ISO/IEC JTC1/SC2/WG2 N4541
L2/14-024
2014-01-22

Proposal to Encode the Zanabazar Square Script in ISO/IEC 10646

Anshuman Pandey
Department of History
University of Michigan
Ann Arbor, Michigan, U.S.A.
pandey@umich.edu

January 22, 2014

1 Introduction

This is a proposal to encode the Zanabazar Square script in the Universal Character Set (ISO/IEC 10646). It supersedes the following documents:

- N3956 L2/10-411 "Preliminary Proposal to Encode the Xawtaa Dorboljin Script in ISO/IEC 10646"
- N4041 L2/11-162 "Preliminary Proposal to Encode the Mongolian Square Script in ISO/IEC 10646"
- N4160 L2/11-379 "Revised Preliminary Proposal to Encode the Mongolian Square Script"
- N4413 L2/13-068 "Proposal to Encode the Mongolian Square Script in ISO/IEC 10646"
- N4471 L2/13-198 "Revised Proposal to Encode the Mongolian Square Script in ISO/IEC 10646"

The 'Zanabazar Square' script was previously referred to as the 'Mongolian Square' script (see section 3.1). Other changes introduced after N4471 L2/13-198 include: the replacement of the generic 'subjoiner' with a VIRAMA character for representing conjunct stacks (see sections 4.8 and 4.10); new names for several characters (see section 3.3); new head marks and ornaments (see section 4.11); the reordering of characters within the block; and the reallocation of the script block to a new range within the Supplementary Multilingual Plane (SMP). These changes are based upon feedback from experts.

The Zanabazar Square font used here is based upon the font developed by Oliver Corff in November 2001 for his "Xäwtää Dörböljin for LaTeX 2_ε" package. The proposal author has made several modifications to Corff's original font, including changes to the original glyphs and the addition of new glyphs.

2 Background

The Zanabazar Square Script is a syllabic alphabet based upon Tibetan and inspired by the Brahmi model. It was used for writing Mongolian, Sanskrit, and Tibetan. The script was invented by Zanabazar (1635–1723), one of the most important Buddhist leaders in Mongolia, who also developed the Soyombo script. The precise date is unknown, but it is believed that its creation preceded that of Soyombo. Named after its creator, the Zanabazar Square script bears similarities to Tibetan and Phags-pa (see tables 7–9). It is actively studied

by scholars and academic works on the script continue to be published (see Shagdarsürüng (2001); Ragchaa (2005); Bareja-Starzyńska and Ragchaa (2012)).

3 Proposal Details

3.1 Script Name

The name for the script block is 'Zanabazar Square'. It is known in Mongolian as 'Хэвтээ Дөрвөлжин бичиг' *xewtee dörböljin bicig* or 'Хэвтээ Дөрвөлжин Үсэг' *xewtee dörböljin üseg*, both of which translate into English as "Horizontal Square Script", but this is a technical name and is not commonly used. The name normalized Latin transliteration 'Xewtee Dorboljin' was used for the script in the preliminary proposal, but it was deemed too localized for usage in an international context. Therefore, the descriptor 'Mongolian' was added to the English translation to produce 'Mongolian Horizontal Square'. However, this name proved a bit long and it was curtailed to 'Mongolian Square'. Yet, 'Mongolian Square' is a generic name as it can also refer to Phags-pa, another Mongolian, or rather Central Asian, script that is also known as 'Дөрвөлжин Үсэг' *dörböljin üseg*, or 'Square Script'.

The script is commonly refered to as 'Занабазарын Дөрвөлжин Үсэг' *zanabazarin dörböljin üseg* "Zanabazar Square Script" in the academic community (see Byambaa Ragchaagiin 2005). This name is used because it differentiates the two Central Asian 'square' scripts on the basis of the names of their inventors: 'Phags-pa' for the vertical and 'Zanabazar' for the horizontal script. For these reasons, 'Zanabazar Square' is a suitable and unique identifier for the script block in the UCS. The Mongolian and alternate English names have been added as aliases for the script in the names list.

3.2 Character Repertoire

The Zanabazar Square block contains 69 characters. A code chart and names list are attached. The proposed repertoire contains elements that are not enumerated as separate characters in traditional arrangements of Zanabazar Square and in secondary sources. In some cases, the repertoire treats textual elements as sequences of characters instead of as atomic characters. For example, vowels such as 𑨀 *ā*, 𑨁 *i*, 𑨂 *ī*, etc. are enumerated as independent characters in traditional charts, but they are to be represented in encoded text as combinations of a base letter and combining vowel sign. Mongolian syllable-final consonants such as 𑨀𑨯 *ag*, 𑨀𑨲 *ang*, etc. are also treated similarly. Redundant elements have been eliminated, such as the occurrence of 𑨀 *ā* among both the vowels and codas, as well as the two instances of 𑨆 *ba* / *va*. These details are explained in section 4.7.1. The deviation from traditional perspectives regarding Zanabazar Square, or any other script, is a necessity for developing a character-encoding standard for the script in the UCS.

3.3 Character Names

An attempt has been made to align names for Zanabazar Square characters with those in the Tibetan block of the UCS. The names also correspond to characters of the Phags-pa block and those of the proposed encoding for Soyombo (see N4414 L2/13-069). The names also align with transliterated values given for Zanabazar Square characters in secondary sources, such as Tseveliin Shagdarsürüng (2001) and Byambaa Ragchaagiin (2005), which have been normalized according to UCS naming conventions. In this proposal, names for proposed characters are given in small capitals and transliterated values in italics, eg. the character 𑨄 is referred to as KA and its transliterated values as Sanskrit / Tibetan *ka* and Mongolian *ga* (see section 4.7.3 for language-specific transliteration).

In previous versions of this proposal the names for consonant letters were based upon Mongolian sound values as given in the available secondary sources. However, the majority of these sources are Mongolian texts, which provide an analysis of the script from a Mongolian perspective. While this perspective is certainly valid, it does not provide distinctive values for all characters. For instance, in Mongolian sources 𑨒 is transliterated as *ga* and is used for writing Mongolian *g* and *γ*, as well as Sanskrit and Tibetan *ka*; however, 𑨓 is also named *ga*, but this letter is used only in Sanskrit and Tibetan contexts. In order to accommodate this perspective, in previous versions of the proposal, letters used specifically for Tibetan and Sanskrit were distinguished using the descriptor '*galig*' (Mongolian: гали), a term applied to letters used for the transcription of non-Mongolian sounds. As a result, 𑨒 and 𑨓 were named GA and GALIG GA, respectively. However, as the Zanabazar Square script is based upon the Tibetan script and is intended for representing Mongolian as well as Sanskrit and Tibetan, the Mongolian analysis is somewhat limiting. Based upon feedback from experts such as Agata Bareja-Starzyńska and Andrew West, the names for Zanabazar Square consonants are now aligned with those of Tibetan. This approach removes the need to use the descriptor 'GALIG' and provides consonant letters with distinctive names that index their values according to the Tibetan model.

3.4 Encoding Order

The encoding order follows the pattern of the Tibetan block in the UCS. This order differs from that given in previous versions of the proposal, which was based upon the traditional Mongolian arrangement of the script. The Tibetan order is preferred because, as the script is modeled upon Tibetan, it offers a more natural order for the letters. In the Mongolian ordering system, '*galig*' letters representing voiced and aspirated consonants were separated from their unvoiced and unaspirated counterparts. The new order provides for contiguous placement of related characters. Moreover, there are four letters (𑨔 CA, 𑨕 CHA, 𑨖 JA, 𑨗 SSA) that are not found in traditional charts of the script, so it is unclear where they should be placed in an encoding order based upon the traditional arrangement. The Tibetan order provides a method of accommodating these characters. Moreover, basing the encoding order of Zanabazar Square with that of Tibetan in the UCS will facilitate aligned encodings for these related scripts.

4 Script Details

4.1 Structure

The Zanabazar Square script is written from left to right. As indicated by its Mongolian name, *xewtee dörböljin bicig*, the script is written horizontally, but in some instances occurs in vertical environments.

Independent vowels are written using a vowel-carrier letter to which vowel signs are attached. Vowel length is indicated by a sign that is attached to a base letter or to a combination of a base letter and one or more dependent vowel signs.

Consonant letters possess the inherent vowel *a*. The phonetic value of a consonant letter is changed by the attachment of a vowel sign. In Mongolian contexts, the inherent vowel is suppressed by a final-consonant mark, which indicates both a syllable-final consonant and a syllabic boundary. In Sanskrit and Tibetan contexts, the inherent vowel of a consonant is silenced using the VIRAMA. There are no consonant clusters in Mongolian, but those of Sanskrit and Tibetan are rendered as conjuncts. Similar to other scripts in the UCS that are based upon the Brahmi model, a sequence of bare consonants marked by VIRAMA forms a cluster, which is represented as a conjunct and rendered as a vertical stack with non-initial letters placed beneath the initial letter. The consonants YA, RA, LA, VA have different representations when they occur in Sanskrit and Tibetan conjuncts, therefore, contextual forms of these letters are provided as separate characters in order to facilitate the encoding model.

The graphical structure of a Zanabazar Square vowel syllable may be described as

> LETTER A [*vowel sign*]* [*vowel length mark*] [CANDRABINDU | ANUSVARA] [VISARGA]

A consonantal syllable may be described as follows:

> [*cluster-initial letter*] consonant [*cluster-final letter** | [*consonant** *cluster-final letter**]] [*vowel sign*]* [*vowel length mark*] [CANDRABINDU | ANUSVARA] [VISARGA]

where there is one base consonant, which may occur in a conjunct with several other consonants. The sources show conjuncts containing at least three consonants, but theoretically the number may be greater. One or more vowel signs may be used with a base consonant or conjunct. The vowel-length mark may attach to a consonant or vowel sign. One of either the CANDRABINDU or ANUSVARA may occur with a consonant or vowel sign. The VISARGA may follow last and can attach to a consonant, vowel sign, or the vowel-length mark.

4.2 Vowel Letter

The 𑨀 LETTER A represents the vowel *a* and a zero vowel depending upon phonotactical conditions. When it occurs independently it has the value *a*. When a vowel sign is attached to it, the LETTER A is a zero vowel and assumes the value of the vowel sign.

4.3 Vowel Length Mark

A long vowel is represented by placing the ◌𑨿 VOWEL LENGTH MARK after a consonant or vowel sign. When combined with the letter 𑨀 A or a consonant letter it represents the lengthening of the inherent vowel *a* to *ā*.

4.4 Vowel Signs

There are 9 dependent vowel signs:

◌𑨁	VOWEL SIGN I	◌𑨄	VOWEL SIGN OE	◌𑨅	VOWEL SIGN REVERSED I
◌𑨂	VOWEL SIGN UE	◌𑨆	VOWEL SIGN O		
◌𑨃	VOWEL SIGN U	◌𑨇	VOWEL SIGN AI		
◌𑨄	VOWEL SIGN E	◌𑨈	VOWEL SIGN AU		

The signs are written with base letters, ie. 𑨀 A and consonants. Multiple vowel signs may combine with a single base letter. Independent vowels are represented by attaching vowel signs to the carrier 𑨀 LETTER A. The independent forms of ◌𑨅 VOWEL SIGN REVERSED I are written according to a different pattern.

The first 8 vowel signs, in conjunction with the VOWEL LENGTH MARK, are used for writing the basic 16 vowels given in traditional script charts:

a 𑨀 <𑨀 LETTER A>

ā 𑨁 <𑨀 LETTER A, ◌𑨿 VOWEL LENGTH MARK>

i <𑨀 LETTER A, ◌𑨯 VOWEL SIGN I>

ī <𑨀 LETTER A, ◌𑨯 VOWEL SIGN I, ◌𑨿 VOWEL LENGTH MARK>

ü, *u* <𑨀 LETTER A, ◌𑨰 VOWEL SIGN UE>

ǖ <𑨀 LETTER A, ◌𑨰 VOWEL SIGN UE, ◌𑨿 VOWEL LENGTH MARK>

u <𑨀 LETTER A, ◌𑨱 VOWEL SIGN U>

ū <𑨀 LETTER A, ◌𑨱 VOWEL SIGN U, ◌𑨿 VOWEL LENGTH MARK>

e <𑨀 LETTER A, ◌𑨲 VOWEL SIGN E>

ē <𑨀 LETTER A, ◌𑨲 VOWEL SIGN E, ◌𑨿 VOWEL LENGTH MARK>

ö <𑨀 LETTER A, ◌𑨳 VOWEL SIGN OE>

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

o <𑨀 LETTER A, ◌𑨴 VOWEL SIGN O>

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

ai <𑨀 LETTER A, ◌𑨵 VOWEL SIGN AI>

au <𑨀 LETTER A, ◌𑨶 VOWEL SIGN AU>

4.4.1 Diphthongs

The signs ◌𑨵 VOWEL SIGN AI and ◌𑨶 VOWEL SIGN AU represent the diphthongs *ai* and *au*, respectively. They also function as secondary vowel signs for *i* and *u* for producing additional diphthongs for Mongolian (see figures 43–45). These diphthongs are represented using combinations of signs:

āi <𑨀 LETTER A, ◌𑨿 VOWEL LENGTH MARK, ◌𑨵 VOWEL SIGN AI>

ii <𑨀 LETTER A, ◌𑨯 VOWEL SIGN I, ◌𑨵 VOWEL SIGN AI>

iu <𑨀 LETTER A, ◌𑨯 VOWEL SIGN I, ◌𑨶 VOWEL SIGN AU>

üi <𑨀 LETTER A, ◌𑨰 VOWEL SIGN UE, ◌𑨵 VOWEL SIGN AI>

ui <𑨀 LETTER A, ◌𑨱 VOWEL SIGN U, ◌𑨵 VOWEL SIGN AI>

ūi <𑨀 LETTER A, ◌𑨱 VOWEL SIGN U, ◌𑨿 VOWEL LENGTH MARK, ◌𑨵 VOWEL SIGN AI>

ei <𑨀 LETTER A, ◌𑨲 VOWEL SIGN E, ◌𑨵 VOWEL SIGN AI>

ēi <𑨀 LETTER A, ◌𑨲 VOWEL SIGN E, ◌𑨿 VOWEL LENGTH MARK, ◌𑨵 VOWEL SIGN AI>

eǖ <𑨀 LETTER A, ◌𑨲 VOWEL SIGN E, ◌𑨰 VOWEL SIGN UE, ◌𑨿 VOWEL LENGTH MARK>

oi 𑨁𑨯 <𑨀 LETTER A, ◌𑨯 VOWEL SIGN O, ◌𑨰 VOWEL SIGN AI>

ōi 𑨁𑨯𑨵 <𑨀 LETTER A, ◌𑨯 VOWEL SIGN O, ◌𑨵 VOWEL LENGTH MARK, ◌𑨰 VOWEL SIGN AI>

ou 𑨁𑨱 <𑨀 LETTER A, ◌𑨯 VOWEL SIGN O, ◌𑨱 VOWEL SIGN AU>

4.4.2 Vocalic Letters

The ◌𑨳 VOWEL SIGN REVERSED I is used for writing the four Sanskrit vocalic letters (eg. Devanagari ऋ *ṛ*, ॠ *ṝ*, ऌ *ḷ*, ॡ *ḹ*. They are represented by combining ◌𑨳 VOWEL SIGN REVERSED I with the consonant letters 𑨪 RA and 𑨫 LA; the ◌𑨵 VOWEL LENGTH MARK is added for the long forms (see figure 17):

ṛ 𑨪𑨳 <𑨪 RA, ◌𑨳 VOWEL SIGN REVERSED I>

◌𑨳 <*consonant*, ◌𑩇 CLUSTER FINAL RA, ◌𑨳 VOWEL SIGN REVERSED I>

ṝ 𑨪𑨳𑨵 <𑨪 RA, ◌𑨳 VOWEL SIGN REVERSED I, ◌𑨵 VOWEL LENGTH MARK>

◌𑨵 <*consonant*, ◌𑩇 CLUSTER FINAL RA, ◌𑨳 VOWEL SIGN REVERSED I, ◌𑨵 VOWEL LENGTH MARK>

ḷ 𑨫𑨳 <𑨫 LA, ◌𑨳 VOWEL SIGN REVERSED I>

◌𑨳 <*consonant*, ◌𑩈 CLUSTER FINAL LA, ◌𑨳 VOWEL SIGN REVERSED I>

ḹ 𑨫𑨳𑨵 <𑨫 LA, ◌𑨳 VOWEL SIGN REVERSED I, ◌𑨵 VOWEL LENGTH MARK>

◌𑨵 <*consonant*, ◌𑩈 CLUSTER FINAL LA, ◌𑨳 VOWEL SIGN REVERSED I, ◌𑨵 VOWEL LENGTH MARK>

As indicated by the independent forms, the Sanskrit vocalic sounds are approximated using the syllables *ri*, *rī*, *li*, *lī*. The dependent forms are actually conjuncts with *ra* and *la* as C_2, occuring as subjoined forms ◌𑩇 CLUSTER FINAL RA and ◌𑩈 CLUSTER FINAL LA (see section 4.10). The ◌𑨳 VOWEL SIGN REVERSED I is necessary for properly distinguishing the vowel *ṛ* in 𑨑𑨳 *kṛ* from the consonant-vowel syllable *ri* in 𑨑𑨳 *kri*.

Kara (1972) shows the vocalic letters written using the full-arched variant form ◌𑨴 instead of the half arch ◌𑨳 of VOWEL SIGN I: 𑨪𑨴 *ṛ*, 𑨪𑨴𑨵 *ṝ*, 𑨫𑨴 *ḷ*, 𑨫𑨴𑨵 *ḹ*. The full-arched forms do not allow for distinctive representations of vocalic letters and consonant-*ra* or consonant-*la* syllables, eg. 𑨑𑨳 could be either *kṛ* or *kri*, and 𑨑𑨴 could be either *kḷ* or *kli*. The Zanabazar Square orthography for Sanskrit vocalic sounds is parallel to that of Tibetan in the UCS, in which ◌ི (U+0F72 TIBETAN VOWEL SIGN I) is reversed as ◌ྀ (U+0F80 TIBETAN VOWEL SIGN REVERSED I) when writing independent ཪྀ *ṛ* and ཬྀ *ḷ*, as well as the dependent forms ◌ྲྀ *-ṛ* and ◌ླྀ *-ḷ*.

4.4.3 Variant Forms of Vowel Signs

The following vowel signs have alternate representations, which are to be handled as glyphic variants:

	Regular	Variant	Usage
VOWEL SIGN I	◌𑨁	◌𑨁	𑨀𑨁, 𑨀𑨊𑨁 instead of 𑨀𑨁 *i*, 𑨀𑨊𑨁 *ī*
VOWEL SIGN U	◌𑨃	◌𑨃	𑨀𑨃, 𑨀𑨊𑨃 instead of 𑨀𑨃 *u*, 𑨀𑨊𑨃 *ū*
VOWEL SIGN OE	◌𑨅	◌𑨅	𑨀𑨅, 𑨀𑨊𑨅 instead of 𑨀𑨅 *ö*, 𑨀𑨊𑨅 *ȫ*
VOWEL SIGN AI	◌𑨇	◌𑨇, ◌𑨇	𑨀𑨇, 𑨀𑨇 instead of 𑨀𑨇 *ai*
VOWEL SIGN AU	◌𑨈	◌𑨈, ◌𑨈	𑨀𑨈, 𑨀𑨈 instead of 𑨀𑨈 *au*
VOWEL LENGTH MARK	◌𑨊	◌𑨊, ◌𑨊	𑨀𑨊 and 𑨀𑨊 instead of 𑨀𑨊 *ā*

The ◌𑨁 VOWEL SIGN I is often written using a full arch ◌𑨁 (see figure 9), which is the most common form for the vowel sign.

The variant form of ◌𑨅 VOWEL SIGN OE is ◌𑨅, which is a reversal of ◌𑨄 VOWEL SIGN E.

The alternate form ◌𑨃 of ◌𑨃 VOWEL SIGN U is shown in several records (see figure 38).

The variant forms ◌𑨇 and ◌𑨈 of ◌𑨇 VOWEL SIGN AI and ◌𑨈 VOWEL SIGN AU occur in a single manuscript and show the influence of Tibetan (see figures 48 and 49). The forms ◌𑨇 and ◌𑨈 are variants produced by turning the hook towards the right and left instead of facing them upwards.

The forms 𑨀𑨊 and 𑨀𑨊 are shown in secondary sources at the end of a list of syllable-final consonants (see figure 19). These are not distinctive letters, but variant forms of 𑨀𑨊 *ā* in which the ◌𑨊 VOWEL LENGTH MARK is written using the glyphic variants ◌𑨊 and ◌𑨊.

4.5 Vowel Modifiers

The following combining signs are used for representing Sanskrit:

◌𑨸 SIGN ANUSVARA indicates nasalization in Sanskrit words. It is transliterated as *ṃ*, eg. 𑨀𑨸 *aṃ*. The sign is not shown in traditional charts, but occurs in the word 𑨭𑨃𑨠𑨸 *śubhaṃ*, which is written at the end of charts (see figure 2). When ANUSVARA occurs in a sequence with a vowel sign it is placed last, eg. 𑨰𑨃𑨸 *huṃ* <𑨰 HA, ◌𑨃 VOWEL SIGN U, ◌𑨸 ANUSVARA>. It corresponds to ◌ཾ U+0F7E TIBETAN SIGN RJES SU NGA RO.

◌𑨹 SIGN VISARGA indicates post-vocalic aspiration in Sanskrit words. It is generally transliterated as *ḥ*, eg. 𑨀𑨹 *aḥ*. When the VISARGA occurs in a sequence with a vowel sign it is placed last, eg. 𑨀𑨊𑨹 *āḥ* <𑨀 A, ◌𑨊 VOWEL LENGTH MARK, ◌𑨹 VISARGA>. It corresponds to ◌ཿ U+0F7F TIBETAN SIGN RNAM BCAD.

4.6 Candrabindu and Candra Ornaments

The following combining signs are used as nasalization marks and ornaments for the head mark:

◌𑨵 SIGN CANDRABINDU indicates nasalization in Sanskrit words. It is transliterated as *ṃ*, eg. 𑨀𑨵 *aṃ*. The CANDRABINDU does not appear in script charts and manuals, but it is attested in various manuscripts, eg. in the word 𑨰𑨃𑨵 *huṃ* (see figure 29). It also occurs frequently in combination with 𑨿 INITIAL HEAD MARK (see

section 4.11). When the CANDRABINDU occurs with a vowel sign, it is always placed after it, eg. 𑨱𑨂𑨵 *huṃ* <𑨱 HA, ◌𑨂 VOWEL SIGN UE, ◌𑨵 CANDRABINDU>. The sign corresponds to ◌ྃ U+0F83 TIBETAN SIGN SNA LDAN.

The ◌𑨶 SIGN CANDRABINDU WITH ORNAMENT and ◌𑨷 SIGN CANDRA WITH ORNAMENT are used in combination with the 𑨿 INITIAL HEAD MARK (see figure 40). They also occur as the variant, reversed forms ◌𑨶 and ◌𑨷. The SIGN CANDRA WITH ORNAMENT corresponds to ◌ྂ U+0F82 TIBETAN SIGN NYI ZLA NAA DA.

4.7 Consonant Letters

There are 40 consonant letters:

𑨋	KA	𑨕	TTHA	𑨟	PHA	𑨩	SMALL A
𑨌	KHA	𑨖	DDA	𑨠	BA	𑨪	YA
𑨍	GA	𑨗	DDHA	𑨡	BHA	𑨫	RA
𑨎	GHA	𑨘	NNA	𑨢	MA	𑨬	LA
𑨏	NGA	𑨙	TA	𑨣	TSA	𑨭	VA
𑨐	CA	𑨚	THA	𑨤	TSHA	𑨮	SHA
𑨑	CHA	𑨛	DA	𑨥	DZA	𑨯	SSA
𑨒	JA	𑨜	DHA	𑨦	DZHA	𑨰	SA
𑨓	NYA	𑨝	NA	𑨧	ZHA	𑨱	HA
𑨔	TTA	𑨞	PA	𑨨	ZA	𑨲	KSSA

4.7.1 Consonant Order

The original Mongolian arrangement and values of the consonant letters are as follows:

𑨋	𑨌	𑨏	𑨣	𑨤	𑨓	𑨙	𑨚	𑨝	𑨞	𑨟	𑨢	𑨪	𑨫	𑨠
ga	*ka*	*ṅa*	*ja*	*ca*	*ña*	*da*	*ta*	*na*	*ba*	*pa*	*ma*	*ya*	*ra*	*va*
𑨬	𑨮	𑨰	𑨱	𑨲	𑨔	𑨕	𑨖	𑨗	𑨘	𑨧	𑨨	𑨩	𑨍	𑨎
la	*śa*	*sa*	*ha*	*kṣa*	*ṭa*	*ṭha*	*ḍa*	*ḍha*	*ṇa*	*zha*	*za*	*’a*	*ga*	*gha*
𑨥	𑨦	𑨭	𑨛	𑨜	𑨠	𑨡								
ja	*jha*	*va*	*da*	*dha*	*ba*	*bha*								

This order does not contain letters that are attested in various sources, but that are not found in charts, eg. 𑨐 CA, 𑨑 CHA, 𑨒 JA, 𑨯 SSA. It is not clear where these four letters would fit into the above order. The traditional order also includes two instances of the character 𑨠 *ba*, an issue which must be resolved for the proposed encoding. For this reason an arrangement that follows the Tibetan order has been adopted. Letters that are not part of the traditional repertoire are highlighted in red:

𑨋	𑨌	𑨍	𑨎	𑨏	𑨐	𑨑	𑨒	𑨓	𑨔	𑨕	𑨖	𑨗	𑨘	𑨙
ka	*kha*	*ga*	*gha*	*ṅa*	*ca*	*cha*	*ja*	*ña*	*ṭa*	*ṭha*	*ḍa*	*ḍha*	*ṇa*	*ta*
𑨚	𑨛	𑨜	𑨝	𑨞	𑨟	𑨠	𑨡	𑨢	𑨣	𑨤	𑨥	𑨦	𑨧	𑨨
tha	*da*	*dha*	*na*	*pa*	*pha*	*ba*	*bha*	*ma*	*tsa*	*tsha*	*dza*	*dzha*	*zha*	*za*
𑨩	𑨪	𑨫	𑨬	𑨭	𑨮	𑨯	𑨰	𑨱	𑨲					
'a	*ya*	*ra*	*la*	*va*	*śa*	*ṣa*	*sa*	*ha*	*kṣa*					

4.7.2 Notes on Consonants

𑨐 CA, 𑨑 CHA, 𑨒 JA These letters do not occur in traditional script charts, but they occur in manuscripts and secondary sources. It is unclear when they were introduced into the script or by whom. They correspond to the Tibetan characters ཅ CA, ཆ CHA, ཇ JA.

𑨠 BA Traditional charts and secondary sources show two instances of the letter 𑨠. It occurs first after 𑨬 LA and secondly before 𑨡 BHA. Based upon the occurrences, it is clear that the first 𑨠 represents *va* in Mongolian, while the second 𑨠 is used for writing *ba* in Sanskrit and Tibetan (see figures 15 and 17). As *va* is represented distinctively using 𑨭 VA, the letter 𑨠 is named BA. It should be noted that 𑨠 is often used in place of 𑨭 VA in Tibetan contexts.

𑨦 DZHA The letter DZHA represents Sanskrit *jha*. It is named DZHA instead of JHA because of its graphical similarity to related characters: 𑨣 TSA, 𑨤 TSHA, 𑨥 DZA as opposed to 𑨐 CA, 𑨑 CHA, 𑨒 JA. Also, in the traditional order of the script, 𑨦 DZHA occurs between 𑨥 DZA and 𑨭 VA, which imitates the Tibetan order and further suggests its value as DZHA, corresponding to Tibetan ཛྷ U+0F60 TIBETAN LETTER DZHA.

𑨩 SMALL A The letter SMALL A corresponds to Tibetan *'a chung* (འ U+0F60 TIBETAN LETTER -A). It is used in some instances in place of ◌𑨊 VOWEL LENGTH MARK for denote a long vowel (see figures 22 and 36). The name of the letter is based upon the corresponding ꡖ U+A856 PHAGS-PA LETTER SMALL A instead of the Tibetan name, because usage of a '-' (hyphen) as the Tibetan name for the character -A is no longer permitted according to UCS naming conventions.

𑨯 SSA The letter SSA is a reversed form of 𑨮 SHA and represents the Sanskrit *ṣa*. It does not appear in script charts, but it is attested in manuscripts, such as in the Sanskrit invocation written above script charts: 𑨝𑨢𑨆 𑨍𑨃𑨫𑨃 𑨢𑨒𑨃𑨍𑨆𑨯𑨊𑨪 *namo guru mañjughoṣāya* "praise to the preceptor Mañjughoṣa" (see figures 10 and 26). The SSA was introduced into the script by Ugalzyn Lama; it is curious that Zanabazar did not provide a distinctive letter for writing *ṣa* in his square script as he did for Soyombo.

𑨲 KSSA The letter 𑨲 KSSA represents the Sanskrit cluster क्ष *kṣa* (/kşa/). While in some Indic scripts the written form for /kşa/ has an encoded representation as a character sequence, such an approach would not be consistent with the model for Zanabazar Square. In this script, this letter represents a phoneme that is phonetically a consonant cluster, but it has the graphical structure of an atomic letter. It is encoded as a consonant letter because in all cases consonant conjunct forms are written as stacks in Zanabazar Square, not as ligatures. This letter corresponds to ཀྵ U+0F69 TIBETAN LETTER KSSA.

4.7.3 Representation of Mongolian, Sanskrit, and Tibetan Consonants

Zanabazar Square consonants possess different values based upon linguistic context. For Mongolian the letters for voiceless sounds (𑨋 KA, 𑨐 CA, 𑨙 TA, 𑨞 PA) are used for voiced stops, while letters for voiceless

aspirated sounds (𑨌 KHA, 𑨑 CHA, 𑨚 THA, 𑨟 PHA) are used for voiceless stops. The values of consonants for Mongolian (‘M’), Sanskrit (‘S’), and Tibetan (‘T’) are given below:

		M	S	T
𑨋	KA	*ga*, *γa*	*ka*	*ka*
𑨌	KHA	*ka*, *qa*, *xa*	*kha*	*kha*
𑨍	GA		*ga*	*ga*
𑨎	GHA		*gha*	*gha*
𑨏	NGA	*nga*	*ṅa*	*ṅa*
𑨐	CA			*ca*
𑨑	CHA			*cha*
𑨒	JA			*ja*
𑨓	NYA		*ña*	*ña*
𑨔	TTA		*ṭa*	*ṭa*
𑨕	TTHA		*ṭha*	*ṭha*
𑨖	DDA		*ḍa*	*ḍa*
𑨗	DDHA		*ḍha*	*ḍha*
𑨘	NNA		*ṇa*	*ṇa*
𑨙	TA	*da*	*ta*	*ta*
𑨚	THA	*ta*	*tha*	*tha*
𑨛	DA		*da*	*da*
𑨜	DHA		*dha*	*dha*
𑨝	NA	*na*	*na*	*na*
𑨞	PA	*ba*	*pa*	*pa*

		M	S	T
𑨟	PHA	*pa*	*pha*	*pha*
𑨠	BA	*va*	*ba*	*ba*
𑨡	BHA		*bha*	*bha*
𑨢	MA	*ma*	*ma*	*ma*
𑨣	TSA	*ǰa*	*ca*	*tsa*
𑨤	TSHA	*ča*	*cha*	*tsha*
𑨥	DZA		*ja*	*dza*
𑨦	DZHA		*jha*	
𑨧	ZHA			*zha*
𑨨	ZA			*za*
𑨩	SMALL A			*’a*
𑨪	YA	*ya*	*ya*	*ya*
𑨫	RA	*ra*	*ra*	*ra*
𑨬	LA	*la*	*la*	*la*
𑨭	VA		*va*	*wa*
𑨮	SHA	*ša*	*śa*	*śa*
𑨯	SSA		*ṣa*	*ṣa*
𑨰	SA	*sa*	*sa*	*sa*
𑨱	HA	*ha*	*ha*	*ha*
𑨲	KSSA		*kṣa*	*kṣa*

4.7.4 Alternate Representations of Consonants

Alternate forms and glyphic variants are attested for some consonant letters:

	Regular	Variant
GHA	𑨎	𑨎
DHA	𑨜	𑨜
TTA	𑨔	𑨔
NNA	𑨘	𑨘

	Regular	Variant
NA	𑨝	𑨝
TSA	𑨣	𑨣
SMALL A	𑨩	𑨩, 𑨩
VA	𑨭	𑨭, 𑨭, 𑨭

The letters 𑨔 and 𑨘 are reversed forms of 𑨙 TA and 𑨝 NA, which are used for representing the Sanskrit retroflex sounds *ṭa* and *ṇa*. Figure 26 shows a subjoined form of 𑨔 in the conjunct 𑨯𑩇𑨔 *ṣṭa*; figure 27 shows 𑨘 in the word 𑨢𑨘𑨁 *maṇi*. There exist distinctive letters for *ṭa* and *ṇa* — 𑨔 TTA and 𑨘 NNA — however, the practice

of reversing 𑨤 TA and 𑨦 NA for 𑨟 *ṭa* and 𑨣 *ṇa* is borrowed from Tibetan, in which the letters for the dental consonants 𑨤 *ta*, 𑨥 *tha*, 𑨦 *da*, 𑨧 *dha*, 𑨦 *na* are reversed for producing the retroflex consonants 𑨟 *ṭa*, 𑨠 *ṭha*, 𑨡 *ḍa*, 𑨢 *ḍha*, 𑨣 *ṇa*; and the palatal sibilant 𑨪 *śa* is reversed for the retroflex sibilant 𑨩 *ṣa*. The use of reversed letters for representing sounds for which distinctive letters exist may be interpreted as scribal idiosyncrasies, which result from imprecise knowledge of the script by the writer, who nonetheless has knowledge of Tibetan. Although 𑨟 *ṭa* and 𑨣 *ṇa* are distinct from the regular letters, their occurrence is limited to a single source. Until additional evidence of their usage is found, they are not proposed for encoding.

Scribal idiosyncrasy may also explain the use of the form 𑨎 for writing 𑨍 SMALL A in figure 35. This form of SMALL A is based upon the Tibetan *dbu med* shape 𑨎 of the regular *dbu can* form 𑨁 *'a chung*. Similarly, the reversed form 𑨏 of SMALL A is used in figure 36.

Similarities between character glyphs may account for the usage of 𑨯 for 𑨰 TSA for representing Mongolian *ǰa* in one manuscript. The form 𑨯 is derived by attaching a hook to 𑨮 BHA. It is likely the result of the scribe writing the left vertical stroke of 𑨰 as a curve from the left terminals instead of as a straight line.

Misinterpretation of glyph boundaries in traditional charts may explain the alternate form 𑨭𑨿 of 𑨭 DHA and the form 𑨶𑨿 for 𑨶 VA (see figure 38). In the traditional ordering of the script, the 𑨭 DHA and 𑨶 VA occur at the end of sequences of letters, eg. "𑨤𑨭𑨿" *da dha* | and "𑨴𑨵𑨶𑨿" *dza dzha va* |. In some manuscripts that show the repertoire of the script, the spacing between letters and the punctuation mark 𑨿 SHAD is rather tight, so "𑨭𑨿" and "𑨶𑨿" may be read incorrectly as "𑨭" and "𑨶", in which the SHAD is seen not as punctuation, but as part of the glyph (see figure 37).

Explanations for other alternate forms require further research. The source for the variant 𑨨𑨿 for 𑨨 GHA as shown in figure 9 is unidentified. The form 𑨧 for 𑨦 NA is shown in figure 28, where both forms are used simultaneously. There is no semantic distinction between 𑨧 and 𑨦. It is unclear why the scribe used two different forms of NA in such close promixity in a single dcument.

4.8 Consonant Modifiers

4.8.1 Final Consonant Mark

The 𑨁 FINAL CONSONANT MARK is used in Mongolian contexts for marking syllable-final consonants. On account of this function, it also serves as a syllable-boundary mark, similar to ་ U+0F0B TIBETAN MARK INTERSYLLABIC TSHEG. It has no control properties. Although it can combine only with a single consonant, there is a single record written in Tibetan in which the FINAL CONSONANT MARK is used below conjuncts and after vowels. This usage is irregular and is likely the result of a scribe using the mark, not as a vowel silencer, as a generic syllable mark instead of specifically as a final-consonant mark. The FINAL CONSONANT MARK occurs in one Mongolian record as 𑨁 (see figure 34), which is considered a glyphic variant of 𑨁.

4.8.2 Virama

The 𑨁 VIRAMA is used in Sanskrit and Tibetan contexts for silencing the inherent vowel of a consonant letter. It can occur only with a consonant. It also specifies that if the bare consonant is followed by another consonant, then the bare consonant is part of a cluster involving the following consonant (see section 4.10). It does not mark syllabic boundaries.

Although VIRAMA is not part of the traditional repertoire of Zanabazar Square, it is attested in texts (see figure 30). It corresponds to the Lantsa, Wartu, and Tibetan sign *halanta*, as shown in figure 30. In addition to silencing the inherent vowel of a consonant, the Zanabazar Square VIRAMA is used for representing conjuncts

in encoded text: a sequence of bare consonants marked by the sign will be displayed as conjunct stacks. Given the form and function of the character in Zanabazar Square, VIRAMA is an appropriate name for the sign as it corresponds to other VIRAMA characters in scripts in the UCS, such as ◌् U+094D DEVANAGARI SIGN VIRAMA. In previous versions of the proposal, the VIRAMA was called *subjoiner*. It was proposed as a character for controling conjunct formation using a generic glyph. While the *subjoiner* was suitable, it was more appropriate to unify its function with that of the VIRAMA, which fits the graphical model of the script.

4.8.3 Necessity for Two Vowel-Silencing Marks

The FINAL CONSONANT MARK and VIRAMA may be considered graphical variants of a single vowel-silencing mark. However, as explained in the above descriptions of each mark, it is necessary to treat them as separate characters in the encoding on account of their language-specific semantics and behavior. The distinct semantics of both characters is is exhibited by the following encoded sequence, in which represents a single *silencer* mark for purposes of illustration:

> <SA, VOWEL SIGN E, THA, *silencer*, KHA, VOWEL SIGN I, LA, *silencer*, PA, VOWEL SIGN E, RA, *silencer*>

The above sequence would be rendered in a Mongolian context as *sed-kil-ber* and in a Sanskrit or Tibetan context as *setkhilper*.

From a character-encoding perspective, although the *silencer* indicates bare consonants in both of the above cases, it also conveys additional details regarding the display of the consonants based upon linguistic context. In the Mongolian example, the *silencer* indicates that THA, LA, and RA are bare consonants and that they occur at the end of a syllable. In the Sanskrit example, the *silencer* indicates that THA, LA, and RA are bare consonants and are part of a cluster if followed by a consonant, and that the cluster should be rendered a as conjunct.

In plain-text environments, there is no means of instructing the *silencer* to behave as would be expected in different linguistic contexts. For this reason, it is necessary to separate the two functions of the *silencer* into separate characters: one that marks bare consonants in syllable-final position and the other that marks bare consonants in a consonant cluster: the FINAL CONSONANT MARK and VIRAMA, respectively. These two characters offer a feasible method of differentiating the representation of the example sequence in Mongolian and Sanskrit contexts:

	sed-kil-ber	<SA, VOWEL SIGN E, THA, FINAL CONSONANT MARK, KHA, VOWEL SIGN I, LA, FINAL CONSONANT MARK, PA, VOWEL SIGN E, RA, FINAL CONSONANT MARK >
	setkhilper	<SA, VOWEL SIGN E, THA, VIRAMA, KHA, VOWEL SIGN I, LA, VIRAMA, PA, VOWEL SIGN E, RA, VIRAMA >

4.9 Mongolian Final Consonants

Mongolian words may end with the following codas: *g*, *k*, *ng*, *d*, *n*, *b*, *m*, *r*, *l*, *š*, *s*. These are shown in charts as *ag*, *ak*, *aṅ*, *ad*, *an*, *ab*, *am*, *ar*, *al*, *aš*, *as*; the initial

𑨀 A is used for illustrating a basic syllable. Syllable-final consonants are indicated by placing the ◌𑨳 FINAL CONSONANT MARK beneath a letter, as follows:

g	𑨋𑨳	<𑨋 KA, ◌𑨳 FINAL CONSONANT MARK>
k	𑨌𑨳	<𑨌 KHA, ◌𑨳 FINAL CONSONANT MARK>
ng	𑨏𑨳	<𑨏 NGA, ◌𑨳 FINAL CONSONANT MARK>
d	𑨙𑨳	<𑨙 TA, ◌𑨳 FINAL CONSONANT MARK>
n	𑨝𑨳	<𑨝 NA, ◌𑨳 FINAL CONSONANT MARK>
b	𑨞𑨳	<𑨞 PA, ◌𑨳 FINAL CONSONANT MARK>
m	𑨢𑨳	<𑨢 MA, ◌𑨳 FINAL CONSONANT MARK>
r	𑨫𑨳	<𑨫 RA, ◌𑨳 FINAL CONSONANT MARK>
l	𑨬𑨳	<𑨬 LA, ◌𑨳 FINAL CONSONANT MARK>
š	𑨮𑨳	<𑨮 SHA, ◌𑨳 FINAL CONSONANT MARK>
s	𑨰𑨳	<𑨰 SA, ◌𑨳 FINAL CONSONANT MARK>

The 𑨩 *ā* is shown at the end of the syllable-final letters. It is glossed as ཨཱ *ā* in several manuscripts. Kara (1972) represents 𑨩 as 𑨩 'aⁿ', while Shagdarsürüng (2001) represents it as 𑨩 'āṉ'. These transliterations indicate that Kara and Shagdarsürüng assign a nasal value to 𑨩, but they do not offer evidence for their claims. Moreover, such usage is not attested in the available sources. Byambaa Ragchaagin (2005) equates the 𑨩 that occurs among the syllable-final letters with the vowel *ā*. The manuscript sources suggest that the occurrence of 𑨩 here simply indicates that a Mongolian syllable may end with this long vowel.

4.10 Consonant Conjuncts

Consonant clusters are written as conjuncts, which are rendered as vertical stacks with non-initial letters descending sequentially beneath the initial letter, eg. 𑨝 NA + 𑨛 DA is written as 𑨝𑩇𑨛 *nda*. Letters are displayed using their regular shape, with the exception of four letters: 𑨪 YA, 𑨫 RA, 𑨬 LA, 𑨭 VA. The forms of these letters are determined by their position in a cluster and the linguistic context in which they occur. In general, the full forms of these letters are used for Sanskrit conjuncts, while contextual alternates are used for Tibetan, but variation does occur (see figures 24):

	Initial		Medial / Final	
	Sanskrit	Tibetan	Sanskrit	Tibetan
𑨪 YA	𑨪	𑨪	◌𑩇𑨪	◌𑨻
𑨫 RA	𑨫	𑨺	◌𑩇𑨫	◌𑨼
𑨬 LA	𑨬	𑨬	◌𑩇𑨬	◌𑨽
𑨭 VA	𑨭	𑨭	◌𑩇𑨭	◌𑨾

Given that only four consonants exhibit special shaping behavior in stacks, the encoded representation of conjuncts requires a subjoining model that utilizes VIRAMA for controling stacking behavior, as well as 5 characters for the contextual forms of YA, RA, LA, VA:

CLUSTER INITIAL RA CLUSTER FINAL YA CLUSTER FINAL LA

CLUSTER FINAL RA CLUSTER FINAL VA

The CLUSTER INITIAL RA can occur only at the head of a cluster and must be followed by consonant letter. By implication, it is the logical base letter of a conjunct and any following consonant letter is always subjoined, therefore VIRAMA is not to be used in conjunction with it.

The CLUSTER FINAL YA, CLUSTER FINAL RA, CLUSTER FINAL LA, CLUSTER FINAL VA may occur only in the penultimate and final positions in a cluster. They are combining characters and logically combine with the preceding consonant. As inherently subjoined forms, the VIRAMA is not used with these letters, either in a preceding or following position. These cluster-final letters may occur after each other in the penultimate and final positions in a cluster, but as VIRAMA cannot combine with them, these letters cannot be used medially.

4.10.1 Rationale for the Encoding Model for Conjuncts

The visual model for encoding Zanabazar Square conjuncts provides for all stacks that may occur in Sanskrit and Tibetan. It also eliminates the need to adopt the Tibetan subjoined-letter model for Zanabazar Square, which would require the independent encoding of a full set of subjoined letters for each consonant letter, in addition to context-specific forms of YA, RA, LA, VA. The decision to encode a CLUSTER INITIAL RA instead of relying on the font to change the shape of RA from full-form to head-position is made in order to eliminate the need to encode two characters that have the same nominal appearance, but different conjoining behaviors, as is the case for Tibetan in the UCS, eg. U+0F62 TIBETAN LETTER RA and U+0FBC TIBETAN LETTER FIXED-FORM RA. The proposed model simply utilizes separate characters for each semantically distinctive form of a consonant. All other consonants are rendered in stacks using the VIRAMA, which simply subjoins one letter beneath the other and does not initiate a shape change of a consonant.

4.10.2 Encoded Representation of Conjuncts

A conjunct stack is represented in encoded text by placing a VIRAMA after each consonant in a cluster except for the last: <*consonant*, VIRAMA, [*consonant*, VIRAMA,]* *consonant*>. The VIRAMA specifies that the consonant to which it is attached is a bare consonant and that the following letter is to be placed beneath it; the same rule applies to each subsequent consonant to which VIRAMA is attached. As VIRAMA behaves as a control character when it occurs between two consonants, it is not displayed visibly in the stack (see sections 4.10.3 and 4.17.5 for additional details on the rendering of stacks):

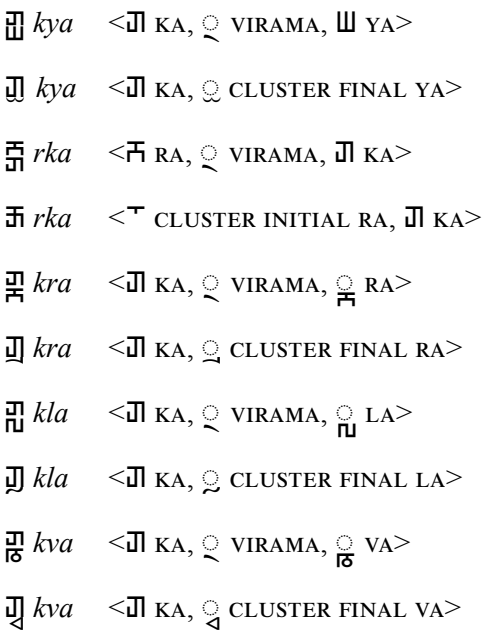
n <NA, VIRAMA>

nda <NA, VIRAMA, DA>

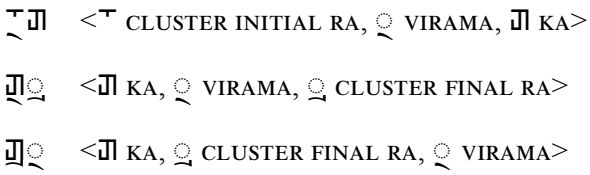
ndra <NA, VIRAMA, DA, VIRAMA, RA>

The exceptions to the usage of VIRAMA in producing conjuncts are CLUSTER INITIAL RA, CLUSTER FINAL YA, CLUSTER FINAL RA, CLUSTER FINAL LA, CLUSTER FINAL VA. The usage of these letters in stacks is

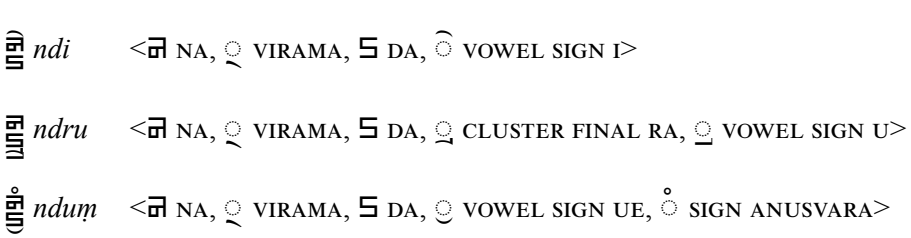
shown in the examples below:



Using VIRAMA with the 5 conjunct-specific letters will yield meaningless sequences:



Naturally, in a conjunct the only letter that can possess a vowel is the final letter in the cluster. Any vowel or other combining signs must be placed after the final letter in a conjunct:



4.10.3 Controlling Conjunct Formation

As is the convention for conjunct models based upon VIRAMA, conjunct formation may be broken by placing the generic non-printing, control character ZWNJ U+200C ZERO WIDTH NON-JOINER (abbreviated as ZWNJ) after VIRAMA. This results in the display of a visible form of the sign, compare the following:

𑨞𑨴𑨜 *nda* <𑨞 NA, ◌𑨴 VIRAMA, 𑨜 DA>

𑨞𑨴‌𑨜 *nda* <𑨞 NA, ◌𑨴 VIRAMA, ‌ ZWNJ, 𑨜 DA>

𑨞𑨴𑨜𑨴𑨬 *ndra* <𑨞 NA, ◌𑨴 VIRAMA, 𑨜 DA, ◌𑨴 VIRAMA, 𑨬 RA>

𑨞𑨴‌𑨜𑨴‌𑨬 *ndra* <𑨞 NA, ◌𑨴 VIRAMA, ‌ ZWNJ, 𑨜 DA, ◌𑨴 VIRAMA, ‌ ZWNJ, 𑨬 RA>

The use of ZWNJ is necessary for representing clusters like 𑨣𑨴‌𑨱𑨴𑨚 *msta*, which occurs in the phrase 𑨯𑨃𑨢𑨣𑨴‌𑨱𑨴𑨚𑨃 *śubhamstu* and 𑨬𑨴‌𑨱 *rsa*, which occurs in the phrase 𑨋𑨱𑨻𑨡𑨬𑨴‌𑨱𑨬𑨴𑨮 *kasyabar sarva* in figure 30. As shown below, if ZWNJ is not used, the VIRAMA will produce a stack instead of a consonant with a visible VIRAMA:

𑨣𑨴𑨱𑨴𑨚 *msta* <𑨣 MA, ◌𑨴 VIRAMA, 𑨱 SA, ◌𑨴 VIRAMA, 𑨚 TA>

𑨣𑨴‌𑨱𑨴𑨚 *msta* <𑨣 MA, ◌𑨴 VIRAMA, ‌ ZWNJ, 𑨱 SA, ◌𑨴 VIRAMA, 𑨚 TA>

𑨬𑨴𑨱 *rsa* <𑨬 RA, ◌𑨴 VIRAMA, 𑨱 SA>

𑨬𑨴‌𑨱 *rsa* <𑨬 RA, ◌𑨴 VIRAMA, ‌ ZWNJ, 𑨱 SA>

As there are no half-forms of consonants in Zanabazar Square, the usage of ‍ U+200D ZERO WIDTH JOINER does not produce any valid output.

4.10.4 Conjuncts Shown in Traditional Script Charts

The following conjuncts are shown in traditional script charts: 𑨋𑨼 *kra*, 𑨌𑨻 *khya*, 𑨍𑨽 *gla*, 𑨺𑨋 *rka*, 𑨱𑨴𑨋 *ska*, 𑨭𑨴𑨋 *lka*. They are not independent characters, but are ligatures that represent consonant conjuncts. They are represented in encoded text as follows:

𑨋𑨼 *kra* <𑨋 KA, ◌𑨼 CLUSTER FINAL RA>

𑨌𑨻 *khya* <𑨌 KHA, ◌𑨻 CLUSTER FINAL YA>

𑨍𑨽 *gla* <𑨍 GA, ◌𑨽 CLUSTER FINAL LA>

𑨺𑨋 *rka* <𑨺 CLUSTER INITIAL RA, 𑨋 KA>

𑨱𑨴𑨋 *ska* <𑨱 SA, ◌𑨴 VIRAMA, 𑨋 KA>

𑨭𑨴𑨋 *lka* <𑨭 LA, ◌𑨴 VIRAMA, 𑨋 KA>

It is likely that *kra*, *khya*, *gla* are presented for illustrating the cluster-final forms of YA, RA, LA and the cluster-initial form of RA. The presence of *ska* and *lka* in this list is intended for illustrating the representation of Tibetan ལ་མགོ་ *la-mgo* and ས་མགོ་ *sa-mgo* letters. The glyphs 𑨱𑨴𑨋 *ska* and 𑨭𑨴𑨋 *lka* are stylized ligatures of 𑨱𑩇𑨋 and 𑨭𑩇𑨋, respectively.

4.10.5 Depth of Conjunct Stacks

Zanabazar Square stacks may consist of numerous consonants. The deepest stack shown in the available sources contains three consonants: 𑨡𑩇𑨝𑩇𑨪𑨊 *mprā*. However, as Zanabazar Square is used for representing Tibetan, the rendering engine must be able to manage stacks consisting of greater than three consonants.

4.11 Head Marks

Sources show 𑨿𑨶 — and the variants 𑨿𑨵, 𑨿𑨷, 𑨿𑨸 — as a head mark used at the beginning of texts (see figure 40 for more variants). Rather than encode one of these as the normative form or each as a separate head-mark character, it is sensible to analyze the mark as consisting of a base character and a combining sign. This approach offers the possibility of representing various forms simply by establishing a bare head mark to which signs proposed for encoding in section 4.6 may be attached. Moreover, as figure 40 illustrates, the head mark may be used without an ornament, even if such usage is rare.

The plain form is proposed for encoding as 𑨿 INITIAL HEAD MARK. It corresponds to ࿓ U+0FD3 TIBETAN MARK INITIAL BRDA RNYING YIG MGO MDUN MA. The various forms of the initial head mark may be represented in encoded text as follows:

- 𑨿𑨵 <𑨿 INITIAL HEAD MARK, ◌𑨵 SIGN CANDRABINDU>
- 𑨿𑨶 <𑨿 INITIAL HEAD MARK, ◌𑨶 SIGN CANDRABINDU WITH ORNAMENT>
- 𑨿𑨷 <𑨿 INITIAL HEAD MARK, ◌𑨷 SIGN CANDRA WITH ORNAMENT>
- 𑨿𑨸 <𑨿 INITIAL HEAD MARK, ◌𑨸 SIGN ANUSVARA>

Additionally, the 𑩀 CLOSING HEAD MARK is proposed for encoding in order to produce a double head mark. It corresponds to ࿔ U+0FD4 TIBETAN CLOSING MARK BRDA RNYING YIG MGO MDUN MA. It is placed after 𑨿 INITIAL HEAD MARK for producing 𑨿𑩀 <𑨿 INITIAL HEAD MARK, 𑩀 CLOSING HEAD MARK>, cf. Tibetan ࿓࿔. It is possible to encode 𑨿𑩀 as a separate character, as was done for ꡵ U+A875 PHAGS-PA DOUBLE HEAD MARK; however, analyzing this mark as consisting of 'initial' and 'closing' parts enables the representation of extended head marks, such as 𑨿𑩀𑩀, which aligns with Tibetan ࿓࿔࿔.

Figure 41 shows a mark 𑨿𑩀, which corresponds to Tibetan ༄༅. This mark is to be considered a glyphic variation of 𑨿𑩀, where 𑨿 = 𑨿 INITIAL HEAD MARK and 𑩀 = 𑩀 CLOSING HEAD MARK. Although these forms are encoded as distinct characters for Tibetan (༄ U+0F04 TIBETAN MARK INITIAL YIG MGO MDUN MA and ༅ U+0F05 TIBETAN MARK CLOSING YIG MGO SGAB MA), for Zanabazar Square it is appropriate to treat them as variants because of their limited occurrence. As such, their display is to be managed through fonts.

The 𑨿 INITIAL HEAD MARK, and a following 𑩀 CLOSING HEAD MARK, is typically followed by a 𑩁 SHAD or 𑩂 DOUBLE SHAD, eg. 𑨿𑨶𑩁, 𑨿𑨷𑩂, 𑨿𑩀𑩀𑩂. In such cases, the SHAD and DOUBLE SHAD are to be placed at the end of the sequence, eg. 𑨿𑨷𑩂 is represented as <𑨿 INITIAL HEAD MARK, ◌𑨷 SIGN CANDRA WITH ORNAMENT, 𑩂 DOUBLE SHAD>.

4.12 Punctuation

The following characters are used for punctuation:

TSHEG is used for indicating the end of a syllable in Tibetan contexts. It corresponds to U+0F0B TIBETAN MARK INTERSYLLABIC TSHEG. Although the FINAL CONSONANT MARK is generally used for marking syllabic boundaries, the TSHEG is proposed for inclusion in the script block because of attested usage (see figures 46 and 47).

SHAD indicates the end of a phrase or sentence. It corresponds to U+0F0D TIBETAN MARK SHAD.

DOUBLE SHAD marks the end of a text section (see figure 23). It corresponds to U+0F0E TIBETAN MARK NYIS SHAD.

LONG TSHEG behaves as a comma (see figure 41). It corresponds to U+0F0E TIBETAN MARK GTER TSHEG.

4.13 Digits

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

4.14 Collation

The default sort order for Zanabazar Square is based for the most part upon the order for Tibetan:

KA < KSSA < KHA < GA < GHA < NGA < CA < CHA <
JA < NYA < TTA < TTHA < DDA < DDHA < NNA < TA <
THA < DA < DHA < NA < PA < PHA < BA < BHA <
MA < TSA < TSHA < DZA < DZHA < ZHA < ZA <
SMALL A < YA << CLUSTER FINAL YA < CLUSTER INITIAL RA << RA <
CLUSTER FINAL RA < LA << CLUSTER FINAL LA < VA <<
CLUSTER FINAL VA < SHA < SSA < SA < HA < A <
VOWEL SIGN I < VOWEL SIGN UE < VOWEL SIGN U < VOWEL SIGN E <
VOWEL SIGN OE < VOWEL SIGN O < VOWEL SIGN AI < VOWEL SIGN AU <
VOWEL SIGN REVERSED I < VOWEL LENGTH MARK < FINAL CONSONANT MARK

The following characters have secondary weights:

CANDRABINDU, CANDRABINDU WITH ORNAMENT, CANDRA WITH ORNAMENT, ANUSVARA, VISARGA

4.15 Vertical Text

Although Zanabazar Square was designed to be written horizontally, there some instances in the available sources in which the script is oriented vertically, for example, the words *haṃkṣamalavaraya* (see figure 21) and *thalīṃ* (see figures 22 and 36):

𑨱𑨵𑨲𑨢𑨬𑨭𑨫𑨪 (vertical)	𑨱𑨵𑨲𑨢𑨬𑨭𑨫𑨪	*haṃkṣamalavaraya*	<𑨱 HA, ◌𑨵 SIGN CANDRABINDU, 𑨲 KSSA, 𑨢 MA, 𑨬 LA, 𑨭 VA, 𑨫 RA, 𑨪 YA>
𑨚𑨬𑨁𑨸𑨩 (vertical)	𑨚𑨬𑨁𑨸𑨩	*thalīṃ*	<𑨚 THA, 𑨬 LA, ◌𑨁 VOWEL SIGN I, ◌𑨸 ANUSVARA, 𑨩 SMALL A>

The graphical representation of vertical text is identical to that of conjuncts: letters are written in their normal, upright shape and are positioned sequentially beneath the first letter of the word, and signs are positioned as they are in conjuncts (see section 4.17). Given this, *haṃkṣamalavaraya* and *thalīṃ* may be incorrectly parsed as the conjuncts **hkṣmlvryaṃ* and **thl'īṃ*, respectively.

Although the vertical orientation of *haṃkṣamalavaraya* and *thalīṃ* may be produced in encoded text as false conjuncts by placing VIRAMA after each consonant letter, such an approach is not recommended because it obscures the semantic value of the words. For instance, in the correct encoded representation of *haṃkṣa-malavaraya* the ◌𑨵 CANDRABINDU attaches to 𑨱 HA, but an attempt to produce the desired output as a conjunct would require that the ◌𑨵 CANDRABINDU be attached, incorrectly, to 𑨪 YA; naturally, combining signs can attach only to the final consonant of a conjunct. Similarly, for *thalīṃ*, both ◌𑨁 VOWEL SIGN I and ◌𑨸 ANUSVARA combine logically with 𑨬 LA, but if the word is encoded as a false conjunct, the signs would need to be combined with 𑨚 THA.

The word *thalīṃ* is a special case. Byambaa Ragchaagiin states that the word is a Tantric *mantra* and is always written vertically, as shown above, and does not occur horizontally as *𑨚𑨬𑨁𑨸𑨩.

Vertical text is outside of the purview of the encoding, which is intended for the representation of plain text. Therefore, the display of vertical text must be managed at the presentation layer. It may be possible to control vertical orientation in OpenType fonts using the 'vert' feature. The basic rules for Zanabazar Square in vertical environments is that text must be set top-to-bottom, left-to-right, with upright glyphs. The recently published "Unicode Technical Report #50: Unicode Vertical Text Layout" describes the character property `Vertical_Orientation` (`vo`) for specifying default character orientation (Ishii 2013). For Zanabazar Square, the property would be defined as `Vertical_Orientation=U` (`vo=U`), where the value '`U`' indicates that the glyphs remain upright in both horizontal and vertical text layout, as shown in the code chart.

4.16 Combining Behavior

Multiple signs may combine with a base letter. The example shown below is theoretical and does not occur in attested texts, but it is nonetheless a valid encoded sequence for Zanabazar Square and illustrates the textual and graphical possibilities that rendering engines and fonts must be able to process and display.

<CLUSTER INITIAL RA, KA, VIRAMA, TA, VIRAMA, BA, CLUSTER FINAL RA, CLUSTER FINAL YA, VOWEL SIGN E, VOWEL LENGTH MARK, VOWEL SIGN AU, ANUSVARA, VISARGA>

The above example, which could be transliterated as *rktvryēuṃḥ*, is produced using 13 characters. See section 4.17.3 for rules regarding the placement of vowel signs with conjuncts.

4.17 Glyph Interactions

4.17.1 Size of Vowel Signs for Use With LETTER A

The widths of glyphs for Zanabazar Square consonant letters are uniform, however, the A is wider than the consonants. This width difference requires a separate set of extra-wide vowel signs for use with A, compare, and , and . With width variant forms, the combinations with A should resemble and .

4.17.2 Placement of Multiple Combining Signs

When multiple signs occur combine with a base letter in the same position, it may be necessary to adjust the glyphs in order to prevent clashing. This may be achieved in various ways (signs are marked in red). One is to horizontally extend the anchor of the sign:

<KA, VOWEL SIGN E, VOWEL SIGN AI (variant)> = → *kei*

Another is to horizontally condense the shape of one sign and to place both laterally:

<LA, VOWEL SIGN I, VOWEL LENGTH MARK, SIGN ANUSVARA> = → *līṃ*

A third is to alter the vertical position of a sign:

<A, VOWEL SIGN O, SIGN ANUSVARA> = → *oṃ*

<NA, VIRAMA, DA, CLUSTER FINAL RA, VOWEL SIGN U> = → *ndru*

4.17.3 Placement of Vowel Signs in Conjuncts

Although all vowel signs are combined with the final letter in the encoded representation of a conjunct, the placement of signs upon the stack is dependent upon the combining behavior of the sign (marked in red):

Above-base vowel signs are placed above the initial letter:

ndi <NA, VIRAMA, DA, VOWEL SIGN I>

Below-base vowel signs (including VOWEL LENGTH MARK) are placed beneath the final letter:

 *ndu* <𑨦 NA, 𑨵 VIRAMA, 𑨤 DA, 𑨳 VOWEL SIGN U>

 *ndru* <𑨦 NA, 𑨵 VIRAMA, 𑨤 DA, 𑨴 CLUSTER FINAL RA, 𑨳 VOWEL SIGN U>

 *ndru* <𑨦 NA, 𑨵 VIRAMA, 𑨤 DA, 𑨴 CLUSTER FINAL RA, 𑨳 VOWEL SIGN U, 𑨹 VOWEL LENGTH MARK>

Right and left spacing marks are positioned on the respective sides of the initial letter:

 *ndai* <𑨦 NA, 𑨵 VIRAMA, 𑨤 DA, 𑨷 VOWEL SIGN AI>

 *ndau* <𑨦 NA, 𑨵 VIRAMA, 𑨤 DA, 𑨸 VOWEL SIGN AU>

The above rules apply to cases where multiple vowel signs and modifiers occur at the end of a conjunct:

 *ndī* <𑨦 NA, 𑨵 VIRAMA, 𑨤 DA, 𑨲 VOWEL SIGN I, 𑨹 VOWEL LENGTH MARK>

 *ndauḥ* <𑨦 NA, 𑨵 VIRAMA, 𑨤 DA, 𑨸 VOWEL SIGN AU, 𑨾 SIGN VISARGA>

 *ndōṃ* <𑨦 NA, 𑨵 VIRAMA, 𑨤 DA, 𑨶 VOWEL SIGN O, 𑨹 VOWEL LENGTH MARK, 𑨿 ANUSVARA>

When 𑨪 CLUSTER INITIAL RA occurs in a conjunct, above-base marks attach to it, while all other marks attach to the following consonants based upon the rules stated above:

 *rki* <𑨪 CLUSTER INITIAL RA, 𑨐 KA, 𑨲 VOWEL SIGN I>

 *rkai* <𑨪 CLUSTER INITIAL RA, 𑨐 KA, 𑨷 VOWEL SIGN AI>

When 𑨹 VOWEL LENGTH MARK occurs with 𑨶 CLUSTER FINAL YA, 𑨴 CLUSTER FINAL RA, 𑨽 CLUSTER FINAL LA, 𑨼 CLUSTER FINAL VA, the mark attaches to the letter which the cluster-final letter combines:

 *krā* <𑨐 KA, 𑨴 CLUSTER FINAL RA, 𑨹 VOWEL LENGTH MARK>

4.17.4 Positioning of CLUSTER INITIAL RA

The 𑨪 CLUSTER INITIAL RA is positioned at the normal head height, not above it. For this reason, it may be necessary to adjust the height of the following consonant letter in order to accommodate fit:

<𑨪 CLUSTER INITIAL RA, 𑨐 KA> = 𑨐 → 𑨑

4.17.5 Subjoined Glyphs for Rendering Conjuncts

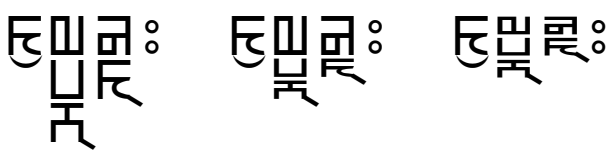
A Zanabazar Square font must contain a full set of subjoined forms for each consonant letter. The font will produce a consonant stack by substituting each <𑨵 VIRAMA, *consonant*> pair with a subjoined form of the

consonant letter. If the subjoined glyph is not available in the font, the VIRAMA will be displayed visibly along with the regular glyph of the letter whose subjoined form is missing. For example, if the subjoined form ◌ of 𑨤 DA is unavailable, then a sequence such as <𑨦 NA, ◌ VIRAMA, 𑨤 DA> will be rendered as 𑨦𑨯𑨤 instead of the expected 𑨥.

4.17.6 Positioning and Sizing of Letters in Conjunct Stacks

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 scribes adjust the size of letters in stacks for visual uniformity with surrounding characters. Shown below is the word *dhumprāndhāḥ* rendered in three different ways:

A B C



The default method involves no size changes and uses the regular forms of letters (column ‘A’). In some sources, the regular glyph size of the initial letter is used, while the glyph of a non-initial consonant is compressed along the vertical axis such that their x-height is similar to that of below-base signs (see figure 4 and column ‘B’). A third practice is to vertically condense the glyphs for the base and subjoined letters such that the height of the stack matches the height of the surrounding letters (see figure 23 and column ‘C’). Depending on the head-height of surrounding letters, such size adjustments may be practical only for stacks consisting of two letters.

5 Character Data

5.1 Character Properties

Character properties given in the data format of `UnicodeData.txt`:

```
11A00;ZANABAZAR SQUARE LETTER A;Lo;0;L;;;;;N;;;;;
11A01;ZANABAZAR SQUARE VOWEL SIGN I;Mn;0;NSM;;;;;N;;;;;
11A02;ZANABAZAR SQUARE VOWEL SIGN UE;Mn;0;NSM;;;;;N;;;;;
11A03;ZANABAZAR SQUARE VOWEL SIGN U;Mn;0;NSM;;;;;N;;;;;
11A04;ZANABAZAR SQUARE VOWEL SIGN E;Mn;0;NSM;;;;;N;;;;;
11A05;ZANABAZAR SQUARE VOWEL SIGN OE;Mn;0;NSM;;;;;N;;;;;
11A06;ZANABAZAR SQUARE VOWEL SIGN O;Mn;0;NSM;;;;;N;;;;;
11A07;ZANABAZAR SQUARE VOWEL SIGN AI;Mc;0;L;;;;;N;;;;;
11A08;ZANABAZAR SQUARE VOWEL SIGN AU;Mc;0;L;;;;;N;;;;;
11A09;ZANABAZAR SQUARE VOWEL SIGN REVERSED I;Mn;0;NSM;;;;;N;;;;;
11A0A;ZANABAZAR SQUARE VOWEL LENGTH MARK;Mc;0;L;;;;;N;;;;;
11A0B;ZANABAZAR SQUARE LETTER KA;Lo;0;L;;;;;N;;;;;
11A0C;ZANABAZAR SQUARE LETTER KHA;Lo;0;L;;;;;N;;;;;
11A0D;ZANABAZAR SQUARE LETTER GA;Lo;0;L;;;;;N;;;;;
11A0E;ZANABAZAR SQUARE LETTER GHA;Lo;0;L;;;;;N;;;;;
11A0F;ZANABAZAR SQUARE LETTER NGA;Lo;0;L;;;;;N;;;;;
11A10;ZANABAZAR SQUARE LETTER CA;Lo;0;L;;;;;N;;;;;
11A11;ZANABAZAR SQUARE LETTER CHA;Lo;0;L;;;;;N;;;;;
11A12;ZANABAZAR SQUARE LETTER JA;Lo;0;L;;;;;N;;;;;
11A13;ZANABAZAR SQUARE LETTER NYA;Lo;0;L;;;;;N;;;;;
11A14;ZANABAZAR SQUARE LETTER TTA;Lo;0;L;;;;;N;;;;;
11A15;ZANABAZAR SQUARE LETTER TTHA;Lo;0;L;;;;;N;;;;;
11A16;ZANABAZAR SQUARE LETTER DDA;Lo;0;L;;;;;N;;;;;
11A17;ZANABAZAR SQUARE LETTER DDHA;Lo;0;L;;;;;N;;;;;
```

```
11A18;ZANABAZAR SQUARE LETTER NNA;Lo;0;L;;;;;N;;;;;
11A19;ZANABAZAR SQUARE LETTER TA;Lo;0;L;;;;;N;;;;;
11A1A;ZANABAZAR SQUARE LETTER THA;Lo;0;L;;;;;N;;;;;
11A1B;ZANABAZAR SQUARE LETTER DA;Lo;0;L;;;;;N;;;;;
11A1C;ZANABAZAR SQUARE LETTER DHA;Lo;0;L;;;;;N;;;;;
11A1D;ZANABAZAR SQUARE LETTER NA;Lo;0;L;;;;;N;;;;;
11A1E;ZANABAZAR SQUARE LETTER PA;Lo;0;L;;;;;N;;;;;
11A1F;ZANABAZAR SQUARE LETTER PHA;Lo;0;L;;;;;N;;;;;
11A20;ZANABAZAR SQUARE LETTER BA;Lo;0;L;;;;;N;;;;;
11A21;ZANABAZAR SQUARE LETTER BHA;Lo;0;L;;;;;N;;;;;
11A22;ZANABAZAR SQUARE LETTER MA;Lo;0;L;;;;;N;;;;;
11A23;ZANABAZAR SQUARE LETTER TSA;Lo;0;L;;;;;N;;;;;
11A24;ZANABAZAR SQUARE LETTER TSHA;Lo;0;L;;;;;N;;;;;
11A25;ZANABAZAR SQUARE LETTER DZA;Lo;0;L;;;;;N;;;;;
11A26;ZANABAZAR SQUARE LETTER DZHA;Lo;0;L;;;;;N;;;;;
11A27;ZANABAZAR SQUARE LETTER ZHA;Lo;0;L;;;;;N;;;;;
11A28;ZANABAZAR SQUARE LETTER ZA;Lo;0;L;;;;;N;;;;;
11A29;ZANABAZAR SQUARE LETTER SMALL A;Lo;0;L;;;;;N;;;;;
11A2A;ZANABAZAR SQUARE LETTER YA;Lo;0;L;;;;;N;;;;;
11A2B;ZANABAZAR SQUARE LETTER RA;Lo;0;L;;;;;N;;;;;
11A2C;ZANABAZAR SQUARE LETTER LA;Lo;0;L;;;;;N;;;;;
11A2D;ZANABAZAR SQUARE LETTER VA;Lo;0;L;;;;;N;;;;;
11A2E;ZANABAZAR SQUARE LETTER SHA;Lo;0;L;;;;;N;;;;;
11A2F;ZANABAZAR SQUARE LETTER SSA;Lo;0;L;;;;;N;;;;;
11A30;ZANABAZAR SQUARE LETTER SA;Lo;0;L;;;;;N;;;;;
11A31;ZANABAZAR SQUARE LETTER HA;Lo;0;L;;;;;N;;;;;
11A32;ZANABAZAR SQUARE LETTER KSSA;Lo;0;L;;;;;N;;;;;
11A33;ZANABAZAR SQUARE FINAL CONSONANT MARK;Mn;0;NSM;;;;;N;;;;;
11A34;ZANABAZAR SQUARE SIGN VIRAMA;Mn;9;NSM;;;;;N;;;;;
11A35;ZANABAZAR SQUARE SIGN CANDRABINDU;Mn;0;NSM;;;;;N;;;;;
11A36;ZANABAZAR SQUARE SIGN CANDRABINDU WITH ORNAMENT;Mn;0;NSM;;;;;N;;;;;
11A37;ZANABAZAR SQUARE SIGN CANDRA WITH ORNAMENT;Mn;0;NSM;;;;;N;;;;;
11A38;ZANABAZAR SQUARE SIGN ANUSVARA;Mn;0;NSM;;;;;N;;;;;
11A39;ZANABAZAR SQUARE SIGN VISARGA;Mc;0;L;;;;;N;;;;;
11A3A;ZANABAZAR SQUARE LETTER CLUSTER INITIAL RA;Lo;0;L;;;;;N;;;;;
11A3B;ZANABAZAR SQUARE LETTER CLUSTER FINAL YA;Mn;0;NSM;;;;;N;;;;;
11A3C;ZANABAZAR SQUARE LETTER CLUSTER FINAL RA;Mn;0;NSM;;;;;N;;;;;
11A3D;ZANABAZAR SQUARE LETTER CLUSTER FINAL LA;Mn;0;NSM;;;;;N;;;;;
11A3E;ZANABAZAR SQUARE LETTER CLUSTER FINAL VA;Mn;0;NSM;;;;;N;;;;;
11A3F;ZANABAZAR SQUARE INITIAL HEAD MARK;Po;0;ON;;;;;N;;;;;
11A40;ZANABAZAR SQUARE CLOSING HEAD MARK;Po;0;ON;;;;;N;;;;;
11A41;ZANABAZAR SQUARE TSHEG;Po;0;L;;;;;N;;;;;
11A42;ZANABAZAR SQUARE SHAD;Po;0;L;;;;;N;;;;;
11A43;ZANABAZAR SQUARE DOUBLE SHAD;Po;0;L;;;;;N;;;;;
11A44;ZANABAZAR SQUARE LONG TSHEG;Po;0;L;;;;;N;;;;;
```

5.2 Linebreaking Properties

Linebreaking properties given in the data format of `LineBreak.txt`:

```
11A00;        AL     # ZANABAZAR SQUARE LETTER A
11A01..11A0A; CM     # ZANABAZAR SQUARE VOWEL SIGN I .. VOWEL LENGTH MARK
11A0B..11A32; AL     # ZANABAZAR SQUARE LETTER KA .. LETTER KSSA
11A33;        CM     # ZANABAZAR SQUARE FINAL CONSONANT MARK
11A34;        CM     # ZANABAZAR SQUARE SIGN VIRAMA
11A35..11A39; CM     # ZANABAZAR SQUARE SIGN CANDRABINDU .. SIGN VISARGA
11A3A;        AL     # ZANABAZAR SQUARE LETTER CLUSTER INITIAL RA
11A3B..11A3E; CM     # ZANABAZAR SQUARE LETTER CLUSTER FINAL YA .. LETTER CLUSTER FINAL VA
11A3F..11A40; BB     # ZANABAZAR SQUARE INITIAL HEAD MARK .. CLOSING HEAD MARK
11A41..11A44; BA     # ZANABAZAR SQUARE TSHEG .. LONG TSHEG
```

5.3 ‘Confusable’ Characters

Some Zanabazar Square letters resemble those found in other scripts encoded in the UCS:

```
11A0C ZANABAZAR SQUARE LETTER KHA               ;  0F41 TIBETAN LETTER KHA
```

```
11A0F ZANABAZAR SQUARE LETTER NGA                ;  A843 PHAGS-PA LETTER NGA
11A12 ZANABAZAR SQUARE LETTER JA                 ;  A846 PHAGS-PA LETTER JA
11A12 ZANABAZAR SQUARE LETTER DZA                ;  0045 LATIN CAPITAL LETTER E
11A28 ZANABAZAR SQUARE LETTER ZA                 ;  018E LATIN CAPITAL LETTER REVERSED E
11A28 ZANABAZAR SQUARE LETTER ZA                 ;  A855 PHAGS-PA LETTER ZA
11A29 ZANABAZAR SQUARE LETTER SMALL A            ;  A855 PHAGS-PA LETTER SMALL A
11A2B ZANABAZAR SQUARE LETTER RA                 ;  A858 PHAGS-PA LETTER RA
11A2C ZANABAZAR SQUARE LETTER LA                 ;  A859 PHAGS-PA LETTER LA
11A2F ZANABAZAR SQUARE LETTER SA                 ;  004E LATIN CAPITAL LETTER N
11A33 ZANABAZAR SQUARE FINAL CONSONANT MARK      ;  093C DEVANAGARI SIGN NUKTA
11A34 ZANABAZAR SQUARE SIGN VIRAMA               ;  0F84 TIBETAN MARK HALANTA
11A35 ZANABAZAR SQUARE SIGN CANDRABINDU          ;  0F83 TIBETAN SIGN SNA LDAN
11A3A ZANABAZAR SQUARE LETTER CLUSTER INITIAL RA  ;  A872 PHAGS-PA SUPERFIXED LETTER RA
11A3C ZANABAZAR SQUARE LETTER CLUSTER FINAL RA    ;  A871 PHAGS-PA SUBJOINED LETTER RA
11A3E ZANABAZAR SQUARE LETTER CLUSTER FINAL VA    ;  A867 PHAGS-PA SUBJOINED LETTER WA
11A3F ZANABAZAR SQUARE INITIAL HEAD MARK          ;  A874 PHAGS-PA SINGLE HEAD MARK
11A41 ZANABAZAR SQUARE TSHEG                     ;  0F0B TIBETAN MARK INTERSYLLABIC TSHEG
11A42 ZANABAZAR SQUARE SHAD                      ;  0F0D TIBETAN MARK SHAD
11A43 ZANABAZAR SQUARE DOUBLE SHAD               ;  0F0E TIBETAN MARK NYIS SHAD
11A44 ZANABAZAR SQUARE LONG TSHEG                ;  0F14 TIBETAN MARK GTER TSHEG
```

There are internal ‘confusable’ characters:

```
11A06 ZANABAZAR SQUARE VOWEL SIGN OE  ;  11A3A ZANABAZAR SQUARE LETTER CONJUNCT INITIAL RA
11A20 ZANABAZAR SQUARE LETTER PHA     ;  11A3E ZANABAZAR SQUARE INITIAL HEAD MARK
```

5.4 Syllabic Categories

Syllabic categories given in the data format of `IndicSyllabicCategory.txt`:

```
# Indic_Syllabic_Category=Bindu
11A35..11A38  ; Bindu               # Mn   [4] SIGN CANDRABINDU .. SIGN ANUSVARA

# Indic_Syllabic_Category=Visarga
11A39         ; Visarga             # Mc       SIGN VISARGA

# Indic_Syllabic_Category=Virama
11A33         ; Virama              # Mn       FINAL CONSONANT MARK
11A34         ; Virama              # Mn       SIGN VIRAMA

# Indic_Syllabic_Category=Vowel_Independent
11C00..11C0D  ; Vowel_Independent   # Lo       LETTER A

# Indic_Syllabic_Category=Vowel_Dependent
11A01..11A06  ; Vowel_Dependent     # Mn   [6] VOWEL SIGN I .. VOWEL SIGN O
11A07..11A08  ; Vowel_Dependent     # Mc   [2] VOWEL SIGN AI .. VOWEL SIGN AU
11A09         ; Vowel_Dependent     # Mn       VOWEL SIGN REVERSED I
11A0A         ; Vowel_Dependent     # Mc       VOWEL LENGTH MARK

# Indic_Syllabic_Category=Consonant
11A0B..11A32  ; Consonant           # Lo   40] LETTER KA .. LETTER KSSA
11A3A         ; Consonant           # Lo       LETTER CLUSTER INITIAL RA

# Indic_Syllabic_Category=Consonant_Subjoined
11A3B..11A3E  ; Consonant_Subjoined  # Mn   [4] LETTER CLUSTER FINAL YA .. LETTER CLUSTER FINAL VA
```

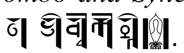
5.5 Matra Categories

Matra categories given in the data format of `IndicMatraCategory.txt`:

```
# Indic_Matra_Category=Right
11A07          ; Right  # Mc        VOWEL SIGN AI
11A0A          ; Right  # Mc        VOWEL LENGTH MARK

# Indic_Matra_Category=Left
```

```
11A08          ; Left   # Mc        VOWEL SIGN AU

# Indic_Matra_Category=Top
11A01          ; Top    # Mn        VOWEL SIGN I
11A04..11A06   ; Top    # Mn    [3] VOWEL SIGN E .. VOWEL SIGN O
11A09          ; Top    # Mn        VOWEL SIGN REVERSED I

# Indic_Matra_Category=Bottom
11A02..11A03   ; Bottom # Mn    [2] VOWEL SIGN UE .. VOWEL SIGN U
11A34          ; Bottom # Mn        SIGN VIRAMA
11A39          ; Bottom # Mn        FINAL CONSONANT MARK
```

6 References

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

Bareja-Starzyńska, Agata. Forthcoming. "Some Notes on the Use of the Horizontal Square Script of Zanabazar". *Mongolo-Tibetica Pragensia*, 2013.

Bareja-Starzyńska, Agata and Byambaa Ragchaa. 2012. "Notes on the Pre-existences of the First Khalkha Jetsundampa Zanabazar according to His Biography Written in the Horizontal Square Script". *Rocznik Orientalistyczny*, vol. LXV, no. 1, pp. 24–40.

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

Corff, Oliver. 2001. "Xäwtää Dörböljin, or The Mongolian Horizontal Square Alphabet for LaTeX 2ε", Release 0.4 alpha. November 5, 2001. http://www.tex.ac.uk/CTAN/languages/mongolian/mxd/

Fynn, Christopher. [nd]. "Encoding Model of the Tibetan Script in the UCS". http://www.thlib.org/tools/scripts/wiki/Encoding%20model%20of%20the%20Tibetan%20script%20in%20the%20UCS.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

Кара, Д. [Kara, György]. 1972. *Книги Монгольских Кочевников* [*Knigi mongol'skikh kochevnikov*]. Moscow.

Pandey, Anshuman. 2010. "Preliminary Proposal to Encode the Xawtaa Dorboljin Script in ISO/IEC 10646". N3956 L2/10-411. October 31, 2010. http://std.dkuug.dk/jtc1/sc2/wg2/docs/n3956.pdf

———. 2010. "Preliminary Proposal to Encode the Soyombo Script in ISO/IEC 10646". ISO/IEC JTC1/SC2/WG2 N3949 L2/10-399. October 30, 2010. http://std.dkuug.dk/jtc1/sc2/wg2/docs/n3949.pdf

———. 2011a. "Determining the Encoding Model for Soyombo Vowels" N3986 L2/11-054. February 5, 2011. http://std.dkuug.dk/jtc1/sc2/wg2/docs/n3986.pdf

———. 2011a. “Revised Preliminary Proposal to Encode Soyombo in the UCS”. N4026 L2/11-125. April 25, 2011. http://std.dkuug.dk/jtc1/sc2/wg2/docs/n4026.pdf

———. 2011b. “Preliminary Proposal to Encode the Mongolian Square Script in ISO/IEC 10646”. N4041 L2/11-162. May 3, 2011. http://std.dkuug.dk/jtc1/sc2/wg2/docs/n4041.pdf

———. 2011c. “Revised Preliminary Proposal to Encode the Mongolian Square Script”. N4160 L2/11-379. October 24, 2011. http://std.dkuug.dk/jtc1/sc2/wg2/docs/n4160.pdf

———. 2013a. “Proposal to Encode the Mongolian Square Script in ISO/IEC 10646”. N4413 L2/13-068. April 22, 2013. http://std.dkuug.dk/jtc1/sc2/wg2/docs/n4413.pdf

———. 2013a. “Revised Proposal to Encode the Soyombo Script in ISO/IEC 10646”. N4414 L2/13-069. April 22, 2013. http://std.dkuug.dk/jtc1/sc2/wg2/docs/n4414.pdf

———. 2013b. “Proposal to Encode the Mongolian Square Script in ISO/IEC 10646”. N4471 L2/13-198. October 27, 2013. http://std.dkuug.dk/jtc1/sc2/wg2/docs/n4471.pdf

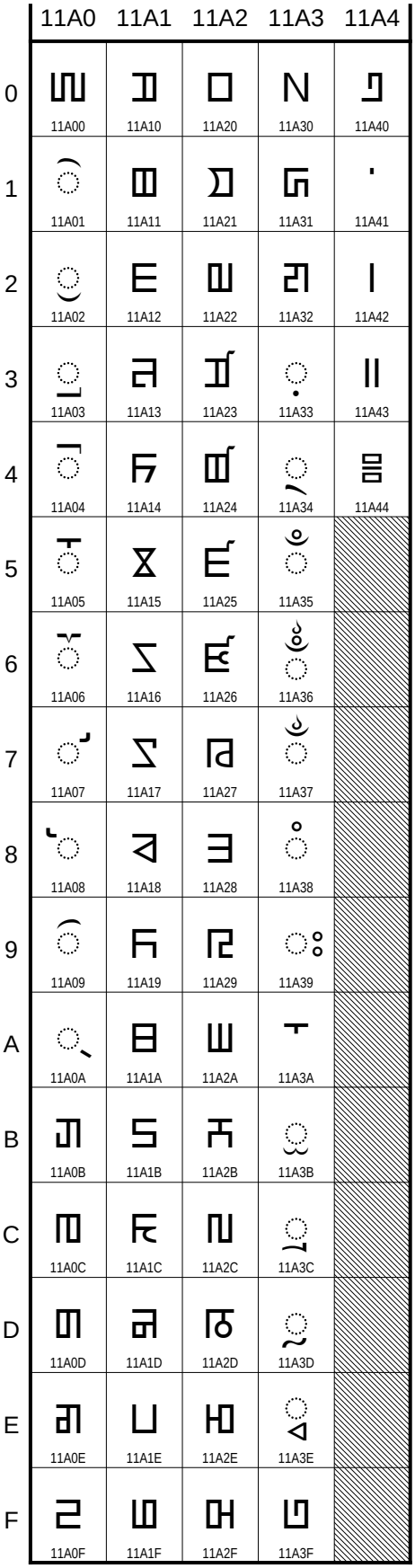
Tseveliin Shagdarsürüng. 2001. *Study of Mongolian Scripts* (Graphic Study or Grammatology). Enlarged 2nd edition. Bibliotheca Mongolica: Monograph 1. Edited by Sharaviin Choimaa. Ulaanbaatar: Center for Mongol Studies, National University of Mongolia.

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 also 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 script. He also granted permission for usage of images of the folios shown in figure 6.

I also owe much thanks to Shriramana Sharma, Peter Constable (Microsoft), and Andrew Glass (Microsoft) for providing detailed comments on the encoding model for consonant conjuncts and various other aspects of the encoding for the script. Laurentiu Iancu (Microsoft) and Ken Whistler (SAP) explained the representation of vertical text. Andrew West and Christopher Fynn provided information on the orthography of conjunct stacks in Tibetan. György Kara (Indiana University, Bloomington) patiently answered several questions during the preliminary stages of my research. Biligsaikhan Batjargal (Ritsumeikan University, Japan) provided information on the Mongolian name of the script for the preliminary proposal.

This project was made possible in part by an earlier grant from the United States National Endowment for the Humanities, which funded the Universal Scripts Project (part of the Script Encoding Initiative at the University of California, Berkeley). Any views, findings, conclusions or recommendations expressed in this publication do not necessarily reflect those of the National Endowment for the Humanities.



	11A0	11A1	11A2	11A3	11A4
0	11A00	11A10	11A20	11A30	11A40
1	11A01	11A11	11A21	11A31	11A41
2	11A02	11A12	11A22	11A32	11A42
3	11A03	11A13	11A23	11A33	11A43
4	11A04	11A14	11A24	11A34	11A44
5	11A05	11A15	11A25	11A35	
6	11A06	11A16	11A26	11A36	
7	11A07	11A17	11A27	11A37	
8	11A08	11A18	11A28	11A38	
9	11A09	11A19	11A29	11A39	
A	11A0A	11A1A	11A2A	11A3A	
B	11A0B	11A1B	11A2B	11A3B	
C	11A0C	11A1C	11A2C	11A3C	
D	11A0D	11A1D	11A2D	11A3D	
E	11A0E	11A1E	11A2E	11A3E	
F	11A0F	11A1F	11A2F	11A3F	

The common Mongolian name for the script is Zanabazarin Dörböljin Useg. It is also known as Xewtee Dörböljin Bicig or in English as the Horizontal Square Script.

Vowel letter

11A00 𑨀 ZANABAZAR SQUARE LETTER A
- used for representing independent vowels in combination with vowel signs

Vowel signs

11A01 ◌𑨁 ZANABAZAR SQUARE VOWEL SIGN I
11A02 ◌𑨂 ZANABAZAR SQUARE VOWEL SIGN UE
11A03 ◌𑨃 ZANABAZAR SQUARE VOWEL SIGN U
11A04 ◌𑨄 ZANABAZAR SQUARE VOWEL SIGN E
11A05 ◌𑨅 ZANABAZAR SQUARE VOWEL SIGN OE
11A06 ◌𑨆 ZANABAZAR SQUARE VOWEL SIGN O
11A07 ◌𑨇 ZANABAZAR SQUARE VOWEL SIGN AI
- also represents secondary vowel i

11A08 ◌𑨈 ZANABAZAR SQUARE VOWEL SIGN AU
- also represents secondary vowel u

Reversed vowel sign

11A09 ◌𑨉 ZANABAZAR SQUARE REVERSED VOWEL SIGN I
- used for Sanskrit vocalic sounds

Vowel length mark

11A0A ◌𑨊 ZANABAZAR SQUARE VOWEL LENGTH MARK

Consonants

11A0B 𑨋 ZANABAZAR SQUARE LETTER KA
- Mongolian ga, gamma

11A0C 𑨌 ZANABAZAR SQUARE LETTER KHA
- Mongolian ka, qa, xa

11A0D 𑨍 ZANABAZAR SQUARE LETTER GA
11A0E 𑨎 ZANABAZAR SQUARE LETTER GHA
11A0F 𑨏 ZANABAZAR SQUARE LETTER NGA
11A10 𑨐 ZANABAZAR SQUARE LETTER CA
11A11 𑨑 ZANABAZAR SQUARE LETTER CHA
11A12 𑨒 ZANABAZAR SQUARE LETTER JA
11A13 𑨓 ZANABAZAR SQUARE LETTER NYA
11A14 𑨔 ZANABAZAR SQUARE LETTER TTA
11A15 𑨕 ZANABAZAR SQUARE LETTER TTHA
11A16 𑨖 ZANABAZAR SQUARE LETTER DDA
11A17 𑨗 ZANABAZAR SQUARE LETTER DDHA
11A18 𑨘 ZANABAZAR SQUARE LETTER NNA
11A19 𑨙 ZANABAZAR SQUARE LETTER TA
- Mongolian da

11A1A 𑨚 ZANABAZAR SQUARE LETTER THA
- Mongolian ta

11A1B 𑨛 ZANABAZAR SQUARE LETTER DA
11A1C 𑨜 ZANABAZAR SQUARE LETTER DHA
11A1D 𑨝 ZANABAZAR SQUARE LETTER NA
11A1E 𑨞 ZANABAZAR SQUARE LETTER PA
- Mongolian ba

11A1F 𑨟 ZANABAZAR SQUARE LETTER PHA
- Mongolian pa

11A20 𑨠 ZANABAZAR SQUARE LETTER BA
- Mongolian va
- also used instead of non-initial 11A2D 𑨭 in conjuncts

11A21 𑨡 ZANABAZAR SQUARE LETTER BHA
11A22 𑨢 ZANABAZAR SQUARE LETTER MA
11A23 𑨣 ZANABAZAR SQUARE LETTER TSA
- Mongolian ja, Sanskrit ca

11A24 𑨤 ZANABAZAR SQUARE LETTER TSHA
- Mongolian ca, Sanskrit cha

11A25 𑨥 ZANABAZAR SQUARE LETTER DZA
- Sanskrit ja

11A26 𑨦 ZANABAZAR SQUARE LETTER DZHA
- Sanskrit jha

11A27 𑨧 ZANABAZAR SQUARE LETTER ZHA
11A28 𑨨 ZANABAZAR SQUARE LETTER ZA
11A29 𑨩 ZANABAZAR SQUARE LETTER SMALL A
= Tibetan 'a-chung
→ 0F60 ཀ tibetan letter -a

11A2A 𑨪 ZANABAZAR SQUARE LETTER YA
11A2B 𑨫 ZANABAZAR SQUARE LETTER RA
11A2C 𑨬 ZANABAZAR SQUARE LETTER LA
11A2D 𑨭 ZANABAZAR SQUARE LETTER VA
11A2E 𑨮 ZANABAZAR SQUARE LETTER SHA
11A2F 𑨯 ZANABAZAR SQUARE LETTER SSA
- created by reversal of 11A2E 𑨮

11A30 𑨰 ZANABAZAR SQUARE LETTER SA
11A31 𑨱 ZANABAZAR SQUARE LETTER HA
11A32 𑨲 ZANABAZAR SQUARE LETTER KSSA

Final Consonant Mark

11A33 ◌𑨳 ZANABAZAR SQUARE FINAL CONSONANT MARK
- used for marking a syllable-final consonant in Mongolian
- indicates a syllabic boundary in Mongolian

Virama

11A34 ◌𑨴 ZANABAZAR SQUARE SIGN VIRAMA
→ 0F84 ◌྄ tibetan mark halanta
- indicates a bare consonant in Sanskrit and Tibetan
- used for producing subjoining conjuncts

Candrabindu and Candra Ornaments

11A35 ◌𑨵 ZANABAZAR SQUARE SIGN CANDRABINDU
- indicates nasalization

→ 0F83 ◌ྃ tibetan sign sna ldan
→ 1880 ᠀ mongolian letter ali gali anusvara one

11A36 ◌𑨶 ZANABAZAR SQUARE SIGN CANDRABINDU WITH ORNAMENT
- used primarily with 11A3F 𑨿

11A37 ◌𑨷 ZANABAZAR SQUARE SIGN CANDRA WITH ORNAMENT
- used primarily with 11A3F 𑨿
→ 0F82 ◌ྂ tibetan sign nyi zla naa da

Signs for Sanskrit

11A38 ◌𑨸 ZANABAZAR SQUARE SIGN ANUSVARA
- indicates nasalization
→ 0F7E ◌ཾ tibetan sign rjes su nga ro

11A39 ◌𑨹 ZANABAZAR SQUARE SIGN VISARGA
- indicates post-vocalic aspiration
→ 0F7F ◌ཿ tibetan sign rnam bcad

Cluster-initial consonant

11A3A 𑨺 ZANABAZAR SQUARE LETTER CLUSTER INITIAL RA
- initial form of 11A2B 𑨫 in Tibetan conjuncts

Cluster-final consonants

11A3B 𑨺𑨿 ZANABAZAR SQUARE LETTER CLUSTER FINAL YA
- final form of 11A2A 𑨪 in Tibetan conjuncts

11A3C 𑨺𑩀 ZANABAZAR SQUARE LETTER CLUSTER FINAL RA
- final form of 11A2B 𑨫 in Tibetan conjuncts

11A3D 𑨺𑩁 ZANABAZAR SQUARE LETTER CLUSTER FINAL LA
- final form of 11A2C 𑨬 in Tibetan conjuncts

11A3E 𑨺𑩂 ZANABAZAR SQUARE LETTER CLUSTER FINAL VA
- final form of 11A2D 𑨭 in Tibetan conjuncts

Head marks

11A3F 𑨿 ZANABAZAR SQUARE INITIAL HEAD MARK
→ 0FD3 ྌ tibetan mark initial brda rnying yig mgo mdun ma
→ 0F04 ྄ tibetan mark initial yig mgo mdun ma
→ 1800 ᠀ mongolian birga

11A40 𑩀 ZANABAZAR SQUARE CLOSING HEAD MARK
→ 0FD4 ྍ tibetan mark closing brda rnying yig mgo sgab ma
→ 0F05 ྅ tibetan mark closing yig mgo sgab ma

Punctuation

11A41 · ZANABAZAR SQUARE TSHEG
→ 0F0B ་ tibetan mark intersyllabic tsheg

11A42 ꡵ ZANABAZAR SQUARE SHAD
→ 0F0D ། tibetan mark shad

11A43 ꡶ ZANABAZAR SQUARE DOUBLE SHAD
→ 0F0E ༎ tibetan mark nyis shad

11A44 𑩄 ZANABAZAR SQUARE LONG TSHEG
→ 0F14 ༔ tibetan mark gter tsheg

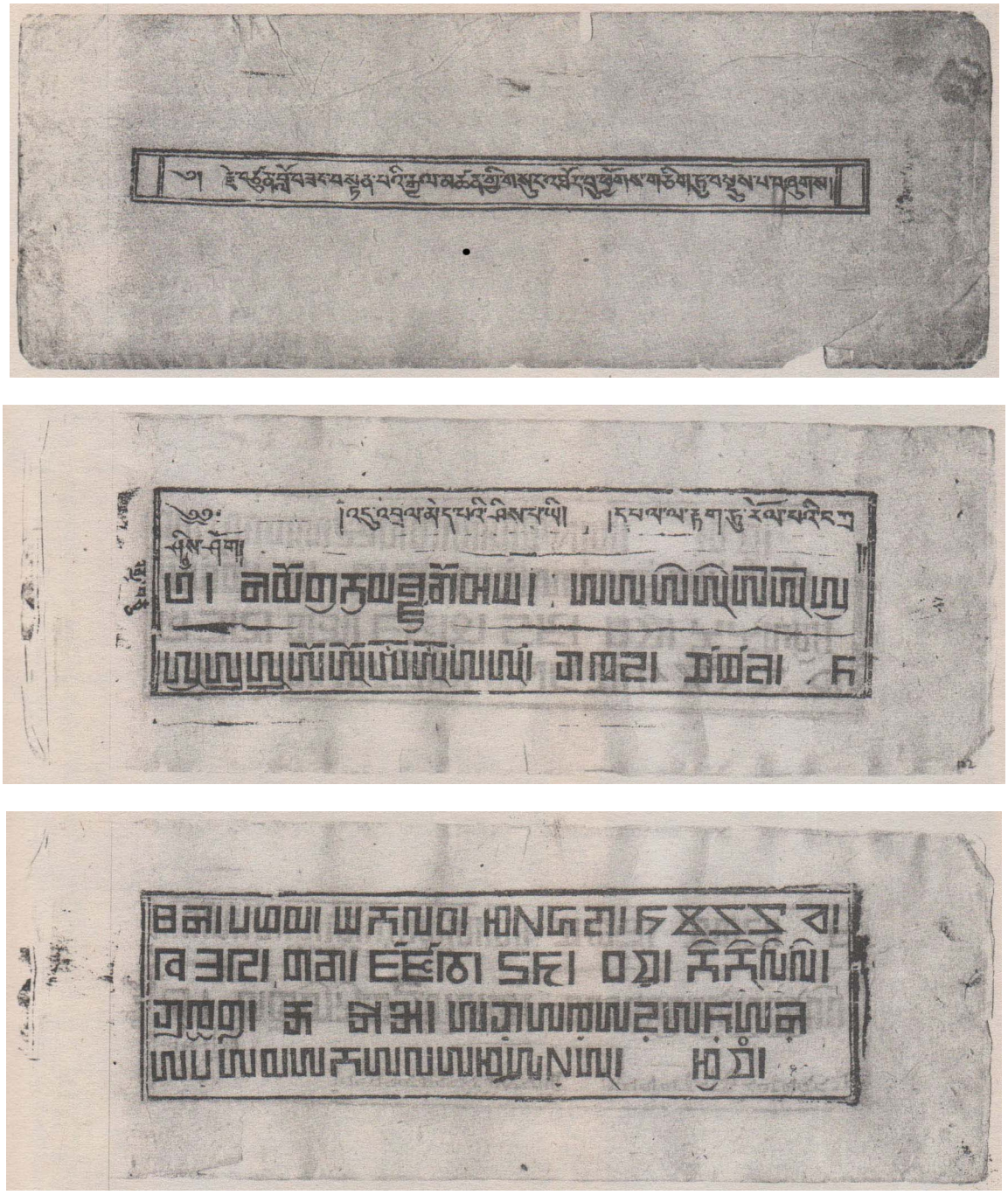


Figure 1: A block print of Zanabazar's square script, from his collected works edited by the Khalkha Zaya Pandita Luvsanprinlei (1642–1715) (from Byambaa Ragchaagiin 2005: 33).

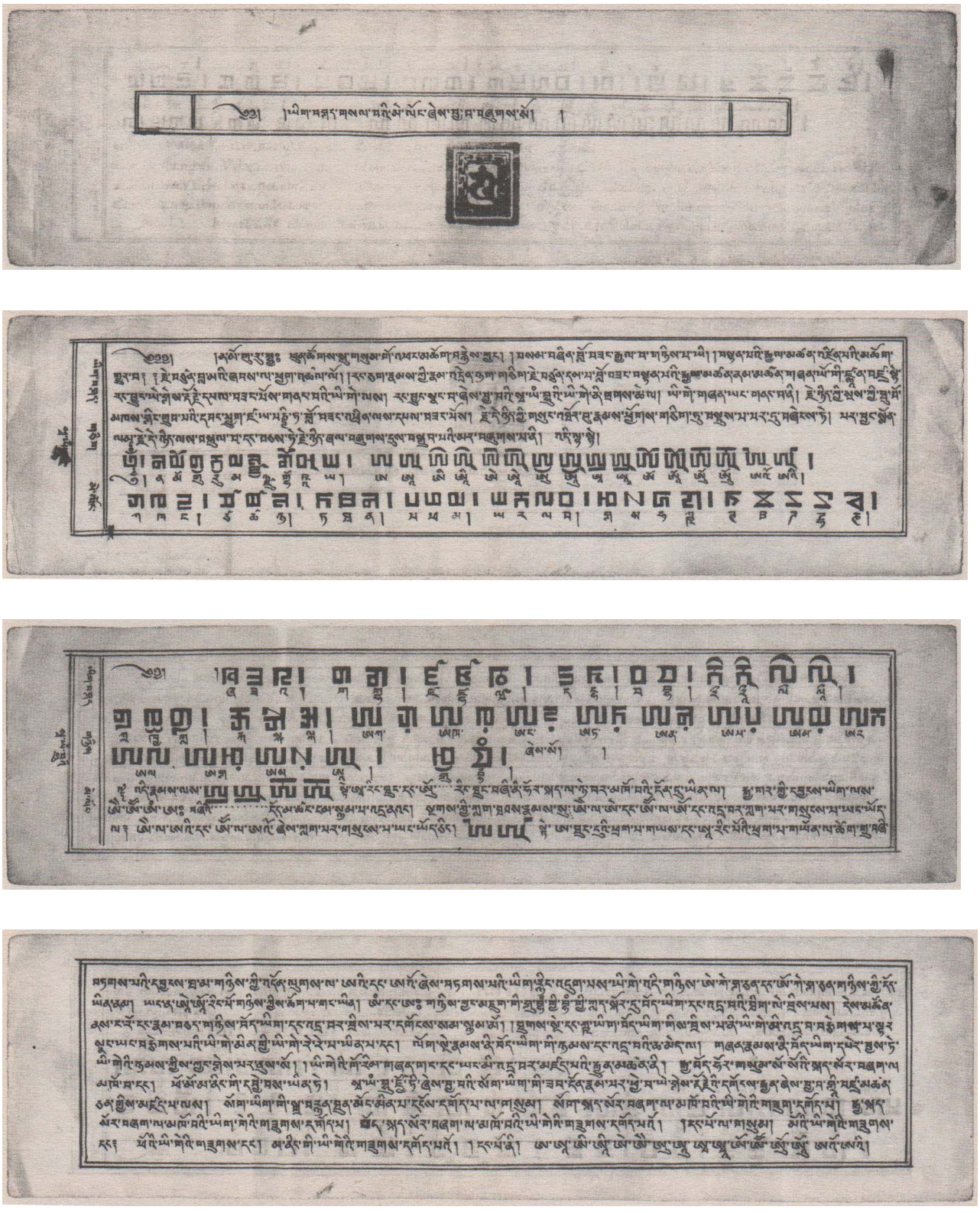


Figure 2: Folios 1a–2a of a treatise on Zanabazar Square by Ugalzyn Lama (from Byambaa Ragchaagiin 2005: 49). Continued in figure 3.

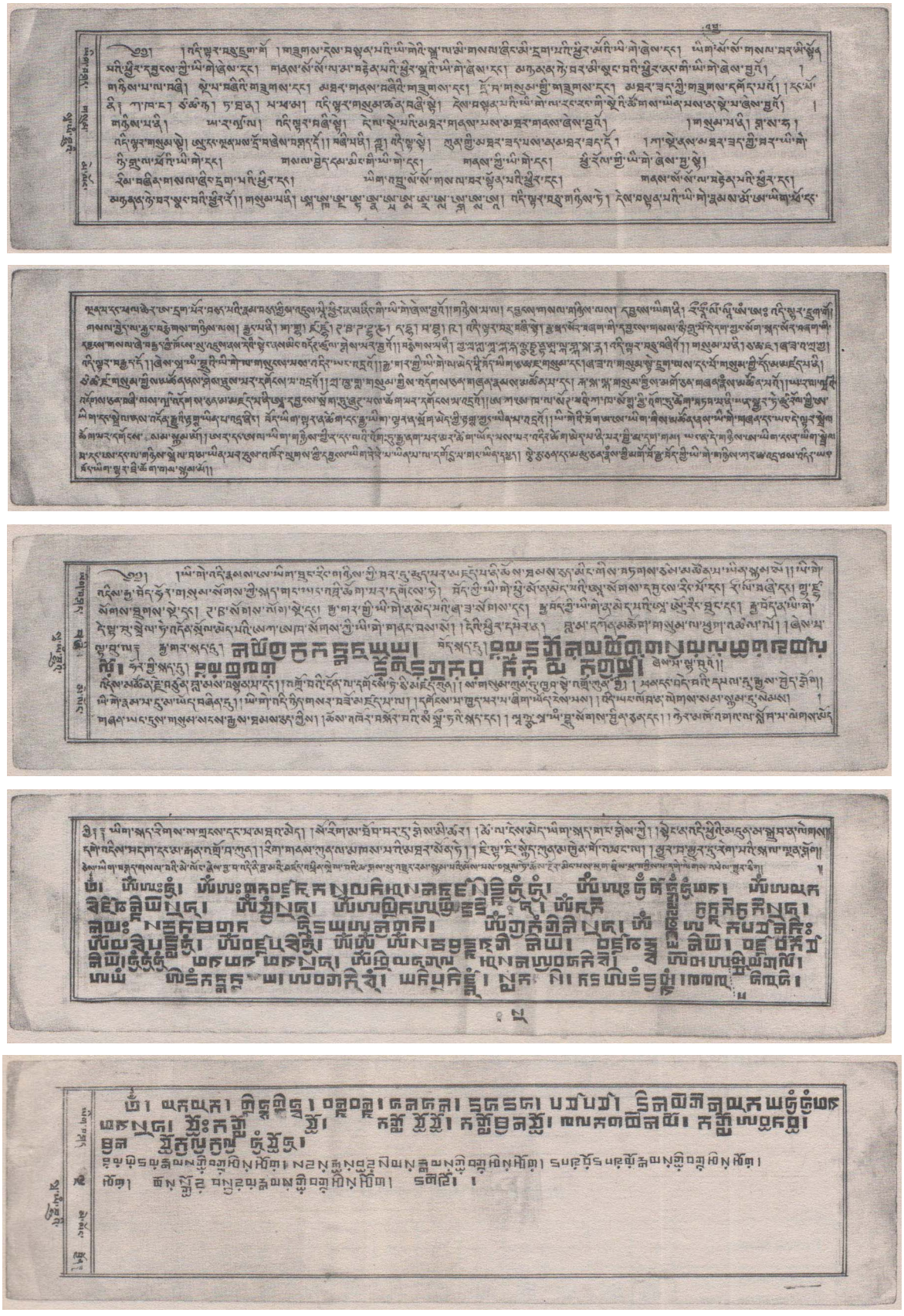


Figure 3: Folios 3a–5a of a treatise on the Zanabazar Square script by Ugalzyn Lama (from Byambaa Ragchaagiin 2005: 50–51). Continued from figure 2.

Figure 4: A manuscript fragment containing text written in Ranjana (Lantsa), Soyombo, Zanabazar Square, and Tibetan scripts (from Tseveliin Shagdarsüüng 2001: 174).

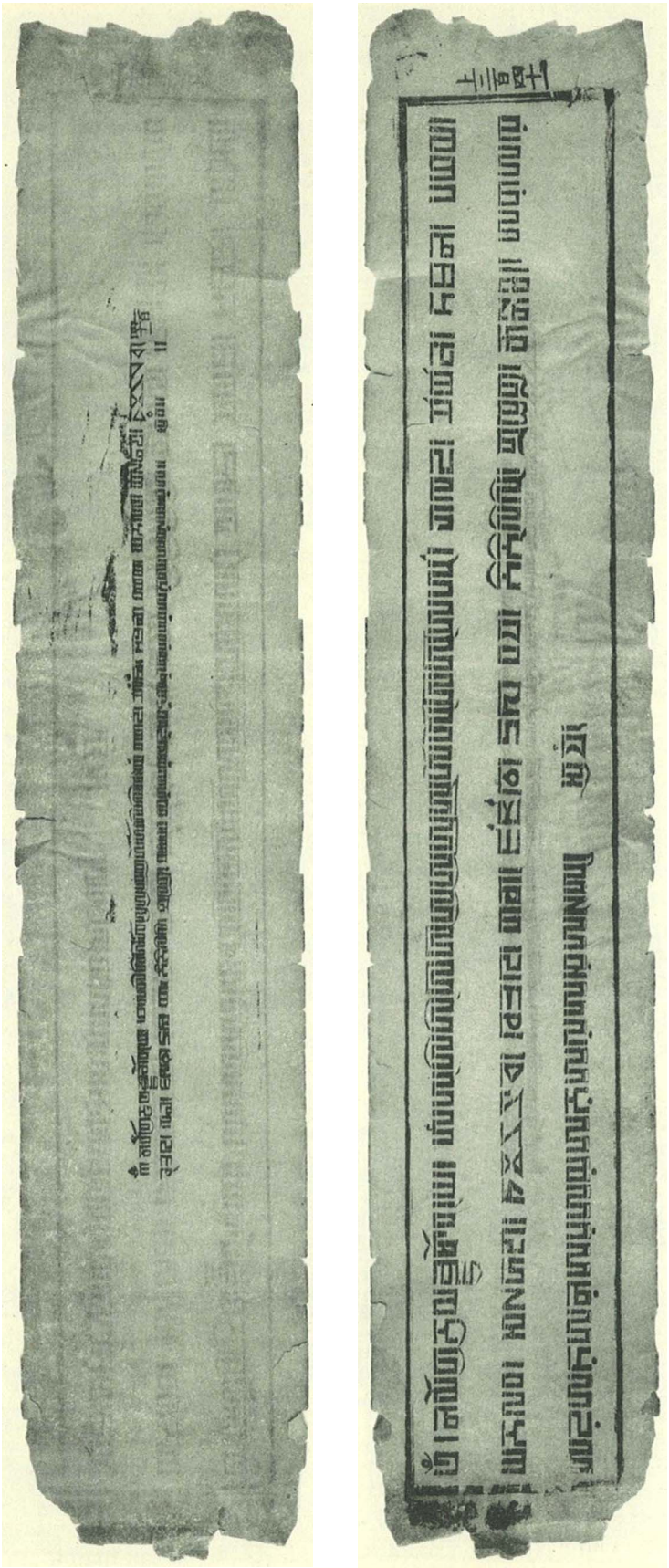


Figure 5: A manuscript showing the Zanabazar Square script arranged according to the traditional Mongolian order (from Byambaa Ragchaagiin 2005: 36).

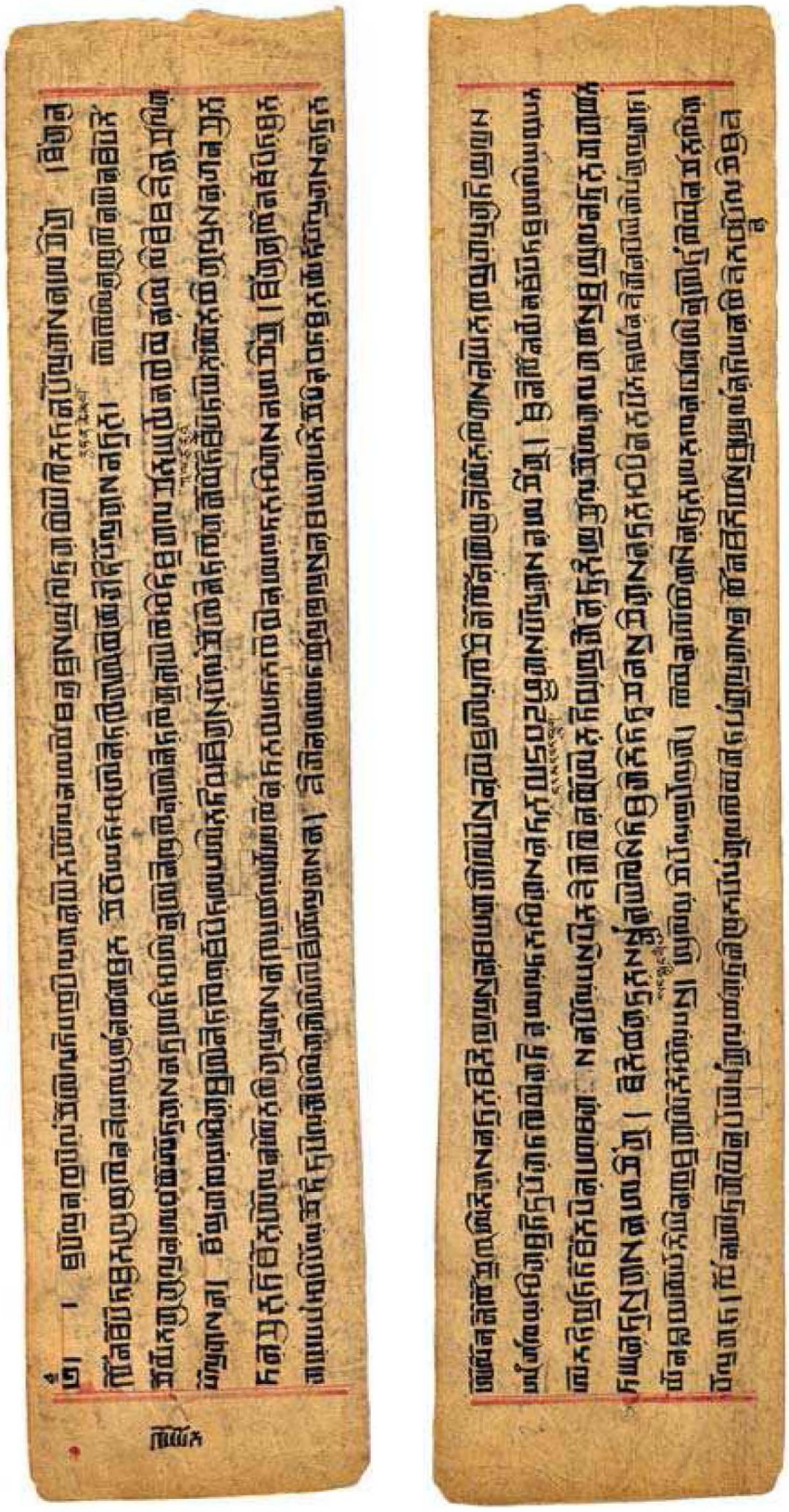


Figure 6: Folios 2a (top) and 2a (bottom) of a manuscript of the biography of Zanabazar (from Bareja-Starzyńska and Ragchaa 2012; images used with permission of Byambaa Ragchaa.

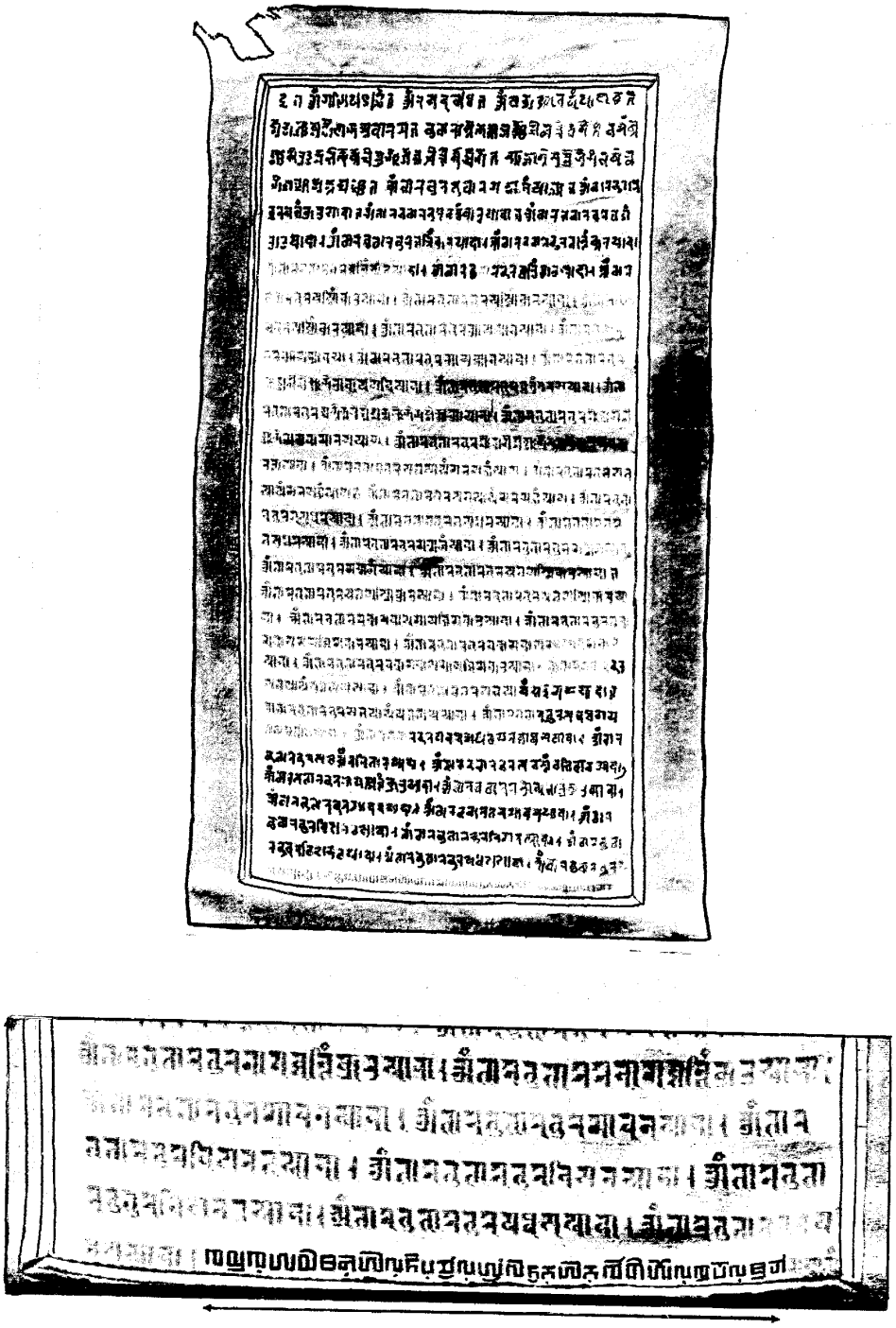


Figure 7: Zanabazar Square text at the bottom of a record written in Ranjana (from Tseveliin Shagdarsürüng 2001: 172).

Figure 8: Xylograph (block print) of a book cover with text in the Soyombo, Zanabazar Square, Mongolian, and Cyrillic scripts (from B. Boldsaikhan, et al. 2005: 330). The title is Sanskrit written in Soyombo: *Mongal-svayaṃbhu-jyoti-varṇa-lipiḥ*. The Zanabazar Square represents Tibetan, the Mongolian represents Mongolian, and the Cyrillic represents Modern (Khalkha) Mongolian.

Figure 9: Characters of the Zanabazar Square script (from Kara 1972: 96). Note the variant form ◌͡◌ for ◌͡◌ VOWEL SIGN I and the variant form 𑨗 of 𑨘 GHA.

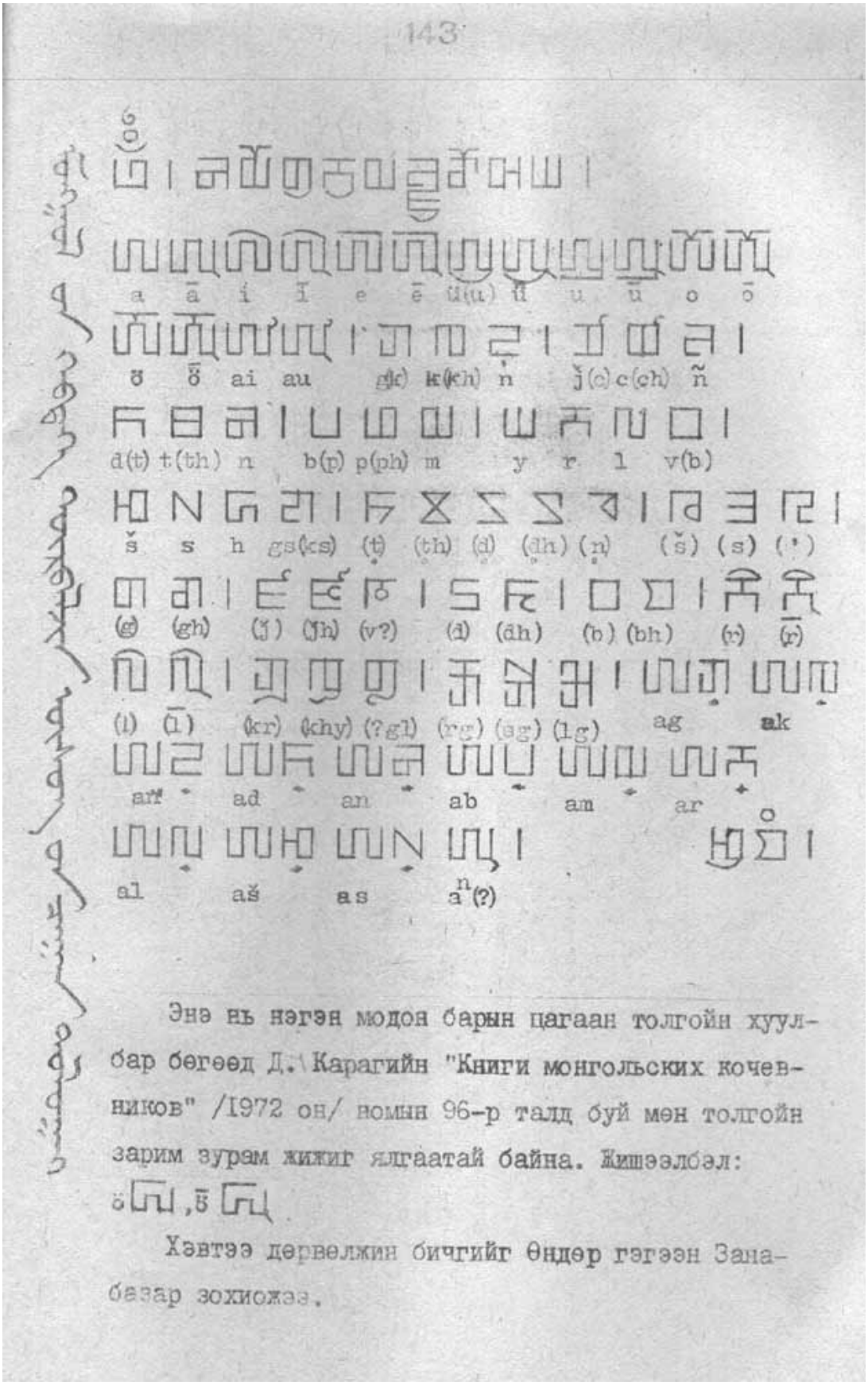


Figure 10: Characters of the Zanabazar Square script (from Kapaj 2002).

гэдэг өгүүлэлдээ тодруулан өгчээ. Академич Ринчен ийнхүү тодруулахдаа Угалзын лам хэмээн олонд алдаршсан Лувсансодовжамц (1878-1961)-ын *"Yig-bçad gsal bai'i me-long žes bya ba bzugs-so"* буюу хэвтээ дөрвөлжин бичгийн тайлбар болгож зохиосон *"Үсэгийн номлол тодорхой толь хэмээх оршивой"* гэдэг гар бичмэл номын[10] мэдээнд үндэслэсэн буй *за*.

Хэвтээ дөрвөлжин бичгээр үлдсэн дурсгал гэвэл 1972 оныг хүртэл хэдэн зүйл хэсэг бусаг цагаан толгой, нэгэн зүйл тарнийн үсгээс өөр тоймтой баримт олдоогүй байсан гэж хэлж болно. 1972 онд проф. Д. Кара *"Книги монгольских кочевников"* номдоо хэвтээ дөрвөлжин бичгийн цагаан толгойд үсэгүйн ажиглалт хийж, дуудалгын латин галигийн хамт анхлан судлагааны хүрээнд танилцуулж Жамсраны Цэвээн авгайн цуглуулгаас олдсон монгол хэлээр, хэвтээ дөрвөлжин бичгээр буй "А-му-гу-ла-н-ту та-ма-га"-ын дардасыг хавсаргажээ[11] (3-р хавсралтаас үзмүү).

Хэдэн жилийн дараа энэ номын зохиогч Гандан хийдийн ламтан Данзан-осор гуайн цуглуулгад байсан самгард хэл, бичгээрхи ханын чимэгийн эцэс дэхи хэвтээ дөрвөлжин бичгээр, монгол хэлээр буй бичвэрийг олж судлагааны эргэлтэнд оруулсан билээ[12]. Монгол хэлээрхи эл дурсгалын талаар хойно арай дэлгэрүүлэн өгүүлэх болно.

1997 онд судлагч Р. Бямбаа хэвтээ дөрвөлжин бичгээр төвөд, монгол, самгард хэлээр буй дурсгалуудыг нэгтгэн судлаж *"Хэвтээ дөрвөлжин үсэг, түүний дурсгалууд"* гэдэг бие даасан тусгай ном нийтлүүлсэн бөгөөд үүндээ уг бичигт холбогдох мэдээ баримтыг багтаан оруулжээ[13]. Энэ жишилэнгээр сүүлийн үес монгол, төвөд, самгард хэлээр хэвтээ дөрвөлжин үсгээр бичсэн дурсгалын зүйл мэр сэр нэмэгдсээр байна.

Хэвтээ дөрвөлжин бичиг, түүгээр үлдсэн монгол хэлний дурсгалын ач холбогдолын тухайд гэвэл түрүүчийн бөлөгт соёмбо бичгийн баримт дурсгал монгол хэлний түүхэнд хэрхэн холбогдох талаар Л. Лигети академичийн хэлсэнтэй агаар нэгэн мөр тул дахин нурших хэрэггүй. Харин үсэгзүйн үүднээс төвөд, самгард үсэг бичигтэй харьцуулан тодруулах зүйл багагүйгээр барахгүй бас 1444/1446 оны солонгос бичигийн зарчимтай төстэй зүйл харагддаг[14] нь шууд буюу эсбөгөөс Төв Ази дахины бусад бичиг үсгийн уламжлалтай дам холбоотойн алин болохыг энэ хир шийдээгүй боловч бас анхааралгүй орхиж болохгүй гэж санаж байна.

Хэвтээ дөрвөлжин бичигийн цагаан толгой

Хуудас эхлэсний буюу хуудасны өвөр талын тэмдэг. Бярга буюу эгчимтэй адил үүрэг гүйцэтгэнэ.

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

tl. a_0;
tc. a. Энэ нь а эгшигийн гол буюу (MF) хэлбэр. Үг буюу үеийн дунд, адаг тохиолдох нууц буюу тэг хэлбэр.

[10] Р. Бямбаа, *Хэвтээ дөрвөлжин үсэг, түүний дурсгалууд,* Улаанбаатар, 1997, х.22-38.
[11] Д. Кара, *"Книги монгольских кочевников",* Москва, 1972, стр. 93-96.
[12] Ц. Шагдарсүрэн, *Монгол үсэг зүй,* Тэргүүн дэвтэр, Улаанбаатар, 1981, 108-110; Ц. Шагдарсүрэн, *Об одном новонайденном памятнике горизонтально-квадратного письма,* - Монгольский лингвистический сборник, Москва, 1985, стр. 150 - 154.
[13] Р. Бямбаа, *Хэвтээ дөрвөлжин үсэг, түүний дурсгалууд,* Улаанбаатар, 1997, 90 х.
[14] Ц. Шагдарсүрэн, *Монгол солонгос бичиг үсгийн харилцаа холбооны асуудалд,* - Mongolian Studies (The Korean Association for Mongol Studies), N. 4 (1996), Soeul, 1997, 169-192 х; Ts. Shagdarsurung, *A Study of Relation between the Korean and Mongolian Scripts,* The Research Paper to The Korea Foundation, Seoul, 1998, pp. 1-27.

Figure 11: Description of Zanabazar Square letters (from Tseveliin Shagdarsürüng 2001: 160).

1.a. 

tl. Ā;
tc. mong., sans.: ā. Урт а эгшигийн бие даасан буюу (IF) хэлбэр.

tl. $ā_0$;
tc. ā. Энэ нь урт ā эгшигийн нууц буюу тэг (Ø) хэлбэр.

2. 

tl. I;
tc. mong., tib., sans.: i. Энэ нь i эгшигийн бие даасан буюу (IF) хэлбэр. Үг буюу үеийн эхинд тохиолдоно.

tl. i;
tc. i. i эгшигийн гол буюу (MF) хэлбэр. Үг буюу үеийн дунд, адагт тохиолдоно.

2.a. 

tl. Ī;
tc. mong., sans.: ī. Урт ī эгшигийн бие даасан буюу (IF) хэлбэр. Үг буюу үеийн эхинд тохиолдоно.

tl. ī.
tc. ī. Урт ī эгшигийн гол буюу (MF) хэлбэр. Үг буюу үеийн дунд, адагт тохиолдоно.

3. 

tl. E;
tc. mong., tib.: e; sans. (diphtongue) e. Энэ нь монгол, төвөд хэлний е эгшиг, самгард хэлний хос е эгшигийн бие даасан буюу (IF) хэлбэр. Үг буюу үеийн эхинд тохиолдоно.

tl. e;
tc. e. Энэ нь е эгшигийн гол буюу (MF) хэлбэр. Үг буюу үеийн дунд, адагт тохиолдоно.

3.a. 

tl. Ē;
tc. mong.: ē; sans: (diphtongue) ai. Монгол хэлний урт ē, самгард хэлний хос ai эгшигийн бие даасан буюу (IF) хэлбэр. Үг буюу үеийн эхинд тохиолдоно.

tl. ē;
tc. ē / ai; Монгол хэлний урт ē, самгард хэлний хос ai эгшигийн гол буюу (MF) хэлбэр. Үг буюу үеийн дунд, адагт тохиолдоно.

4. 

tl. mong., tib.: Ü; sans.: U;
tc. mong., tib.: ü; sans.: u. Монгол, төвөд хэлний ü эгшиг, самгард хэлний u эгшигийн бие даасан буюу (IF) хэлбэр. Үг буюу үеийн эхинд тохиолдоно.

tl. mong., tib.: ü; sans. u;
tc. mong., tib.: ü; sans. u; Монгол, төвөд хэлний ü эгшиг, самгард хэлний u эгшигийн гол буюу (MF) хэлбэр. Үг буюу үеийн дунд, адагт тохиолдоно.

4.a. 

tl. mong. Ū; sans.: ū;
tc. mong. ū; sans. ū; Монгол хэлний урт ü эгшиг, самгард хэлний урт ū эгшигийн бие даасан буюу (IF) хэлбэр. Үг буюу үеийн эхинд тохиолдоно.

Figure 12: Description of Zanabazar Square letters (from Tseveliin Shagdarsürüng 2001: 161).

tl. mong. ū; sans. ū;
tc. mong. ū; sans. ū. Монгол, хэлний урт ӣ эгшиг, самгард хэлний урт ӣ эгшигийн гол буюу (MF) хэлбэр. Үг буюу үеийн дунд, адагт тохиолдоно.

5. 

tl. mong. U;
tc. mong. u. Монгол хэлний u эгшигийн бие даасан буюу (IF) хэлбэр. Үг буюу үеийн эхинд тохиолдоно.

tl. mong. u;
tc. mong. u. U эгшигийн гол буюу (MF) хэлбэр. Үг буюу үеийн дунд болоод адагт тохиолдоно.

5.a. 

tl. mong. ū;
tc. mong. ū. Монгол хэлний урт ū эгшигийн бие даасан буюу (IF) хэлбэр. Үг буюу үеийн эхинд тохиолдоно.

tl. mong. ū;
tc. mong. ū. Монгол хэлний урт ū эгшигийн гол буюу (MF) хэлбэр. Үг буюу үеийн дунд болоод адагт тохиолдоно.

6. 

tl. mong., tib., O; sans. (diphtongue) O;
tc. mong., tib. o; sans. (diphtongue) o. Энэ нь монгол, төвөд хэлний o эгшиг, самгард хэлний хос o эгшигийн бие даасан буюу (IF) хэлбэр. Үг буюу үеийн эхинд тохиолдоно.

tl. mong., tib. o; sans. (diphtongue) o;
tc. mong., tib. o; sans. (diphtongue) o. Энэ нь монгол, төвөд хэлний o эгшиг, самгард хэлний хос o эгшигийн гол буюу (MF) хэлбэр. Үг буюу үеийн дунд болоод адагт тохиолдоно.

6.a. 

tl. mong. ō; sans. (diphtongue) AU;
tc. mong. ō; sans. (diphtongue) au. Энэ нь монгол хэлний урт ō эгшиг, самгард хэлний хос au эгшигийн бие даасан буюу (IF) хэлбэр. Үг буюу үеийн эхинд тохиолдоно.

tl. mong. ō; sans. (diphtongue) au;
tc. mong. ō; sans. (diphtongue) au. Энэ нь монгол хэлний урт ō эгшиг, самгард хэлний хос au эгшигийн гол буюу (MF) хэлбэр. Үг буюу үеийн дунд болоод адагт тохиолдоно.

7. 

tl. mong. Ö;
tc. mong. ö. Энэ нь монгол хэлний ö эгшигийн бие даасан буюу (IF) хэлбэр. Үг буюу үеийн эхинд тохиолдоно.

tl. mong. ö;
tc. mong. ö. Энэ нь монгол хэлний ö эгшигийн гол буюу (MF) хэлбэр. Үг буюу үеийн дунд болоод адагт тохиолдоно.

Figure 13: Description of Zanabazar Square letters (from Tseveliin Shagdarsürüng 2001: 162).

7.a. 

tl. mong. Ō;
tc. mong. ō. Энэ нь монгол хэлний урт ö эгшигийн бие даасан буюу (IF) хэлбэр хэлбэр. Yr буюу үеийн эхинд тохиолдоно.

tl. mong. ō;
tc. mong. ō. Энэ нь монгол хэлний урт ö эгшигийн гол буюу (MF) хэлбэр. Yr буюу үеийн дунд болоод адагт тохиолдоно

8. 

tl. Au_1;
tc. mong. au (буюу diphtongue waw). Монгол хэлнээ тохиолдоно. (Энэ хэлбэрийг жагсаалтын No. 5-д дурдсан u эгшигтэй харьцуулагтун).

9. 

tl. Ai_1;
tc. mong. ai (буюу diphtongue yod). Монгол хэлнээ тохиолдоно. (Энэ хэлбэрийг жагсаалтын No. 2-т дурдсан i эгшигтэй харьцуулагтун).

10. 

tl. mong. G(a); tib., sans.: K(a);
tc. mong. ga; tib., sans. ka. Монгол хэлний ga, төвөд, самгард хэлний ka гийгүүлэгчийн бие даасан буюу (IF) хэлбэр.

11. 

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

12. 

tl. (mong. n̲), tib., sans.: Ṅ(a);
tc. mong., tib., sans.: n̲a. Монгол хэлнээ үгийн буюу үеийн эхинд тохиолддоггүй. Төвөд, самгард хэлний ṅa гийгүүлэгчийн бие даасан (IF) хэлбэр. Монгол хэлний тухайд зөвхөн гол буюу (MF) хэлбэрээр тохиолдоно.

13. 

tl. mong. J(a), tib., sans.: C(a);
tc. mong. ja, tib., sans.: ca. Монгол хэлний ja, төвөд, самгард хэлний ca гийгүүлэгчийн бие даасан буюу (IF) хэлбэр.

14. 

tl. mong. C(a); tib., sans.: CH(a);
tc. mong. ca, tib., sans.: cha. Монгол хэлний ca, төвөд, самгард хэлний cha гийгүүлэгчийн бие даасан буюу (IF) хэлбэр.

15. 

tl. tib., sans.: N̄(a);
tc. tib., sans.: n̄a. Төвөд, самгард хэлний n̄a 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.

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.

Figure 14: Description of Zanabazar Square letters (from Tseveliin Shagdarsürüng 2001: 163).

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

19. 
tl. mong. ba, tib., sans.: P(a);
tc. mong. ba, tib., sans.: pa. Монгол хэлний ba, төвөд, самгард хэлний pa гийгүүлэгчийн бие даасан буюу (IF) хэлбэр.

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.

21. 
tl. mong., tib., sans. M(a);
tc. mong., tib., sans. ma. Монгол, төвөд, самгард хэлний ma гийгүүлэгчийн бие даасан буюу (IF) хэлбэр.

22. 
tl. mong., tib.: Y(a), sans. (semi-vowel) Y(a);
tc. mong., tib.: ya, sans. (semi-vowel) ya. Монгол, төвөд хэлний ya, самгард хэлний тал (буюу хагас) ya гийгүүлэгчийн бие даасан буюу (IF) хэлбэр.

23. 
tl. mong., tib.: R(a), sans. (semi-vowel) R(a);
tc. mong., tib.: ra, sans.: (semi-vowel) ra. Монгол, төвөд хэлний ra, самгард хэлний тал (буюу хагас) ra гийгүүлэгчийн бие даасан буюу (IF) хэлбэр.

24. 
tl. mong., tib., L(a), sans.; (semi-vowel) L(a);
tc. mong., tib., la, sans.: (semi-vowel) la. Монгол, төвөд хэлний la, самгард хэлний тал (буюу хагас) la гийгүүлэгчийн бие даасан буюу (IF) хэлбэр.

25. 
tl. mong., tib., V(a), sans.; (semi-vowel) V(a);
tc. mong., tib., va, sans.: (semi-vowel) va. Монгол, төвөд хэлний va, самгард хэлний тал (буюу хагас) va гийгүүлэгчийн бие даасан буюу (IF) хэлбэр.

26. 
tl. mong. Š(a), tib., sans.: Ç(a) / Ś(a).
tc. mong. ša, tib., sans.: ça / sa. Монгол хэлний ša, төвөд, самгард хэлний ça / sa гийгүүлэгчийн бие даасан буюу (IF) хэлбэр.

27. 
tl. mong., tib., sans. S(a);
tc. mong., tib., sans. sa. Монгол, төвөд, самгард хэлний sa гийгүүлэгчийн бие даасан буюу (IF) хэлбэр.

28. 
tl. mong., tib. H(a), sans.: (semi-vowel) H(a);
tc. mong., tib., ha, sans: (semi-vowel) ha. Монгол, төвөд хэлний ha, самгард хэлний тал (буюу хагас) ha гийгүүлэгчийн бие даасан буюу (IF) хэлбэр.

29. 
tl. sans. KS(a);
tc. sans. ksa. Самгард хэлний ksa гийгүүлэгчийн бие даасан буюу (IF) хэлбэр.

Figure 15: Description of Zanabazar Square letters (from Tseveliin Shagdarsürüng 2001: 164).

30.  **tl.** sans. T(a);
tc. sans. ta. Самгард хэлний ta гийгүүлэгчийн бие даасан буюу (IF) хэлбэр.

31.  **tl.** sans. TH(a);
tc.sans. tha. Самгард хэлний tha гийгүүлэгчийн бие даасан буюу (IF) хэлбэр.

32.  **tl.** sans. D(a);
tc. sans. da. Самгард хэлний da гийгүүлэгчийн бие даасан буюу (IF) хэлбэр.

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

34.  **tl.** sans. N(a);
tc. sans. na. Самгард хэлний na гийгүүлэгчийн бие даасан буюу (IF) хэлбэр.

35.  **tl.** tib. Z(a);
tc. tib. za. Төвөд хэлний za гийгүүлэгчийн бие даасан буюу (IF) хэлбэр.

36.  **tl.** tib. Z(a);
tc. tib. za. Төвөд хэлний za гийгүүлэгчийн бие даасан буюу (IF) хэлбэр.

37.  **tl.** tib. '(a);
tc. tib. 'a. Төвөд хэлний 'a-čhuṅ буюу шанаган "a" гийгүүлэгчийн бие даасан буюу (IF) хэлбэр.

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

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

40.  **tl.** tib., sans.: J(a);
tc. tib., sans.: ja. Төвөд, самгард хэлний ja гийгүүлэгчийн бие даасан буюу (IF) хэлбэр.

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

42.  **tl.** sans. S(a);
tc. sa (?). Самгард хэлний sa (?) гийгүүлэгчийн бие даасан буюу (IF) хэлбэр.

Figure 16: Description of Zanabazar Square letters (from Tseveliin Shagdarsürüng 2001: 165).

43. 
tl. sans. D(a);
tc. sans. da. Самгард хэлний da гийгүүлэгчийн бие даасан буюу (IF) хэлбэр.

44. 
tl. sans. DH(a);
tc. sans. dha. Самгард хэлний dha гийгүүлэгчийн бие даасан буюу (IF) хэлбэр.

45. 
tl. sans. B(a) / P(a);
tc. sans. ba / pa. Самгард хэлний ba / pa гийгүүлэгчийн бие даасан буюу (IF) хэлбэр.

46. 
tl. sans. BH(a);
tc. sans. bha. Самгард хэлний bha гийгүүлэгчийн бие даасан буюу (IF) хэлбэр.

47. 
tl. sans. R / Ri;
tc. sans. r / ri. Самгард хэлний r / ri эгшигийн бие даасан буюу (IF) хэлбэр.


tl. sans. r / ri;
tc. sans. r /ri. Самгард хэлний r / ri эгшигийн гол буюу (MF) хэлбэр. Yr буюу үеийн дунд, адагт тохиолдоно.

47.a. 
tl. sans. R / Ri;
tc. sans. r / ri. Самгард хэлний урт r / ri эгшигийн бие даасан буюу (IF) хэлбэр.


tl. sans. R / Ri;
tc. sans. r / ri. Самгард хэлний урт r / ri эгшигийн гол буюу (MF) хэлбэр. Yr буюу үеийн дунд, адагт тохиолдоно.

48. 
tl. sans. L / Li;
tc. sans. l / li. Самгард хэлний l / li эгшигийн бие даасан буюу (IF) хэлбэр.


tl. sans. l / li;
tc. sans. l / li. Самгард хэлний l / li эгшигийн гол буюу (MF) хэлбэр. Yr буюу үеийн дунд, адагт тохиолдоно.

48.a. 
tl. sans. L / Li;
tc. sans. l / li. Самгард хэлний урт l / li эгшигийн бие даасан буюу (IF) хэлбэр.


tl. sans. L / Li;
tc. sans. l / li. Самгард хэлний урт l / li эгшигийн гол буюу (MF) хэлбэр. Yr буюу үеийн дунд, адагт тохиолдоно.

49. 
tl. tib., sans. Kr(a);
tc. tib., sans. kra. Төвөд, самгард хэлний нийлмэл (буюу га зүүлтэт) kra гийгүүлэгчийн бие даасан буюу (IF) хэлбэр.

Figure 17: Description of Zanabazar Square letters (from Tseveliin Shagdarsürüng 2001: 166). Note the use of ◌͗ VOWEL SIGN REVERSED I for writing *r̥, r̥̄, l̥, l̥̄*.

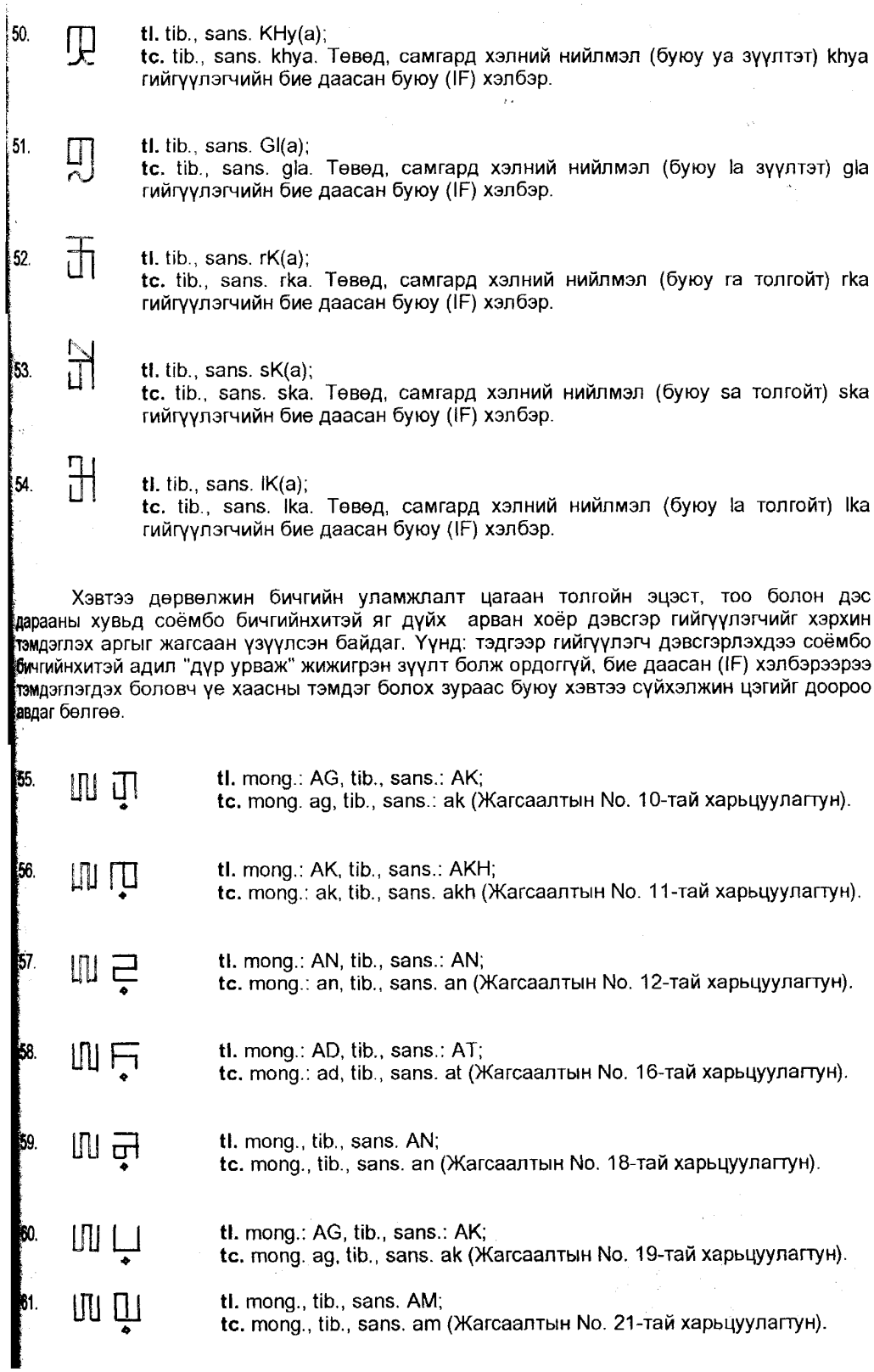
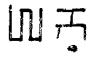


Figure 18: Description of Zanabazar Square letters (from Tseveliin Shagdarsürüng 2001: 167).

62. 
tl. mong., tib., sans. AR;
tc. mong., tib., sans. ar (Жагсаалтын No. 23-тай харьцуулагтун).

63. 
tl. mong., tib., sans. AL;
tc. mong., tib., sans. al (Жагсаатын No. 24-тай харьцуулагтун).

64. 
tl. mong. Aš, tib., sans.: Aç / Aś;
tc.mong. aš, tib., sans.: aç / aś (Жагсаалтын No. 26-тай харьцуулагтун).

65.
tl. mong., tib., sans. AS;
tc. mong., tib., sans. as (Жагсаалтын No. 27-тай харьцуулагтун).

66.
tl. mong.Ān;
tc. mong. ān. Үүнийг соёмбо бичигийн цагаан толгойн 41-рт дурдсан тайлбар сэлттэй харьцуулан үзнэ үү ().

Хэвтээ дөрвөлжин бичигт Монгол, төвөд, самгард хэлнээ дэвсгэрлэж орсон гийгүүлэгчийг ийнхүү дор нь тусгайлан тэмдэглэдэг уламжлал Төв Азийн бусад үндэстний бичиг үсгийн тогтолцоонд ч харагддаг бөлгөө. Тухайлбал: солонгос бичигт гол төлөв харь үгийн дэвсгэр гийгүүлэгчийг иймэрхүү байдлаар тэмдэглэдэг тухай энэ номын зохиогчийн бичсэн зүйл буй[15].

Дээрхи жагсаалтаас үзэхүл, хэвтээ дөрвөлжин бичигийг төвөд үсэг болон түүнээ үндэслэн зохиосон монгол дөрвөлжин бичигт тулгуурлаж, тэр цагийнхаа номын гурван хэл болж байсан монгол, төвөд, самгард хэлний үгийг тэмдэглэхэд зориулан таацуулж зохиосон болох нь тодорхой харагдана.

Одоо энэ хир хэвтээ дөрвөлжин бичигээрхи дурсгалын зүйл гэвэл тоо ширхэгийн хувьд тийм ч цөөнгүй, хэмжээний хувьд харьцангуй янз бүр, зарим нь тамгын дардас төдий байхад зарим нь 7-8 хуудас ар өвөргүй байх жишээтэй. Хэлний хувьд, төвөд болон самгардаар бие даалган бичсэн буюу хадсан дурсгал харьцангуй илүү боловч сүүлийн үес монгол хэлээр буй дурсгал нэмэгдэн олдсоор буй бөгөөд эдгээр дурсгалуудыг цуглуулах, судлах, хэвлэн нийтлэхэд Р. Бямбаа онцгой үүрэг гүйцэтгэснийг энэ ташрамд дурдалгүй орхих аргагүй.[16] Үүнээс гадна Р. Бямбаа номдоо соёмбо бичгийн тайлбар болгон тусгай ном зохиож байсан "Угалзын лам" хэмээн алдаршсан Лувсансодовжамцын *"Yig-bçad gsal ba'i me-long žes bya ba bzugs-so"* буюу *"Үсэгийн номлол тодорхой толь хэмээх оршивой"* гэдэг нэртэй бүтээлийг монгол орчуулгын хамт эрдэм шинжилгээний гүйлгээнд оруулсан нь хэвтээ дөрвөлжин бичгийн талаар энэ хир бидний үетэй золгосон цорын ганц уламжлалт тайлбар зохиол болж өгсөн ач холбогдолтой юм. Энэ номын мэдээнээс үзвэл нэлээд зүйл тодорхой болж өгнө. Тухайлбал:

> 1.б.: (3)... Ранжүн Ишдоржбалсамбуу-бээр зохиосон үсгээс өөрөө аяндаа гарсан "Соёмбо" хэмээх үсэг нь их алдаршсан бөгөөд <u>үсэг бүсдыг зохиосон</u> нь энэ богдын шавийн ахмад

[15] Ц. Шагдарсүрэн, *Монгол солонгос бичиг үсгийн харилцаа холбооны асуудалд*, - Mongolian Studies (The Korean Association for Mongol Studies), N. 4 (1996), Soeul, 1997, 169-192 х; Ts. Shagdarsurung, *A Study of Relation between the Korean and Mongolian Scripts,* The Research Paper to The Korea Foundation, Seoul, 1998, pp. 1-27.
[16] Р. Бямбаа, *Хэвтээ дөрвөлжин үсэг, түүний дурсгалууд*, Улаанбаатар, 1997, 90 х.

Figure 19: Description of Zanabazar Square letters (from Tseveliin Shagdarsürüng 2001: 168).

63.  **tl.** mong. AL
tc. mong. al

64.  **tl.** mong. AŠ
tc. mong. aš

65.  **tl.** mong. AS
tc. mong. as

66.  **tl.** mong. Ā;
tc. mong. ā.

Эдгээр үсгүүд дээр нэмэлт болгон цагаан толгойд бичигдээгүй ч бусад сурвалжуудад гарсан зарим нэгэн санскрит, төвд хэлний үсгийн зүйлүүдийг жагсаавал:

67.  **tl.** sans. (anusvāra) AṂ
tc. sans. (anusvāra) aṁ. Энэ нь санскрит хэлний хос aṁ эгшгийн бие даасан буюу (IF) хэлбэр. Үг буюу үеийн эхэнд тохиолдно.

 **tl.** sans. (anusvāra) aṁ
tc. sans. (anusvāra) aṁ. Энэ нь санскрит хэлний хос aṁ эгшгийн гол буюу (MF) хэлбэр.

68.  **tl.** sans. (visarga) Aḥ
tc. sans. (visarga) aḥ. Энэ нь санскрит хэлний aḥ эгшгийн бие даасан буюу (IF) хэлбэр.

 **tl.** sans. (visarga) Aḥ
tc. sans. (diphtongueue) aḥ. Энэ нь санскрит хэлний aḥ эгшгийн гол буюу (MF) хэлбэр.

69.  **tl.** sans. Ṣ(a)
tc. sans. ṣa. Санскрит хэлний ṣa гийгүүлэгчийн бие даасан буюу (IF) хэлбэр.

70.  **tl.** tib. C(a)
tc. Төвд хэлний ca гийгүүлэгчийн бие даасан буюу (IF) хэлбэр.

71.  **tl.** tib. CH(a)
tc. tib. cha. Төвд хэлний cha гийгүүлэгчийн бие даасан буюу (IF) хэлбэр.

72.  **tl.** tib. J(a);
tc. tib. ja. Төвд хэлний ja гийгүүлэгчийн бие даасан буюу (IF) хэлбэр.

Figure 20: Description of the letters 𑩐 SSA, 𑩑 CA, 𑩒 CHA, 𑩓 JA, which are not enumerated in traditional charts (from Byambaa Ragchaagiin 2005: 32).

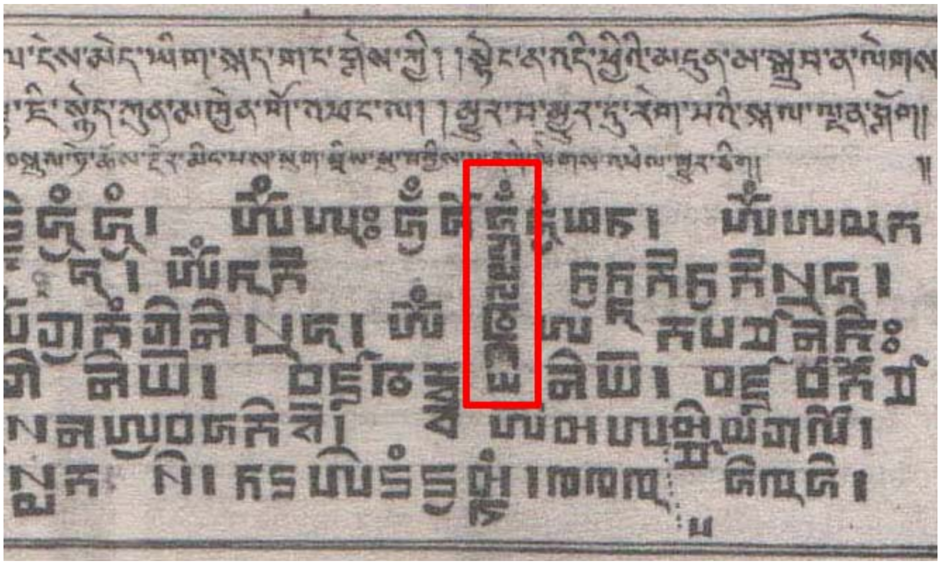


Figure 21: Excerpt from a folio shown in figure 3 illustrating Zanabazar Square text written vertically. The text is *haṃkṣamalavaraya*, which may be incorrectly interpreted as a conjunct representing *hkṣmlvryaṃ*.

Figure 22: Seal showing the word *thalīṃ* written vertically in the Zanabazar Square script (from Byambaa Ragchaagiin 2005: 97). Note the use of SMALL A instead of ◌ VOWEL LENGTH MARK for indicating *ī*. Also note the combining signs for LA placed above THA.

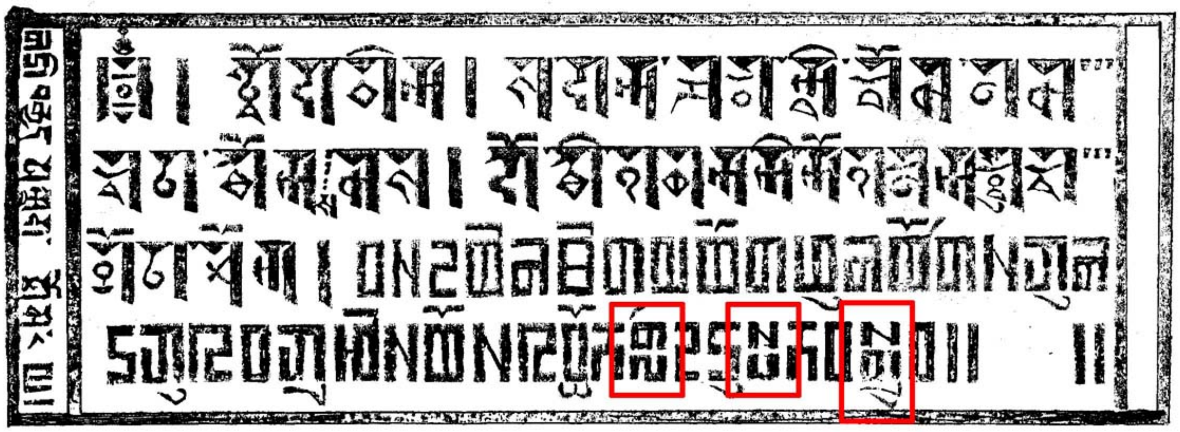


Figure 23: A manuscript containing text in Soyombo and Zanabazar Square (from Tseveliin Shagdarsürüng 2001: 173). Note the sizing of consonant stacks, such that the letters are reduced so that the height of the stack matches the regular letter height.

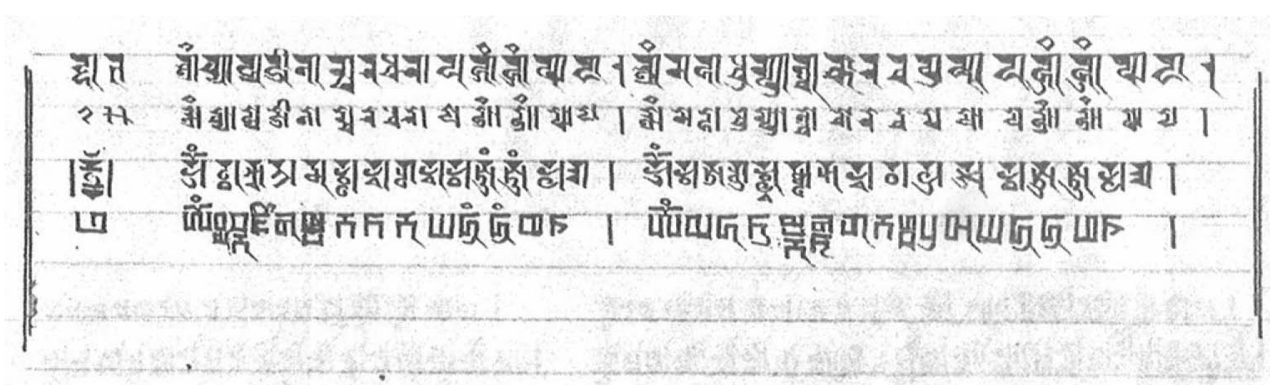


Figure 24: Folio showing Sanskrit written in Zanabazar Square (from Byambaa Ragchaagiin 2005: 103). Conjuncts with the full forms of 𑨩 YA, 𑨪 RA, and 𑨟 BA as C_2 are shown: 𑨬𑩇𑨩 *vya*, 𑨟𑩇𑨪𑨊 *brā*, 𑨡𑩇𑨟 *mba*, 𑨡𑩇𑨝𑩇𑨪𑨊 *mprā*, 𑨫𑩇𑨬 *lva*. Also shown is the conjunct 𑨜𑩇𑨛 *ndha*.

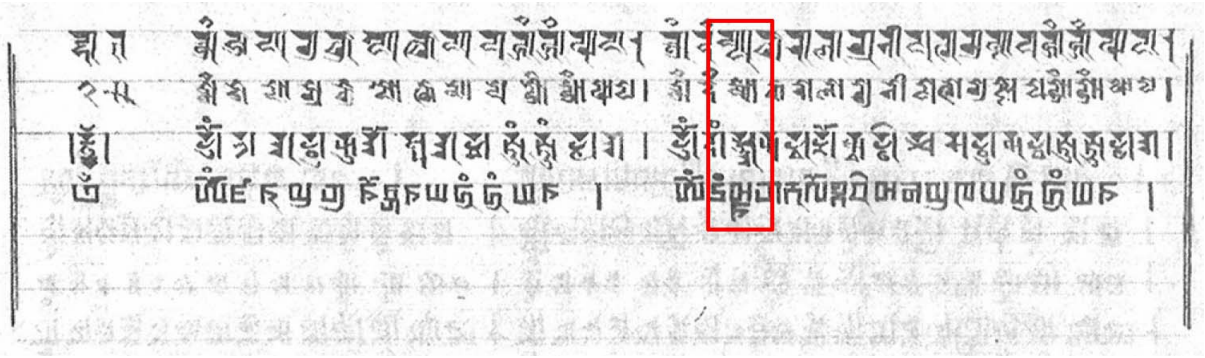


Figure 25: Folio showing the conjunct 𑨹 *ṣṭa* (from Byambaa Ragchaagiin 2005: 103). Compare to the form of the conjunct shown in figure 26.

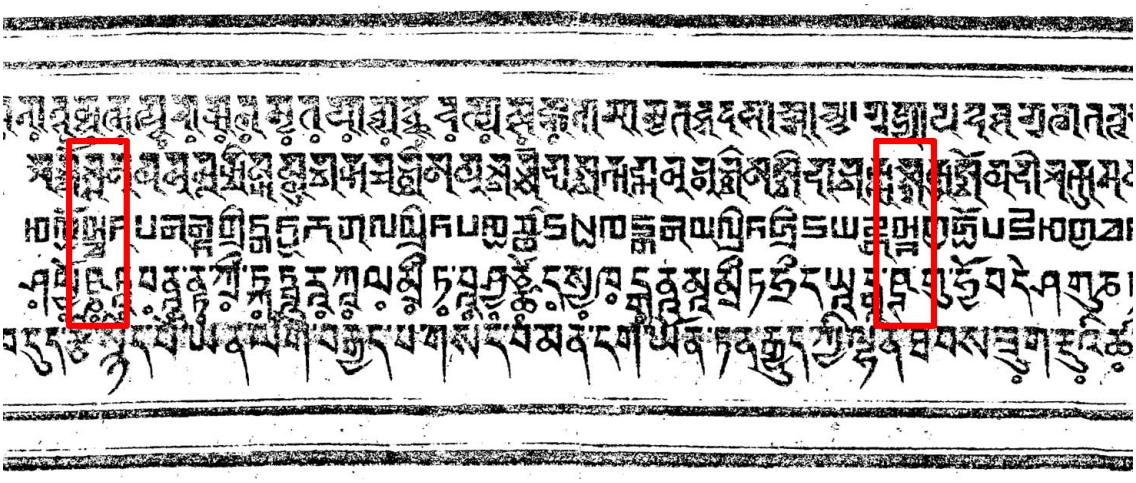


Figure 26: Detail of manuscript from figure 4 showing the conjunct *ṣṭa* written as 𑨹 instead of the expected 𑨹. Here the *ṭa* is represented with a reversed form 𑨄 of 𑨃 TA instead of 𑨂 TTA. See section 4.7.4 for more information on the use of reversed letters in Zanabazar Square. The other conjunct involving SSA in this folio is 𑨹 *ṣṇa*, which is written as expected.

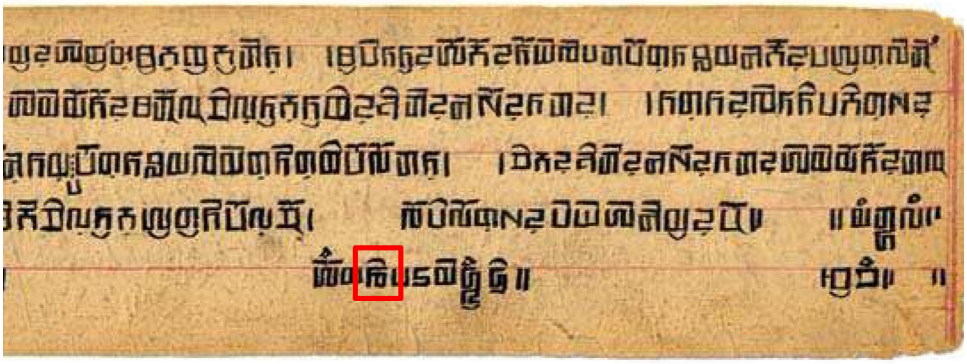


Figure 27: The use of 𑨦, a reversed form of 𑨤 NA, instead of 𑨝 NNA for writing the Sanskrit retroflex *ṇa* in the word 𑨫𑨦 *maṇi* (excerpted from Bareja-Starzyńska and Ragchaa 2012: folio 7a).

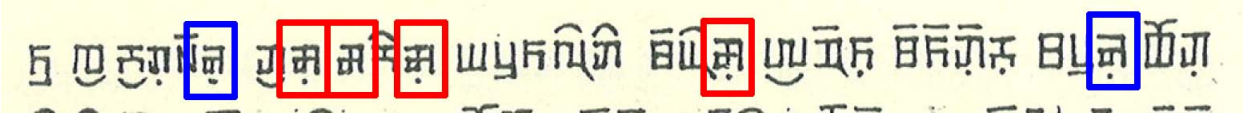


Figure 28: Usage of the variant form 𑨥 (highlighted in red) of 𑨤 NA (highlighted in blue) (from Byambaa Ragchaagiin 2005: 85). There is no semantic distinction between the variant and normative form; they both represent *na*.

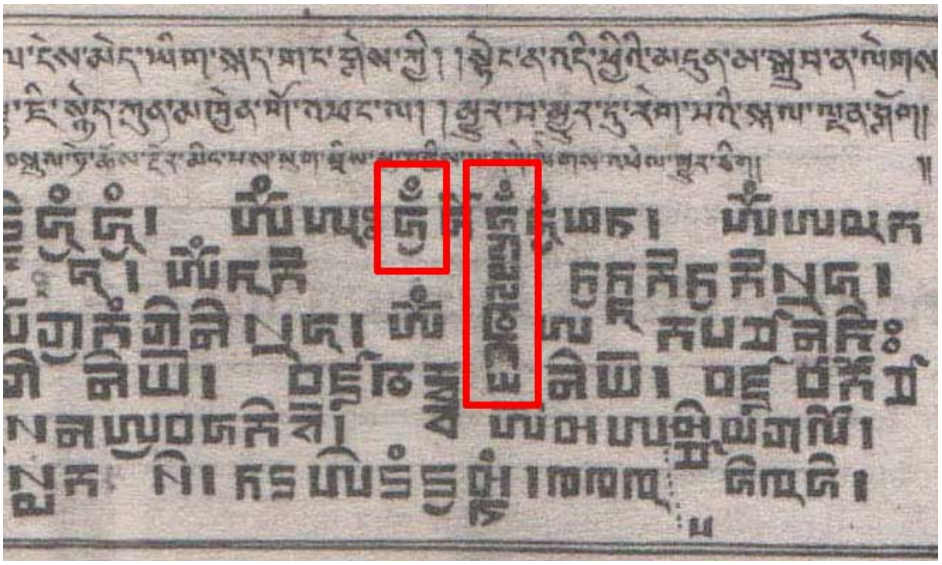


Figure 29: Excerpt from a folio shown in figure 3 illustrating the usage of 𑨁 SIGN CANDRABINDU.

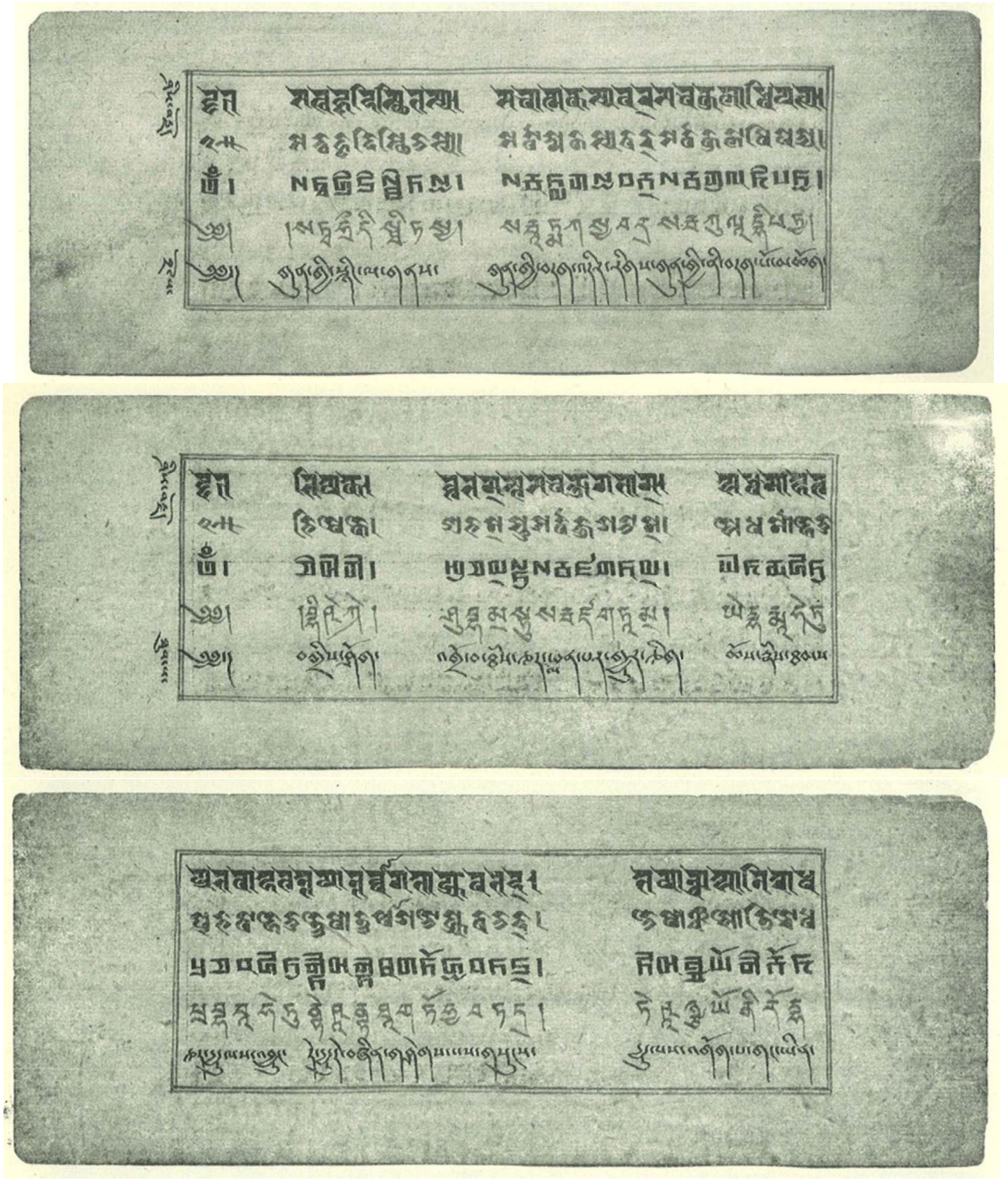


Figure 30: Folios showing usage of the ◌ VIRAMA (from Byambaa Ragchaagiin 2005: 113–114). In the top folio, VIRAMA occurs with RA in *kasyabar* (*kasyapaḥ* "Kasyapa" is likely intended). The expected representation of the phrase is *kasyabaḥ-sarva-kulādhipatya* "Kasyapa, the sovereign of all clans"; the ◌ः *ḥ* is written as *r* due to incorrect application of sandhi). In the middle folio the VIRAMA occurs with MA twice in the phrase *śubhamstu-sarva-jagatam* "may there be auspiciousness in all the world"; the use of VIRAMA in *śubham* is curious because in Sanskrit the word would join with the following *stu* (correct: *astu*) as *śubhamastu*. In the bottom folio the VIRAMA occurs with DA in the word *hyavatad* (correct: *hyavadat*) "thusly explained".

Уг дурсгалын эхийг галиглавал:

1. jir-gu-gan er-ke-te-ne e-je-leg-
2. ci a-nu gag-ca sed-kel bu-yu /
3. sed-kil-ber ko-go-son ke-men ba-
4. rim-ta-lal ü-gi-gü-yi a-gu-lug-
5. san-yir ö-bör-sü-ben al-dar-
6. cü / a-li-madeta-ber ba-rim-tal-
7. ku ü-gīn tu-lad / jir-gu-gan
8. er-ke-ten-ber ö-börun ci-
9. nar-tür al-dar-ku bo-lo-mui /

Зураг 17
Уг бичээсийг тусгайд нь томруулсан байдал

Figure 31: Usage of ◌̣ FINAL CONSONANT MARK for representing Mongolian final consonants (from Byambaa Ragchaagiin 2005: 63).

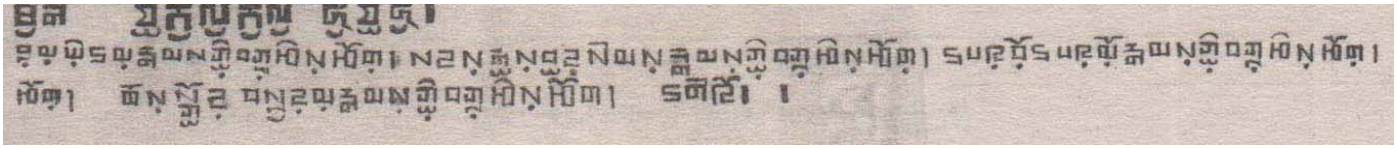


Figure 32: Excerpt from the last folio in figure 3 showing usage of ◌̣ FINAL CONSONANT MARK in Tibetan text. Here, its usage departs from its normative behavior in marking syllable-final consonants and occurs after conjuncts and vowels, functioning more like a below-base ་ TSHEG than a final-consonant mark. For example it occurs with ྦྷ̣ *bla* (first syllable), ཡི̣ *yi* (third syllable), ཀྱི̣ *kyi* (ninth syllable), among others. Agata Bareja-Starzyńska suggests that text (a Tibetan prayer) may have been added to the manuscript by another scribe, who did not fully know the Zanabazar Square script, as the usage of FINAL CONSONANT MARK here differs from that of the primary text of the original author.

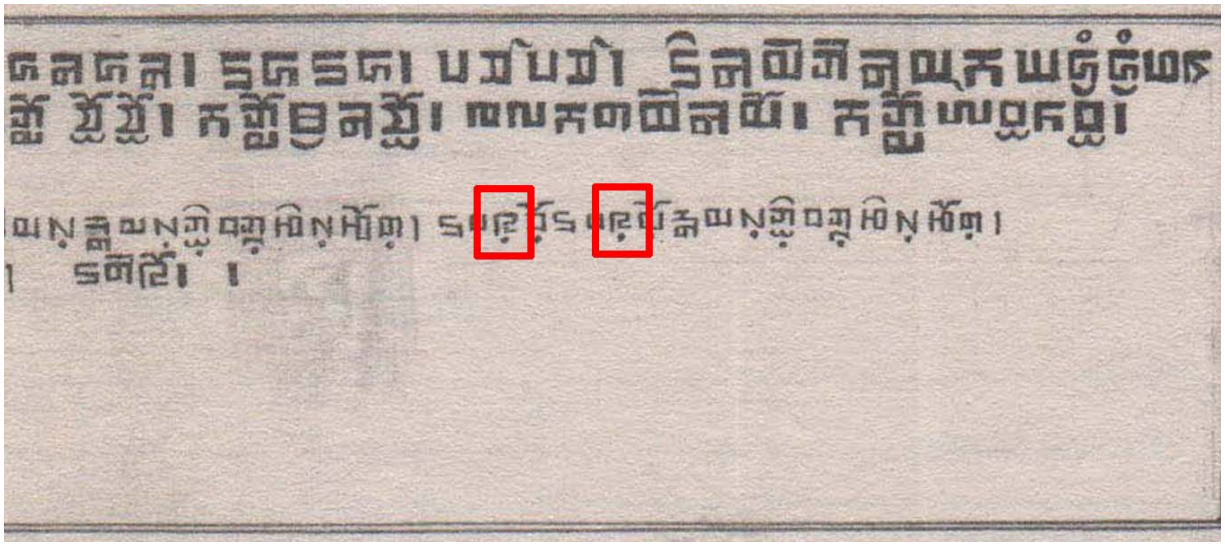


Figure 33: Manuscript excerpt showing usage of the ◌̣ FINAL CONSONANT MARK with 𑨪 SMALL A for indicating word-final 𑨪̣ *'a* (from Byambaa Ragchaagiin 2005: 51).

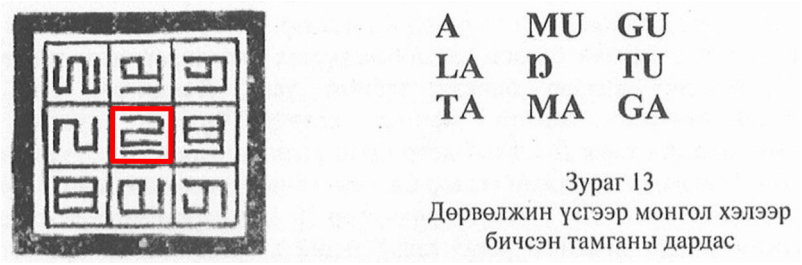
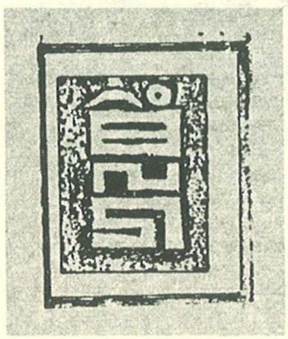


Figure 34: A seal showing the variant form ◌̲ of ◌̣ FINAL CONSONANT MARK for representing the Mongolian final consonant *ng*: 𑩵 for 𑨵̣ (from Byambaa Ragchaagiin 2005: 60).

Figure 35: Portion containing Zanabazar Square text from figure 8 showing usage of the variant form for SMALL A in the words *pa'i* and *'bri*. Details courtesy of Agata Bareja-Starzyńska.



6. Мөн *Цогт нууцын хураангуйн язгуур үндэс*ний 11-р бүлгээс гарсан Язгуурын таван бурханыг бэлгэдсэн нүдний гэмийг амирлуулагч *thalīṁ* "талим" хэмээх бэлгэдэлийн тарнийн үсгийг Занабазарын дөрвөлжин үсгээр бичсэн байдаг. Гэвч энэ тарнийг бичихдээ бар сийлэгч *шанаган* а үсгийг буруу харуулан сийлсэн болонтой.

Зураг 25
"Талим" үсгийн дардас

Figure 36: Zanabazar Square document showing the word *thalīṃ* written vertically with , the reversed form of SMALL A; see also figure 22 (from Byambaa Ragchaagiin 2005: 97).

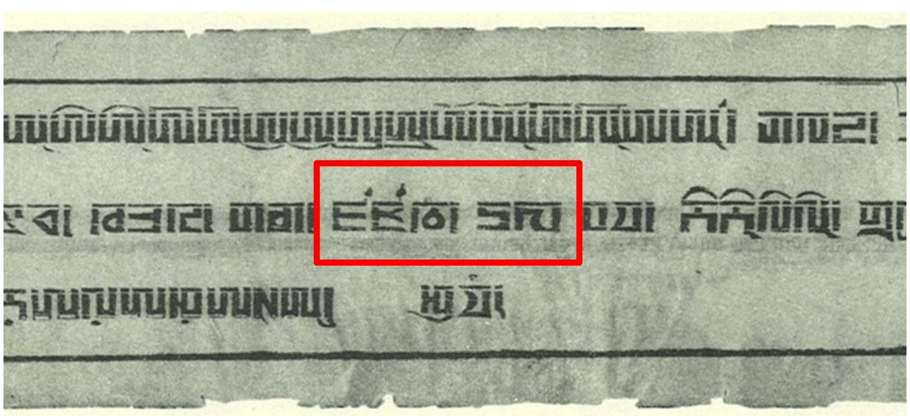


Figure 37: Excerpt of the second folio from figure 5 showing the sequence "" *dza dzha va . da dha .* in the traditional arrangement of the script. Note the appearance of "" <VA, SHAD> as and "" <DHA, SHAD> as .

Figure 38: Manuscript showing the variant form 𑨭 of 𑨭 VA in the phrase 𑨱𑨝𑨃𑨭𑨮𑨭𑨫𑨱𑨁𑨰𑩇𑨭𑨊𑨱𑨊 *ha nu va śa va ra hi svā hā* (from Byambaa Ragchaagiin 2005: 98). Note the variant form ◌𑨃 of ◌𑨃 VOWEL SIGN U in 𑨝𑨃 *nu* and the variant form ◌𑩇𑨭 of the subjoined form ◌𑩇𑨭 of 𑨭 VA in the conjunct 𑨰𑩇𑨭𑨊 *svā*.

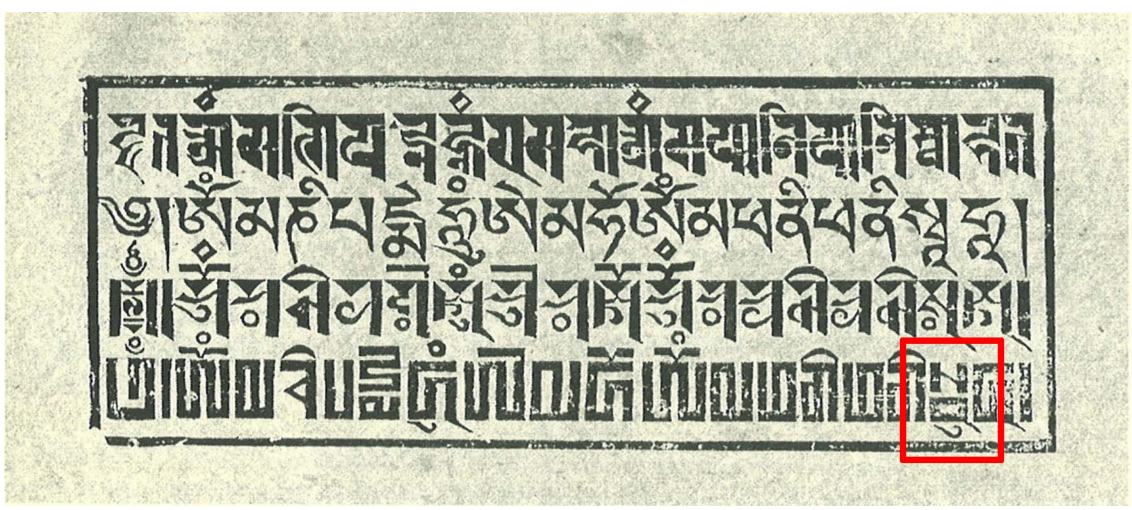


Figure 39: Manuscript showing the word 𑨰𑩇𑨠𑨃𑨱𑨊 *svuhā* (*svāhā*) written with the conjunct 𑨰𑩇𑨠 *sba* (from Byambaa Ragchaagiin 2005: 98).

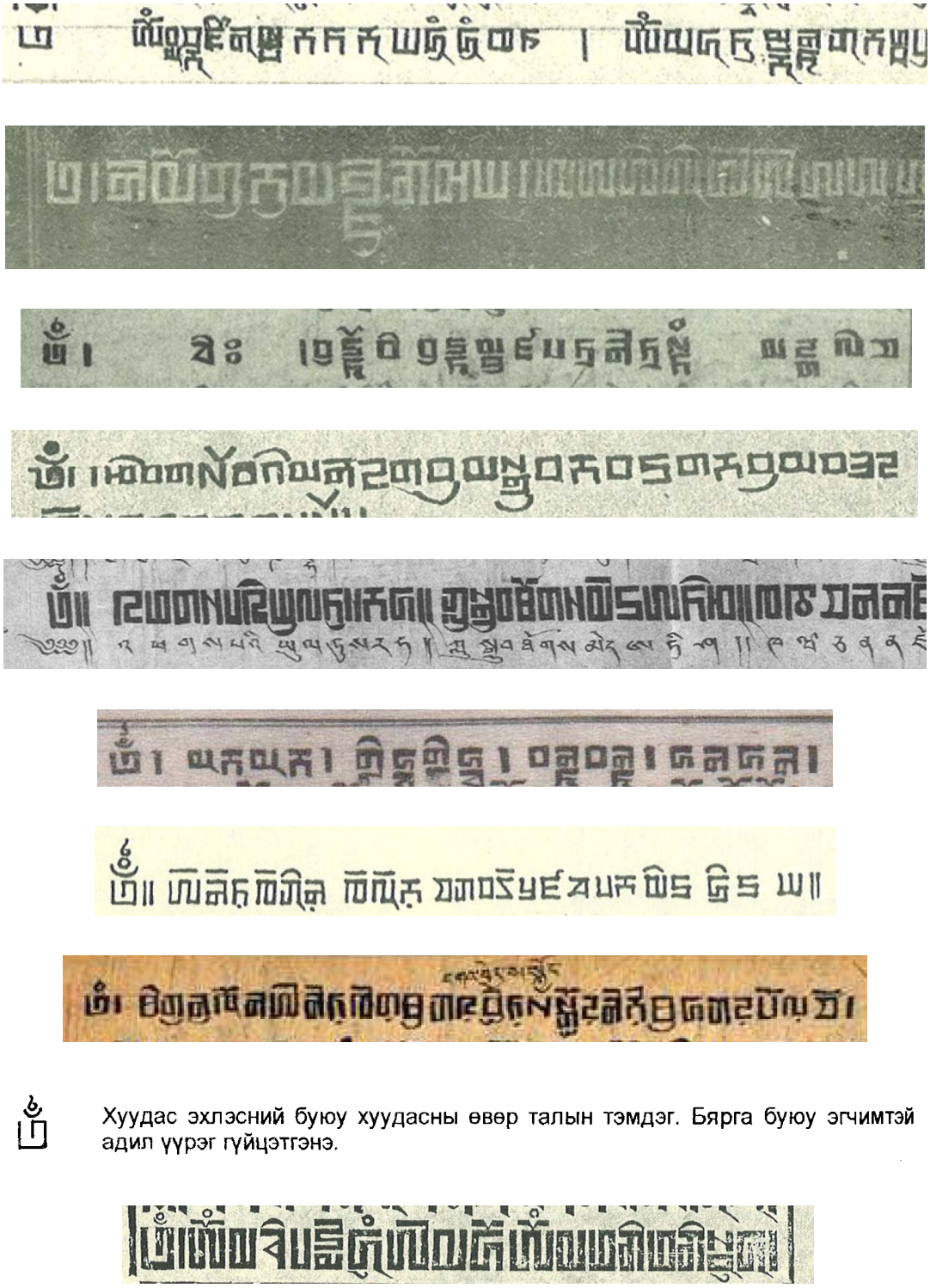


Figure 40: Variations of the 𑩐 INITIAL HEAD MARK written as a bare mark as well as with various ornaments. See section 4.11 for details.

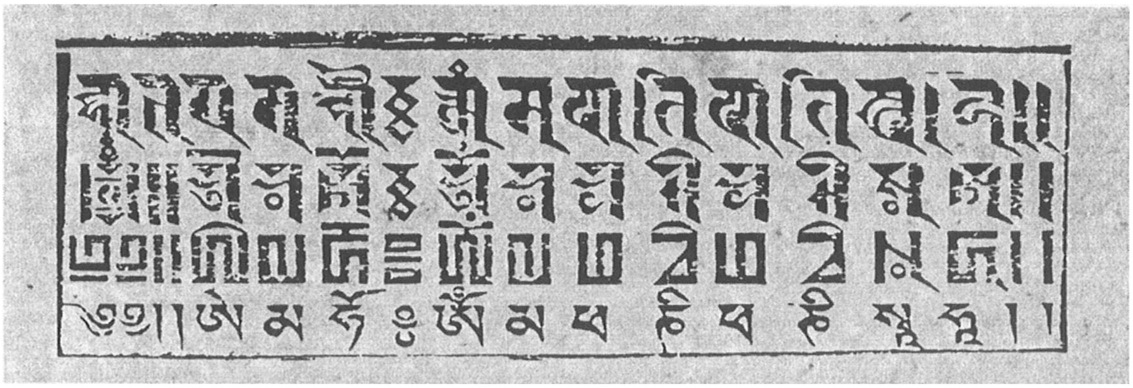


Figure 41: Manuscript showing usage of the head mark 𑨿𑩀 (from Byambaa Ragchaagiin 2005: 97). This mark corresponds to the Tibetan ༄. For purposes of the encoding, it is considered a glyphic variant of 𑨿𑩀 and is to be produced using <𑨿 INITIAL HEAD MARK, 𑩀 CLOSING HEAD MARK>. Note also the usage of the 𑩄 LONG TSHEG, corresponding to the Tibetan ༔ mark.

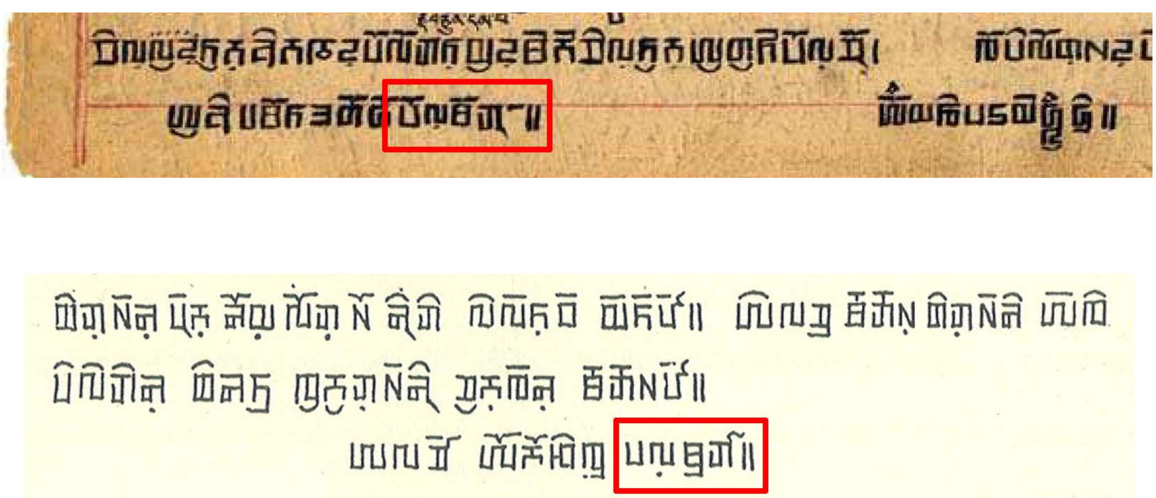


Figure 42: Two manuscript excerpts showing elongation of the ◌𑨇 VOWEL SIGN AI at the end of a text (top from Bareja-Starzyńska and Ragchaa 2012; bottom from Byambaa 2005: 87). In both excerpts, elongation occurs in the Mongolian word *boltugai* "let it be", which is written at top as *boltogai* and at bottom as *baltugai*. This form of VOWEL SIGN AI is a stylistic variation. Details provided by Agata Bareja-Starzyńska.

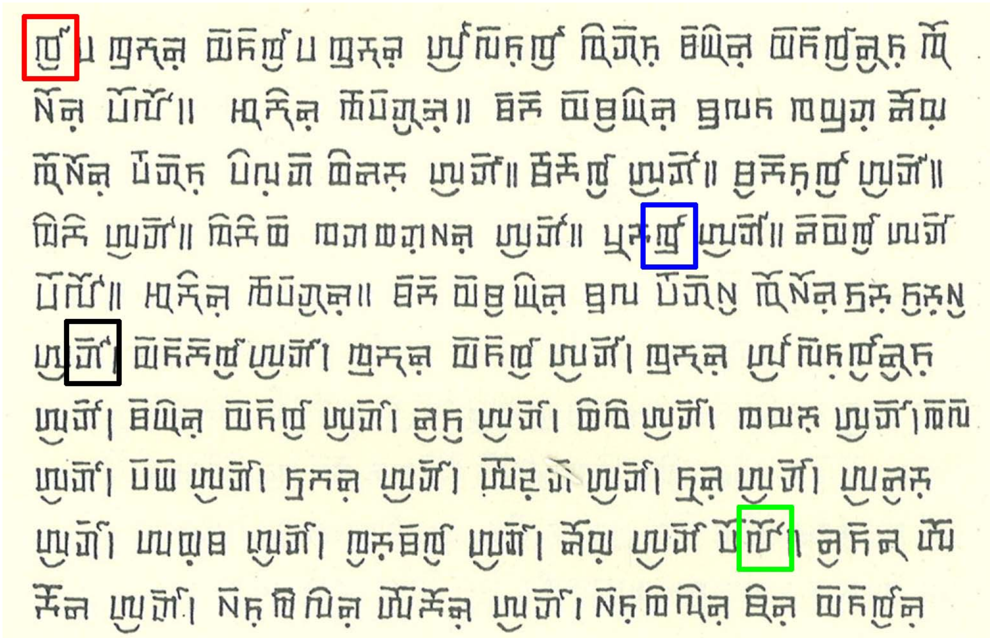


Figure 43: Mongolian diphthongs (from Byambaa Ragchaagiin 2005: 85). Shown in red is *küi*; in blue is *kui*; in black is *gei*; in green is *loi*. See section 4.4 for details on the representation of these diphthongs. Note also the use of the variant form for VOWEL SIGN AI.

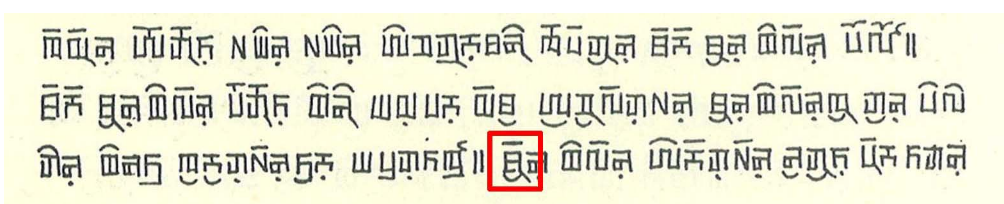


Figure 44: The Mongolian diphthong *eǖ* in *theǖ* (from Byambaa Ragchaagiin 2005: 86).

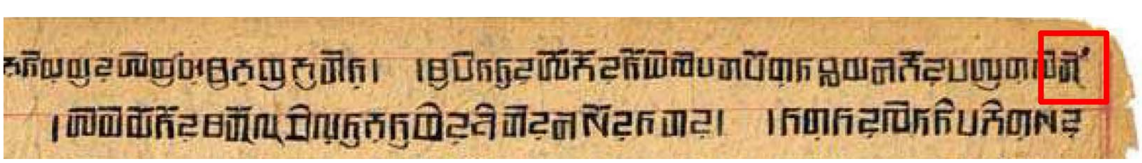


Figure 45: The Mongolian diphthong *ēi* in *gēi* (from Bareja-Starzyńska and Ragchaa 2012).

Figure 46: Detail of the specimen in figure 8 showing usage of ˙ TSHEG for marking syllabic boundaries in the sentence *sog pa'i yi ge sva yam bhu jyo ti 'bri tshul bzhugs so*.

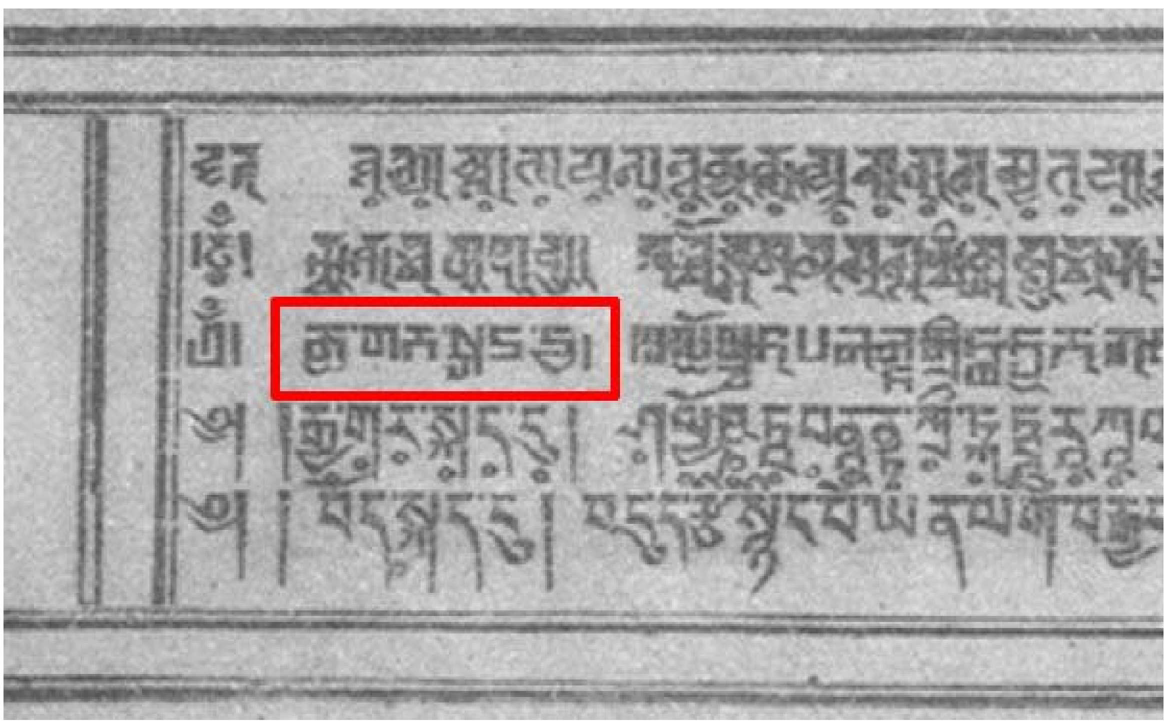


Figure 47: Detail of a folio showing usage of ˙ TSHEG for marking syllabic boundaries in the Zanabazar Square phrase *rgya gar skad du* (from Byambaa Ragchaagiin 2005: 107).

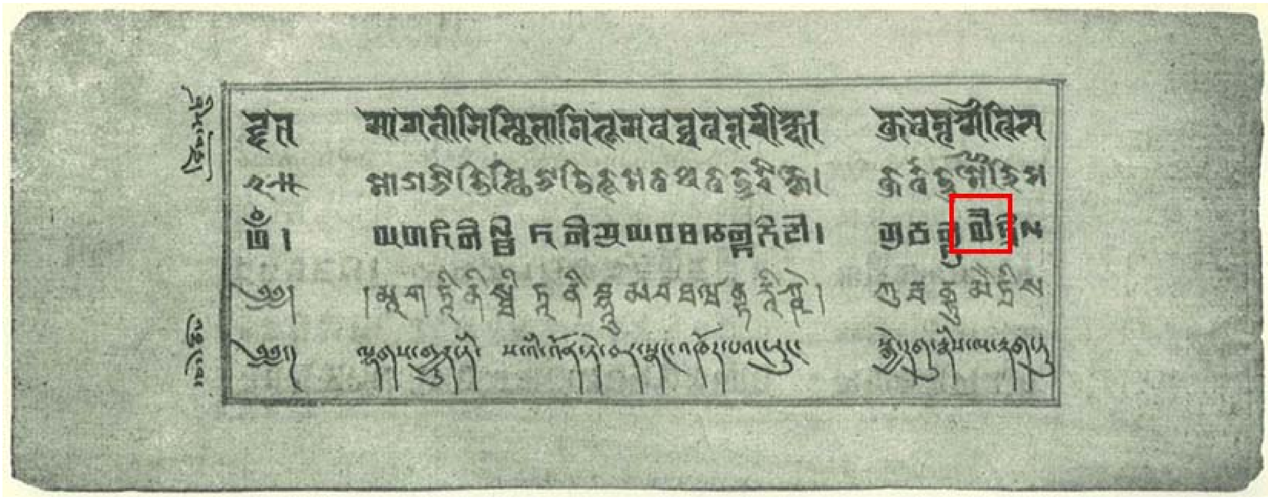


Figure 48: Folio showing usage of a variant form of ◌𑨇 VOWEL SIGN AI used for writing 𑨢𑨇 *mai* (from Byambaa Ragchaagiin 2005: 113).

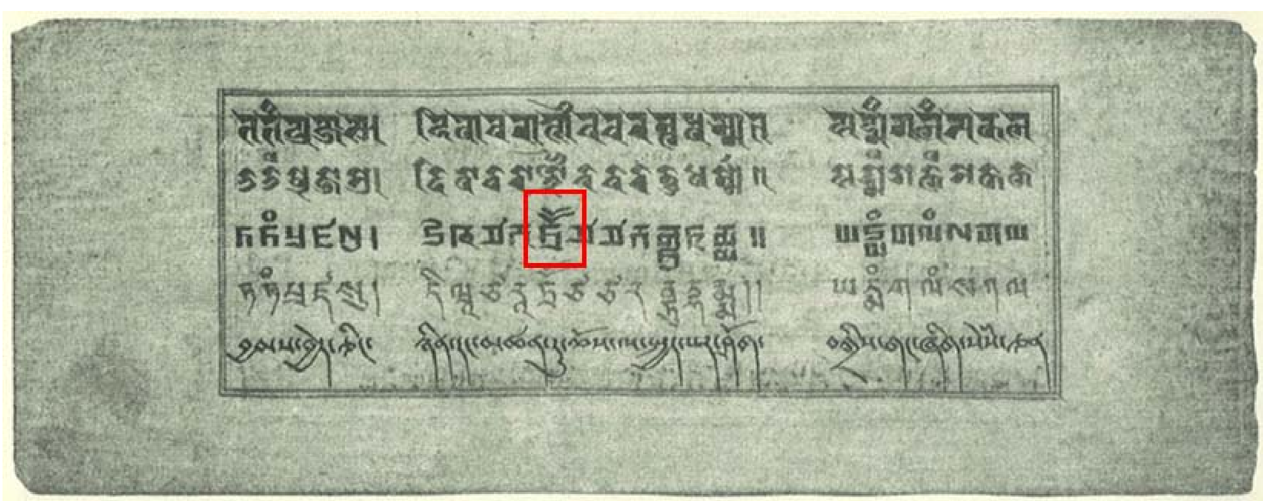


Figure 49: Folio showing usage of a variant form of ◌𑨈 VOWEL SIGN AU used for writing 𑨙𑩇𑨫𑨈 *trau* (from Byambaa Ragchaagiin 2005: 113).

	Zanabazar Square	Phags-pa	Tibetan (*dbu can*)	Tibetan (*dbu med*)
ka	𑨐	ཀ	ཀ	ཀ
kha	𑨑	ཁ	ཁ	ཁ
ga	𑨒	ག	ག	ག
gha	𑨓	—	ཀྷ	ྷ
ṅa	𑨔	ང	ང	ང
ca	𑨕	ཅ	ཅ	ཅ
cha	𑨖	ཆ	ཆ	ཆ
ja	𑨗	ཇ	ཇ	ཇ
ña	𑨘	ཉ	ཉ	ཉ
ṭa	𑨙	ཊ	ཊ	ཊ
ṭha	𑨚	ཋ	ཋ	ཋ
ḍa	𑨛	ཌ	ཌ	ཌ
ḍha	𑨜	—	ཌྷ	ཌྷ
ṇa	𑨝	ཎ	ཎ	ཎ
ta	𑨞	ཏ	ཏ	ཏ
tha	𑨟	ཐ	ཐ	ཐ
da	𑨠	ད	ད	ད
dha	𑨡	—	དྷ	ྷ
na	𑨢	ན	ན	ན

Table 7: Comparison of consonant letters of Zanabazar Square, Phags-pa, and Tibetan.

	Zanabazar Square	Phags-pa	Tibetan (*dbu can*)	Tibetan (*dbu med*)
pa	𑨤	𑩛	པ	པ
pha	𑨥	𑩜	ཕ	ཕ
ba	𑨦	𑩝	བ	བ
bha	𑨧	—	ཙྷ	ཚ
ma	𑨨	𑩞	མ	མ
tsa	𑨩	𑩠	ཙ	ཙ
tsha	𑨪	𑩡	ཚ	ཚ
dza	𑨫	𑩢	ཛ	ཛ
dzha	𑨬	—	ཛྷ	ཛྷ
zha	𑨭	𑩣	ཞ	ཞ
za	𑨮	𑩤	ཟ	ཟ
’a	𑨯	𑩥	འ	འ
ya	𑨰	𑩦	ཡ	ཡ
ra	𑨱	𑩧	ར	ར
la	𑨲	𑩨	ལ	ལ
va	𑨳	𑩩	ཝ	ཝ
śa	𑨴	𑩪	ཤ	ཤ
ṣa	𑨵	—	ཥ	ཥ
sa	𑨶	𑩫	ས	ས
ha	𑨷	𑩬	ཧ	ཧ
kṣa	𑨸	—	ཀྵ	ཀྵ

Table 8: Comparison of consonant letters of Zanabazar, Phags-pa, Soyombo, and Tibetan.

	Zanabazar Square	Phags-pa	Tibetan (*dbu can*)	Tibetan (*dbu med*)
a	𑨀	ꡀ	ཨ	ཨ
ā	𑨀𑨹	—	ཨཱ	ཨཱ
i	𑨁	ꡁ	ཨི	ཨི
ī	𑨁𑨹	—	ཨཱི	ཨཱི
ü	𑨂	—	—	—
ǖ	𑨂𑨹	—	—	—
u	𑨃	ꡂ	ཨུ	ཨུ
ū	𑨃𑨹	—	ཨཱུ	ཨཱུ
e	𑨄	ꡃ	ཨེ	ཨེ
ē	𑨄𑨹	—	—	—
o	𑨅	ꡄ	ཨོ	ཨོ
ō	𑨅𑨹	—	—	—
ö	𑨆	—	—	—
ǒ	𑨆𑨹	—	—	—
ai	𑨇	—	ཨཻ	ཨཻ
au	𑨈	—	ཨཽ	ཨཽ
ṛ	𑨉	—	ཨྲྀ	ཨྲྀ
ṝ	𑨉𑨹	—	ཨཱྲྀ	ཨཱྲྀ
ḷ	𑨊	—	ཨླྀ	ཨླྀ
ḹ	𑨊𑨹	—	ཨཱླྀ	ཨཱླྀ

Table 9: Comparison of vowels of Zanabazar Square, Phags-pa, Soyombo, and Tibetan.

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://www.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://www.dkuug.dk/JTC1/SC2/WG2/docs/summaryform.html**.**
See also http://www.dkuug.dk/JTC1/SC2/WG2/docs/roadmaps.html **for latest *Roadmaps*.**

A. Administrative

1. **Title:** *Proposal to Encode the Zanabazar Square Script in ISO/IEC 10646*
2. Requester's name: *Script Encoding Initiative (SEI) / Anshuman Pandey (pandey@umich.edu)*
3. Requester type (Member body/Liaison/Individual contribution): *Liaison contribution*
4. Submission date: *2014-01-22*
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: *Zanabazar Square*
   b. The proposal is for addition of character(s) to an existing block:
      Name of the existing block:
2. Number of characters in proposal: *69*
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.):
   Anshuman Pandey (pandey@umich.edu)
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: N4102-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 of N4471 L2/13-198; see "Introduction" for list of major 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)*
   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) *Virama; see text of the proposal*
13. Does the proposal contain any Ideographic compatibility characters? *No*
    If YES, are the equivalent corresponding unified ideographic characters identified?
    If YES, reference: