <SOYOMBO LETTER A, SOYOMBO FINAL CONSONANT SIGN N>

ab, *ap* <SOYOMBO LETTER A, SOYOMBO FINAL CONSONANT SIGN B>

am <SOYOMBO LETTER A, SOYOMBO FINAL CONSONANT SIGN M>

ar <SOYOMBO LETTER A, SOYOMBO FINAL CONSONANT SIGN R>

al <SOYOMBO LETTER A, SOYOMBO FINAL CONSONANT SIGN L>

aš <SOYOMBO LETTER A, SOYOMBO FINAL CONSONANT SIGN SH>

as <SOYOMBO LETTER A, SOYOMBO FINAL CONSONANT SIGN S>

'a <SOYOMBO LETTER A, SOYOMBO FINAL CONSONANT SIGN -A>

The final-consonant sign always occurs after a vowel sign or the VOWEL LENGTH MARK in encoded text:

riul <SOYOMBO LETTER RA, SOYOMBO VOWEL SIGN I, SOYOMBO VOWEL SIGN AU, SOYOMBO FINAL CONSONANT SIGN L>

gīg <SOYOMBO LETTER KA, SOYOMBO VOWEL SIGN I, SOYOMBO VOWEL LENGTH MARK, SOYOMBO FINAL CONSONANT SIGN G>

4.6 Consonant conjuncts

In general, a consonant cluster is written as a conjunct. Geminated consonants are an exception (see section 4.7). A conjunct is rendered as a vertical stack that consists of the regular form of the initial letter and the nucleus of each non-initial letter descending sequentially beneath the initial letter: *ṅka*, *ddha*, *sva*, etc. Four consonant letters have alternate representations in conjuncts: LA, SHA, SA, RA. Depending upon the linguistic context, clusters involving these letters may be rendered using either stacked or prefixed letters, eg. *sva* may occur as both and . Soyombo sources show conjuncts that contain three consonants, but theoretically a conjunct could contain any number of consonants.

The proposed representation of Soyombo conjuncts in encoded text is based upon the *virama* model used for Brahmi-based scripts in the UCS. However, as Soyombo does not have a native *virama* character, the following special character is proposed for controling conjunct formation:

SOYOMBO SUBJOINER

Additionally, four cluster-initial letters are proposed for representing pre-fixed letters (see figure 7 for examples of usage):

SOYOMBO CLUSTER-INITIAL LETTER RA

SOYOMBO CLUSTER-INITIAL LETTER LA

SOYOMBO CLUSTER-INITIAL LETTER SHA

SOYOMBO CLUSTER-INITIAL LETTER SA

The SUBJOINER and four cluster-initial letters support all of the conjunct formation requirements for Soyombo.

The dotted box is not part of the glyphic representations of these characters, but is used in the code chart and descriptions in order to convey that these letters have special behaviors.

4.6.1 Cluster-Initial Letters

The cluster-initial forms are small geometric shapes that attach to the left side of the triangle of the following letter in the cluster, in the following positions on the frame:

There is no glyphic correspondence between the cluster-initial and regular forms of these four letters.

4.6.2 Encoded Representation of Conjuncts

A conjunct is represented in encoded text by placing the ◌ SUBJOINER between each consonant in a cluster:

sava	𑩐𑩖	<𑩐 SA, 𑩖 VA>
sva	𑩓	<𑩐 SA, ◌ SUBJOINER, 𑩖 VA>
savaya	𑩐𑩖𑩗	<𑩐 SA, 𑩖 VA, 𑩗 YA>
svya	𑩔	<𑩐 SA, ◌ SUBJOINER, 𑩖 VA, ◌ SUBJOINER, 𑩗 YA>

The SUBJOINER indicates that the following consonant is to be represented in its nuclear form and placed below the nucleus of the base letter. Conceptually, the function of the SUBJOINER is to remove the frame of a letter in order to produce the nucleus and to place the nucleus below the preceding letter:

<𑩌 JA, ◌ SUBJOINER, 𑩍 NYA>	→	𑩌 + (𑩍 - 𑩦 → 𑩎)	→	𑩕 *ñca*
<𑩣 TTA, ◌ SUBJOINER, 𑩣 TTHA>	→	𑩣 + (𑩣 - 𑩦 → 𑩤)	→	𑩥 *ṭṭha*
<𑩘 NA, ◌ SUBJOINER, 𑩙 DA>	→	𑩘 + (𑩙 - 𑩦 → 𑩚)	→	𑩛 *nda*
<𑩨 MA, ◌ SUBJOINER, 𑩩 BHA>	→	𑩨 + (𑩩 - 𑩦 → 𑩬)	→	𑩭 *mbha*

Placing the SUBJOINER after 𑩜 LA, 𑩞 SHA, 𑩐 SA, 𑩗 RA will produce a normal stacked conjunct. The cluster-initial letters must be used in order to produce the alternate conjuncts:

rva	𑩒	<𑩗 RA, ◌ SUBJOINER, 𑩖 VA>
rva	𑩖	<◌ CLUSTER-INITIAL RA, 𑩖 VA>
lva	𑩔	<𑩜 LA, ◌ SUBJOINER, 𑩖 VA>
lva	𑩖	<◌ CLUSTER-INITIAL LA, 𑩖 VA>

śva  <SHA, ◌ SUBJOINER, VA>

śva  <CLUSTER-INITIAL SHA, VA>

sva <SA, ◌ SUBJOINER, VA>

sva <CLUSTER-INITIAL SA, VA>

These four cluster-initial letters are the logical initial consonant in a conjunct, but the consonant to which they attach is to be considered the base letter. All signs and marks attach to the following letter:

rvi  <RA, ◌ SUBJOINER, VA, ◌ VOWEL SIGN I>

rvi  <CLUSTER-INITIAL RA, VA, ◌ VOWEL SIGN I>

4.6.3 Special conjunct behavior

The six letters KA, KHA, GA, GHA, NGA, HA, KSSA use the 'unjoined' frame . The gap between the triangle and frame is preserved when these these letters are initial in a conjunct. When these letters are non-initial in a conjunct, the gap in their frame is preserved by means of a split in the frame at the position where the nucleus is placed.

Hypothetical examples of conjuncts with these six letters in non-initial position are:

tka  <TA, ◌ SUBJOINER, KA>

skha  <SA, ◌ SUBJOINER, KHA>

dga <DA, ◌ SUBJOINER, GA>

bgha  <BA, ◌ SUBJOINER, GHA>

dṅa  <DDA, ◌ SUBJOINER, NGA>

lha  <LA, ◌ SUBJOINER, HA>

ṣkṣa  <SSA, ◌ SUBJOINER, KSSA>

4.6.4 Consonant conjuncts in script charts

Traditional Soyombo charts show 14 letters as independent characters. These letters are consonant conjuncts. They are not the only conjuncts used in Soyombo, as others are attested in manuscripts. These conjuncts are to be represented in encoded text as:

kya <KA, ◌𑩇 SUBJOINER, YA>

kra <KA, ◌𑩇 SUBJOINER, RA>

kla <KA, ◌𑩇 SUBJOINER, LA>

kva <KA, ◌𑩇 SUBJOINER, VA>

kka <KA, ◌𑩈 GEMINATION MARK>

ṅka <NGA, ◌𑩇 SUBJOINER, KA>

ñca <NYA, ◌𑩇 SUBJOINER, CA>

ṇṭa <NNA, ◌𑩇 SUBJOINER, TTA>

nta <NA, ◌𑩇 SUBJOINER, TA>

mpa <MA, ◌𑩇 SUBJOINER, PA>

lka <◌𑩐 CLUSTER-INITIAL LA, KA>

śka <◌𑩑 CLUSTER-INITIAL SHA, KA>

ska <◌𑩒 CLUSTER-INITIAL SA, KA>

rka <◌𑩓 CLUSTER-INITIAL RA, KA>

They are likely included in the traditional arrangement in order to illustrate the manner of representing consonant clusters. These conjuncts are of four types: Non-initial semi-vowels (*kya*, *kra*, *kla*, *kva*); gemination (*kka*); consonants and nasals from the five classes of articulation in Sanskrit phonology (*ṅka*, *ñca*, *ṇṭa*, *nta*, *mpa*); and cluster-initial letters used in Tibetan (*rka*, *lka*, *ska*) and Sanskrit (*śka*).

4.7 Gemination mark

The following character is used for indicating geminated consonants:

◌𑩈 SOYOMBO GEMINATION MARK

It is stacked above the triangle of the frame:

𑩈

It is theoretically possible to represent geminated consonants as conjuncts, as shown below, but such behavior is not attested in the available sources:

kka 𑩘 <𑩖 KA, ◌𑩘 SUBJOINER, 𑩖 KA>

The GEMINATION MARK is placed immediately after the base letter before any combining sign. Other above-base signs are placed above the mark.

kka 𑩖◌̆ <𑩖 KA, ◌̆ GEMINATION MARK>

kkā 𑩖◌̆𑩰 <𑩖 KA, ◌̆ GEMINATION MARK, ◌𑩰 VOWEL LENGTH MARK>

kki 𑩖◌̆◌͡◌ <𑩖 KA, ◌̆ GEMINATION MARK, ◌͡◌ VOWEL SIGN I>

4.8 Punctuation

Three punctuation marks are proposed for encoding:

𑩇 SOYOMBO TSHEG

𑩈 SOYOMBO SHAD

𑩉 SOYOMBO DOUBLE SHAD

The 𑩇 TSHEG is used for marking the end of a syllable; it corresponds to ་ U+0F0B TIBETAN MARK INTERSYLLABIC TSHEG.

The 𑩈 SHAD indicates the end of a phrase or sentence; it corresponds to ། U+0F0D TIBETAN MARK SHAD and । U+0964 DEVANAGARI DANDA.

The 𑩉 DOUBLE SHAD indicates the end of a section; it corresponds to ༎ U+0F0E TIBETAN MARK NYIS SHAD and ॥ U+0965 DEVANAGARI DOUBLE DANDA.

4.9 Head marks

Three head marks are proposed for encoding:

 SOYOMBO HEAD MARK WITH TRIPLE FLAME

 SOYOMBO HEAD MARK WITH SINGLE FLAME

 SOYOMBO HEAD MARK WITH CANDRABINDU

The following glyphic variants have been identified in the available sources (see table 13):

	regular	variants
HEAD MARK WITH TRIPLE FLAME		
HEAD MARK WITH SINGLE FLAME		
HEAD MARK WITH CANDRABINDU		

There is no known semantic distinction between the  ‘black’ and  ‘white’ forms of the head-mark base. The differences between the normative and variant forms are stylistic. The primary point of variation is the style of the  center element, which has a  ‘solid’ variant. There is a  ‘white’ variant that also has a  ‘solid’ form. The ‘solid’ forms lack the ‘hollow’ points found in the halves of the normative center elements. The other point of variation is the  ‘black’ and  ‘white’ forms of the *candrabindu*. These variant forms are not proposed for encoding as separate characters. They are to be unified with the representative forms and display of glyphic variants is to be managed through fonts.

The ‘head marks’ correspond semantically to other characters in the UCS, such as U+1800 MONGOLIAN BIRGA and  U+0FD3 TIBETAN MARK INITIAL BRDA RNYING YIG MGO MDUN MA. In several texts a ‘head mark’ may be followed by a SHAD or DOUBLE SHAD.

The  HEAD MARK WITH TRIPLE FLAME is also known as the ‘Svayambhu’ or ‘Soyombo’ symbol. In addition to its usage in written texts, it is an official symbol in Mongolia and appears on the flag and coat of arms of the country (see figures 45 and 46). It is also used in other language environments; see figure 43 for usage of the HEAD MARK WITH TRIPLE FLAME in a seal bearing the Mongolian script.

4.10 Terminal marks

Various ‘terminal marks’ are used at the end of a text. Three such marks are proposed for encoding:

 SOYOMBO TERMINAL MARK-1

 SOYOMBO TERMINAL MARK-2

 SOYOMBO TERMINAL MARK-3

Forms of these marks as they appear in manuscript sources are shown in table 14.

4.11 Digits

The available sources do not indicate the use of digits or number forms in the script.

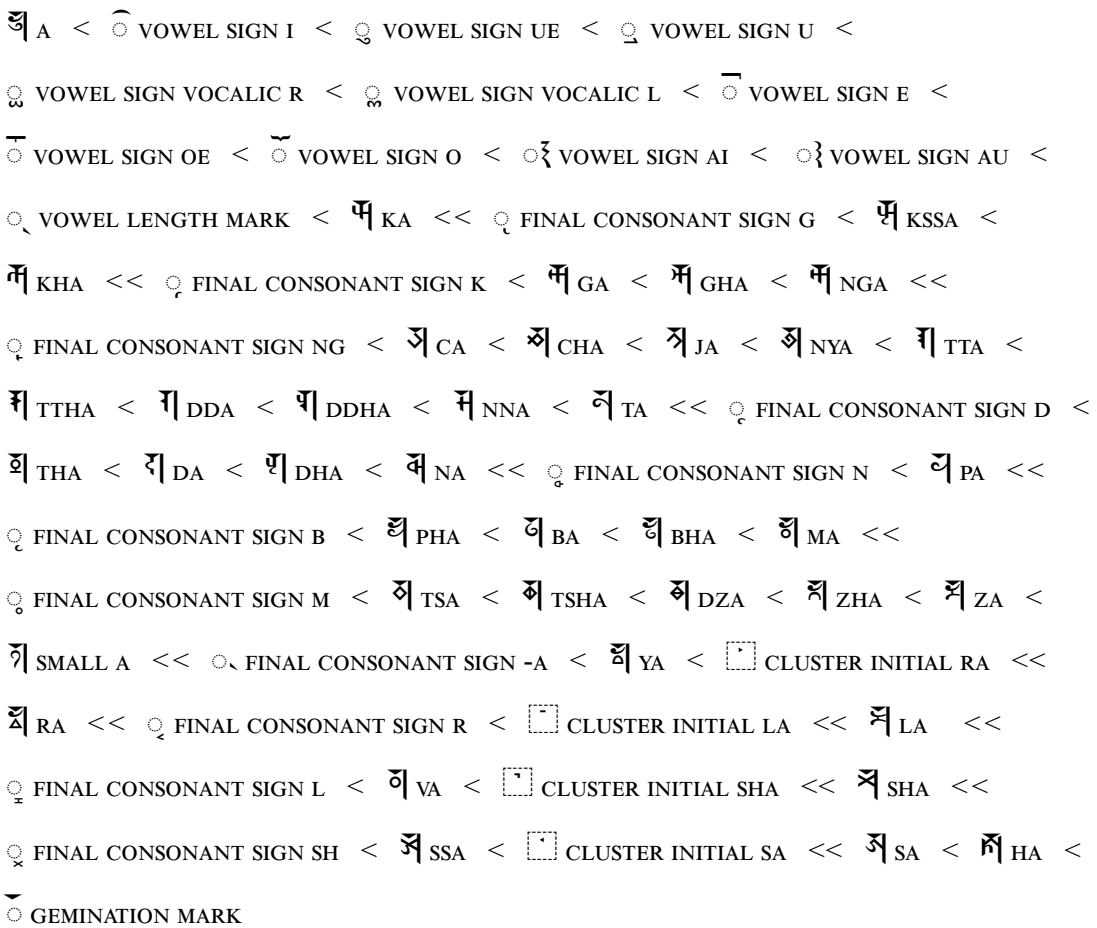
4.12 Vertical text

All the available sources show Soyombo text represented horizontally, left-to-right, top-to-bottom. The seal (*tamga*) in figure 14 may appear to contain vertical text, but it is actually horizonal text with line breaks after every third graphical syllable. It was likely designed as such in order to provide visual similarity to the Mongolian and Phags-pa scripts that also appear on the seal. There is only one instance of Soyombo represented vertically, shown here in figure 15. The text, which contains some errors, has some words written as vertical ligatures, which are to be considered idiosyncratic. These 'word ligatures' appear to be used for facilitating the spacing and fit of vertical text within the physical boundaries of the seal. Therefore, is no real requirement to support vertical orientation for Soyombo in plain text.

As a general rule, if Soyombo is represented in a vertical environment the minimal unit of vertical segmentation should be the graphical syllable. The text should be oriented top-to-bottom, left-to-right, with upright glyphs. See the comments in figure 15 for an example. The "Unicode Technical Report #50: Unicode Vertical Text Layout" describes the character property `Vertical_Orientation` (`vo`) for specifying default character orientation. For Soyombo, the property would be defined as: `Vertical_Orientation=U` or `vo=U`, where the value '`U`' indicates that the glyphs remain upright in both horizontal and vertical text layout.

4.13 Collation

The default sort order for Soyombo is as follows:



The following characters have secondary weights:

HEAD MARK WITH TRIPLE FLAME, HEAD MARK WITH SINGLE FLAME, HEAD MARK WITH CANDRABINDU, TERMINAL MARK-1, TERMINAL MARK-2, TERMINAL MARK-3, ANUSVARA, VISARGA

4.14 Rendering considerations

4.14.1 Stylistic variations of the frame

Variations exist in the shape of the triangle, depth of the vertical bar, and the shape of the 'swoosh':

These are all stylistic variations whose appearances are to be controlled through fonts.

4.14.2 Depth of the vertical bar

The height of the vertical bar varies across Soyombo records. The primary consideration for height is that below marks should sit at a position below the nucleus or attached to the vertical bar such that there is sufficient space between the mark and the nucleus above. A short vertical bar is fine for pure Mongolian, which requires only sufficient clearance beneath the nucleus of a letter and the terminal of the frame for below-base vowel signs and final consonant signs.

The depth of the vertical bar should be lower in Sanskrit and Tibetan contexts, where conjuncts are common. Generally, a conjunct consisting of two consonants should be rendered such that the second nucleus fits within the normal letter height.

In Soyombo sources that show conjuncts with three consonants, the nucleus of the third consonant is simply placed beneath the second without extending the vertical bar:

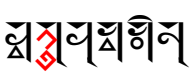
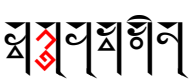
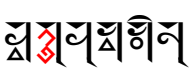
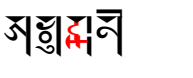
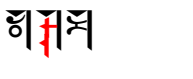
However, depth of the word *grva* breaks the symmetry. A solution is to extend the depth of the vertical bar in order to bring the last nucleus of *grva* within the frame.

This approach, however, disrupts the proportions of the rest of the letters and words. One solution would be to increase the depth of the vertical bar for the entire line.

The above solution maintains the graphical structure of the script, but does not provide the most elegant appearance as the depth of the vertical bar distorts the proportion of the size of the nucleus to the size of the bar. The next section discusses methods of rendering conjuncts in order to reduce alterations to the proportions of Soyombo letter-forms.

4.14.3 Adjustments for Conjuncts

There are no formal rules for sizing character glyphs within a stack. However, based upon an examination of conjunct styles in manuscripts, it is evident that some scribes adjusted the size of letters in stacks in an attempt to ensure proportions with surrounding characters. However, the available sources also reveal that some scribes were content with representing conjuncts simply by placing consonant nucleii below the baseline without consideration for the depth of the vertical bar.

‘A’	‘B’	‘C’
regular letter size normal bar depth	regular letter size greater bar depth	condensed letters normal bar depth
		
		
		
		
		

Column ‘A’ shows conjuncts rendered without any adjustments. The nucleii in the conjuncts are of the normal letter size and the depth of the vertical bar is of the normal letter height. It will be noticed that the final nucleus of the conjuncts in column ‘A’ protrudes slightly beneath the baseline and the bottom edge of the vertical bar. In column ‘B’ the nucleii are also the normal size, but the depth of the vertical bar has been increased such that it sit lower than the lowest nucleus. Column ‘C’ shows nucleii that have been condensed such that they fit completely within the normal letter height.

4.14.4 Frame adjustments for conjuncts

The gap in the 'unjoined' frame is preserved when a letter with such a frame occurs at any depth in a stack:

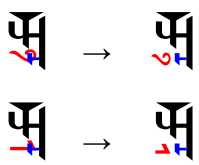
4.14.5 Glyph requirements for rendering conjuncts

A Soyombo font must contain a full set of nuclear forms for each consonant letter. The font will produce a consonant stack by substituting each <◌𑩇 SUBJOINER, *consonant*> pair with the half-form of the consonant. If the half-form glyph is not available, the SUBJOINER will be displayed visibly along with the regular glyph of the letter whose nucleus is missing. For example, if the half-form 𑩦 of 𑩬 KA is unavailable, then a sequence such as <𑩭 NGA, ◌𑩇 SUBJOINER, 𑩬 KA> will be rendered as 𑩭◌𑩇𑩬 instead of the expected 𑩮.

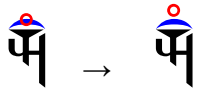
4.14.6 Placement of multiple combining signs

When multiple signs occur combine with a base letter at the same position, it may be necessary to adjust the glyphs to prevent clashing. Such adjustments depend upon the position of the sign.

Below-base signs such as ◌𑩗 VOWEL SIGN UE and ◌𑩘 VOWEL SIGN U commonly occur alongside final-consonant signs. In such cases the size and horizontal position of the vowel signs are modified.



Above-base marks such as ◌𑩀 ANUSVARA commonly occur with other above-base signs. A typical adjustment is to prevent clashing by placing one sign above the other according to rules for ordering:



4.14.7 Adjustments for positioning of the gemination mark

There are different ways of rendering the ◌𑩁 GEMINATION MARK. It may protude above the normal letter height or may align with the letter height. This second method requires vertical compression of the base triangle.

Compressing the GEMINATION MARK and base triangle, as shown to the right in the above allows uniformity in the positioning of above-base vowel signs and other marks.

5 Character Data

5.1 Character properties

In the format of `UnicodeData.txt`:

```
11A50;SOYOMBO LETTER A;Lo;0;L;;;;N;;;;;
11A51;SOYOMBO VOWEL SIGN I;Mn;0;NSM;;;;N;;;;;
11A52;SOYOMBO VOWEL SIGN UE;Mn;0;NSM;;;;N;;;;;
11A53;SOYOMBO VOWEL SIGN U;Mn;0;NSM;;;;N;;;;;
11A54;SOYOMBO VOWEL SIGN E;Mn;0;NSM;;;;N;;;;;
11A55;SOYOMBO VOWEL SIGN OE;Mn;0;NSM;;;;N;;;;;
11A56;SOYOMBO VOWEL SIGN O;Mn;0;NSM;;;;N;;;;;
11A57;SOYOMBO VOWEL SIGN AI;Mc;0;L;;;;N;;;;;
11A58;SOYOMBO VOWEL SIGN AU;Mc;0;L;;;;N;;;;;
11A59;SOYOMBO VOWEL SIGN VOCALIC R;Mn;0;NSM;;;;N;;;;;
11A5A;SOYOMBO VOWEL SIGN VOCALIC L;Mn;0;NSM;;;;N;;;;;
11A5B;SOYOMBO VOWEL LENGTH MARK;Mn;0;NSM;;;;N;;;;;
11A5C;SOYOMBO LETTER KA;Lo;0;L;;;;N;;;;;
11A5D;SOYOMBO LETTER KHA;Lo;0;L;;;;N;;;;;
11A5E;SOYOMBO LETTER GA;Lo;0;L;;;;N;;;;;
11A5F;SOYOMBO LETTER GHA;Lo;0;L;;;;N;;;;;
11A60;SOYOMBO LETTER NGA;Lo;0;L;;;;N;;;;;
11A61;SOYOMBO LETTER CA;Lo;0;L;;;;N;;;;;
11A62;SOYOMBO LETTER CHA;Lo;0;L;;;;N;;;;;
11A63;SOYOMBO LETTER JA;Lo;0;L;;;;N;;;;;
11A64;SOYOMBO LETTER JHA;Lo;0;L;;;;N;;;;;
11A65;SOYOMBO LETTER NYA;Lo;0;L;;;;N;;;;;
11A66;SOYOMBO LETTER TTA;Lo;0;L;;;;N;;;;;
11A67;SOYOMBO LETTER TTHA;Lo;0;L;;;;N;;;;;
11A68;SOYOMBO LETTER DDA;Lo;0;L;;;;N;;;;;
11A69;SOYOMBO LETTER DDHA;Lo;0;L;;;;N;;;;;
11A6A;SOYOMBO LETTER NNA;Lo;0;L;;;;N;;;;;
11A6B;SOYOMBO LETTER TA;Lo;0;L;;;;N;;;;;
11A6C;SOYOMBO LETTER THA;Lo;0;L;;;;N;;;;;
11A6D;SOYOMBO LETTER DA;Lo;0;L;;;;N;;;;;
11A6E;SOYOMBO LETTER DHA;Lo;0;L;;;;N;;;;;
11A6F;SOYOMBO LETTER NA;Lo;0;L;;;;N;;;;;
11A70;SOYOMBO LETTER PA;Lo;0;L;;;;N;;;;;
11A71;SOYOMBO LETTER PHA;Lo;0;L;;;;N;;;;;
11A72;SOYOMBO LETTER BA;Lo;0;L;;;;N;;;;;
11A73;SOYOMBO LETTER BHA;Lo;0;L;;;;N;;;;;
11A74;SOYOMBO LETTER MA;Lo;0;L;;;;N;;;;;
11A75;SOYOMBO LETTER TSA;Lo;0;L;;;;N;;;;;
11A76;SOYOMBO LETTER TSHA;Lo;0;L;;;;N;;;;;
11A77;SOYOMBO LETTER DZA;Lo;0;L;;;;N;;;;;
11A78;SOYOMBO LETTER ZHA;Lo;0;L;;;;N;;;;;
11A79;SOYOMBO LETTER ZA;Lo;0;L;;;;N;;;;;
11A7A;SOYOMBO LETTER -A;Lo;0;L;;;;N;;;;;
11A7B;SOYOMBO LETTER YA;Lo;0;L;;;;N;;;;;
11A7C;SOYOMBO LETTER RA;Lo;0;L;;;;N;;;;;
11A7D;SOYOMBO LETTER LA;Lo;0;L;;;;N;;;;;
11A7E;SOYOMBO LETTER VA;Lo;0;L;;;;N;;;;;
11A7F;SOYOMBO LETTER SHA;Lo;0;L;;;;N;;;;;
11A80;SOYOMBO LETTER SSA;Lo;0;L;;;;N;;;;;
11A81;SOYOMBO LETTER SA;Lo;0;L;;;;N;;;;;
11A82;SOYOMBO LETTER HA;Lo;0;L;;;;N;;;;;
11A83;SOYOMBO LETTER KSSA;Lo;0;L;;;;N;;;;;
11A84;SOYOMBO CLUSTER-INITIAL LETTER RA;Lo;0;L;;;;N;;;;;
11A85;SOYOMBO CLUSTER-INITIAL LETTER LA;Lo;0;L;;;;N;;;;;
```

```
11A86;SOYOMBO CLUSTER-INITIAL LETTER SHA;Lo;0;L;;;;;N;;;;;
11A87;SOYOMBO CLUSTER-INITIAL LETTER SA;Lo;0;L;;;;;N;;;;;
11A88;SOYOMBO FINAL CONSONANT SIGN G;Mn;0;NSM;;;;;N;;;;;
11A89;SOYOMBO FINAL CONSONANT SIGN K;Mn;0;NSM;;;;;N;;;;;
11A8A;SOYOMBO FINAL CONSONANT SIGN NG;Mn;0;NSM;;;;;N;;;;;
11A8B;SOYOMBO FINAL CONSONANT SIGN D;Mn;0;NSM;;;;;N;;;;;
11A8C;SOYOMBO FINAL CONSONANT SIGN N;Mn;0;NSM;;;;;N;;;;;
11A8D;SOYOMBO FINAL CONSONANT SIGN B;Mn;0;NSM;;;;;N;;;;;
11A8E;SOYOMBO FINAL CONSONANT SIGN M;Mn;0;NSM;;;;;N;;;;;
11A8F;SOYOMBO FINAL CONSONANT SIGN R;Mn;0;NSM;;;;;N;;;;;
11A90;SOYOMBO FINAL CONSONANT SIGN L;Mn;0;NSM;;;;;N;;;;;
11A91;SOYOMBO FINAL CONSONANT SIGN SH;Mn;0;NSM;;;;;N;;;;;
11A92;SOYOMBO FINAL CONSONANT SIGN S;Mn;0;NSM;;;;;N;;;;;
11A93;SOYOMBO FINAL CONSONANT SIGN -A;Mn;0;NSM;;;;;N;;;;;
11A94;SOYOMBO SIGN ANUSVARA;Mn;0;NSM;;;;;N;;;;;
11A95;SOYOMBO SIGN VISARGA;Mc;0;L;;;;;N;;;;;
11A96;SOYOMBO GEMINATION MARK;Mn;7;NSM;;;;;N;;;;;
11A97;SOYOMBO SUBJOINER;Mn;9;NSM;;;;;N;;;;;
11A98;SOYOMBO TSHEG;Po;0;L;;;;;N;;;;;
11A99;SOYOMBO SHAD;Po;0;L;;;;;N;;;;;
11A9A;SOYOMBO DOUBLE SHAD;Po;0;L;;;;;N;;;;;
11A9B;SOYOMBO HEAD MARK WITH TRIPLE FLAME;Po;0;ON;;;;;N;;;;;
11A9C;SOYOMBO HEAD MARK WITH SINGLE FLAME;Po;0;ON;;;;;N;;;;;
11A9D;SOYOMBO HEAD MARK WITH CANDRABINDU;Po;0;ON;;;;;N;;;;;
11A9E;SOYOMBO TERMINAL MARK-1;Po;0;ON;;;;;N;;;;;
11A9F;SOYOMBO TERMINAL MARK-2;Po;0;ON;;;;;N;;;;;
11AA0;SOYOMBO TERMINAL MARK-3;Po;0;ON;;;;;N;;;;;
```

5.2 Linebreaking

In the format of `LineBreak.txt`:

```
11A50; AL              # LETTER A
11A51..11A5B; CM       # VOWEL SIGN I .. VOWEL LENGTH MARK
11A5C..1181F; AL       # LETTER KA .. LETTER KSSA
11A84..11A87; CM       # CLUSTER-INITIAL LETTER RA .. CLUSTER-INITIAL LETTER SA
11A88..11A93; CM       # FINAL CONSONANT SIGN G .. FINAL CONSONANT SIGN -A
11A94..11A95; CM       # SIGN ANUSVARA .. SIGN VISARGA
11A96; CM              # GEMINATION MARK
11A97; CM              # SUBJOINER
11A98..11A9A; BA       # TSHEG .. DOUBLE SHAD
11A9B..11A9D; BB       # HEAD MARK WITH TRIPLE FLAME .. HEAD MARK WITH CANDRABINDU
11A9E..11AA0; BA       # TERMINAL MARK-1 .. TERMINAL MARK-3
```

5.3 Syllabic categories

In the format of `IndicSyllabicCategory.txt`:

```
# Indic_Syllabic_Category=Bindu
11A94         ; Bindu               # Mn         SIGN ANUSVARA

# Indic_Syllabic_Category=Visarga
11A95         ; Visarga             # Mc         SIGN VISARGA

# Indic_Syllabic_Category=Virama
11A97         ; Virama              # Mn         SUBJOINER

# Indic_Syllabic_Category=Vowel_Independent
```

```
11A50         ; Vowel_Independent    # Lo         LETTER A

# Indic_Syllabic_Category=Vowel_Dependent
11A51..11A56  ; Vowel_Dependent      # Mn     [6] VOWEL SIGN I .. VOWEL SIGN O
11A57..11A58  ; Vowel_Dependent      # Mc     [2] VOWEL SIGN AI .. VOWEL SIGN AU
11A59..11A5A  ; Vowel_Dependent      # Mn     [2] VOWEL SIGN VOCALIC R ..
                                                  VOWEL SIGN VOCALIC L
11A5B         ; Vowel_Dependent      # Mc         VOWEL LENGTH MARK

# Indic_Syllabic_Category=Consonant
11A5C..1181F  ; Consonant            # Lo    [40] LETTER KA .. LETTER KSSA
11A84..11A87  ; Consonant            # Lo     [4] CLUSTER INITIAL RA ..
                                                  CLUSTER INITIAL SA

# Indic_Syllabic_Category=Consonant_Final
11A88..11A93  ; Consonant_Final      # Mc    [12] FINAL CONSONANT SIGN G ..
                                                  FINAL CONSONANT SIGN -A
```

A new category should be established for the four Soyombo cluster-initial letters. This category is tentatively named `Consonant_Prefixed`:

```
# Indic_Syllabic_Category=Consonant_Prefixed
11A84..11A87  ; Consonant_Prefixed   # Lo    [4] CLUSTER-INITIAL LETTER RA ..
                                                 CLUSTER-INITIAL LETTER SA
```

5.4 Positional categories

In the format of `IndicPositionalCategory.txt`:

```
# Indic_Positional_Category=Right
11A57..11A58   ; Right  # Mc      [2] VOWEL SIGN AI .. VOWEL SIGN AU

# Indic_Positional_Category=Top
11A51          ; Top    # Mn          VOWEL SIGN I
11A54..11A54   ; Top    # Mn      [4] VOWEL SIGN E .. VOWEL SIGN O

# Indic_Positional_Category=Bottom
11A52..11A53   ; Bottom # Mn      [2] VOWEL SIGN UE .. VOWEL SIGN U
11A59..11A5A   ; Bottom # Mn      [2] VOWEL SIGN VOCALIC R .. VOWEL SIGN VOCALIC L
11A5B          ; Bottom # Mc          VOWEL LENGTH MARK
```

6 References

Atwood, Christopher P. 2004. *Encyclopedia of Mongolia and the Mongol Empire*. New York: Facts on File.

Boldsaikhan, B., B. Batsana, and Ts. Oyuntsetseg. 2005. *Соёмбо Нууц ба Синергетик* [= *Secret of Soyombo and Synergetic*]. (Түвэд, монгол бичгийн эхийг орчуулан хавсаргав). [illegible]. Ed. by T. Bulgan. Ulaanbaatar: Shambala Association, System Science Research Institute, Mongolian University of Science and Technology.

Рагчаайн Бямбаа [Byambaa Ragchaagiin]. 2005. *Занабазарын Дөрвөлжин Үсэг* [Zanabazarin Dörböljin Üseg].

Corff, Oliver. "Soyombo for LaTeX", v. 0.9. September 2, 1996. `http://userpage.fu-berlin.de/~corff/im/Soyombo/overview.Soyombo.html`

Dás, Sarat Chandra, 1888. "The Sacred and Ornamental Characters of Tibet". *Journal of the Asiatic Society of Bengal*, vol. LVII, no. II, pp. 41–48.

"Histoire du livre". 2010. "Histoire du livre et histoire de l'écriture: l'écriture, entre pratique, symbole et économie" (26 August 2010). http://histoire-du-livre.blogspot.com/2010/08/lecriture-entre-le-symbole-et-la.html

Ishii, Koji. 2013. "Unicode Technical Report #50: Unicode Vertical Text Layout". http://www.unicode.org/reports/tr50/

Kapaj, Luigi. 2002. "Mongol Scripts". http://www.viahistoria.com/SilverHorde/main.html?research/MongolScripts.html

Kara, György. 1959. "Un texte mongol en écriture soyombo". *Acta Orientalia Academiae Scientiarum Hungaricae*, vol. IX., pp. 1–38.

Mongolwiki. 2008. "Сайн чанарын зураг байхгүй байна". http://wiki.ecm-outsourcing.com/index.php?title=Дүрс:Soyombo_usgiin_bar.gif

Mongolia and Japan. 1988 "Addition of Soyombo Script". ISO/IEC JTC1/SC2/WG2 N1855 L2/98-358. September 21, 1998. http://std.dkuug.dk/jtc1/sc2/wg2/docs/n1855.pdf

Pandey, Anshuman. 2014. "Proposal to Encode the Zanabazar Square Script in ISO/IEC 10646". ISO/IEC JTC1/SC2/WG2 N4541R L2/14-024R. January 22, 2014 (revised: December 3, 2015). http://www.unicode.org/L2/L2014/14024r-zanabazar-square.pdf

———. 2015. "Revised code chart and names list for Soyombo". L2/15-247. October 20, 2015. http://www.unicode.org/L2/L2015/15247-soyombo-chart-revised.pdf

Rintschen, B. 1952. "Zwei unbekannte mongolische Alphabete aus dem XVII. Jahrhundert". *Acta Orientalia Academiae Scientiarum Hungaricae*, vol. II., pp. 63–71.

———. 1953. "A propos de la sigillographie mongole". *Acta Orientalia Academiae Scientiarum Hungaricae*, vol. III., pp. 25–31.

Sato, Takayuki K. "Soyombo and Pagba (old Mongol scripts)". ISO/IEC JTC1/SC2/WG2 N2163 L2/00-055. January 6, 2000. http://std.dkuug.dk/jtc1/sc2/wg2/docs/n2163.doc

Shagdarsürüng, Tseveliin. 2001. *Study of Mongolian Scripts* (Graphic Study or Grammatology). Enl. 2nd ed. Bibliotheca Mongolica: Monograph 1. Ed. by Sharaviin Choimaa. Ulaanbaatar: Center for Mongol Studies, National University of Mongolia.

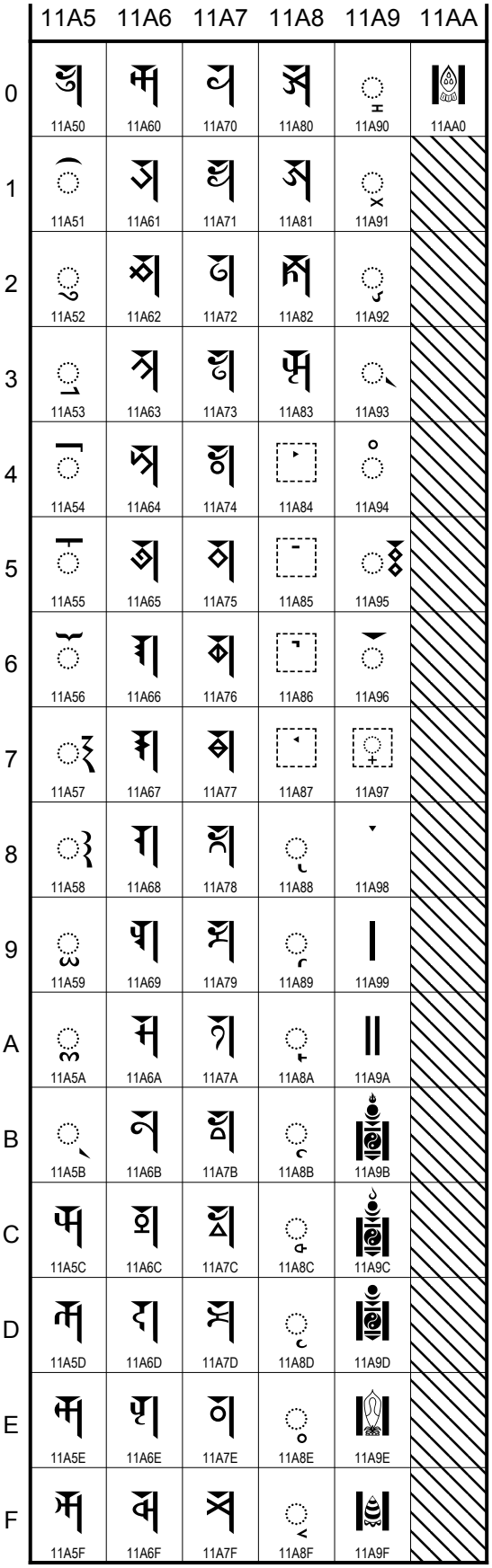
7 Acknowledgments

I express my gratitude to Agata Bareja-Starzyńska (University of Warsaw, Poland) for sharing her knowledge of the script, for her diligent review of the information I have presented, and for enlightening discussions on the script and comments regarding the character reptoire, character names, encoding order, and language-specific transliterations. I am thankful to Byambaa Ragchaagiin, who provided a copy of his book *Занабазарын Дөрвөлжин Үсэг*, which contains several manuscripts and other records which facilitated my analysis of the Soyombo script. I also owe thanks to Shriramana Sharma, Peter Constable (Microsoft), and György Kara (Indiana University, Bloomington), John Hudson (Tiro Typeworks), and Andrew Glass

(Microsoft) for reviewing previous proposals and earlier drafts of this proposal and for providing detailed comments on various aspects of the encoding.

This project was made possible in part through a Google Research Award, granted to Deborah Anderson for the Script Encoding Initiative. Research for an earlier phase of 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 Google or the National Endowment for the Humanities.

The author would also like to acknowledge the Henry Luce Foundation, which provided financial support for the October 2015 meeting with experts in Tokyo on the Soyombo and Zanabazar Square scripts.



Vowel carrier

11A50 𑩐 SOYOMBO LETTER A

Vowel signs

11A51 ◌𑩑 SOYOMBO VOWEL SIGN I
11A52 ◌𑩒 SOYOMBO VOWEL SIGN UE
11A53 ◌𑩓 SOYOMBO VOWEL SIGN U
11A54 ◌𑩔 SOYOMBO VOWEL SIGN E
11A55 ◌𑩕 SOYOMBO VOWEL SIGN O
11A56 ◌𑩖 SOYOMBO VOWEL SIGN OE
11A57 ◌𑩗 SOYOMBO VOWEL SIGN AI
11A58 ◌𑩘 SOYOMBO VOWEL SIGN AU
11A59 ◌𑩙 SOYOMBO VOWEL SIGN VOCALIC R
11A5A ◌𑩚 SOYOMBO VOWEL SIGN VOCALIC L

Vowel length mark

11A5B ◌𑩛 SOYOMBO VOWEL LENGTH MARK

Consonants

11A5C 𑩜 SOYOMBO LETTER KA
• Mongolian g
11A5D 𑩝 SOYOMBO LETTER KHA
• Mongolian k
11A5E 𑩞 SOYOMBO LETTER GA
11A5F 𑩟 SOYOMBO LETTER GHA
11A60 𑩠 SOYOMBO LETTER NGA
11A61 𑩡 SOYOMBO LETTER CA
• Mongolian j
11A62 𑩢 SOYOMBO LETTER CHA
• Mongolian c
11A63 𑩣 SOYOMBO LETTER JA
11A64 𑩤 SOYOMBO LETTER JHA
11A65 𑩥 SOYOMBO LETTER NYA
11A66 𑩦 SOYOMBO LETTER TTA
11A67 𑩧 SOYOMBO LETTER TTHA
11A68 𑩨 SOYOMBO LETTER DDA
11A69 𑩩 SOYOMBO LETTER DDHA
11A6A 𑩪 SOYOMBO LETTER NNA
11A6B 𑩫 SOYOMBO LETTER TA
• Mongolian d
11A6C 𑩬 SOYOMBO LETTER THA
• Mongolian t
11A6D 𑩭 SOYOMBO LETTER DA
11A6E 𑩮 SOYOMBO LETTER DHA
11A6F 𑩯 SOYOMBO LETTER NA
11A70 𑩰 SOYOMBO LETTER PA
• Mongolian b
11A71 𑩱 SOYOMBO LETTER PHA
• Mongolian p
11A72 𑩲 SOYOMBO LETTER BA
11A73 𑩳 SOYOMBO LETTER BHA
11A74 𑩴 SOYOMBO LETTER MA
11A75 𑩵 SOYOMBO LETTER TSA
11A76 𑩶 SOYOMBO LETTER TSHA
11A77 𑩷 SOYOMBO LETTER DZA
11A78 𑩸 SOYOMBO LETTER ZHA
11A79 𑩹 SOYOMBO LETTER ZA
11A7A 𑩺 SOYOMBO LETTER -A
11A7B 𑩻 SOYOMBO LETTER YA
11A7C 𑩼 SOYOMBO LETTER RA
11A7D 𑩽 SOYOMBO LETTER LA
11A7E 𑩾 SOYOMBO LETTER VA
11A7F 𑩿 SOYOMBO LETTER SHA
11A80 𑪀 SOYOMBO LETTER SSA
11A81 𑪁 SOYOMBO LETTER SA
11A82 𑪂 SOYOMBO LETTER HA
11A83 𑪃 SOYOMBO LETTER KSSA

Cluster-initial letters

11A84 ◌𑪄 SOYOMBO CLUSTER-INITIAL LETTER LA
11A85 ◌𑪅 SOYOMBO CLUSTER-INITIAL LETTER SHA
11A86 ◌𑪆 SOYOMBO CLUSTER-INITIAL LETTER SA
11A87 ◌𑪇 SOYOMBO CLUSTER-INITIAL LETTER RA

Final consonant signs

11A88 ◌𑪈 SOYOMBO FINAL CONSONANT SIGN G
11A89 ◌𑪉 SOYOMBO FINAL CONSONANT SIGN K
11A8A ◌𑪊 SOYOMBO FINAL CONSONANT SIGN NG
11A8B ◌𑪋 SOYOMBO FINAL CONSONANT SIGN D
11A8C ◌𑪌 SOYOMBO FINAL CONSONANT SIGN N
11A8D ◌𑪍 SOYOMBO FINAL CONSONANT SIGN B
11A8E ◌𑪎 SOYOMBO FINAL CONSONANT SIGN M
11A8F ◌𑪏 SOYOMBO FINAL CONSONANT SIGN R
11A90 ◌𑪐 SOYOMBO FINAL CONSONANT SIGN L
11A91 ◌𑪑 SOYOMBO FINAL CONSONANT SIGN SH
11A92 ◌𑪒 SOYOMBO FINAL CONSONANT SIGN S
11A93 ◌𑪓 SOYOMBO FINAL CONSONANT SIGN -A

Various signs

11A94 ◌𑪔 SOYOMBO SIGN ANUSVARA
11A95 ◌𑪕 SOYOMBO SIGN VISARGA

Gemination mark

11A96 ◌𑪖 SOYOMBO GEMINATION MARK

Subjoiner

11A97 ◌𑪗 SOYOMBO SUBJOINER
• used for producing consonant conjuncts

Punctuation

11A98 𑪘 SOYOMBO TSHEG
11A99 𑪙 SOYOMBO SHAD
11A9A 𑪚 SOYOMBO DOUBLE SHAD

Head marks

11A9B 𑪛 SOYOMBO HEAD MARK WITH TRIPLE FLAME
11A9C 𑪜 SOYOMBO HEAD MARK WITH SINGLE FLAME
11A9D 𑪝 SOYOMBO HEAD MARK WITH CANDRABINDU

Terminal Marks

11A9E 𑪞 SOYOMBO TERMINAL MARK-1
11A9F 𑪟 SOYOMBO TERMINAL MARK-2
11AA0 𑪠 SOYOMBO TERMINAL MARK-3

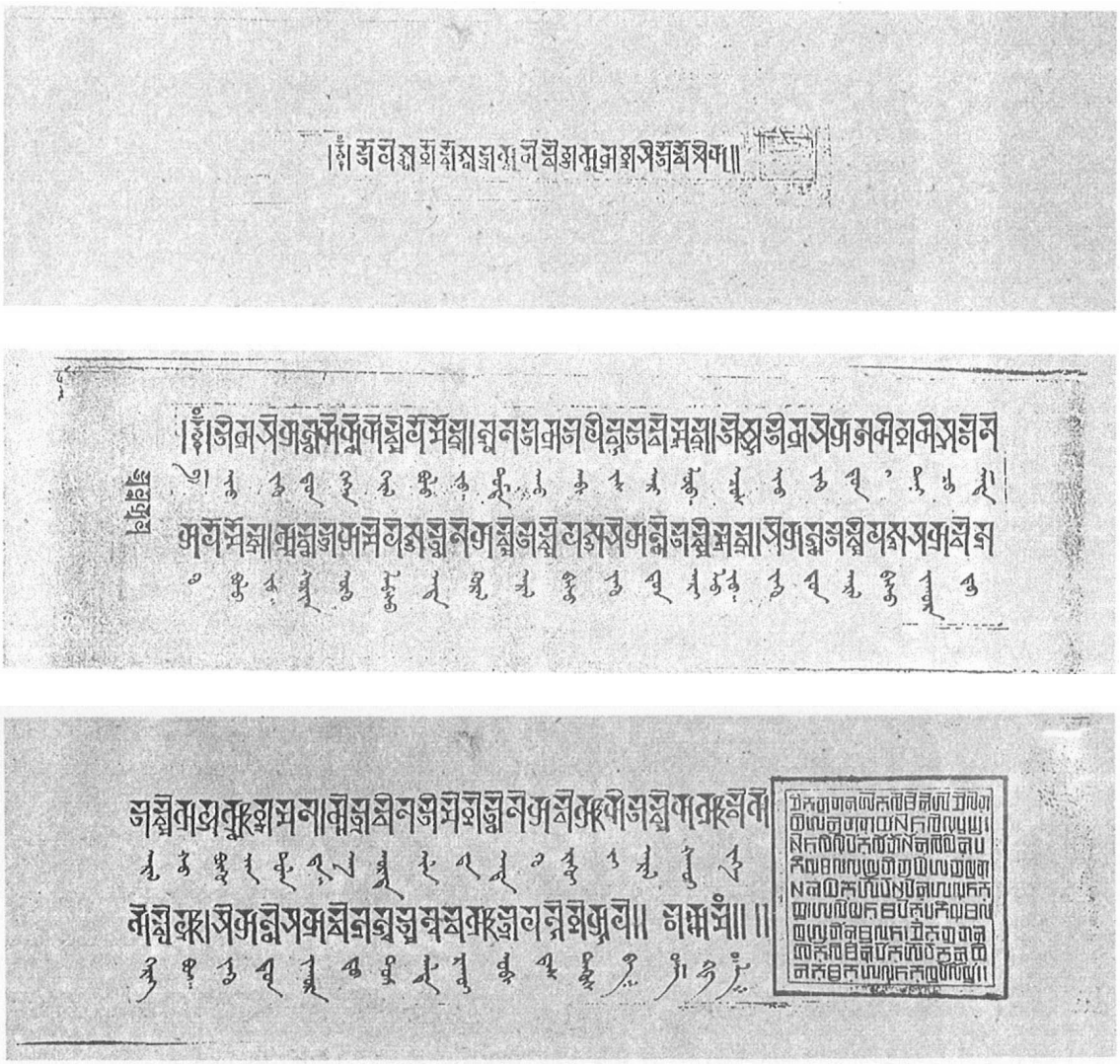


Figure 1: A manuscript containing Mongolian text in the Soyombo and Mongolian scripts (from Byambaa Ragchaagiin 2005: 63, 64). The last folio contains text in the Zanabazar Square Script.

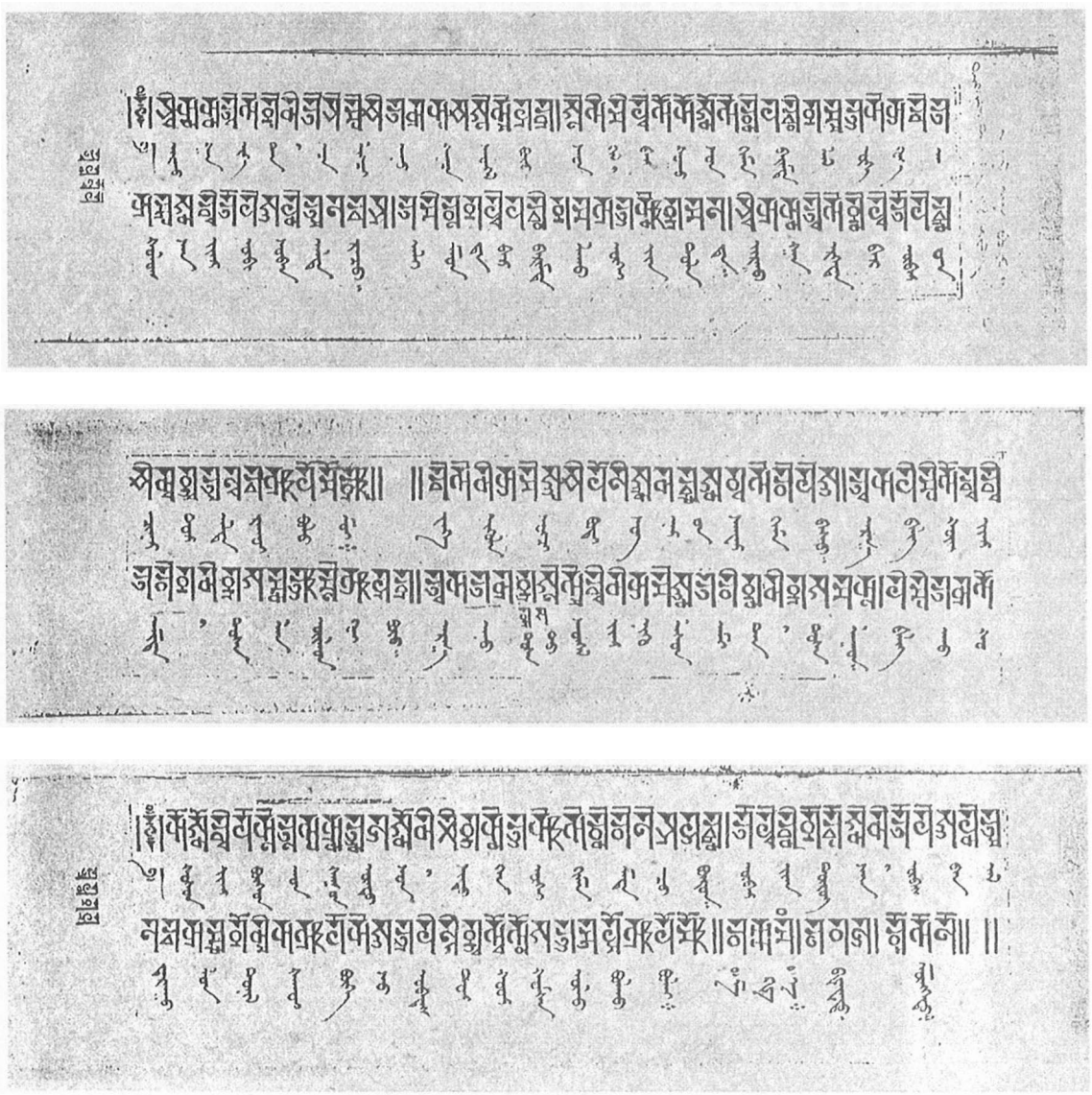


Figure 2: A manuscript containing Mongolian text in the Soyombo and Mongolian scripts (from Byambaa Ragchaagiin 2005: 64). Continued from figure 1.

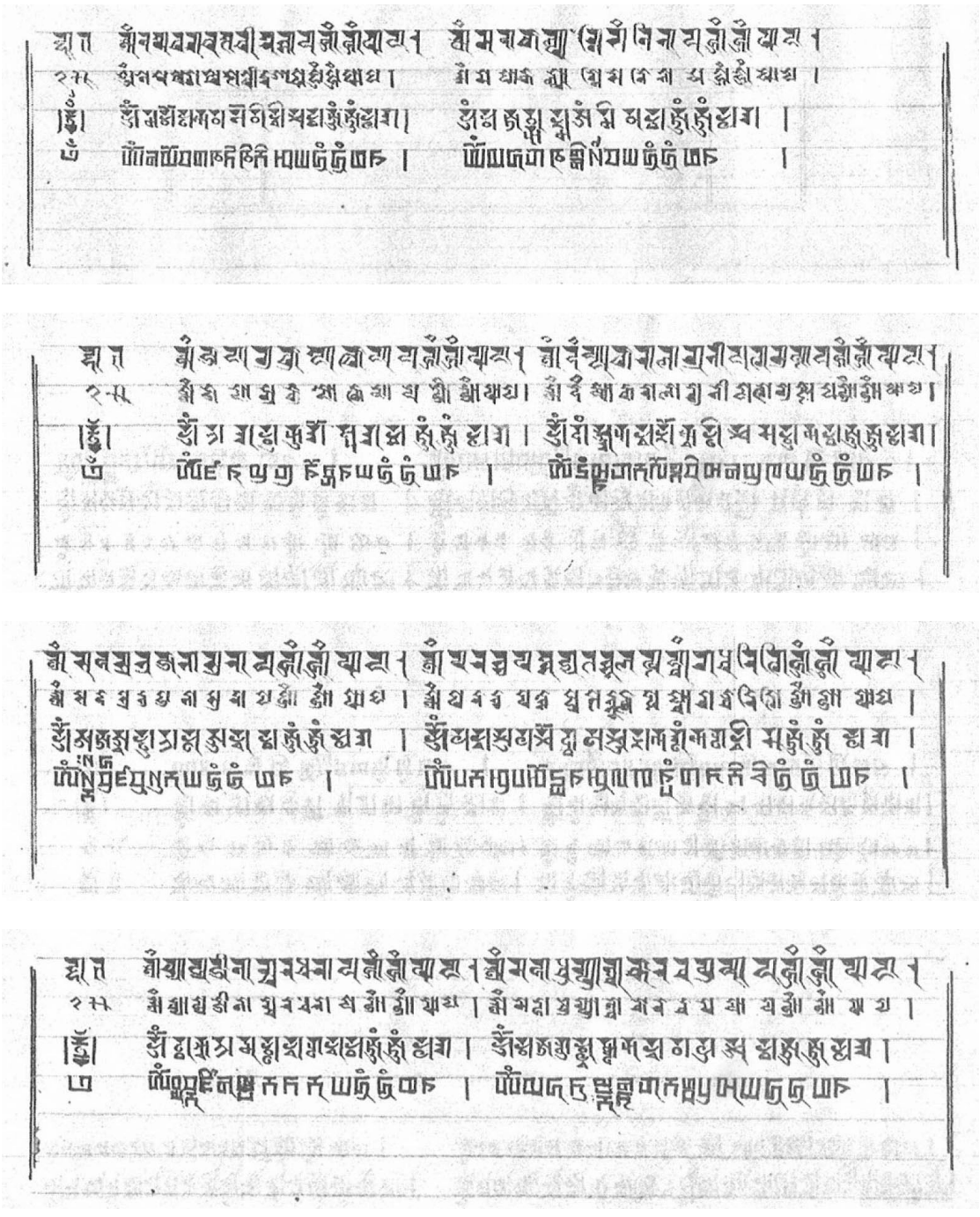


Figure 3: A manuscript containing text in the Ranjana, Vartu, Soyombo, and Zanabazar Square scripts (from Byambaa Ragchaagiin 2005: 103).

Figure 4: A manuscript from Mongolia with Sanskrit text written in Ranjana, Soyombo, Zanabazar Square, and Tibetan scripts (from Byambaa Ragchaagiin 2005: 97).

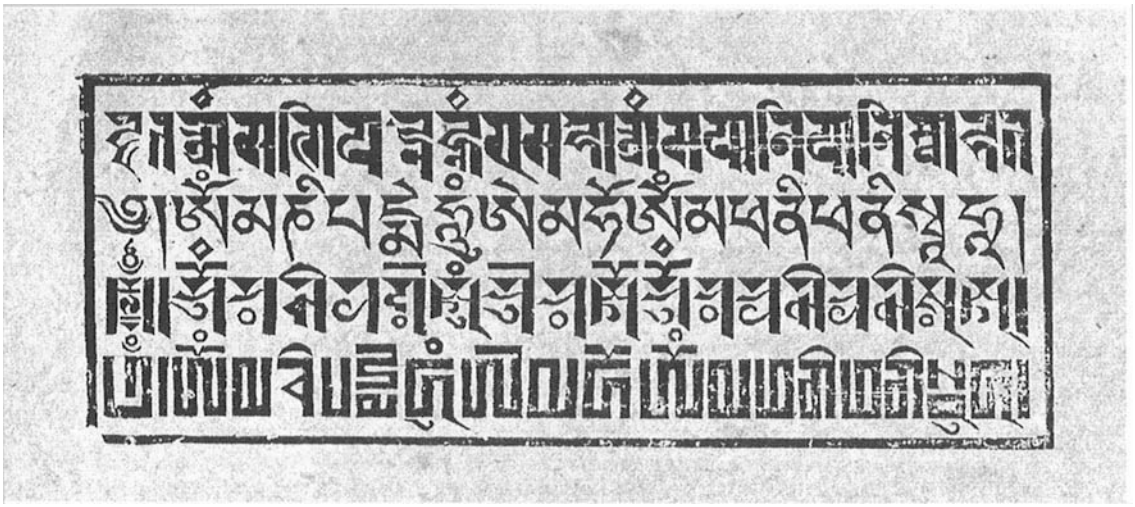


Figure 5: A manuscript from Mongolia with Sanskrit text written in the Ranjana, Tibetan, Soyombo, and Zanabazar Square scripts (from Byambaa Ragchaagiin 2005: 98).

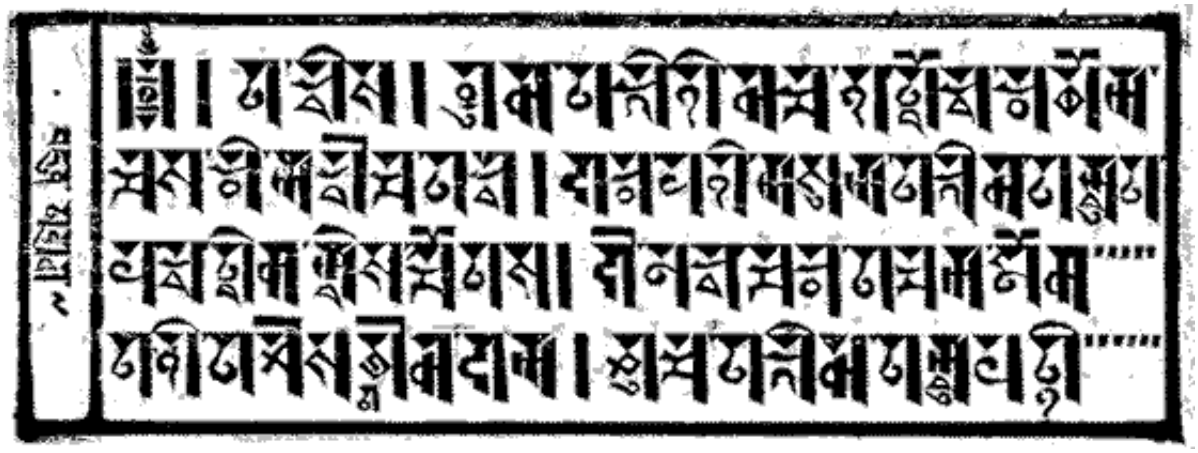


Figure 6: A Soyombo manuscript showing use of ˙ TSHEG (from Mongolwiki 2008).

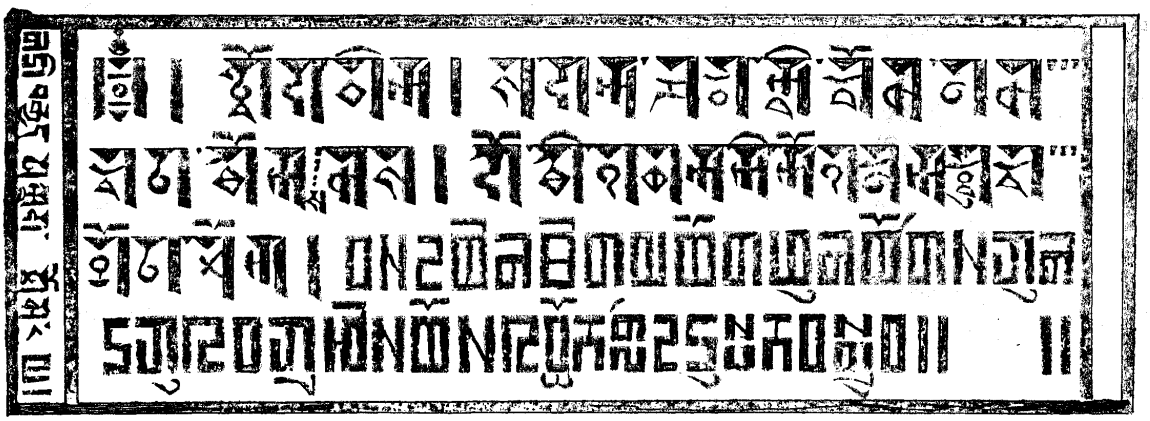


Figure 7: A manuscript in Soyombo and Zanabazar Square showing use of 𑩐 CLUSTER INITIAL LETTER SA and 𑩑 CLUSTER INITIAL LETTER RA (from Shagdarsürüng 2001: 173).

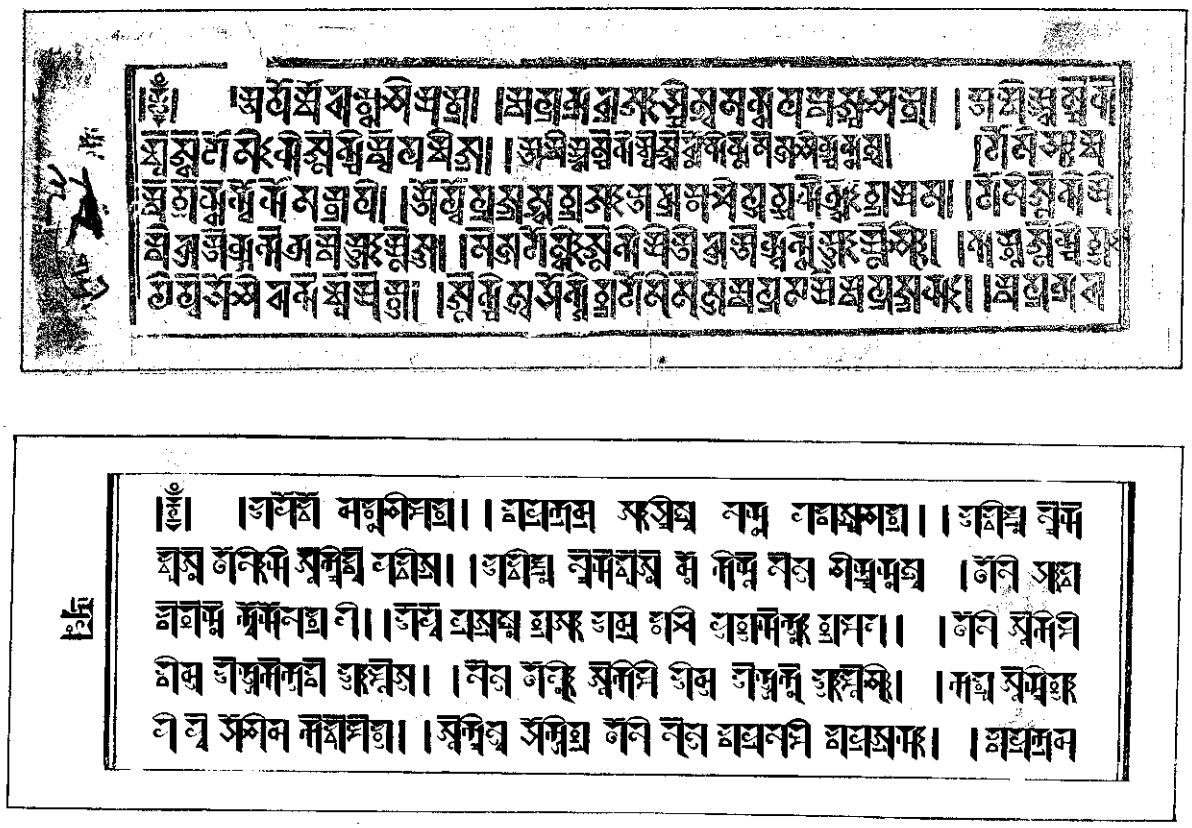


Figure 8: Soyombo manuscript (above) with what appears to be a digitized reproduction (from Shagdarsürüng 2001: 155).

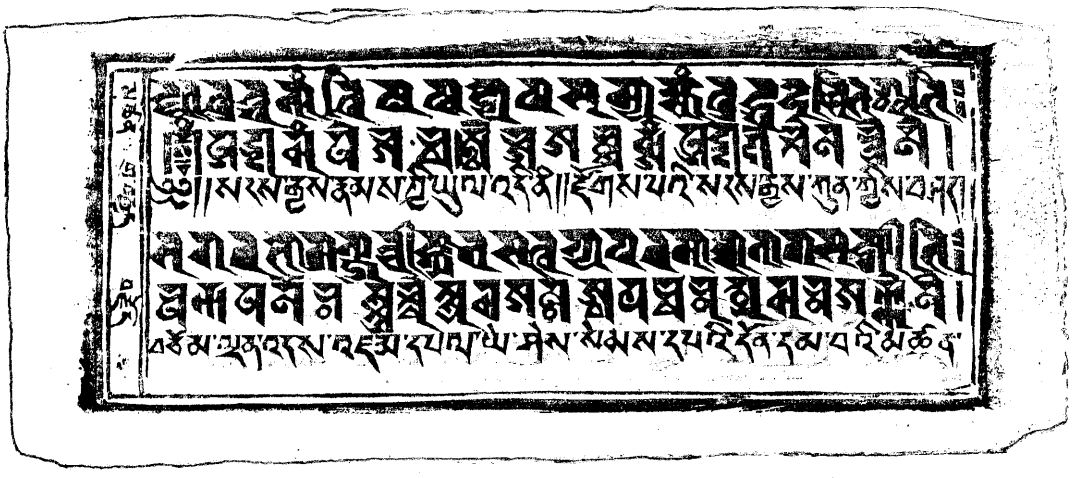


Figure 9: Manuscript folio in Ranjana, Soyombo, and Tibetan (from Shagdarsürüng 2001: 156).

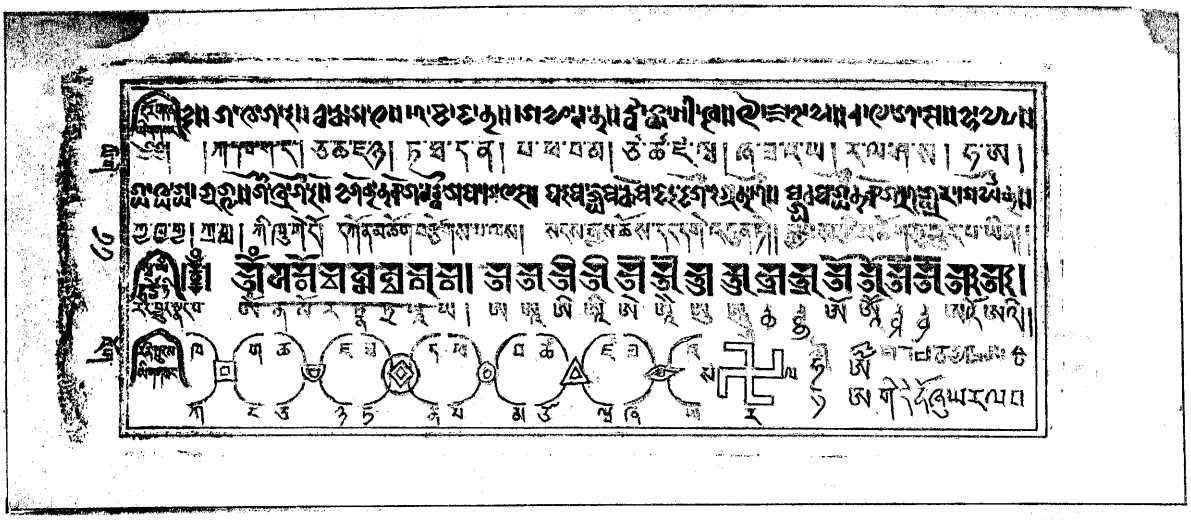


Figure 10: Manuscript folio in various Tibetan scripts and Soyombo (from Shagdarsürüng 2001: 154).

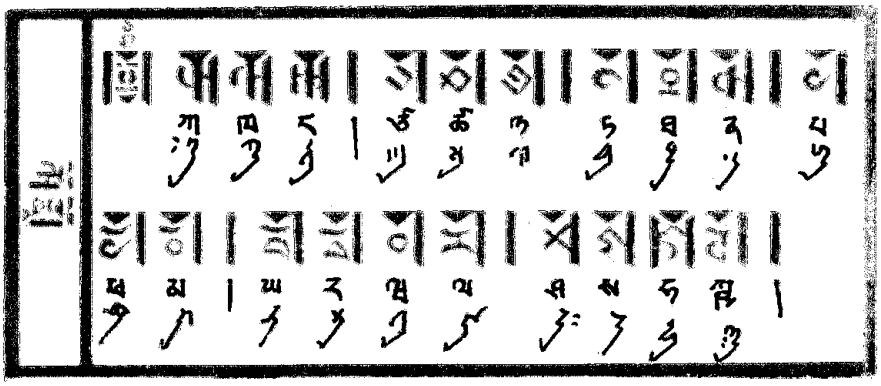


Figure 11: Manuscript folio showing corresponding Tibetan and Mongolian letters for Soyombo (from Shagdarsürüng 2001: 154).

Figure 12: Sample of text in Soyombo (from Rintschen 1952: 65).

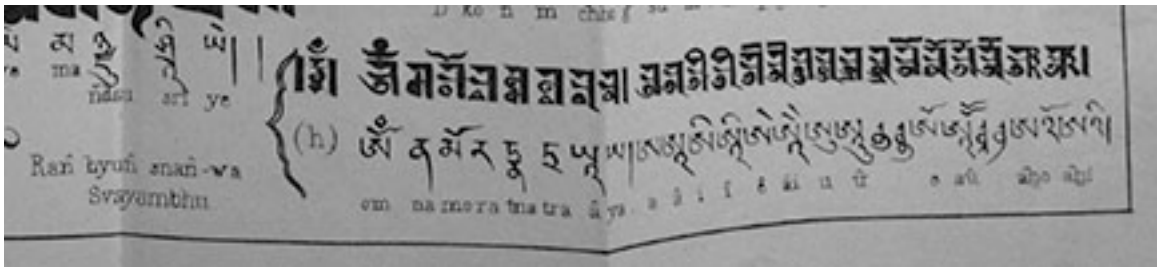


Figure 13: Excerpt showing Soyombo on an illustration of Tibetan sacred and ornamental scripts (from Dás 1888: plate V.)

Figure 14: The *tamga* of Bogd Khan from 1911. The seal contains text written in, from left to right, the Soyombo, Mongolian, and Phags-pa scripts

‘A’	‘B’	‘C’
source text	normative	normative
vertical top-to-bottom left-to-right	vertical top-to-bottom left-to-right	horizontal left-to-right top-to-bottom

Figure 15: Seal showing Soyombo text in idiosyncratic vertical orientation (from Shagdarsürüng 2001: 150). The seal reads: *naya selel kör yenu dzur kha grva chanun temedeg*. Agata Bareja-Starzyńska has related through personal communication that it is a seal of one of the Buddhist monastic colleges in the capital city before 1924. She notes that the text contains several errors and has suggested that it might more accurately be read as *neyislel küriyen-ü jiruqai grva-tshang-un temdeg* “The seal of the astrological college of the capital city”.

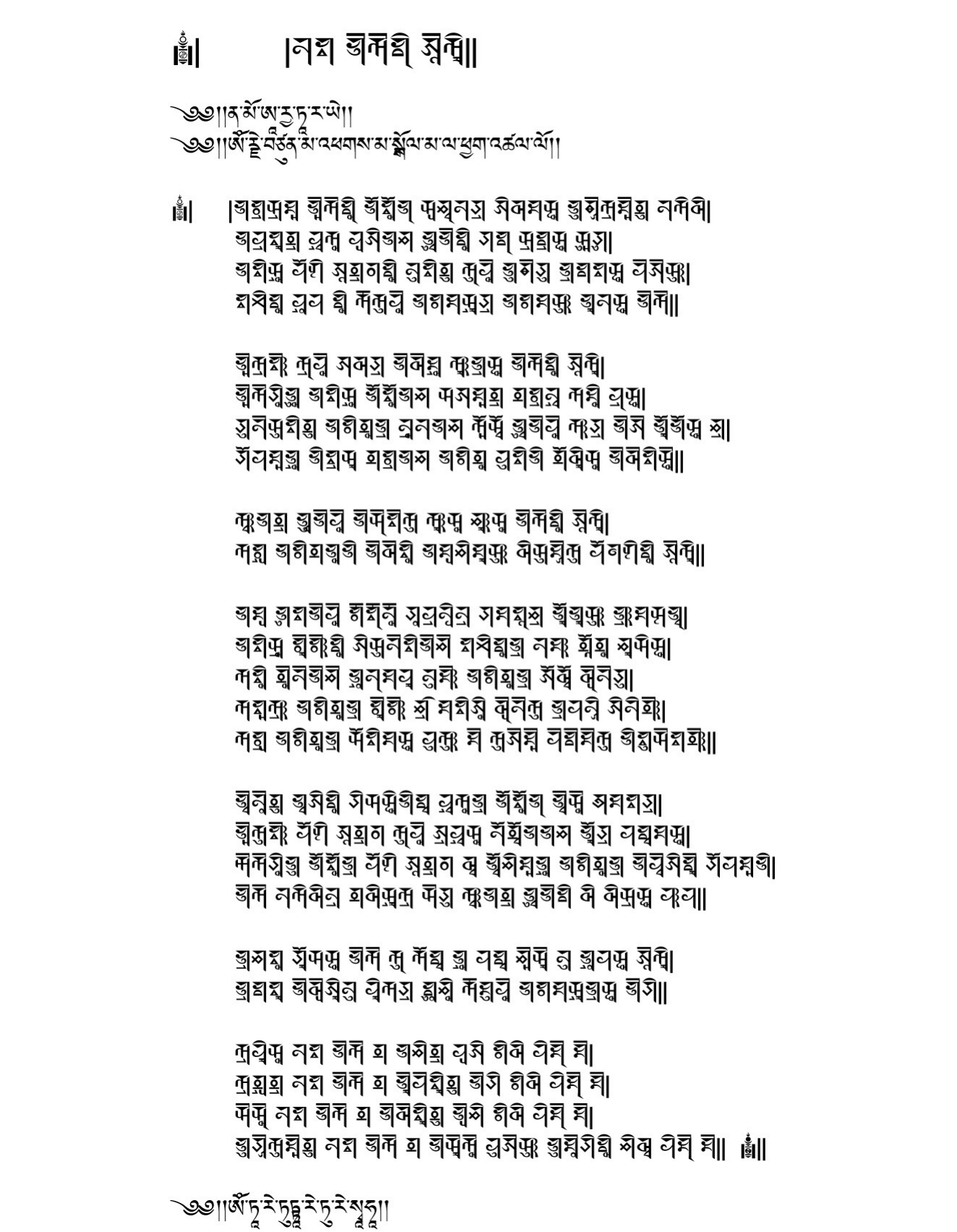


Figure 16: Poem for the Green Tara typeset in the digitized ‘JG Soyombo’ font. Source: `http://en.wikipedia.org/wiki/File:Green_Tara_Poem_Soyombo.jpg`.

Figure 17: Xylograph (block print) of a book cover in Soyombo, Zanabazar Square, Mongolian, and Cyrillic (from Boldsaikhan 2005: 330). The Zanabazar Square text represents Tibetan, the Mongolian represents Mongolian, and the Cyrillic represents Modern (Khalkha) Mongolian.

Figure 18: A pot from a Buddhist temple bearing the Soyombo inscription *oṃ mani padme huṃ*. Source `http://4.bp.blogspot.com/-L_GW-20upYY/UZGz7c8QPfI/AAAAAAAABRQ/Aqxeg0agJl8/s1600/IMG_2535.JPG`

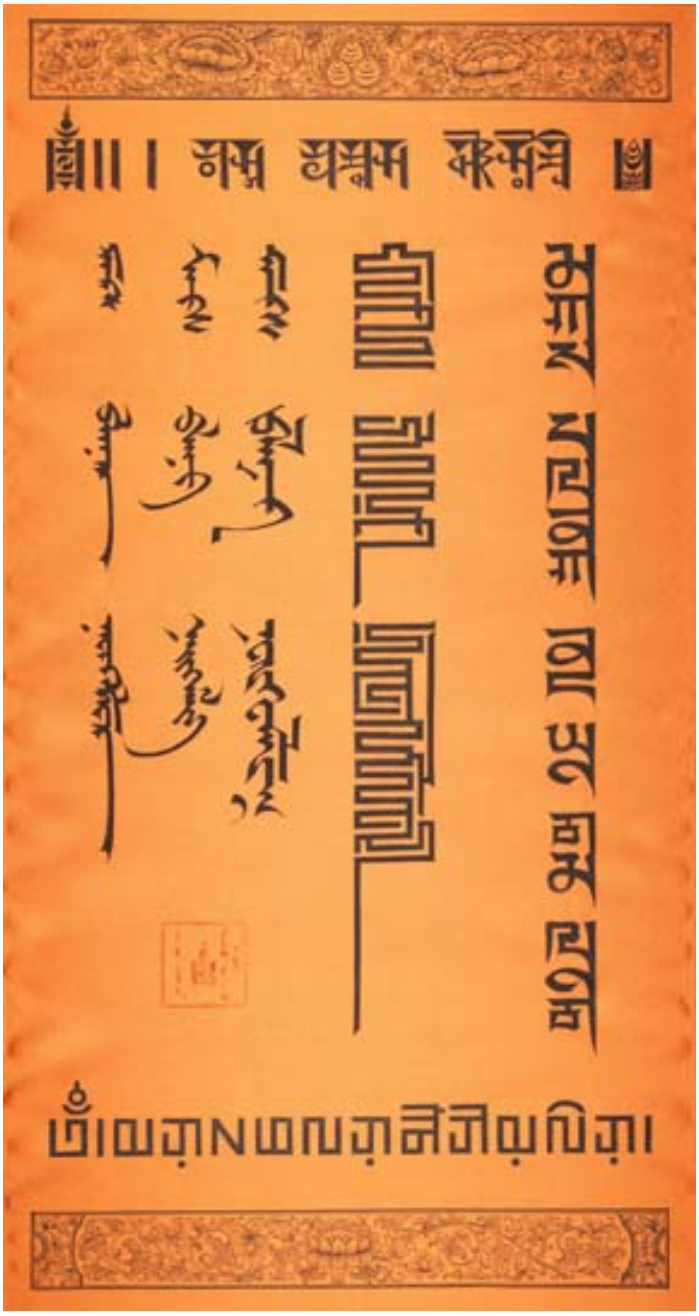


Figure 19: A souvenir for the Max Planck Institute containing seven Mongolian scripts.

Table 13: Variations of head marks used in Soyombo.

Table 14: Variations of terminal marks used in Soyombo.

Figure 20: Photograph of a chart of the Soyombo script (from "Histoire du livre" 2010).

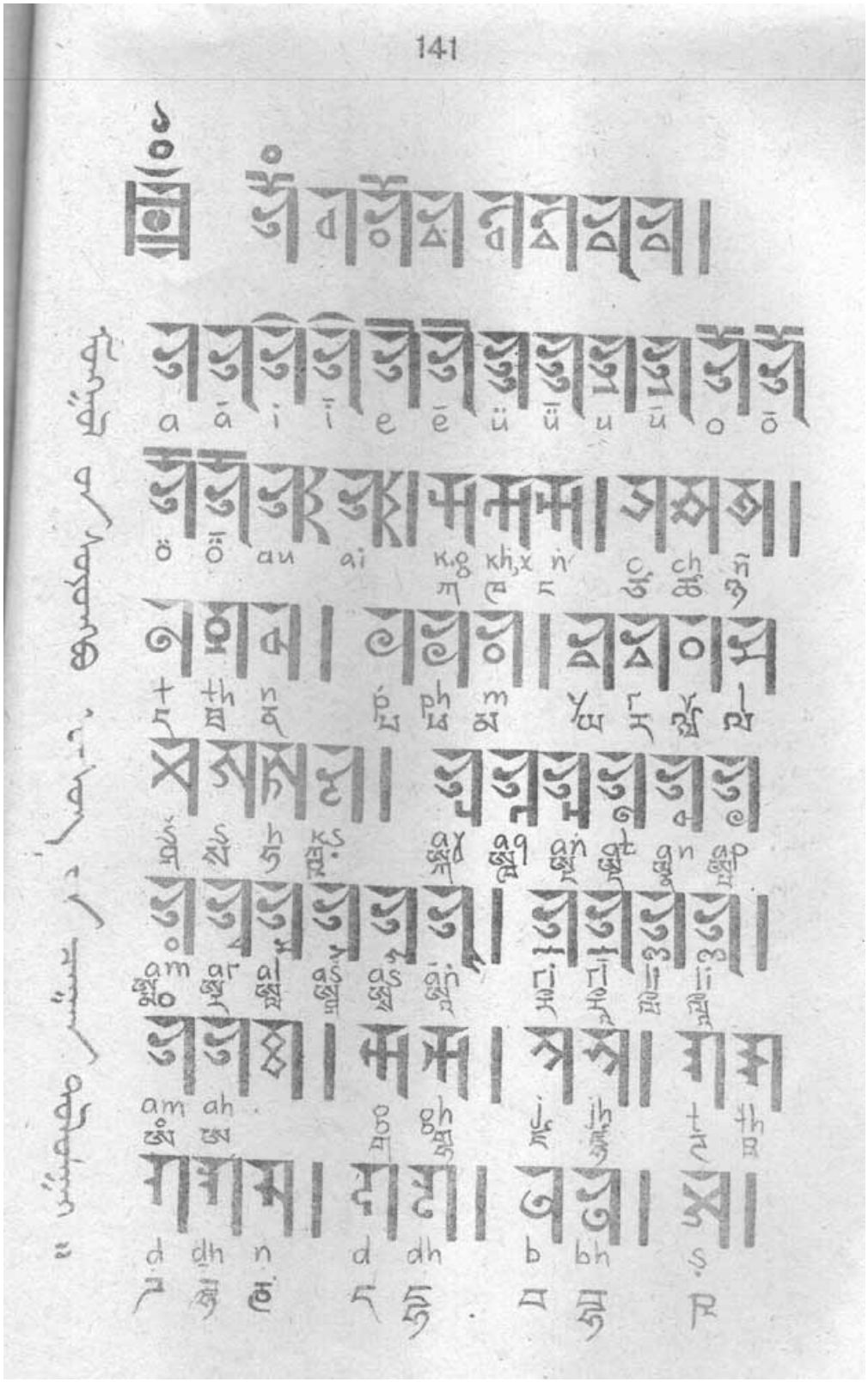


Figure 21: Characters of the Soyombo script (from Kapaj 2002).

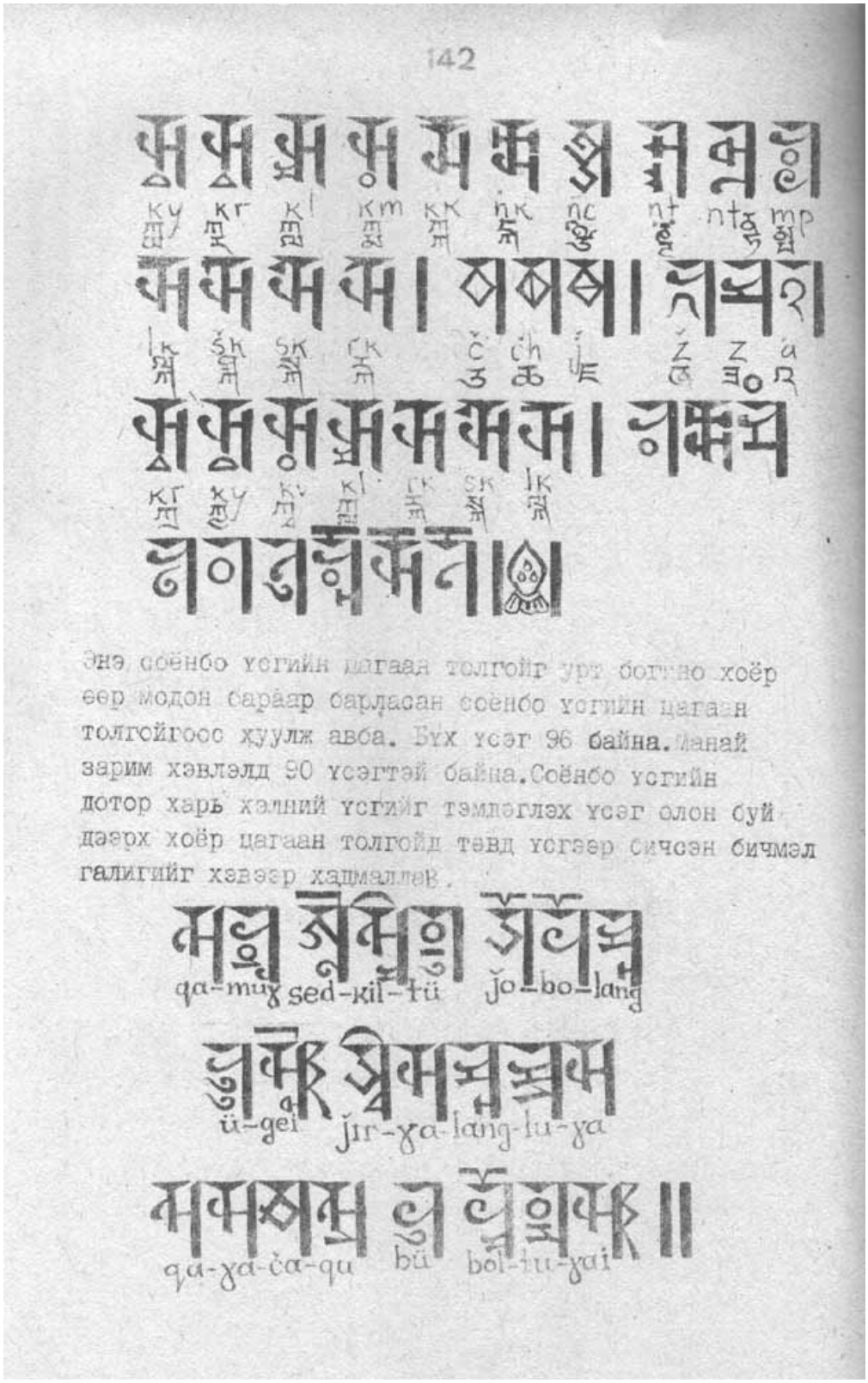


Figure 22: Characters of the Soyombo script (from Kapaj 2002).

Figure 23: Traditional chart of Soyombo (from Shagdarsürüng 2001: 152).

Хуудас эхлэсний тэмдэг. Номд бярга буюу эгчим ( ) -тэй адил үүрэгтэй. Соёмбо бичгийн энэ тэмдэг монголчуудын хувьд тусгаар тогтнолын бэлгэдэл болжээ.

Figure 24: Description of a Soyombo head mark (from Shagdarsürüng 2001: 133).

1. 

tl. A; **tc.** mong., tib., sans. a
Энэ хэлбэр нь а эгшигийн бие даасан (IF) буюу үгийн (зарим тохиолдолд үеийн) эхинд тохиолдоно. Жишээлбэл:

tl. A-li$_1$;
tc. a-li.

(Ø) **tl.** a$_o$; **tc.** a. Энэ нь а эгшигийн гол хэлбэр (MF) буюу үгийн (зарим тохиолдолд үеийн) дунд болоод адагт тохиолдох нууц буюу тэг (Ø) хэлбэр. Жишээлбэл:

tl. Ga$_o$-Ja$_o$r;
tc. ga-jar

1.a. 

tl. ā; **tc.** ā. Урт а эгшигийн бие даасан буюу (IF) хэлбэр. Жишээ нь:

tl. A-Da$_o$r;
tc. ā-dar.

tl. ¯a$_o$; **tc.** ā. Энэ нь угтаа эгшигийн уртын тэмдэг. а эгшигтэй тохиолдвол, тэрхүү а эгшиг нь нууц буюу "тэг" (Ø) хэлбэртэй байдаг. Жишээлбэл:

tl. Kā$_o$n;
tc. kān.

2. 

tl. I; **tc.** mong, tib., sans: i. Энэ нь і эгшигийн (IF) хэлбэр. Жишээлбэл:

tl. I-Te-Gel;
tc. i-te-gel.

tl. i$_1$; **tc.** i. Энэ нь і эгшигийн гол хэлбэрийн нэг буюу (MF$_1$) хэлбэр. Зөвхөн гийгүүлэгч (C)-ийн дараа буюу (C + i$_1$) нөхцөлд л тохиолдоно. Жишээлбэл:

tl. A-Ci$_1$-tu$_1$;
tc. a-či-tu.

tl. i$_2$; **tc.** i. Энэ нь і эгшигийн гол хэлбэрийн нэг буюу (MF$_2$) хэлбэр. Зөвхөн эгшиг (V)-ийн дараа буюу (V + i$_2$) нөхцөлд тохиолдолдоно. Жишээлбэл:

tl. Bol-Tu$_1$-Ga$_o$i$_2$;
tc. bol-tu-gai.

2.a. 

tl. I; **tc.** mong., sans: ī. Урт і эгшигийн бие даасан (IF) хэлбэр. Жишээлбэл:

tl. Ī-Me;
tc. ī-me.

Figure 25: Description of Soyombo vowels (from Shagdarsürüng 2001: 134).

tl. $\bar{\imath}_1$; **tc.** ī. Урт і эгшигийн гол буюу (MF) хэлбэр. Жишээ нь:

tl. Ü-Sü-G$\bar{\imath}_1$g;
tc. ü-sü-gīg.

tl. i_1i_2; **tc.** ī. Урт і эгшигийн гол буюу (MF) хэлбэр. Жишээ нь:

tl. Ma_0-Ni_1i_2n;
tc. ma-nīn 'харъяалахын тийн ялгалд буй *маани* '.

tl. $\bar{\imath}_1i_2$; **tc**. ī. "Маш урт"-аар дуудагдах і эгшигийн (MF) хэлбэр. Энд тусхайлан тэмдэглэх нь: үүний өмнө, "Тод бичиг"-ийн бөлөгт самгард хэлний эгшигийн *охор, урт, маш урт* гэсэн гурван янзын үргэлжлэц бүхий урт эшигийг хэрхэн тэмдэглэх тухай арай дэлгэрэнгүй өгүүлсэн билээ. Иймд энэхүү "маш урт" і нь зөвхөн тарнийн хэлнээ тохиолдоно. Жишээ нь:

tl. Bo-D$\bar{\imath}_1i_2$;
tc. bo-dī.

3. 

tl. E; **tc.** mong., tib.: e / sans.; (diphtongue) e. Энэ нь монгол, төвөд хэлний ерийн е, самгардын хос е эгшигийн бие даасан (IF) хэлбэр. Жишээлбэл:

tl. E-Ne;
tc. e-ne.

tl. e; **tc**. e. Энэ нь е эгшигийн гол буюу (MF) хэлбэр. Жишээлбэл:

tl. Sei-Te;
tc. sei-te.

3.a. 

tl. $\bar{E}$; **tc.** mong., ē; sans. (dipthongue) ai. Энэ нь монгол хэлний ерийн урт е, самгардын хос ai эгшигийн бие даасан буюу (IF) хэлбэр. Жишээлбэл:

tl. $\bar{E}$-ri_1 g-Ci_1;
tc. ē-rig-ci.

tl. ē; **tc**. ē. Энэ нь урт е эгшигийн гол буюу (MF) хэлбэр. Жишээ нь:

tl. Ke-Jē;
tc. ke-jē.

4. 

tl. Ü; **tc.** mong., tib. ü; sans. u. Энэ нь монгол, төвөд хэлний ü, самгард хэлний u эгшигийн бие даасан (IF) хэлбэр. Жишээлбэл:

tl. Ü-Nēr;
tc. ü -nēr.

Figure 26: Description of Soyombo vowels (from Shagdarsürüng 2001: 135).

tl. ü; **tc.** ü. Энэ нь ü эгшигийн гол буюу (MF) хэлбэр. Жишээ нь:

tl. Ün-Dü-Sü;
tc. ün-dü-sü.

4.a. 

tl. Ū; **tc.** mong. ū, sans. ū. Монгол хэлний урт ū, самгардын урт ū эгшигийн бие даасан буюу (IF) хэлбэр. Жишээ нь:

tl. Ūr;
tc. ūr.

tl. ū; **tc.** ū. Монгол хэлний урт ū, самгардын урт ū эгшигийн гол буюу (MF) хэлбэр. Жишээ нь:

tl. Ü-Jū-Lüg-Sen;
tc. ü-jū-lüg-sen.

5. 

tl. U; **tc.** mong. u. Монгол хэлний u эгшигийн бие даасан буюу (IF) хэлбэр. Жишээ нь:

tl. U-sun;
tc. u-sun.

tl. u_1; **tc.** u. Энэ нь u эгшигийн гол буюу (MF) хэлбэрийн нэг. Зөвхөн гийгүүлэгчийн дараа буюу (C + u_1) нөхцөлд тохиолдоно. Жишээ нь:

tl. Ya_0-Bu_1-Da_0l;
tc. ya-bu-dal

tl. u_2; **tc.** u. Энэ нь u эгшигийн гол буюу (MF) хэлбэрийн нэг. Зөвхөн эгшигийн дараа буюу (V + u_2) нөхцөлд тохиолдоно. Жишээ нь:

tl. Jo-Ri_1u_2l-Ji_1u_2;
tc. jo-riul-ju.

tl. Ya_0-Ga_0u_2l;
tc. ya-gaul 'шалтгаан, үндэс '[15]

5.a. 

tl. ū; **tc.** mong., sans. ū. Энэ нь монгол болон самгардын урт ū эгшигийн бие даасан буюу (IF) хэлбэр. Жишээ нь:

---

[15] Соёмбо бичгийн *"Итгэл"*-д төвөд хэлний tib. rgyu гэдэг үгийг ya-gaul (<Mo. *yaγaγul) хэмээн орчуулсан нь буй. Энэ үгийг Ойродын Зая Бандидын орчуулсан тод *"Итгэл"*-д ündüsün гэж, буриад бичмэлд siltaγan хэмээн орчуулжээ. Энэ тухай G. Kara, *Un texte mongol en écriture soyombo*, - AOH, Tomus. IX, Fasc. 1, Budapest, 1959, pp. 1-38 болон Chagdarsureng, *Sur quelques traductions mongoles du "Natha"*, - Studia Mongolica, Tom. 2 (10), Fasc. 11, Ulan-Bator, 1975, p. 183 (N. 86) -д үзмүү.

Figure 27: Description of Soyombo vowels (from Shagdarsürüng 2001: 136).

tl. ū-Da$_0$m;
tc. ū-dam.

tl. ū; **tc.** ū. Урт ū эгшигийн (MF) хэлбэр. Жишээ нь:

tl. A-Rūl;
tc. ā-rūl.

6. 

tl. O; **tc.** mong., tib. o, sans. diphtongue o. Энэ нь монгол, төвөд хэлний ерийн о эгшиг, самгард хэлний хос о эгшигийн бие даасан (IF) хэлбэр. Жишээлбэл:

tl. O-Ron;
tc.o-ron.

tl. o; **tc.** o. Энэ нь о эгшигийн гол буюу (MF) хэлбэр. Жишээ нь:

tl. Oi-Ro;
tc.oi-ro.

6.a. 

tl. Ō; **tc.** mong. ō, sans. diphtongue au. Энэ нь монгол хэлний урт о болон самгард хэлний au хос эгшигийн бие даасан (IF) хэлбэр. Жишээлбэл:

tl. ō-sor;
tc.ō-sor.

tl. ō; **tc.** ō. Энэ нь урт о эгшигийн гол буюу (MF) хэлбэр. Жишээ нь:

tl. Kō-son;
tc. kō-son.

7. 

tl. Ö; **tc.** mong. ö. Энэ нь ö эгшигийн бие даасан (IF) хэлбэр бөгөөд зөвхөн монгол хэлнээ тохиолдоно. Жишээлбэл:

tl. Ö-mög;
tc. ö-mög.

tl. ö; **tc.** ö. Монгол хэлний ö эгшигийн гол буюу (MF) хэлбэр. Жишээ нь:

tl. Ö-mö-nö;
tc. ö-mö-nö.

7.a. 

tl. Ō; **tc.** mong. ō. Монгол хэлний урт ö эгшигийн бие даасан (IF) хэлбэр. Жишээлбэл:

tl. Ōr;
tc. ōr.

Figure 28: Description of Soyombo vowels (from Shagdarsürüng 2001: 137).

tl. ō; **tc.** ō. Урт ö эгшигийн гол буюу (MF) хэлбэр. Жишээ нь:

tl. Bōr;
tc. bōr.

8. 

tl. Au_2. **tc.** mong. au буюу diphtongue 'waw '. Энэхүү хос эгшигийн талаар цагаан толгойн жагсаалтын 5-ын u_2-аас лавшруулан үзнэ үү.

9. 

tl. Ai_2; **tc.** mong. ai буюу diphtongue 'yod'. Энэхүү хос эгшигийн талаар цагаан толгойн жагсаалтын 2-ын i_2-аас лавшруулан үзнэ үү.

10. 

tl. mong. G(a), tib., sans: K(a); **tc.** mong. ga, tib., sans; ka. Энэ нь монгол хэлний g(a), төвөд, самгардын к(а) гийгүүлэгчийн бие даасан (IF) хэлбэр.

tl. Ga_0-Ja_0r;
tc. ga-jar.

tl. g; **tc.** g. Энэ нь монгол хэлний гийгүүлэгчийн гол (MF) хэлбэр бөгөөд амьгүй буюу дэвсгэрлэж ордог хэлбэр. Жишээ нь:

tl. Ka_0-mu_1g;
tc. ka-mug.

tl. sans. k; **tc.** sans. k. Зөвхөн самгард хэлнээ тохиолдоно. Жишээ нь:

tl. kKa_0;
tc. kka.

11. 

tl. mong. K(a), tib., sans.: KH(a); **tc.** mong. ka, tib., sans.: kha.
Монгол хэлний ка; төвөд, самгард хэлний гийгүүлэгчийн бие даасан (IF) хэлбэр. Жишээлбэл:

tl. Ka_0-mu_1g;
tc. ka-mug.

tl. (k) / kh; **tc.** (k) / kh. Энэ нь (к) / kh гийгүүлэгчийн гол буюу (MF) хэлбэр. Монгол бичгийн хэлний зөв бичих зүйд энэ гийгүүлэгч дэвсгэрлэж ордоггүй. Харин Тод бичигт дэвсгэрлэж ордог ёсон буй.

12. 

tl. tib., sans: N(a); **tc.** tib., sans.: na. Энэ нь төвөд, самгард хэлний гүн угийн гийгүүлэгчийн бие даасан буюу (IF) хэлбэр. Монгол хэлнээ эл гийгүүлэгч нь үгийн болон үеийн эхинд тохиолддоггүй учир монгол бичвэрт ийм хэлбэр хэрэглэгдэхгүй.

tl. mong. n̄ / tib., sans.: ṅ; **tc.** mong. n̄ / tib., sans.: ṅ. Энэ нь n̄/ṅ гийгүүлэгчийн гол буюу (MF) хэлбэр. Монгол хэлнээ зөвхөн үеийн эцэст тохиолдоно.

tl. Ji_1 r-Ga_0-La_0 n̄;
tc. jir-ga-lan̄.

Figure 29: Description of Soyombo consonants (from Shagdarsürüng 2001: 138).

13. **tl.** mong. J(a), tib., sans.: C(a).
tc. mong. ja (i -гээс бусад эгшигийн өмнө) / ja(i эгшигийн өмнө), tib., sans.: ca. Энэ нь монгол хэлний ja/ ja, төвөд, самгардын са гийгүүлэгчийн бие даасан (IF) хэлбэр. Жишээлбэл:


tl. Ja_0r-Li_1g;
tc.jar-lig.


tl. Ji_1 r-Ga_0 -La_0 n;
tc. jir-ga-lan.

14. **tl.** mong. C(a), tib., sans.: CH(a);
tc. mong. ca (i -гээс бусад эгшигийн өмнө) / ča (зөвхөн i эгшигийн өмнө); tib., sans.: cha. Энэ нь монгол хэлний са / ča, төвөд, самгард хэлний cha гийгүүлэгчийн бие даасан (IF) хэлбэр. Жишээлбэл:


tl. Ca_0γ -ga_0-ca_0;
tc. ca-ga-ca (<Mo. čaγ-ača).


tl. Ci_1-Na_0r;
tc. či-nar.


tl. Ci_1u_2l-Ga_0-Nu_1;
tc. čiul-ga-nu
(<Mo. ciγulγan-u).

15. **tl.** tib., sans.: $\bar{N}$(a);
tc. tib., sans.: ña. Төвөд, самгард хэлний гийгүүлэгчийн бие даасан (IF) хэлбэр.

16. **tl.** mong. D(a), tib., sans.: t(a);
tc. mong. da, tib., sans. ta. Монгол хэлний da, төвөд самгард хэлний ta гийгүүлэгчийн бие даасан буюу (IF) хэлбэр. Жишээлбэл:


tl. Da_0-ga_0n;
tc. da-gan (< Mo. daγa-).

tl. mong. d, tib., sans.: t;
tc.mong. d, tib., sans.: t. Монгол хэлний d, төвөд, самгард хэлний t гийгүүлэгчийн гол буюу (MF) хэлбэр. Жишээлбэл:


tl. Ki-Ged;
tc.ki-ged (< Mo. kiged).

17. **tl.** mong. T(a); tib., sans.: TH(a);
tc. mong. ta, tib., sans.: tha. Монгол хэлний ta, төвөд, самгард хэлний tha гийгүүлэгчийн бие даасан буюу (IF) хэлбэр. Жишээлбэл:


tl. Ta_0-Ra_0g-Sa_0n;
tc. ta-rag-san (< Mo. tara-).

18. **tl.** mong., tib., sans.: N(a);
tc. mong., tib., sans.: na. Монгол, төвөд, самгард хэлний na гийгүүлэгчийн бие даасан (IF) хэлбэр. Жишээ нь:


tl. Na_0-Ra_0;
tc. na-ra.

Figure 30: Description of Soyombo consonants (from Shagdarsürüng 2001: 139).

tl. n.
tc. n. Монгол, төвөд, самгард хэлний n гийгүүлэгчийн гол буюу (MF) хэлбэр. Жишээ нь:

tl. Na_0-Ya_0n;
tc. na-yan.

19. 

tl. mong. B(a), tib., sans.: P(a);
tc. mong. ba, tib., sans.: pa. Монгол хэлний ba, төвөд, самгард хэлний pa гийгүүлэгчийн бие даасан буюу (IF) хэлбэр. Жишээлбэл:

tl. Ba_0-Ri_1-
tc. ba-ri- (Mo. bari-).

tl. mong. b, tib., sans.: p;
tc. mong. b, tib., sans.: p.
Монгол хэлний b, төвөд, самгард хэлний p гийгүүлэгчийн гол буюу (MF) хэлбэр. Жишээлбэл:

tl. El-Deb;
tc. el-deb.

20. 

tl. mong. P(a), tib., sans.: PH(a);
tc. mong. pa, tib., sans.: pha. Монгол хэлний pa, төвөд, самгард хэлний pha гийгүүлэгчийн бие даасан буюу (IF) хэлбэр нь.

21. 

tl. mong., tib., sans.: M(a);
tc. mong., tib., sans.: ma. Монгол, төвөд, самгард хэлний ma гийгүүлэгчийн бие даасан (IF) хэлбэр. Жишээ нь:

tl. Ma_0-$Ši_1$;
tc. ma-si.

tl. m;
tc. m. Энэ нь m гийгүүлэгчийн гол буюу (MF) хэлбэр. Жишээлбэл:

tl. Er-Dem;
tc. er-dem.

tl. tib., sans.: (O)m;
tc. tib., sans. : (o)m. Энэ нь төвөд, самгард бичвэрт, тарнийн хэлнээ тохиолдоно. Уугуул монгол хэлнээ тохиолдохгүй.

22. 

tl. mong., tib.: Y(a), sans.: semi-vowel Y(a);
tc. mong., tib., ya, sans.: semi-vowel ya. Монгол, төвөд хэлний y гийгүүлэгч, самгард хэлний заримдаг (тал) y эгшигийн бие даасан (IF) хэлбэр.

tl. Ya_0-Ga_0u_2l;
tc. ya-gaul (Энэ бөлөгийн 15-р зүүлтээс тодруулна уу.)

Figure 31: Description of Soyombo consonants (from Shagdarsürüng 2001: 140).

tl. tib., sans.: y(a);
tc. tib., sans.: ya. Хэдийгээр энэ нь монгол, төвөд хэлний хувьд у гийгүүлэгч, самгард хэлний тухайд заримдаг (буюу тал) у эгшигийн гол (MF) хэлбэр мөн боловч хэрэглээ нь хязгаартай. Үүнд: соёмбо бичигийн цагаан толгойгоос үзэхэд нэгд, монгол хэлний дэвсгэр гийгүүлэгчид багтаагаагүй ажээ; хоёрт, төвөд болон самгард хэлнээ зүүлтэт үсэгт тохиолдохоор тэмдэглэжээ.

tl. Kya_0;
tc.kya.

23. 

tl. mong., tib., sans.: R(a);
tc. mong., tib.: ra, sans.: semi-vowel ra. Энэ нь монгол, төвөд хэлний ra гийгүүлэгч, самгард хэлний ra хэмээх заримдаг (буюу тал) эгшигийн бие даасан (IF) хэлбэр. Жишээлбэл:

tl. Ja_0u_2-Ra_0-Da_0;
tc.jau-ra-da (< Mo. jaγuradu ~ jaγurada)

tl. mong., tib.: r, sans.: (semi-vowel) r;
tc. mong., tib.: r, sans. (semi-vowel) r. Монгол, төвөд хэлний r гийгүүлэгч, самгард хэлний r хэмээх заримдаг (буюу тал) эгшигийн гол (MF) хэлбэр нь. Монгол хэлнээ дэвсгэрлэж ороходоо энэхүү гол хэлбэрийг дундуур нь таллаж, тал гурвалжин хэлбэрээр тэмдэглэнэ. Жишээ нь:

tl. Er-Dem;
tc. er-dem.

Харин төвөд, самгард хэлний давхар үсгийн тухайд, r гийгүүлэгч (буюу заримдаг эгшиг)-ийн гол хэлбэрийг хялбарчлахгүй, яг хэвээр нь толгой, зүүлт болгож бичдэг. Жишээлбэл:

tl. Kra_0;
tc. kra.

tl. rKa_0;
tc.rka.

24. 

tl. tib., V(a), sans.: (semi-vowel): V(a);
tc. tib. va, sans.: (semi-vowel): va. Төвөдийн va гийгүүлэгч, самгардын заримдаг (буюу тал) va эгшигийн бие даасан (IF) хэлбэр нь.

tl. tib., sans.: v(a);
tc. tib., sans.: va.
Энэ нь төвөд, самгардын давхар үсэгт v(a) зүүлт болж ордог гол (MF) хэлбэр. Үүнийг эл бөлөгийн No. 63-аас тодруулан үзмүү.

25. 

tl. mong., tib.: L(a), sans.: (semi-vowel): L(a);
tc. mong., tib.: la. sans.: (semi-vowel): la. Энэ нь Монгол, төвөд хэлний l гийгүүлэгч, самгард хэлний l хэмээх заримдаг (буюу тал) эгшигийн бие даасан (IF) хэлбэр нь. Жишээлбэл:

tl. Nom-La_0g-Sa_0n;
tc. nom-lag-san.

Figure 32: Description of Soyombo consonants (from Shagdarsürüng 2001: 141).

tl. mong., tib.: l, sans.: (semi-vowel): l;
tc. mong., tib.: l, sans.: (semi-vowel): l.
Энэ нь монгол, төвөд хэлний l гийгүүлэгч, самгардын заримдаг (буюу тал) ᠯ эгшигийн гол (MF) хэлбэр. Жишээлбэл:

tl. Bel-ge;
tc. bel-ge.

Харин энэхүү (MF) хэлбэр нь төвөд, самгардийн давхар үсэгт толгой болох ороходоо ганц хөндлөн зураас болж ордог. Тухайлбал:

tl. lKa$_0$;
tc. lka.

26. 

tl. mong. Š(a), tib., sans. Ç(a) / Ś(a).
tc. mong. ša, tib., sans.: ça / śa. Энэ нь монгол хэлний ša, төвөд, самгард хэлний ça буюу śa гийгүүлэгийн бие даасан (IF) хэлбэр. Жишээлбэл:

tl. Teg-Še;
tc. teg-še.

tl. mong. š;
tc. mong. š. Энэ нь š гийгүүлэгчийн монгол хэлэнд тохиолдох гол буюу (MF) хэлбэр. Соёмбо бичигийн цагаан толгойноос үзэхэд үүнийг монгол хэлний дэвсгэр үсэгт багтаажээ (Тод бичигт ч бас ийм буй). Тухайлбал:

tl. Aš;
tc. aš.

Самгард хэлнээ, давхар үсэгт энэ гийгүүлэгчийг толгой болгож залгахдаа (MF) хэлбэрийг таллаж арай хялбарчлан тэмдэглэдэг бөлгөө.

tl. çKa$_0$;
tc. çka.

27. 

tl. mong., tib., sans. S(a);
tc. mong., tib., sans. sa. Sa гийгүүлэгчийн бие даасан (IF) хэлбэр.

tl. Sa$_0$-Ra$_0$;
tc. sa-ra.

tl. s; **tc.** s. Энэ нь sa гийгүүлэгчийн гол буюу (MF) хэлбэр. Монгол хэлний тухайд дэвсгэрлэж орох хэлбэр гэсэн үг. Жишээлбэл:

tl. Te-güs;
tc. te-güs.

Төвөд, самгард хэлний давхар үсэгт sa толгойг бичихдээ бяцхан гурвалжин болгож доорхи байдлаар тэмдэглэнэ.

Figure 33: Description of Soyombo consonants (from Shagdarsürüng 2001: 142).

tl. sKa$_o$;
tc. ska.

28. 

tl. mong., tib., H(a), sans.: (semi-vowel): H(a);
tc. mong., tib. ha, sans.: (semi-vowel): ha. Энэ нь монгол, төвөд хэлний ha гийгүүлэгч, самгард хэлний заримдаг (буюу тал) ha эгшигийн бие даасан (IF) хэлбэр.

29. 

tl. sans. (compound consonant): KS(a);
tc. sans. (compound consonant): ksa.
Самгард хэлний нийлмэл гийгүүлэгч ksa -гийн бие даасан (IF) хэлбэр.

30. 

tl. mong., Ag, tib., sans.: Ak;
tc. mong. ag, tib., sans.: ak. Монгол хэлний дэвсгэр -g гийгүүлэгчийн хэлбэр (Уг үсгийн талаар жагсаалтын No. 10-аас тодруулан үзнэ үү).

31. 

tl. mong. Ak, tib., sans.: Akh;
tc. mong. ak, tib., sans. akh. Дэвсгэрлэж орох -k гийгүүлэгчийн хэлбэр (Уг үсгийн талаар жагсаалтын No. 11-аас тодруулан үзнэ үү).

32. 

tl. mong. An̲, tib., sans.: Aṅ;
tc. mong. an̲, tib., sans. aṅ. Монгол хэлний дэвсгэр -n гийгүүлэгчийн хэлбэр (Уг үсгийн талаар жагсаалтын No. 12-аас тодруулан үзнэ үү).

33. 

tl. mong. Ad, tib., sans.: At;
tc. mong. ad, tib., sans.: at; Монгол хэлний дэвсгэр -d гийгүүлэгчийн хэлбэр (Уг үсгийн талаар жагсаалтын No. 16-аас тодруулан үзнэ үү).

34. 

tl. mong. An, tib., sans.: An;
tc. mong., an, tib., sans.: an. Монгол хэлний дэвсгэр -n гийгүүлэгчийн хэлбэр (Уг үсгийн талаар жагсаалтын No. 18-аас тодруулан үзнэ үү).

35. 

tl. mong. Ab, tib., sans.: Ap;
tc. mong. ab; tib., sans.: ap. Монгол хэлний дэвсгэр -b гийгүүлэгчийн хэлбэр (Уг үсгийн талаар жагсаалтын No. 19-аас тодруулан үзнэ үү).

36. 

tl. mong., tib., sans.: Am;
tc. mong., tib., sans.: am. Монгол хэлний дэвсгэр -m гийгүүлэгчийн хэлбэр (Уг үсгийн талаар жагсаалтын No. 21-аас тодруулан үзнэ үү).

37. 

tl. mong., tib., sans.: Ar;
tc. mong., tib., sans.: ar. Монгол хэлний дэвсгэр -r гийгүүлэгчийн хэлбэр (Уг үсгийн талаар жагсаалтын No. 23-аас тодруулан үзнэ үү).

38. 

tl. mong., tib., sans.: Al;
tc. mong., tib., sans.: al. Монгол хэлний дэвсгэр -l гийгүүлэгчийн хэлбэр (Уг үсгийн талаар жагсаалтын No. 25-аас тодруулан үзнэ үү).

Figure 34: Description of Soyombo consonants (from Shagdarsürüng 2001: 143).

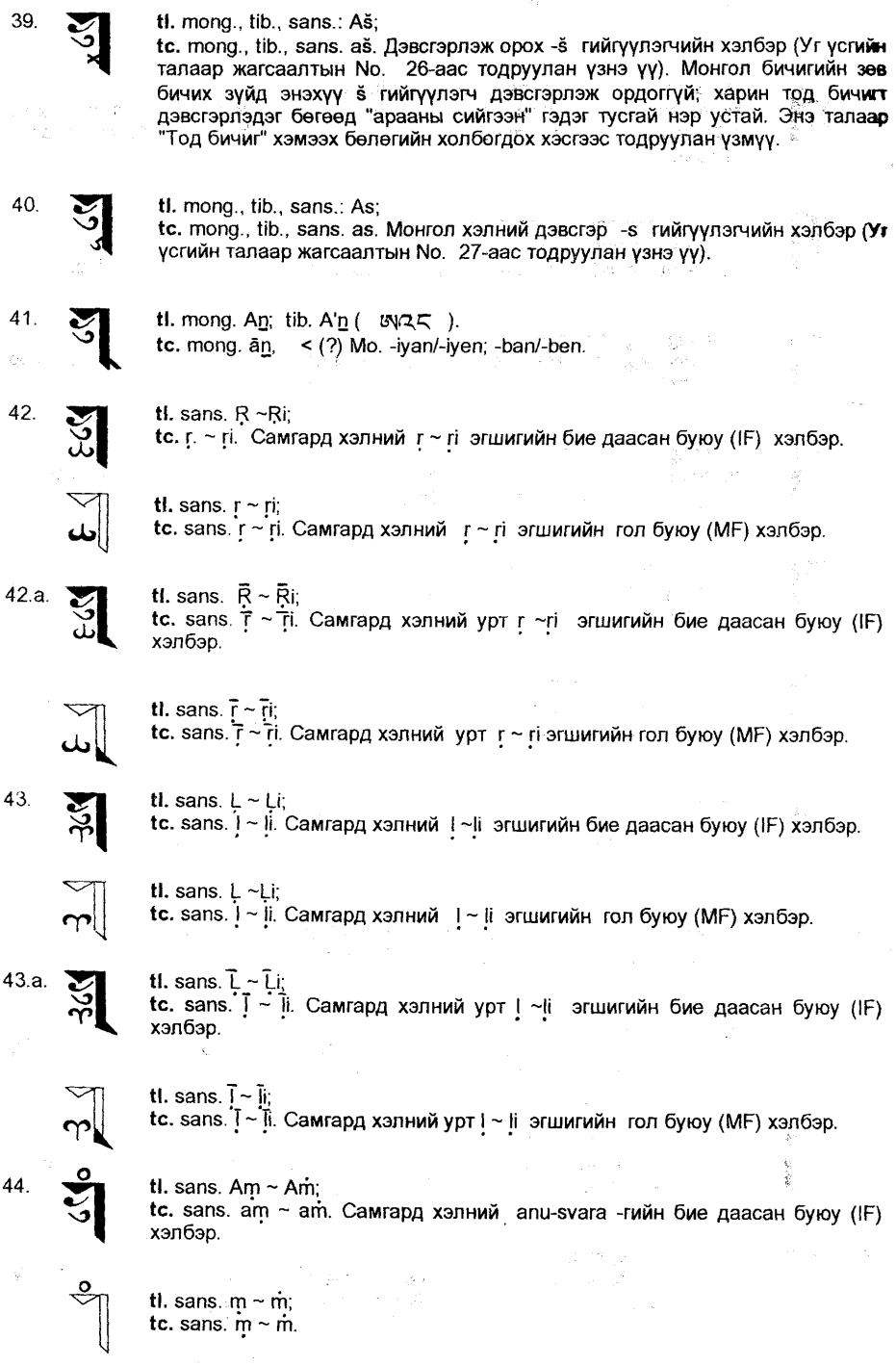


Figure 35: Description of Soyombo consonants (from Shagdarsürüng 2001: 144).

Самгард хэлний anu-svara -гийн гол буюу (MF) хэлбэр. Жишээлбэл:

tl. Sva_0-ya_0ṁ-bhü;
tc. sva-yaṁ-bhü.

45. 
tl. sans. AḤ;
tc. sans. aḥ. Самгард хэлний vi-sarga гийгүүлэгчийн бие даасан буюу (IF) хэлбэр.

46. 
tl. tib., sans.: G(a);
tc. tib., sans.: ga. Төвөд, самгард хэлний ga гийгүүлэгчийн бие даасан буюу (IF) хэлбэр.

47. 
tl. sans. GH(a);
tc. sans. gha. Самгард хэлний gha гийгүүлэгчийн бие даасан буюу (IF) хэлбэр.

48. 
tl. tib., sans.: J(a);
tc. tib., sans.: ja. Самгард хэлний ja гийгүүлэгчийн бие даасан буюу (IF) хэлбэр.

49. 
tl. sans. JH(a);
tc. sans. jha. Самгард хэлний jha гийгүүлэгчийн бие даасан буюу (IF) хэлбэр.

50. 
tl. sans. Ṭ(a);
tc. sans. ṭa. Самгард хэлний ṭa гийгүүлэгчийн бие даасан буюу (IF) хэлбэр.

51. 
tl. sans. ṬH(a);
tc. sans. ṭha. Самгард хэлний ṭha гийгүүлэгчийн бие даасан буюу (IF) хэлбэр.

52. 
tl. sans. Ḍ(a);
tc. sans. ḍa. Самгард хэлний ḍa гийгүүлэгчийн бие даасан буюу (IF) хэлбэр.

53. 
tl. sans. ḌH(a);
tc. sans. ḍha. Самгард хэлний ḍha гийгүүлэгчийн бие даасан буюу (IF) хэлбэр.

54. 
tl. sans. Ṇ(a);
tc. sans. ṇa. Самгард хэлний ṇa гийгүүлэгчийн бие даасан буюу (IF) хэлбэр.

55. 
tl. tib., sans.: D(a);
tc. tib., sans.: da.
Төвөд, самгард хэлний da гийгүүлэгчийн бие даасан буюу (IF) хэлбэр.

Figure 36: Description of Soyombo consonants (from Shagdarsürüng 2001: 145).

56. 
tl. sans. DH(a);
tc. sans. dha. Самгард хэлний dha гийгүүлэгчийн бие даасан буюу (IF) хэлбэр.

57. 
tl. tib., sans.: B(a);
tc. tib., sans.: ba. Төвөд, самгард хэлний ba гийгүүлэгчийн бие даасан буюу (IF) хэлбэр.

58. 
tl. sans. BH(a);
tc. sans. bha. Самгард хэлний bha гийгүүлэгчийн бие даасан буюу (IF) хэлбэр.

59. 
tl. sans. Ṣ(a);
tc. sans. ṣa. Самгард хэлний ṣa гийгүүлэгчийн бие даасан буюу (IF) хэлбэр.

60. 
tl. tib., sans. Ky(a);
tc. tib., sans. kya. Төвөд, самгард хэлний kya гийгүүлэгчийн бие даасан буюу (IF) хэлбэр. (Энэ жагсаалтын No. 22-оос лавлагтун!)

61. 
tl. tib., sans.: Kr(a);
tc. tib., sans.: kra. Төвөд, самгард хэлний kra гийгүүлэгчийн бие даасан буюу (IF) хэлбэр. (Энэ жагсаалтын No. 23-аас лавлагтун!)

62. 
tl. tib., sans.: Kl(a);
tc. tib., sans.: kla. Төвөд, самгард хэлний kla гийгүүлэгчийн бие даасан буюу (IF) хэлбэр. (Энэ жагсаалтын No. 25-аас лавлагтун!)

63. 
tl. tib., sans.: Kv(a);
tc. tib., sans.: kva. Төвөд, самгард хэлний kva гийгүүлэгчийн бие даасан буюу (IF) хэлбэр. (Энэ жагсаалтын No. 24-өөс лавлагтун!)

64. 
tl. sans. kK(a);
tc. sans. kka. Самгард хэлний kka гийгүүлэгчийн бие даасан буюу (IF) хэлбэр. (Энэ жагсаалтын No. 10-аас лавлагтун!)

65. 
tl. sans. Ṅ K(a);
tc. sans. ṅka. Самгард хэлний нийлмэл ṅka гийгүүлэгчийн бие даасан буюу (IF) хэлбэр. (Энэ жагсаалтын No. 10, 12-оос лавлагтун!)

66. 
tl. sans. N̄C(a);
tc. sans. n̄ca. Самгар хэлний нийлмэл nca гийгүүлэгчийн бие даасан буюу (IF) хэлбэр. (Энэ жагсаалтын No. 13, 15-аас лавлагтун!)

67. 
tl. sans. Ṇ Ṭ(a);
tc. sans. ṇṭa. Самгард хэлний нийлмэл ṇṭa гийгүүлэгчийн бие даасан буюу (IF) хэлбэр. (Энэ жагсаалтын No. 50, 54-өөс лавлагтун!)

Figure 37: Description of Soyombo consonants (from Shagdarsürüng 2001: 146).

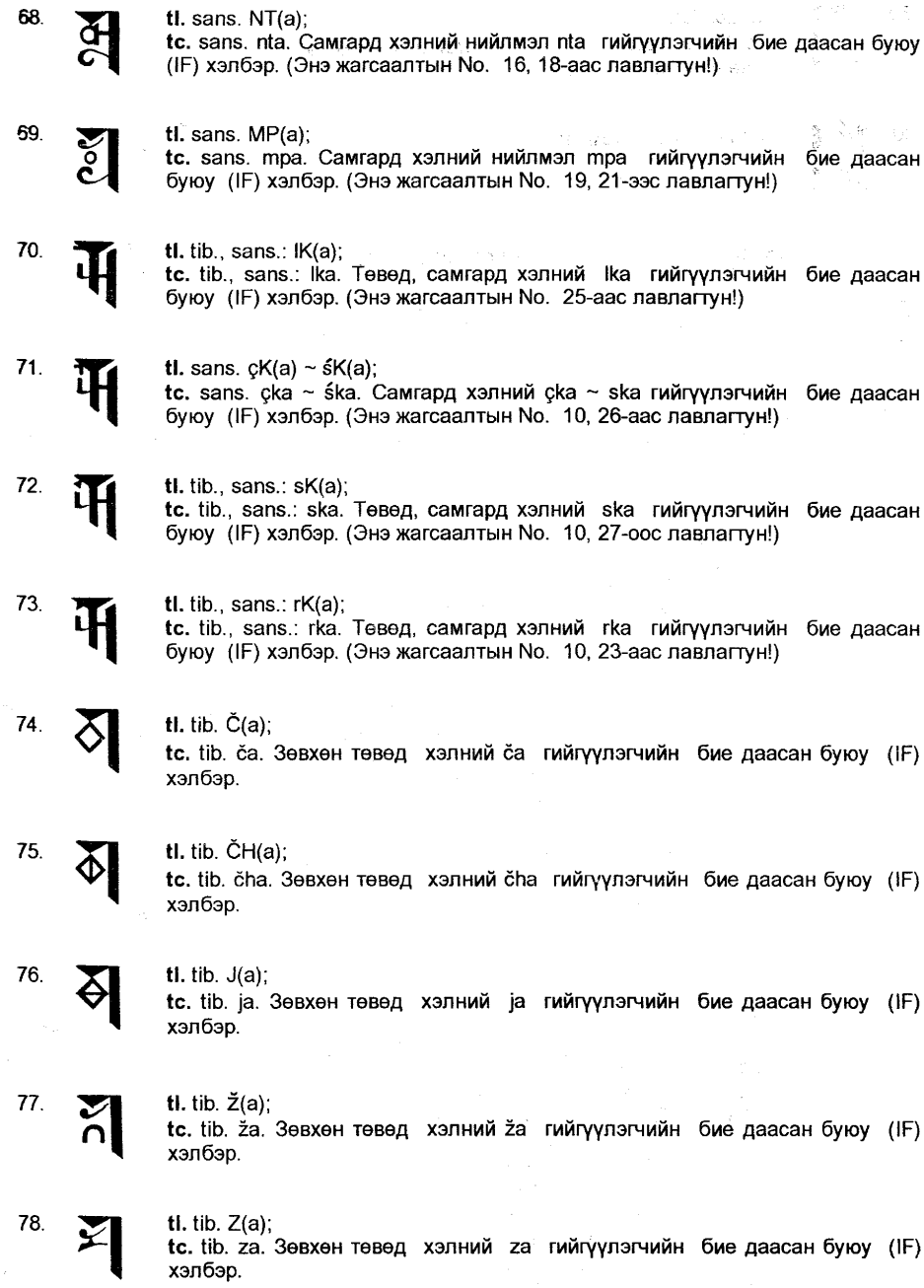


Figure 38: Description of Soyombo consonants (from Shagdarsürüng 2001: 147).

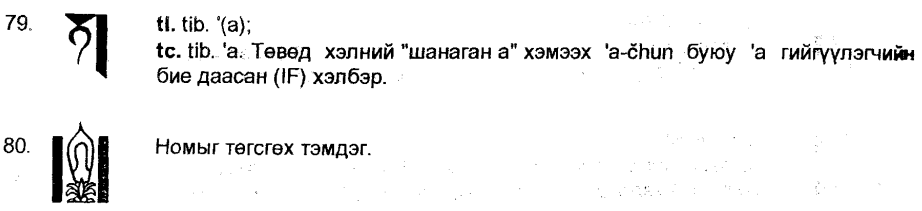


Figure 39: Description of a Soyombo consonant and terminal mark (from Shagdarsürüng 2001: 148).

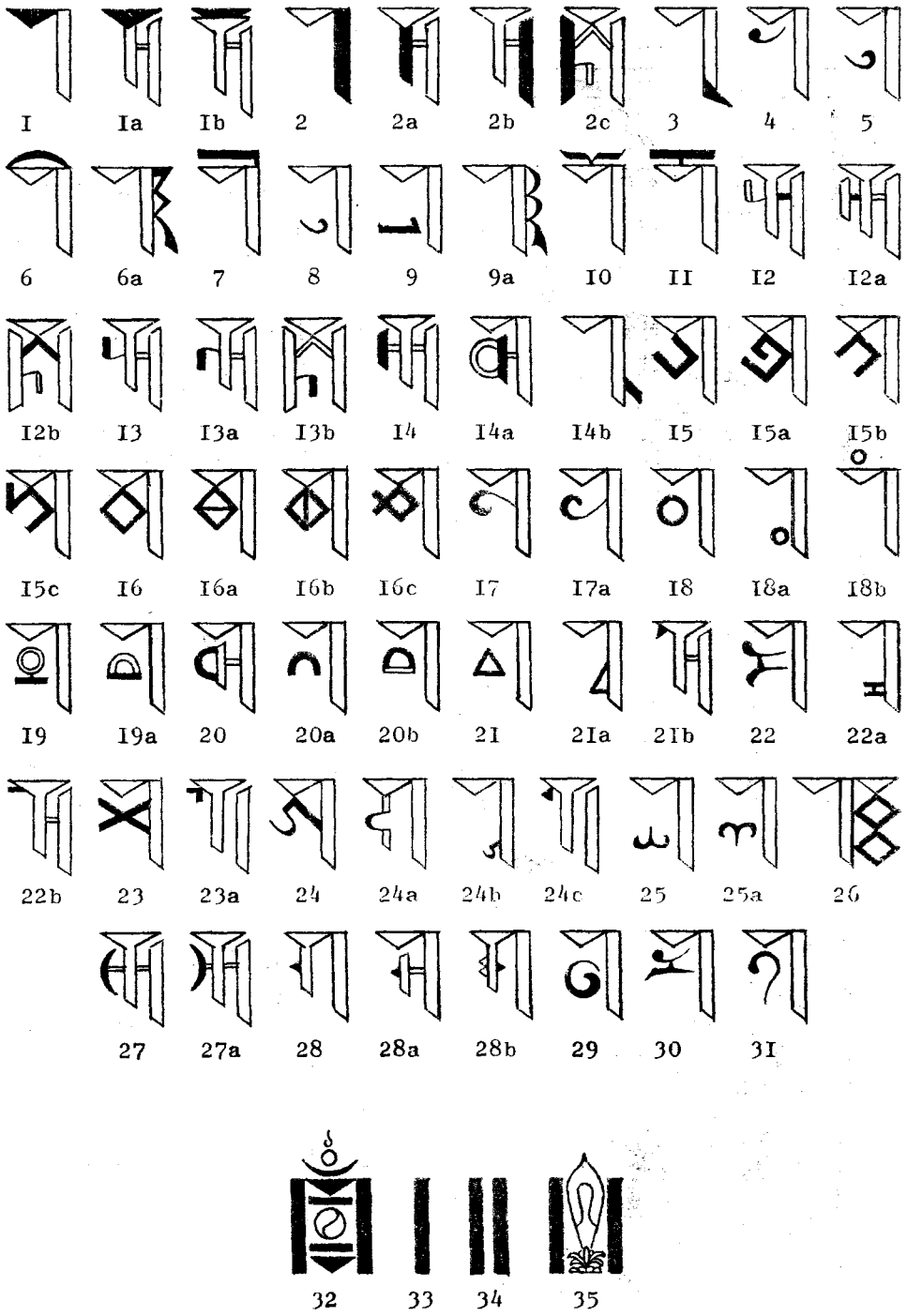
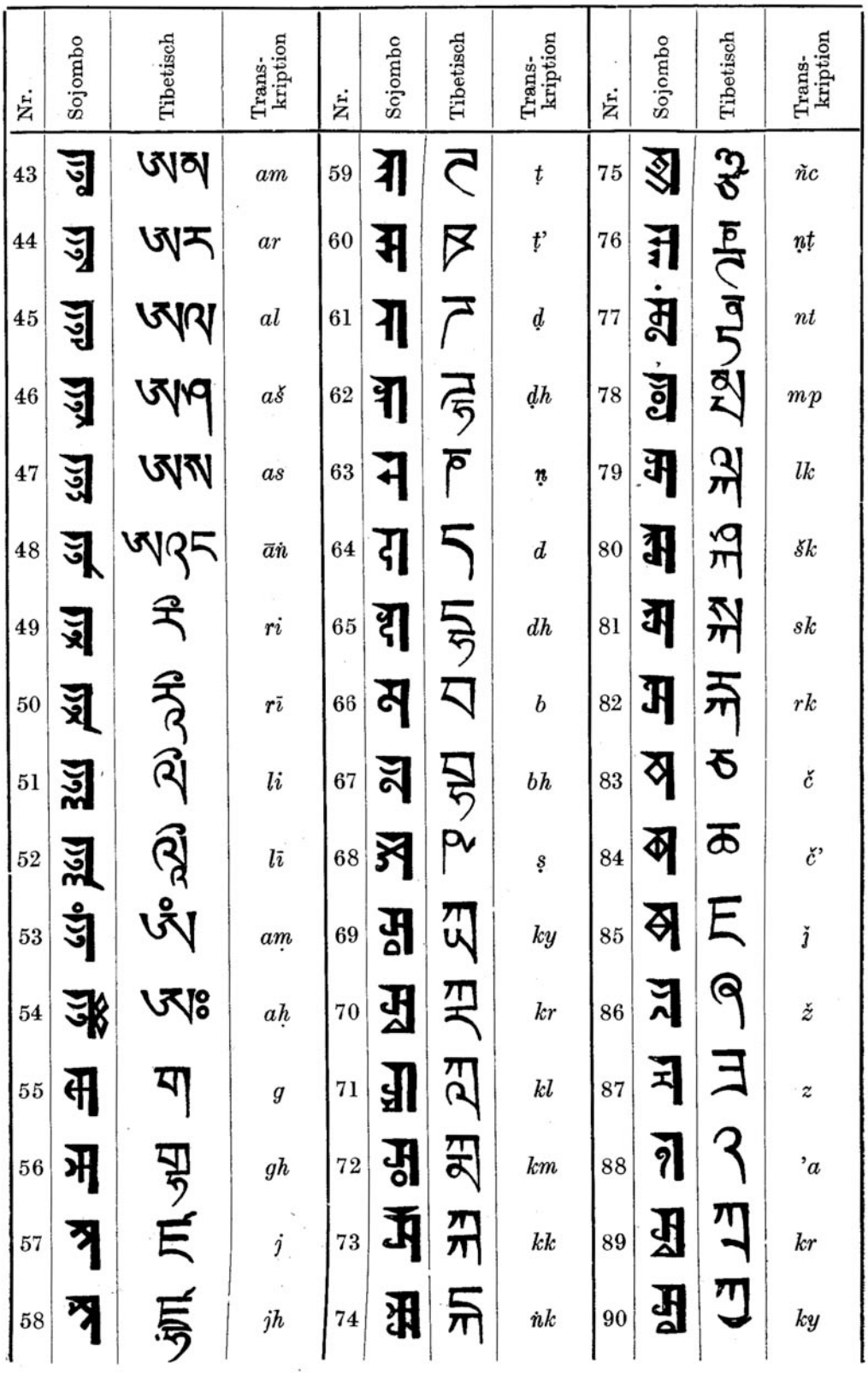


Figure 40: Character elements used in Soyombo characters (from Shagdarsürüng 2001: 153).

Figure 41: Comparison of Soyombo and Tibetan scripts (from Rintschen 1952: 68). Continued in figure 42.



Nr.	Sojombo	Tibetisch	Trans-kription	Nr.	Sojombo	Tibetisch	Trans-kription	Nr.	Sojombo	Tibetisch	Trans-kription
43			*am*	59			*ṭ*	75			*ñc*
44			*ar*	60			*ṭ'*	76			*ṇṭ*
45			*al*	61			*ḍ*	77			*nt*
46			*aš*	62			*ḍh*	78			*mp*
47			*as*	63			*ṇ*	79			*lk*
48			*āṁ*	64			*d*	80			*šk*
49			*ri*	65			*dh*	81			*sk*
50			*rī*	66			*b*	82			*rk*
51			*li*	67			*bh*	83			*č*
52			*lī*	68			*ṣ*	84			*č'*
53			*aṃ*	69			*ky*	85			*ǰ*
54			*aḥ*	70			*kr*	86			*ž*
55			*g*	71			*kl*	87			*z*
56			*gh*	72			*km*	88			*'a*
57			*j*	73			*kk*	89			*kr*
58			*jh*	74			*ṅk*	90			*ky*

Figure 42: Comparison of Soyombo and Tibetan scripts (from Rintschen 1952: 69). Continued from figure 41.

Figure 43: The HEAD MARK WITH TRIPLE FLAME or *svayambhu* symbol with Mongolian text (from Rintschen 1953: 8).

Соёмбо тэмдэг болон норов бадам тамгын тиг

Соёмбо тэмдэгийн тигийн хувьд бие даасан чиглэл болон хөгжөөнийг дурьдах хэрэгтэй юм. Ийм ч учраас түүнийг бүтээх тиг нь харилцан адилгүй хөгжөөр иржээ. Үүнд зассан үзгээр бичих үсэгзүйн тигийн талаар гаргах тул бусад тигээс өөр байхыг анхаарахад илүүдэхгүй юм. Соёмбо тэмдэгийн тигийг үндсэн хоёр хуваана. Үүнд:

1. Өндөрийн хэмжээ
2. Уртын хэмжээ болой.

Өндөрийн хэмжээ нь 16 хувь байна. Үүнд:

1. Гал	3 хувь
2. Нар	2 хувь
3. Сар	2 хувь
4. Дээд талын чойнжүн	1.5 хувь
5. Дээд талын сум	1 хувь
6. Эвам	2 хувь
7. Доод талын сум	1 хувь
8. Доод талын чойнжүн	1.5 хувь
9. Завсарын тэмдэгүүд	2 хувь
Нийт	16 хувь болно

Зураг 50

Норов бадам тамгын өндөрийн хэмжээ 8 хувь, өргөний хэмжээ 6 хувь байна. Хос шадны хоорондох норов бадамын дүрсийн өндөрийг 8 хувь, өргөнийг нь 4 хувьд багтаан зохистой сайхнаар бичнэ. / Зураг 51 /

Зураг 51

Өргөний хэмжээ нь 6 хувь байна.

Үүнд:

1. Гал	1.5 хувь
2. Нар	2.5 хувь
3. Сар	3 хувь
4. Чойнжүн	4 хувь

5. Хос шад тус бүр нэг хувь ба нийт 2 хувь болно.
6. Дээд доод сум болон эвамыг 4 хувьд багтаан аль зохистойгоор бичнэ. / Зураг 50 /

Figure 44: Description of HEAD MARK WITH SINGLE FLAME and TERMINAL MARK-2 (from Boldsaikhan 2005: 357).

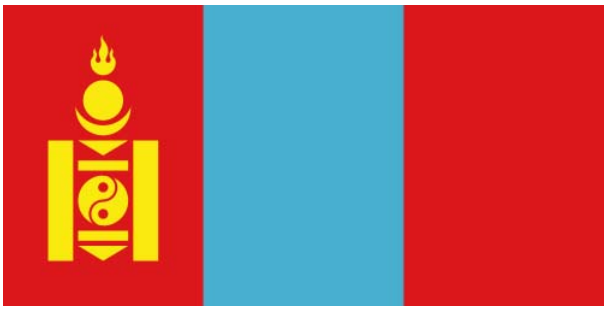


Figure 45: The flag of Mongolia showing the 𑩐 *svayambhu* symbol, which is proposed for encoding as SOYOMBO HEAD MARK WITH TRIPLE FLAME. Source: `http://en.wikipedia.org/wiki/File:Flag_of_Mongolia.svg`

Figure 46: The coat of arms of Mongolia showing the 𑩐 *svayambhu* symbol, which is proposed for encoding as SOYOMBO HEAD MARK WITH TRIPLE FLAME. Source: `http://en.wikipedia.org/wiki/File:Coat_of_Arms_of_Mongolia.svg`

Figure 47: The 𑩐 *svayambhu* symbol on the recto of a 1,000 Mongolian tögrög (tugrik) note. The symbol is proposed for encoding as SOYOMBO HEAD MARK WITH TRIPLE FLAME. Source: `http://en.wikipedia.org/wiki/File:1000_Tugriks_-_Recto.jpg`

		Corff	Glavy
a			
ka			
kha			
ga			
gha			
ṅa			
ca			
cha			
ja			
jha			
ña			
ṭa			
ṭha			
ḍa			
ḍha			
ṇa			
ta			
tha			
da			
dha			
na			
pa			
pha			
ba			
bha			
ma			
tsa			
tsha			
dza			
zha			
za			
'a			
ya			
ra			
la			
va			
śa			
ṣa			
sa			
ha			
kṣa			

Table 15: Comparison of Soyombo fonts by Oliver Corff and Jason Glavy.

ISO/IEC JTC 1/SC 2/WG 2
PROPOSAL SUMMARY FORM TO ACCOMPANY SUBMISSIONS
FOR ADDITIONS TO THE REPERTOIRE OF ISO/IEC 10646[1]
Please fill all the sections A, B and C below.
Please read Principles and Procedures Document (P & P) from http://std.dkuug.dk/JTC1/SC2/WG2/docs/principles.html **for guidelines and details before filling this form.**
Please ensure you are using the latest Form from http://std.dkuug.dk/JTC1/SC2/WG2/docs/summaryform.html.
See also http://std.dkuug.dk/JTC1/SC2/WG2/docs/roadmaps.html **for latest *Roadmaps*.**

A. Administrative

1. **Title:** *Proposal to Encode the Soyombo Script in ISO/IEC 10646*
2. Requester's name: *Script Encoding Initiative / Anshuman Pandey (anshuman.pandey@berkeley.edu)*
3. Requester type (Member body/Liaison/Individual contribution): *Liaison contribution*
4. Submission date: *2015-01-26*
5. Requester's reference (if applicable):
6. Choose one of the following:
   This is a complete proposal: *Yes*
   (or) More information will be provided later:

B. Technical – General

1. Choose one of the following:
   a. This proposal is for a new script (set of characters): *Yes*
      Proposed name of script: *Soyombo*
   b. The proposal is for addition of character(s) to an existing block:
      Name of the existing block:
2. Number of characters in proposal: *81*
3. Proposed category (select one from below - see section 2.2 of P&P document):
   A-Contemporary ( ) B.1-Specialized (small collection) (X) B.2-Specialized (large collection) ( )
   C-Major extinct ( ) D-Attested extinct ( ) E-Minor extinct ( )
   F-Archaic Hieroglyphic or Ideographic ( ) G-Obscure or questionable usage symbols ( )
4. Is a repertoire including character names provided? *Yes*
   a. If YES, are the names in accordance with the "character naming guidelines" in Annex L of P&P document? *Yes*
   b. Are the character shapes attached in a legible form suitable for review? *Yes*
5. Fonts related:
   a. Who will provide the appropriate computerized font to the Project Editor of 10646 for publishing the standard?
      Anshuman Pandey
   b. Identify the party granting a license for use of the font by the editors (include address, e-mail, ftp-site, etc.):
      Awaiting permission from original font designers for use of their glyphs
6. References:
   a. Are references (to other character sets, dictionaries, descriptive texts etc.) provided? *Yes*
   b. Are published examples of use (such as samples from newspapers, magazines, or other sources) of proposed characters attached? *Yes*
7. Special encoding issues:
   Does the proposal address other aspects of character data processing (if applicable) such as input, presentation, sorting, searching, indexing, transliteration etc. (if yes please enclose information)? *Yes*

8. Additional Information:

Submitters are invited to provide any additional information about Properties of the proposed Character(s) or Script that will assist in correct understanding of and correct linguistic processing of the proposed character(s) or script. Examples of such properties are: Casing information, Numeric information, Currency information, Display behaviour information such as line breaks, widths etc., Combining behaviour, Spacing behaviour, Directional behaviour, Default Collation behaviour, relevance in Mark Up contexts, Compatibility equivalence and other Unicode normalization related information. See the Unicode standard at http://www.unicode.org for such information on other scripts. Also see Unicode Character Database (http://www.unicode.org/reports/tr44/) and associated Unicode Technical Reports for information needed for consideration by the Unicode Technical Committee for inclusion in the Unicode Standard.

[1] Form number: N4502-F (Original 1994-10-14; Revised 1995-01, 1995-04, 1996-04, 1996-08, 1999-03, 2001-05, 2001-09, 2003-11, 2005-01, 2005-09, 2005-10, 2007-03, 2008-05, 2009-11, 2011-03, 2012-01)

C. Technical - Justification

1. Has this proposal for addition of character(s) been submitted before? *Yes*
   If YES explain *Complete revision and expansion of N4414 L2/13-069; see proposal for changes*
2. Has contact been made to members of the user community (for example: National Body, user groups of the script or characters, other experts, etc.)? *Yes*
   If YES, with whom? *Agata Bareja-Starzyńska (University of Warsaw, Poland)*
   György Kara (Indiana University,Bloomington)
   If YES, available relevant documents:
3. Information on the user community for the proposed characters (for example: size, demographics, information technology use, or publishing use) is included? *Yes*
   Reference:
4. The context of use for the proposed characters (type of use; common or rare) *Rare*
   Reference:
5. Are the proposed characters in current use by the user community? *Yes*
   If YES, where? Reference: *By scholars of Mongolian culture, history, and linguistics*
6. After giving due considerations to the principles in the P&P document must the proposed characters be entirely in the BMP? *N/A*
   If YES, is a rationale provided?
   If YES, reference:
7. Should the proposed characters be kept together in a contiguous range (rather than being scattered)? *Yes*
8. Can any of the proposed characters be considered a presentation form of an existing character or character sequence? *No*
   If YES, is a rationale for its inclusion provided?
   If YES, reference:
9. Can any of the proposed characters be encoded using a composed character sequence of either existing characters or other proposed characters? *No*
   If YES, is a rationale for its inclusion provided?
   If YES, reference:
10. Can any of the proposed character(s) be considered to be similar (in appearance or function) to, or could be confused with, an existing character? *No*
    If YES, is a rationale for its inclusion provided?
    If YES, reference:
11. Does the proposal include use of combining characters and/or use of composite sequences? *Yes*
    If YES, is a rationale for such use provided? *Yes*
    If YES, reference: *Combining signs*
    Is a list of composite sequences and their corresponding glyph images (graphic symbols) provided?
    If YES, reference:
12. Does the proposal contain characters with any special properties such as control function or similar semantics? *Yes*
    If YES, describe in detail (include attachment if necessary) *Subjoiner*
13. Does the proposal contain any Ideographic compatibility characters? *No*
    If YES, are the equivalent corresponding unified ideographic characters identified?
    If YES, reference: