L2/18-126
2018-04-30

Preliminary proposal to encode Old Uyghur in Unicode

Anshuman Pandey
pandey@umich.edu

April 30, 2018

1 Introduction

This is a preliminary proposal to encode the 'Old Uyghur' script in Unicode. It provides a brief background on the script, a description of the proposed encoding model, a tentative character repertoire and related properties, orthographic details, and specimens of the script. Research in ongoing and a formal proposal is forthcoming. The information presented here is subject to change. Feedback from experts is requested and may be directed to the email address given below the name of the proposal author above.

2 Background

The 'Old Uyghur' script was used between the 8th and 17th centuries primarily in the Tarim Basin of Central Asia, located in present-day Xinjiang, China. It was used for writing Turkic languages, such as Uyghur, as well as Chinese, Mongolian, Sanskrit, Sogdian, and Tibetan. It was used for civil, literary, and religious purposes, and was important for the transmission of Buddhism in Central Asia.

As part of a vibrant multilingual culture in Central Asia, Old Uyghur was used alongside other scripts. There are numerous documents in the Old Uyghur script with intralinear Han characters, and Chinese manuscripts with Turkic translations in Old Uyghur script. Several manuscripts contain the Old Uyghur script with interlinear Sanskrit annotations in 'Turkestani' or Central Asian styles of Brahmi. The Old Uyghur script also occurs in records containing the Phags-pa script, and in annotations accompanying the Khitan large script in a manuscript fragment. Documents containing text in both the Old Uyghur and the Arabic scripts are also extant.

The script is a cursive-joining alphabet with features of an *abjad*. It developed from the 'cursive' style of the Sogdian script during the 8th–9th century (Kara 1996: 539). This style is also known as the 'Sogdian Uyghur' script. It was written horizontally, but at some point the writing direction was rotated 90 degrees counter-clockwise and written vertically from top to bottom in columns that extend from left to right. This orientation became an intrinsic feature of the Old Uyghur script.

The scribal tradition of Old Uyghur may be divided into three styles based upon analysis of letterforms in various records: 'cursive', 'formal', and 'standard'. It may also be periodized into 'early', 'proper', and 'late'. The 'cursive' style occurs in numerous civil and administrative documents from the 9th through 15th

centuries (see fig. 28–29). By the 10th century, scribal refinements resulted in a style that may be called 'formal' Old Uyghur, which was used for religious and literary manuscripts (see fig. 19–26). The script was developed further through printing. The introduction of block printing for producing texts gave rise to what may be considered a 'standard' Old Uyghur script in the 12th century. Numerous folios and fragments of block-printed books have been preserved (see fig. 8–16 for specimens). This 'standard' block-print style is similar to the formal inscriptional type, which appears on the stone walls of the Cloud Platform at Juyong Guan, Beijing, erected in the 14th century (see fig. 30).

Just as Turkic peoples adopted the Sogdian script for writing their language, speakers of other languages in Central Asia adapted the Old Uyghur script for their own writing practices. A popular narrative states that in the 13th century, during the reign of Genghis Khan, the scholar Tata Tonga developed an orthography for writing the Mongolian language using the Old Uyghur script. The Uyghur-based Mongolian script developed into a distinctive script with its own scribal and print culture. The Old Uyghur script flourished through the 15th century and was replaced by a new orthography based upon the Arabic script in the 16th century. It appears that Old Uyghur was used in Gansu through the 17th century.

At the beginning of the 20th century, German and Russian scholars adapted the Old Uyghur script for modern typesetting. Buddhist texts in the Old Uyghur script were edited and published by V. V. Radlov and others (see fig. 31, 32). The metal types used in these editions appear to have been cut to match the letterforms found in Old Uyghur block-prints.

There has been active scholarship on the Old Uyghur script and manuscripts since the early 20th century. It was during this time that European expeditions to Turfan unearthed vast amounts of materials in Old Uyghur and other scripts. The past century has witnessed increasing growth of interest in Old Uyghur sources of the 8th through 15th centuries within studies of cultures, peoples, and polities of the Silk Road. Various institutions that obtained materials from Turfan and other sites have digitized their collections or are in the process of doing so, such as the Berlin-Brandenburgische Akademie der Wissenschaften (BBAW), British Library, and other institutions associated with the International Dunhuang Project (IDP).

3 Script identifier

The proposed Unicode identifier for the script is 'Old Uyghur'. The script is also known as 'Uyghur', which is spelled variously: 'Uighur', 'Uigur', 'Uygur', 'Uyğur', etc. To be sure, neither 'Uyghur' nor 'Old Uyghur' is an entirely accurate designation. Clauson notes that the "name is probably as anachronisic as that name when applied to the language" (1962: 100). The script had been in use before the Uyghur language became prominent in the region in the 8th century (1962: 43). However, Clauson concludes that "no useful purpose would be served by suggesting some other name" (1962: 100–101). For purposes of identifying the script in Unicode, the adjective 'Old' is appended to 'Uyghur' in order to distinguish the script from the later Arabic orthography used for writing the modern Uyghur language (which is not directly related to the Uyghur language of the 8th century). Given the polysemia of 'Uyghur', the term 'Old Uyghur' has become common for referring to the script, even if it is imprecise.

4 Encoding history

A proposal for Old Uyghur was previously submitted to the Unicode Technical Committee (UTC) by Omarjan Osman in 2013. The "Proposal to Encode the Uyghur Script in ISO/IEC 10646" (L2/13-071) provides valuable background on the history and usage of the script, and details about the representation of letterforms and orientations of the script in different manuscripts. Based upon the provenance and attributes of two important sources, Osman identified two major variations of the script along a geographic basis. He describes the 'western' form as being written horizontally from right to left, and an 'eastern' form that is written vertically from top to bottom (p. 11). Osman thought it necessary to accommodate both orientations of the script at the character level. Therefore, he proposed a repertoire that contains upright glyphs for the horizontal form (for right-to-left display), and the same glyphs rotated 90 degrees counter-clockwise for the vertical form.

The model presented in L2/13-071 is ambitious, but it is not practical. It is also incompatible with the Unicode character-glyph model. The encoding of separate characters for horizontal and vertical orientations of a letter results in a model that establishes separate semantic values for glyphic variants of a given letter. Such a repertoire is redundant and prone to complications, for example, errors caused by usage of a horizontal letter in a string of vertical characters, etc. It would be more appropriate to consider such glyphs as directional variants instead of separate characters. Moreover, instead of attempting to accommodate orientations of the script at the character level, it would be practical to use mark-up and layout to achieve the desired display. Nonetheless, Osman's proposal is a useful resource for further investigating the requirements for encoding Old Uyghur. His proposed repertoire includes digits and several diacritics (whose exact provenance is not given), which must be investigated in order to determine a complete character repertoire for representing Old Uyghur texts.

The present proposal provides an encoding for Old Uyghur that is aligned with Unicode principles and that utilizes current technologies for vertical and bidirectional text. The model is described in the next section.

5 Approach to the Encoding

5.1 Plain text representation

The Old Uyghur script is proposed for encoding as a distinctive script in Unicode. At a purely glyphic level it may appear that the script could be unified with Sogdian or Mongolian. But, a separate encoding is required to support the technical, graphical, and orthographic features of the script.

- *Plain text representation* The development of a block-printing standard established a distinctive Old Uyghur script. There is a need to represent content in this script in plain text. In particular, it is necessary to contrast texts in Old Uyghur from Sogdian and Mongolian in order to distinguish texts in these scripts for text processing and digitization on the basis of character identity and semantics.

- *Scope* Over the course of its development the Old Uyghur script developed its own stylistic variants. A distinctive Old Uyghur block provides a means for managing the taxonomy of these script variants by unifying the 'cursive', 'formal', and print 'standard' forms. The representative form of the script is based upon the block print style. Other styles may be displayed using fonts.

- *Character repertoire and semantics* The repertoire, order, and names of Old Uyghur letters is based upon that of Sogdian. The proposed encoding for Old Uyghur retains these attributes. The Mongo-

lian encoding uses different names and ordering for letters, which reflect Mongolian preferences and pronunciations. Mongolian letter names do not correspond to Old Uyghur values.

- *Glyphic distinction* Following from the above, a separate encoding preserves the glyphic distinctions of formal Old Uyghur in multilingual contexts that include Sogdian and Mongolian text. In particular, Mongolian glyphs do not adequately transmit the aesthetic and orthographic features of Old Uyghur. A comparison of Uyghur letters with corresponding letters in the Mongolian block is given in fig. 39.

- *Encoding model* A practical encoding for Old Uyghur is based upon a graphetic model. The base letters of the script would be encoded, while letters with diacritics would be decomposed into a base and a combining sign. The graphetic model for Old Uyghur facilitates implementation. The current model for Mongolian, which has a phonetic basis, presents several issues and is unsuitable for Old Uyghur.

- *Orientation* The default orientation for Old Uyghur is vertical. On the other hand, Sogdian is defined as a horizontal script. Mongolian is defined as a vertical script, but it differs from Old Uyghur in its representation in horizontal layouts.

The proposed model for Uyghur offers a practical implementation for a vertical script that avoids the complications of the Mongolian model.

5.2 Character names

In the scholarly literature, Uyghur letters are known by the names of the original Aramaic letters from where they are derived. This nomenclature is likely derived from the names used for Sogdian letters. This convention is followed for assigning Unicode names to Uyghur letters. Throughout this proposal, italics are used for scholarly names for graphemes, while small capitals indicate Unicode character names, eg. 𐽀 is *aleph* and OLD UYGHUR LETTER ALEPH. For brevity, the descriptor 'OLD UYGHUR' may be dropped, eg. OLD UYGHUR LETTER ALEPH is truncated to ALEPH. Characters of other scripts are designated by their full Unicode names. Latin transliteration of Old Uyghur follows the current scholarly convention.

5.3 Representative glyphs

Representative glyphs and calligraphic features are based upon the script used in block prints, exemplified by the manuscript in fig. 8, and those in 9–16. The block-print style is the most suitable representative of the Old Uyghur script because it may be considered a 'standard'. In general, preparing a script for printing requires making conscientious decisions about letterforms, calligraphic features, and general aesthetics of a script. Moreover, the block prints exemplify a Old Uyghur script that is distinctive. The font used here was designed by the proposal author.

6 Proposed repertoire

The proposed repertoire for Old Uyghur contains 28 characters: 17 letters, 4 combining signs, 6 punctuation signs, and 1 stem-extending sign. The code chart and names list follows p. 12. The encoded set may differ from traditional and scholarly inventories of the script. Such differences naturally arise from the requirements for digitally representing a script in plain text and for preserving the semantics of characters.

6.1 Letters

Character name	Glyph	Joining	Latin
OLD UYGHUR LETTER ALEPH	𐽰	dual	ʾ
OLD UYGHUR LETTER BETH	𐽱	dual	β
OLD UYGHUR LETTER GIMEL	𐽲	dual	γ
OLD UYGHUR LETTER WAW	𐽳	dual	w
OLD UYGHUR LETTER ZAYIN	𐽴	dual	z, ž
OLD UYGHUR LETTER HETH	𐽵	dual	x, q
OLD UYGHUR LETTER YODH	𐽶	dual	y
OLD UYGHUR LETTER KAPH	𐽷	dual	k
OLD UYGHUR LETTER LAMEDH	𐽸	dual	δ
OLD UYGHUR LETTER MEM	𐽹	dual	m
OLD UYGHUR LETTER NUN	𐽺	dual	n
OLD UYGHUR LETTER SAMEKH	𐽻	dual	s, š
OLD UYGHUR LETTER PE	𐽼	dual	p
OLD UYGHUR LETTER SADHE	𐽽	dual	c
OLD UYGHUR LETTER RESH	𐽾	dual	r
OLD UYGHUR LETTER TAW	𐽿	dual	t
OLD UYGHUR LETTER LESH	𐾀	dual	l

Nominal forms Nominal forms are based upon isolated forms or are derived from final forms. Isolated forms of letters are shown in an inventory written in the margin of U 40 (see fig. 22).

Vowels Vowels are indicated using 𐽀 *aleph*, 𐽅 *waw*, and 𐽆 *yodh*, and combinations of these letters in digraphs and trigraphs (see § 9.1).

gimel and ***heth*** These letters are distinguished by the nominal / isolated shapes, eg. 𐽁 and 𐽇, respectively; however, they have the same initial and medial forms. The sound *q* is represented by writing the diacritic ◌𐽐 above 𐽇 *heth* (see § 9.2). In initial and medial positions the same diacritic is used for distinguishing *gimel* from *heth*, eg. medial 𐽃 *gimel* and medial 𐽃◌𐽐 *heth*.

zayin In some sources 𐽄 *zayin* does not connect to a following letter. However, it may connect to the left in other contexts. It has been defined as a dual-joining letter in order to enable the left-joining feature. The control character [ZWNJ] ZWNJ may be placed after ZAYIN to prevent joining with the following letter. When *zayin* is non-connecting the terminal of its final form is slightly truncated from 𐽄 to 𐽉. In some sources *zayin* is distinguished using diacritics as 𐽄◌𐽑 and 𐽄◌𐽒 (see § 9.2).

yodh In some sources, the 𐽆 *yodh* does not connect to a following letter. The control character [ZWNJ] ZWNJ may be placed after YODH to prevent joining with the following letter.

nun The 𐽂 *nun* may be written with a dot above 𐽂◌𐽓 to distinguish it from 𐽀 *aleph* in non-final positions, eg. medial 𐽂◌𐽓 *nun* and medial 𐽊 *aleph* (see § 9.2).

shin The letter 𐽌◌𐽔 is used for representing *shin*, or the sound *š*. It is not a distinctive letter, but is written by applying the diacritic ◌𐽔 to the base 𐽌. This base is identical to 𐽌 *samekh*, and has the same contextual forms. For this reason, a letter *SHIN has not been encoded; instead it is to be represented using a sequence of 𐽌 SAMEKH and the combining sign ◌𐽔 (see § 9.2). Clauson reports that *samekh* and *shin* were distinguished in early manuscripts (1962: 109). Further research is required to confirm the distinctive forms of these letters in early Old Uyghur documents.

lesh The letter 𐽍 is known as 'hooked r' and represents the sound [l]. It is derived from 𐽍 U+10F44 SOGDIAN LETTER LESH, which likely evolved from the practice of indicating [l] by placing a subscript *resh* (◌𐽍 U+10F4F SOGDIAN COMBINING RESH BELOW) beneath 𐽍 U+10F40 SOGDIAN LETTER RESH-AYIN, eg. 𐽍◌𐽍. The letter has been assigned the name 'LESH' based upon the name assigned to the corresponding letter in Sogdian. This is not a historical name, but one suggested by modern scholars as a neologism that aligns with the Aramaic naming convention. The name 'hooked r' has been specified as an alias in the names list.

6.2 Combining signs

Four combining signs are used for disambiguation and representation of new sounds (see § 9.2 for details):

Character name	Glyph
OLD UYGHUR COMBINING DOT RIGHT	◌𐽓
OLD UYGHUR COMBINING TWO DOTS RIGHT	◌𐽔

OLD UYGHUR COMBINING DOT LEFT	◌𐽑
OLD UYGHUR COMBINING TWO DOTS LEFT	◌𐽒

Although these signs resemble small strokes, they originate from Sogdian dot diacritics, and therefore, are named as 'dot' or 'dots'. The descriptors 'right' and 'left' refer to the placement of these signs with respect to the base letter in the traditional vertical orientation of the script. When Old Uyghur occurs in non-vertical contexts, the signs labeled 'right' would be placed below the base, and the signs labeled 'left' would occur above the base letter. For example, in horizontal layout the ◌𐽐 COMBINING DOT RIGHT would appear as ◌𐽐.

Erdal (1984) describes some diacritic signs used for diambiguation and transliteration of Arabic in administrative documents in the Old Uyghur script of the 11th century from Yarkand. Clark (2010) also describes some signs used in the Old Uyghur manuscript of the *Kutadgu Bilig*, an 11th century Karakhanid work by Yusūf Khāṣṣ Ḥājib. Further research is required to determine the complete set of these signs and the method for encoding them.

6.3 Punctuation signs

The following signs of punctuation are proposed for encoding:

Character name	Glyph
OLD UYGHUR PUNCTUATION BAR	𐽑
OLD UYGHUR PUNCTUATION TWO BARS	𐽒
OLD UYGHUR PUNCTUATION TWO DOTS	𐽓
OLD UYGHUR PUNCTUATION FOUR DOTS	𐽔
OLD UYGHUR PUNCTUATION FIVE DOTS	𐽕
OLD UYGHUR SECTION MARK	𐽖

The signs 𐽒 and 𐽓 are common forms of punctuation. They are used for delimiting text segments of various lengths, such as sentences. When these two signs are used together, 𐽒 indicates smaller segments, while 𐽓 marks longer sections (see fig. 10, 15).

The 𐽑 is also used for delimiting short segments (see fig. 16).

The sign 𐽔 is used for indicating the end of a section. While this sign is similar to the ⁘ U+2058 FOUR DOT PUNCTUATION already encoded in Unicode, the Old Uyghur 𐽔 is used in a vertical environment and is, therefore, proposed for encoding as a script-specific character.

Similarly, the 𐽑 is used as a general sign of punctuation and decoration, for example in fig. 12, is also found in Sogdian documents, but is encoded as a script-specific sign because of directional considerations.

The 𐽐 is used in the Juyong Pass inscription as a section mark.

6.4 Stem extender

The following character is used for extending the baseline (see § 9.3 for details). It is used as a typographic filler and also for indicating a suffix that is separated from the stem.

Character name	Glyph
OLD UYGHUR STEM EXTENDER	𐽆

7 Script details

7.1 Structure

The Old Uyghur script is a cursive joining alphabet. Letters have nominal shapes when they occur in isolation and contextual shapes when they occur in initial, medial, or final position. All letters are defined as dual joining. In some sources the connection between letters is suspended (see § 6.1, 8.1). This feature may be supported by usage of the control character U+200C ZERO WIDTH NON-JOINER (abbreviated as ZWNJ).

Although derived from an *abjad*, Old Uyghur possesses the features of an 'alphabet'. Short vowels are generally indicated, and diagraphs and trigraphs are used for denoting the rich vowel repertoire of Turkic languages.

Diacritics are used for diambiguating letters with similar appearances and for representing sounds for which distinctive letters do not exist.

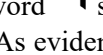
7.2 Directionality

Although Old Uyghur was written horizontally in early and late stages, the proper script was written vertically. Modern printed editions of Old Uyghur texts express fidelity to this orientation (see fig. 31, 32). The default directionality for Old Uyghur in Unicode is vertical, from top to bottom in columns that run from left to right.

When rendering Old Uyghur text in a system that does not support vertical layout, the text should be oriented in horizontal lines that run from right to left, and from top to bottom. This orientation is identical to conventional right-to-left scripts such as Sogdian and Arabic. In such cases, the glyphs of Old Uyghur letters are to be rotated 90 degrees clockwise with respect to their orientation in the code chart.

The above orientation is preferred over that specified for Mongolian in non-vertical environments. Mongolian is also a vertical script that runs from right to left, top to bottom. The Unicode core specification suggests that in non-vertical contexts Mongolian text should be laid out in horizontal lines running from left

to right with glyphs as represented in the code chart. Mongolian glyphs are shown in the code chart rotated 90 degrees clockwise with respect to their conventional vertical orientation. In Latin or other left-to-right contexts, this glyph orientation results in Mongolian text being rendered upside down in relation to the letters of the surrounding script.

Accordingly, the Old Uyghur word  should appear in horizontal contexts as  , not , as in the orientation for Mongolian. As evident in the previous sentence, when Old Uyghur occurs in horizontal layouts, a right-to-left orientation with glyphs upright as in Latin is practical for purposes of legibility. This orientation was preferred by scholars, especially because it was convenient to reference Old Uyghur words and phrases in multilingual contexts that also contain Arabic, Cyrillic, Devanagari, Tibetan, and other scripts (see fig. 38). Given the global distribution of scholars of Old Uyghur and Turkic studies, it is likely that these users will prefer to read the script with glyphs oriented upright when it appears in non-vertical layouts.

Following upon the above, throughout this document the proposed Old Uyghur characters are presented in their conventional vertical forms, while references to letters and signs within a Latin-script environment are given in a right-to-left orientation.

7.3 Line-breaking

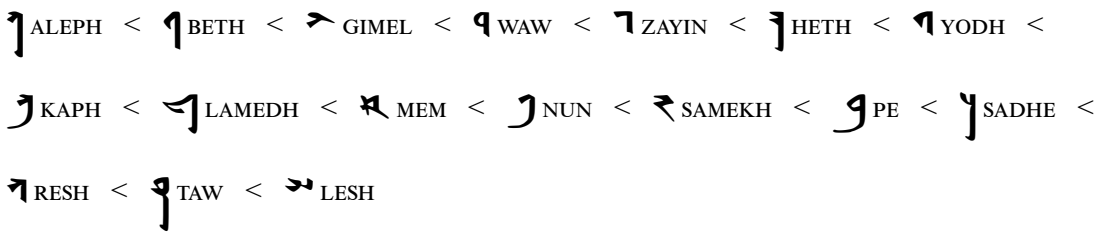
There are no formal rules for the breaking of words at end of line. Moreover, the available sources do not contain text with line-breaks for words. It may be assumed that words were not split at line boundaries. There are no indications of hyphens or other continuation marks. In digital layouts, line-breaks should occur occur after words.

7.4 Spacing and punctuation

Spacing is used for separating words. Signs of punctuation are used for indicating end of sentence and for larger sections of text.

7.5 Collation

The sort order for Old Uyghur follows the encoded order:



8 Joining behavior

The nominal ('X_n'), final ('X_f') medial ('X_m'), and initial ('X_i') forms of letters are shown below. All letters are defined as dual joining.

	X_n	X_f	X_m	X_i
ALEPH	𐽀	𐽀	𐽀	𐽀
BETH	𐽁	𐽁	𐽁	𐽁
GIMEL	𐽂	𐽂	𐽂	𐽂
WAW	𐽃	𐽃	𐽃	𐽃
ZAYIN	𐽄	𐽄	𐽄	𐽄
HETH	𐽅	𐽅	𐽅	𐽅
YODH	𐽆	𐽆	𐽆	𐽆
KAPH	𐽇	𐽇	𐽇	𐽇
LAMEDH	𐽈	𐽈	𐽈	𐽈
MEM	𐽉	𐽉	𐽉	𐽉
NUN	𐽊	𐽊	𐽊	𐽊
SAMEKH	𐽋	𐽋	𐽋	𐽋
PE	𐽌	𐽌	𐽌	𐽌
SADHE	𐽍	𐽍	𐽍	𐽍
RESH	𐽎	𐽎	𐽎	𐽎
TAW	𐽏	𐽏	𐽏	𐽏
LESH	𐽐	𐽐	𐽐	𐽐

The terminal stroke of the final form of some letters may be written in different directions, even within the same source.

	conventional	alternate
final ALEPH	𐽀	𐽁
final HETH	𐽅	𐽆
final KAPH	𐽊	𐽋
final NUN	𐽎	𐽏 , 𐽐

The orientation of terminals vary according to the whim of the scribe or the space available on a page. Terminal variation occurs most often at the end of a line for filling space or for compensating for lack of space at a margin. These stroke variations are stylistic and there is no semantic difference between final forms with different terminals. Alternate final forms may be controlled through fonts.

The shaping engine substitutes the nominal glyph for each letter in the input with the appropriate positional glyph to produce the expected joined output. In order to illustrate the joining properties of letters, representations of words from Old Uyghur records are given below along with their input strings:

ʾʾltwn
‘altun’

𐽀 ALEPH, 𐽀 ALEPH, 𐽙 LESH, 𐽔 TAW, 𐽃 WAW, 𐽍 NUN

bʾmsʾn
‘vamsan’

𐽀 BETH, 𐽀 ALEPH, 𐽂 MEM, 𐽗 SAMEKH, 𐽀 ALEPH, 𐽍 NUN

wyγwr
‘üigür’

𐽃 WAW, 𐽃 YODH, 𐽘 GIMEL, 𐽃 WAW, 𐽄 RESH

qwtlwγ
‘qutlug’

𐽅 HETH, ◌𐽞 COMBINING TWO DOTS LEFT, 𐽔 TAW, 𐽙 LESH, 𐽘 GIMEL

mncwšry
‘mancusari’

𐽂 MEM, 𐽍 NUN, 𐽌 SADHE, 𐽃 WAW, 𐽗 SAMEKH, ◌𐽟 COMBINING TWO DOTS RIGHT, 𐽄 RESH, 𐽃 YODH

swδwr
'sutur'

SAMEKH, WAW, LAMEDH, WAW, RESH

pwδystb
'bodisatav'

PE, WAW, LAMEDH, YODH, SAMEKH, TAW BETH,

pwrγʾn
'burxan'

PE, WAW, RESH, GIMEL, ALEPH, NUN

pylyk
'bilig'

PE, YODH, LESH, YODH, KAPH

twykʾl
'tükäl'

TAW, WAW, YODH, KAPH, ALEPH, LESH

tnkry
'tängri'

TAW, NUN, KAPH, RESH, YODH

8.1 Modification of cursive joining

In some texts certain letters do not join to a following letter in order to distinguish between letters that have similar appearances. The ZWNJ U+200C ZERO WIDTH NON-JOINER (ZWNJ) is to be used for modifying cursive joining. The ZWNJ is placed after the letter whose connection is suspended. The letter is rendered using its final form and the following letter appears in its initial form.

When *zayin* is non-connecting the terminal of its final form is slightly truncated from to .

ʾwyzʾ
'üzä'

ALEPH, WAW, YODH, ZAYIN, ALEPH

ʾwyzʾ
'üz-ä'

ALEPH, WAW, YODH, ZAYIN, ZWNJ, ALEPH

pwlzwn
'bolzun'

PE, WAW, LESH, ZAYIN, WAW, NUN

pwlz-wn
'bolz-un'

PE, WAW, LESH, ZAYIN, ZWNJ, WAW, NUN

8.2 Glyph interactions

The following letters have special behaviors when they interact with other letters.

waw In initial and medial position, the tails of *kaph* and *pe* attach below the baseline of the following letter, eg.  <KAPH, NUN> and  <PE, NUN>. When these letters are followed by  *waw*, their tails curve into the body of the *waw* to produce a ligature:

Character sequence	Ligated	Unligated
<KAPH, WAW>		
<PE, WAW>		

mem The extender of *mem* extends below the baseline in initial  and medial  positions. The extender of medial *mem* is written at an angle that slopes downward. The shaping of a word containing *mem* depends upon the position of the letter within the word:

- *Following a word-initial letter*: When a word-initial letter is followed by *mem*, the letter is enlarged and its baseline connects to the extender of *mem*, while the letter that follows *mem* joins to the body, eg.  <ALEPH, MEM, WAW>.

- *Following a non-initial letter*: When following after a non-word-initial letter, it is shifted towards the baseline and the preceding letter is angled downward in order connect to its extender. In such cases, the following letter is shifted away from the baseline, eg.  <ALEPH, ALEPH, MEM, WAW>.

lesh When  *lesh* follows letters with elements that extend below the baseline, the hook is detached from *lesh* and placed beneath the extension of the previous letter:  <KAPH, LESH>,  <MEM, LESH>,  <PE, LESH>. Even if *lesh* does not immediately follow *kaph*, *mem*, or *pe*, its hook attaches to the terminal of the latter for aesthetic considerations:

	shifted hook	static hook
pylyk ‘bilig’		
k›lm›dwk ‘kälmädük’		

9 Encoded representations

9.1 Vowels

Vowels are indicated using 𐽀 *aleph*, 𐽅 *waw*, and 𐽆 *yodh*, and combinations of these letters in digraphs and trigraphs. Vowels are represented as follows:

	Initial		Medial	
ä		𐽀 ALEPH		𐽀 ALEPH
a, *e*		𐽀 ALEPH, 𐽀 ALEPH		𐽀 ALEPH
i, *ï*		𐽀 ALEPH, 𐽆 YODH		𐽆 YODH
ī, *ī̈*		𐽀 ALEPH, 𐽆 YODH, 𐽆 YODH		𐽆 YODH, 𐽆 YODH
o, *u*		𐽀 ALEPH, 𐽅 WAW		𐽅 WAW
ö, *ü*		𐽀 ALEPH, 𐽅 WAW, 𐽆 YODH		𐽅 WAW
ö, *ü*		𐽅 WAW, 𐽆 YODH		𐽅 WAW, 𐽆 YODH
ō, *ō̈*, *ū*, *ṻ*		𐽀 ALEPH, 𐽅 WAW, 𐽅 WAW		𐽅 WAW, 𐽅 WAW

9.2 Disambiguation and extension of letters

The combining signs enumerated in § 6.2 are written with letters to diambiguate consonants or to represent consonants for which distinctive letters do not exist. The following forms are attested. Combining signs are placed after a letter in encoded text:

		X_n	X_f	X_m	X_i	
dotted *gimel*	*γ*					GIMEL, ◌ COMBINING TWO DOTS LEFT
dotted *zayin*	*ž*					ZAYIN, ◌ COMBINING DOT RIGHT
two dotted *zayin*	*ž*					ZAYIN, ◌ COMBINING TWO DOTS RIGHT
two dotted *heth*	*q*					HETH, ◌ COMBINING TWO DOTS LEFT
dotted *nun*	*n*					NUN, ◌ COMBINING DOT LEFT
shin	*š*					SAMEKH, ◌ COMBINING TWO DOTS RIGHT

9.3 Stem extension

In some texts, a space and a short extension of the baseline is used for indicating suffixes. For such cases the STEM EXTENDER may be used:

tynly lr-r
'tinlag-lar-r'

TAW, YODH, NUN, LESH, GIMEL, SP SPACE,
LESH, RESH, SP SPACE, STEM EXTENDER, RESH

If there is a need to indicate explicitly that the suffix belongs to the preceding word in encoded text, then ZWNJ ZWNJ may be used before the STEM EXTENDER instead of a space.

10 Character Properties

10.1 Core data: `UnicodeData.txt`

```
10F70;OLD UYGHUR LETTER ALEPH;Lo;0;AL;;;;;N;;;;;
10F71;OLD UYGHUR LETTER BETH;Lo;0;AL;;;;;N;;;;;
10F72;OLD UYGHUR LETTER GIMEL;Lo;0;AL;;;;;N;;;;;
```

```
10F73;OLD UYGHUR LETTER WAW;Lo;0;AL;;;;;N;;;;;
10F74;OLD UYGHUR LETTER ZAYIN;Lo;0;AL;;;;;N;;;;;
10F75;OLD UYGHUR LETTER HETH;Lo;0;AL;;;;;N;;;;;
10F76;OLD UYGHUR LETTER YODH;Lo;0;AL;;;;;N;;;;;
10F77;OLD UYGHUR LETTER KAPH;Lo;0;AL;;;;;N;;;;;
10F78;OLD UYGHUR LETTER LAMEDH;Lo;0;AL;;;;;N;;;;;
10F79;OLD UYGHUR LETTER MEM;Lo;0;AL;;;;;N;;;;;
10F7A;OLD UYGHUR LETTER NUN;Lo;0;AL;;;;;N;;;;;
10F7B;OLD UYGHUR LETTER SAMEKH;Lo;0;AL;;;;;N;;;;;
10F7C;OLD UYGHUR LETTER PE;Lo;0;AL;;;;;N;;;;;
10F7D;OLD UYGHUR LETTER SADHE;Lo;0;AL;;;;;N;;;;;
10F7E;OLD UYGHUR LETTER RESH;Lo;0;AL;;;;;N;;;;;
10F7F;OLD UYGHUR LETTER TAW;Lo;0;AL;;;;;N;;;;;
10F80;OLD UYGHUR LETTER LESH;Lo;0;AL;;;;;N;;;;;
10F81;OLD UYGHUR COMBINING DOT RIGHT;Mn;220;NSM;;;;;N;;;;;
10F82;OLD UYGHUR COMBINING TWO DOTS RIGHT;Mn;220;NSM;;;;;N;;;;;
10F83;OLD UYGHUR COMBINING DOT LEFT;Mn;230;NSM;;;;;N;;;;;
10F84;OLD UYGHUR COMBINING TWO DOTS LEFT;Mn;230;NSM;;;;;N;;;;;
10F85;OLD UYGHUR PUNCTUATION BAR;Po;0;AL;;;;;N;;;;;
10F86;OLD UYGHUR PUNCTUATION TWO BARS;Po;0;AL;;;;;N;;;;;
10F87;OLD UYGHUR PUNCTUATION TWO DOTS;Po;0;AL;;;;;N;;;;;
10F88;OLD UYGHUR PUNCTUATION FOUR DOTS;Po;0;AL;;;;;N;;;;;
10F89;OLD UYGHUR PUNCTUATION FIVE DOTS;Po;0;AL;;;;;N;;;;;
10F8A;OLD UYGHUR SECTION MARK;Po;0;AL;;;;;N;;;;;
10F8B;OLD UYGHUR STEM EXTENDER;Po;0;AL;;;;;N;;;;;
```

10.2 Linebreak data: `LineBreak.txt`

```
10F70..10F80;AL # Lo [17] OLD UYGHUR LETTER ALEPH..OLD UYGHUR LETTER LESH
10F81..10F84;CM # Mn  [4] OLD UYGHUR COMBINING DOT RIGHT..
                          OLD UYGHUR COMBINING TWO DOTS LEFT
10F85..10F8A;AL # Po  [6] OLD UYGHUR PUNCTUATION BAR..OLD UYGHUR SECTION MARK
10F8B;AL        # Po      OLD UYGHUR STEM EXTENDER
```

10.3 Property list: `PropList.txt`

```
10F8A           ; Extender # Po       OLD UYGHUR STEM EXTENDER
```

10.4 Shaping properties: `ArabicShaping.txt`

```
10F70; OLD UYGHUR ALEPH; D; No_Joining_Group
10F71; OLD UYGHUR BETH; D; No_Joining_Group
10F72; OLD UYGHUR GIMEL; D; No_Joining_Group
10F73; OLD UYGHUR WAW; D; No_Joining_Group
10F74; OLD UYGHUR ZAYIN; D; No_Joining_Group
10F75; OLD UYGHUR HETH; D; No_Joining_Group
10F76; OLD UYGHUR YODH; D; No_Joining_Group
10F77; OLD UYGHUR KAPH; D; No_Joining_Group
10F78; OLD UYGHUR LAMEDH; D; No_Joining_Group
10F79; OLD UYGHUR MEM; D; No_Joining_Group
10F7A; OLD UYGHUR NUN; D; No_Joining_Group
10F7B; OLD UYGHUR SAMEKH; D; No_Joining_Group
10F7C; OLD UYGHUR PE; D; No_Joining_Group
10F7D; OLD UYGHUR SADHE; D; No_Joining_Group
10F7E; OLD UYGHUR RESH; D; No_Joining_Group
```

```
10F7F; OLD UYGHUR TAW; D; No_Joining_Group
10F80; OLD UYGHUR LESH; D; No_Joining_Group
```

11 References

Clark, Larry. 2010. "The Turkic script and the *Kutadgu Bilig*". *Turcology in Mainz*, Turcologica, Band 82, Hendrik E Boeschoten and Julian Rentzsch (ed.), pp. 89–106. Wiesbaden: Harrassowitz Verlag.

Clauson, Gerard. 1962. *Studies in Turkic and Mongolic Linguistics*. London: Royal Asiatic Society of Great Britain and Ireland.

Coulmas, Florian. 1996. *The Blackwell Encyclopedia of Writing Systems*. Oxford: Blackwell Publishers.

Erdal, Marcel. 1984. "The Turkish Yarkand Documents". *Bulletin of the School of Oriental and African Studies, University of London*, vol. 47, no. 2, pp. 260–301.

Kara, György. 1996. "Aramaic Scripts for Altaic Languages". *The World's Writing Systems*, Peter T. Daniels and William Bright (ed.), pp. 536–558. New York and Oxford: Oxford University Press.

Müller, Friedrich Wilhelm Karl. 1910. *Uigurica*, vol. II. Berlin: Verlag der Königlichen Akademie der Wissenschaften.

Nadeliaev, V. M.; Nasilov, D. M.; Scherbak, A. M.; Tenishev, E. R. 1969. *Drevnetiurkskii slovar*. Akademiia nauk SSSR. Institut iazykoznaniia. Leningrad: Izd-vo "Nauka" Leningradskoe otd-nie.

Olmez, Mehmet. 2016. "Compared transcription system for Old Uyghur Alphabet". Lecture at École des Hautes Études en Sciences Sociales (EHESS), Paris, May 2016. http://www.academia.edu/24939281/Lectures_at_EHESS_1_Compared_transcription_system_for_Old_Uighur_Alphabet

Osman, Omarjan. 2013. "Proposal to Encode the Uyghur Script in ISO/IEC 10646". L2/13-071. http://www.unicode.org/L2/L2013/13071-uyghur.pdf

Pandey, Anshuman. 2016. "Revised proposal to encode the Sogdian script in Unicode". L2/16-371R. http://www.unicode.org/L2/L2016/16371r2-sogdian.pdf

Radloff, W. 1910. *Ṭišastvustik*: ein in Türkischer Sprache bearbeitetes Buddhistisches Sūtra. I. Transcription and Übersetzung; II. Bemerkungen zu den Brāhmīglossen des Ṭišastvustik-Manuscripts (Mus. As. Kr. VII) von Baron A. von Staël-Holstein. Bibliotheca Buddhica, XII. St.-Pétersbourg.

Radlov, Vasiliĭ Vasil′evich; Malov, S. Efimovich. 1913. *Suvarṇaprabhāsa*. (Sutra zolotogo bleska); tekst ujgurskogo redakcii. Bibliotheca Buddhica, XVII. Sanktpeterburg: Imper. Akad. Nauk.

Sims-Williams, Nicholas. 1981. "The Sogdian sound-system and the origins of the Uyghur script". *Journal Asiatique*, pp. 347–360.

von Gabain, Annemarie. 1950. *Alttürkische Grammatik*. Mit Bibliographie, Lesestücken und Wörterverzeichnis, auch Neutürkisch. Mir vier Schrifttafeln under sieben Schriftproben. Porta linguarum orientalium. no. 23. 1. verbesserte Auflage. Leipzig: Otto Harrassowitz.

West, Andrew. 2006. “Phags-pa Script : Old Uyghur Script”.
http://www.babelstone.co.uk/Phags-pa/Uighur.html

———. 2011a. “Khitan Miscellanea 1: Oh, How the Gods Mock Us!”.
http://babelstone.blogspot.com/2011/10/khitan-miscellanea-1.html

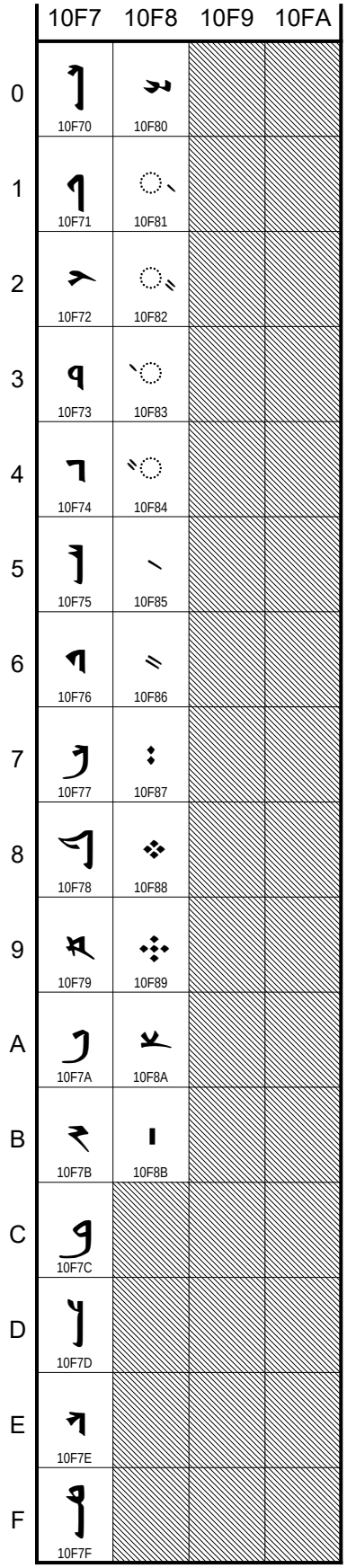
———. 2011b. “Phags-pa Uyghur Seals”.
http://babelstone.blogspot.com/2011/11/phags-pa-uyghur-seals.html

12 Acknowledgments

I would like to thank Andrew West for providing reproductions and transcriptions of the Old Uyghur inscriptions on the walls of the Cloud Platform at Juyong Guan (fig. 30).

This project has been made possible in part by funding from the Adopt-A-Character program of the Unicode Consortium, and has been supervised by Deborah Anderson and Rick McGowan.

It was also made possible in part by a grant from the U.S. National Endowment for the Humanities, which funded the Universal Scripts Project (part of the Script Encoding Initiative at UC Berkeley). Any views, findings, conclusions or recommendations expressed in this publication do not necessarily reflect those of the National Endowment of the Humanities.



Letters

10F70 OLD UYGHUR LETTER ALEPH
10F71 OLD UYGHUR LETTER BETH
10F72 OLD UYGHUR LETTER GIMEL
10F73 OLD UYGHUR LETTER WAW
10F74 OLD UYGHUR LETTER ZAYIN
10F75 OLD UYGHUR LETTER HETH
10F76 OLD UYGHUR LETTER YODH
10F77 OLD UYGHUR LETTER KAPH
10F78 OLD UYGHUR LETTER LAMEDH
10F79 OLD UYGHUR LETTER MEM
10F7A OLD UYGHUR LETTER NUN
10F7B OLD UYGHUR LETTER SAMEKH
10F7C OLD UYGHUR LETTER PE
10F7D OLD UYGHUR LETTER SADHE
10F7E OLD UYGHUR LETTER RESH
10F7F OLD UYGHUR LETTER TAW
10F80 OLD UYGHUR LETTER LESH
• hooked r

Combining signs

10F81 OLD UYGHUR COMBINING DOT RIGHT
10F82 OLD UYGHUR COMBINING TWO DOTS RIGHT
10F83 OLD UYGHUR COMBINING TWO DOTS LEFT
10F84 OLD UYGHUR COMBINING DOT LEFT

Punctuation

10F85 OLD UYGHUR PUNCTUATION BAR
10F86 OLD UYGHUR PUNCTUATION TWO BARS
10F87 OLD UYGHUR PUNCTUATION TWO DOTS
10F88 OLD UYGHUR PUNCTUATION FOUR DOTS
10F89 OLD UYGHUR PUNCTUATION FIVE DOTS
10F8A OLD UYGHUR SECTION MARK

Stem extender

10F8B OLD UYGHUR STEM EXTENDER

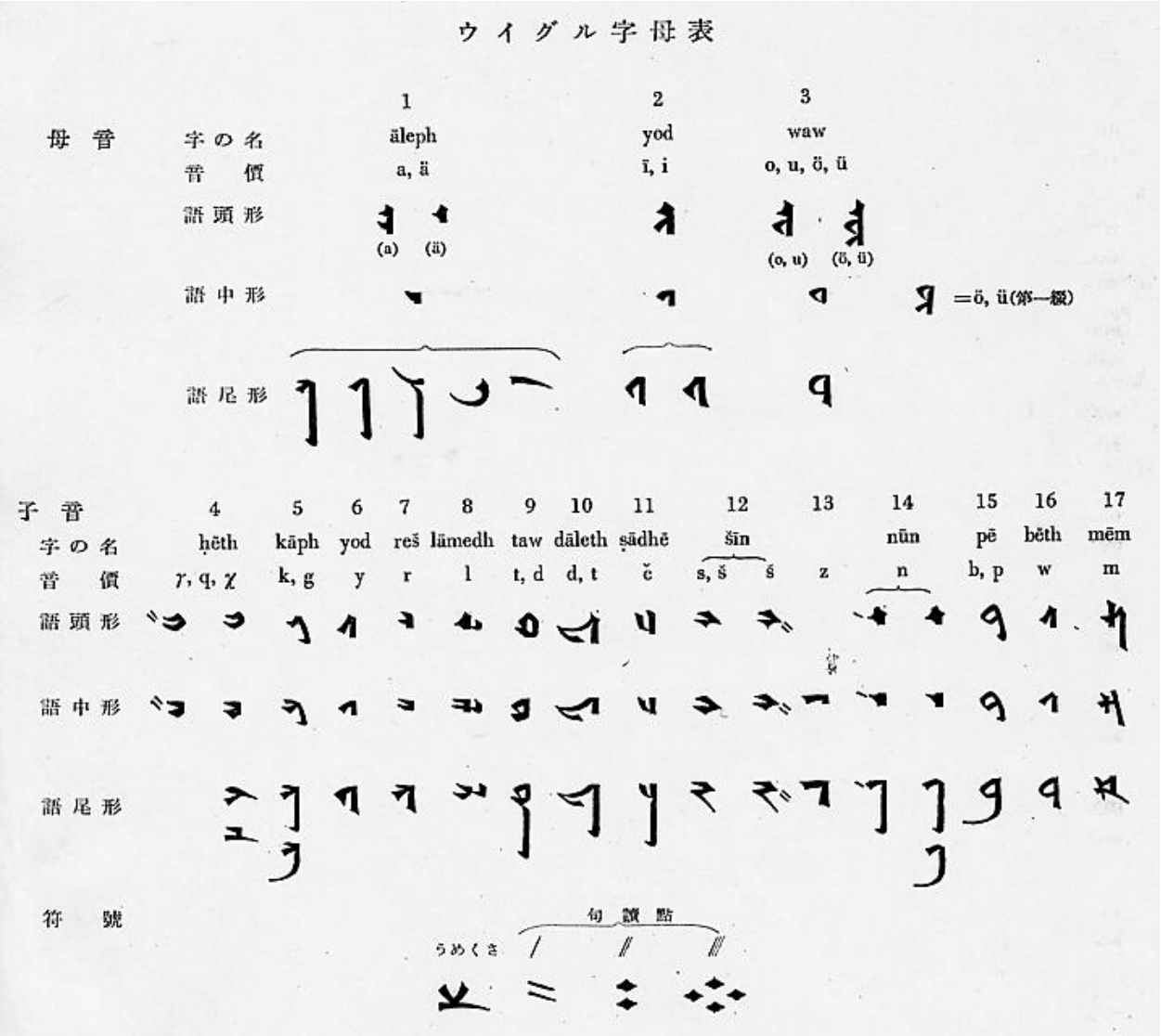
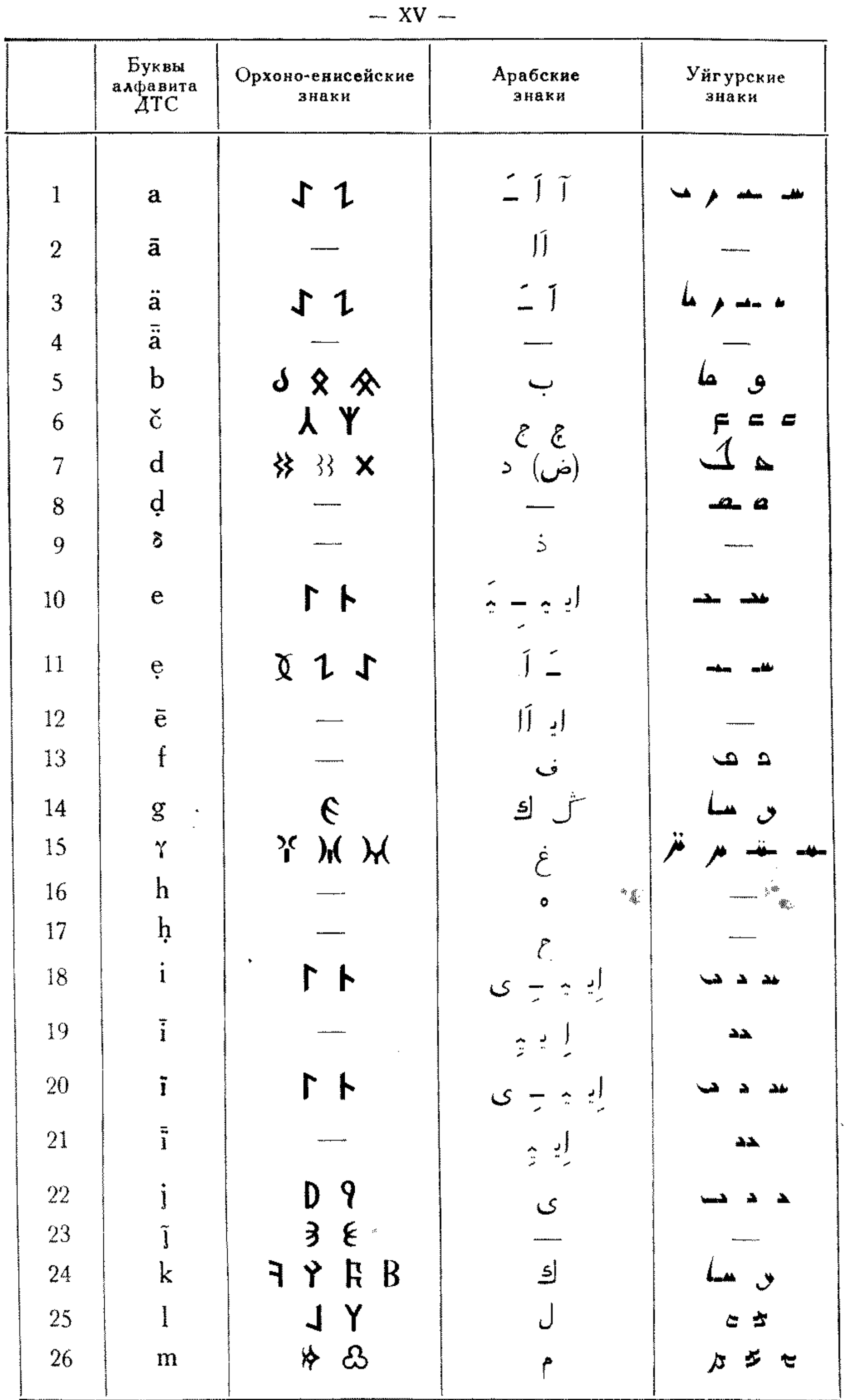


Figure 1: Table of Old Uyghur characters used in the Uyghur inscription in the multi-script Yuan dynasty inscriptions at Juyong Guan 居庸關 pass at the Great Wall northwest of Beijing (from Chü-Yung-Kuan 居庸關, "The Buddhist Arch of the Fourteenth Century A.D. at the Pass of the Great Wall Northwest of Peking", vol. 1, p. 165; reproduced from West 2006). See photograph containing an excerpt of the inscription in fig. 30.



— XV —

	Буквы алфавита ДТС	Орхоно-енисейские знаки	Арабские знаки	Уйгурские знаки
1	a	[illegible]	آ أ ـَ	[illegible]
2	ā	—	آا	—
3	ä	[illegible]	آ ـَ	[illegible]
4	ǟ	—	—	—
5	b	[illegible]	ب	[illegible]
6	č	[illegible]	ج چ	[illegible]
7	d	[illegible]	د (ض)	[illegible]
8	ḍ	—	—	[illegible]
9	δ	—	ذ	—
10	e	[illegible]	اِ ـِ ـ ـِ	[illegible]
11	ẹ	[illegible]	اَ ـَ	[illegible]
12	ē	—	اِي آا	—
13	f	—	ف	[illegible]
14	g	[illegible]	ك گ	[illegible]
15	γ	[illegible]	غ	[illegible]
16	h	—	ه	—
17	ḥ	—	ح	—
18	i	[illegible]	اِي ـيـ ـي ى	[illegible]
19	ī	—	اِ يـ ـيـ	[illegible]
20	ï	[illegible]	اِي ـيـ ـي ى	[illegible]
21	ï̄	—	اِيـ ـيـ	[illegible]
22	j	[illegible]	ى	[illegible]
23	ǰ	[illegible]	—	—
24	k	[illegible]	ك	[illegible]
25	l	[illegible]	ل	[illegible]
26	m	[illegible]	م	[illegible]

Figure 2: Representation of Old Turkic sounds in the Orkhon, Arabic, and Old Uyghur scripts (from Nadeliaev, et al. 1969: xv). Continued in fig. 3.

— XVI —

	Буквы алфавита ДТС	Орхоно-енисейские знаки	Арабские знаки	Уйгурские знаки
27	n	𐰠 𐰢 𐰴	ن	
28	ŋ	𐰚 𐰲	ڭ نك	
29	o	𐰺	اُو ـُـ و	
30	ō	—	—	
31	ö	𐰆 𐰸	اُو ـُـ و	
32	ȫ	—	—	
33	p	𐰃	پ ب	
34	q	𐰇 𐰋 𐰓	ق	
35	r	𐰼 𐰾	ر	
36	s	𐰽 𐰉	س ص	
37	ṣ	𐰿 𐰽	—	
38	š	𐰿 𐰽 𐰍	ش	
39	ṣ̌	𐰽 𐰉	—	—
40	t	𐰔 𐰖 𐰘 𐰘	ة ط ت	
41	ṭ	—	—	
42	ⱦ	—	ت	—
43	u	𐰺	اُو ـُـ و	
44	ū	—	—	
45	ü	𐰆 𐰸	اُو ـُـ و	
46	ǖ	—	—	
47	v	—	ڤ ۋ و ف	
48	w	см. 47	см. 47	см. 47
49	ɣ	—	غ	
50	z	𐰏 𐰢 𐰽	ض ز ظ	
51	ẓ	—	—	
52	ž	—	ژ	
53	ẓ̌	—	—	
54	ǯ	—	ع	
55	ʾ	—	ء	—
56	ʿ	—	ع	—

Figure 3: Representation of Old Turkic sounds in the Orkhon, Arabic, and Old Uyghur scripts (from Nadeliaev, et al. 1969: xvi). Continued from fig. 2.

TABLE 49.2: *Uyghur Script*[a]

Name[b]	*Uyghur*	*Initial*	*Medial*	*Final*	*Separate*	*Ligatures*	*Uyghur*
ʾaleph	e/vowel initial						ka/e
	a/e						pa/e
beth	w/v						
gimel	γ						
waw	o/u						
waw+yodh	ö/ü						
	o/u/ö/ü[c]						ko/u/ö/ü
							po/uö/ü
zain	z		-				
marked z	ž		-				
heth	x						
2-dotted	q						
yodh	y						ki/ï
							pi/ï
kaph	k/g						
lamedh	d/δ						
mem	m						ml
nun	n						
pe	b/p						
tsadi	č						
resh	r						
shin	s						
marked s	š						
tau	t						
hooked r	l						

a. Diacritics are often omitted. Some Uyghur alphabets have shin for samekh before pe; marked *z*, final *m*, and final *q* are added after hooked resh.
b. Hebrew name for the ancestral Aramaic letter.
c. In syllables other than the first.

Figure 4: Table showing letters of the Old Uyghur script (from Kara 1996: 540). See table of Mongolian letters from the same source in fig. 40.

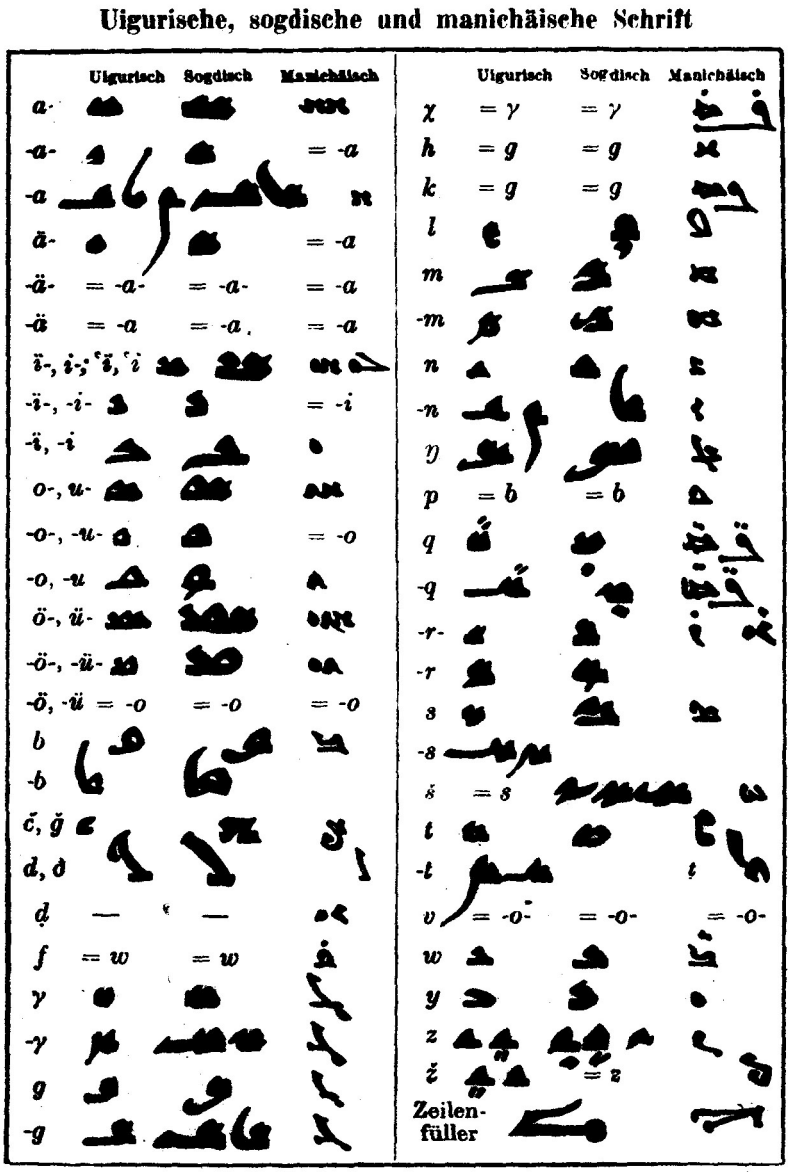


Figure 5: Comparison of Old Uyghur, Sogdian, and Manichaean letters (from von Gabain 1950: 17). For clearer examples of Old Uyghur letterforms referenced by von Gabain see the three Old Uyghur manuscripts, two in the formal script and the third in the cursive script, illustrated and transcribed in her work, reproduced here in fig. 33–37.

Uighur writing

Transliteration	10th C.	10th C.	1072	13th – 14th C.
1 ’				
2 β				
3 γ				
4 w				
5 z				
6 x				
7 y				
8 k				
9 d(δ)				
10 m				
11 n				
12 s				
13 p				
14 č				
15 r				
16 š				
17 t				
18 l				
19 ž				
20 -m				
21 q̈				

Table 2 Various forms of the Old Uighur alphabet from texts dating between the fourteenth and the tenth centuries BCE
Source: *adapted from Zieme 1991*

Figure 6: Comparison of Old Uyghur letterforms (from Coulmas 1996: 526).

Compared transcription system for Old Uighur Alphabet

	Berliner Transkription system	Turkey	transcription at *Uigurisches Wörterbuch*	transliteration at *Uigurisches Wörterbuch*
	a, ą	a, ą	a	’’ / ’
	b	b	b	P
	č	ç	č	Č
	d, ṭ	d, ṭ	d, ḍ	D, T
	ä, ’ä	e, ’e	ä	’
	[e] i	ė / i	e	Y / ’Y
	g	g	g	K
	γ / γ́	g / ġ	g	Q, Ǫ̈, Ǫ̇
	h / χ, x, ẍ	h / ḫ, ḥ	h	H / X
	ï	ı	ı	Y, Y
	i	i	i	Y, ’Y
	ǰ̧, ž	j	ǰ̧, ž	J̧̌, Ž, Ẓ
	k	k	k	K
	[k] q, q̈, q̇	k / ķ	k	K / Q, Ǫ̈, Ǫ̇
	l	l	l	L
	m	m	m	M
	n, ṅ	n, ṅ	n	N, Ṅ
	ng, ñ, ŋ	n͡g, ng, ñ	ŋ	NK
	o	o	o	W / ’W
	ö, o̧	ö, o̧	ö	W / WY / ’WY
	p	p	p	P
	r	r	r	R
	s, ẓ	s, ẓ	s, ṣ	S, Z
	š	ş	š	Ş, Ş
	t, ḍ	t, ḍ	t, ṭ	T, D
	u	u	u	W / ’W
	ü, u̧	ü, u̧	ü	W / WY / ’WY
[]	v	v	v	V
	y	y	y	Y
	z, ṣ	z, ṣ	z, ẓ	Z, S

Figure 7: Comparison of transliteration schemes for Old Uyghur (from Olmez 2016).

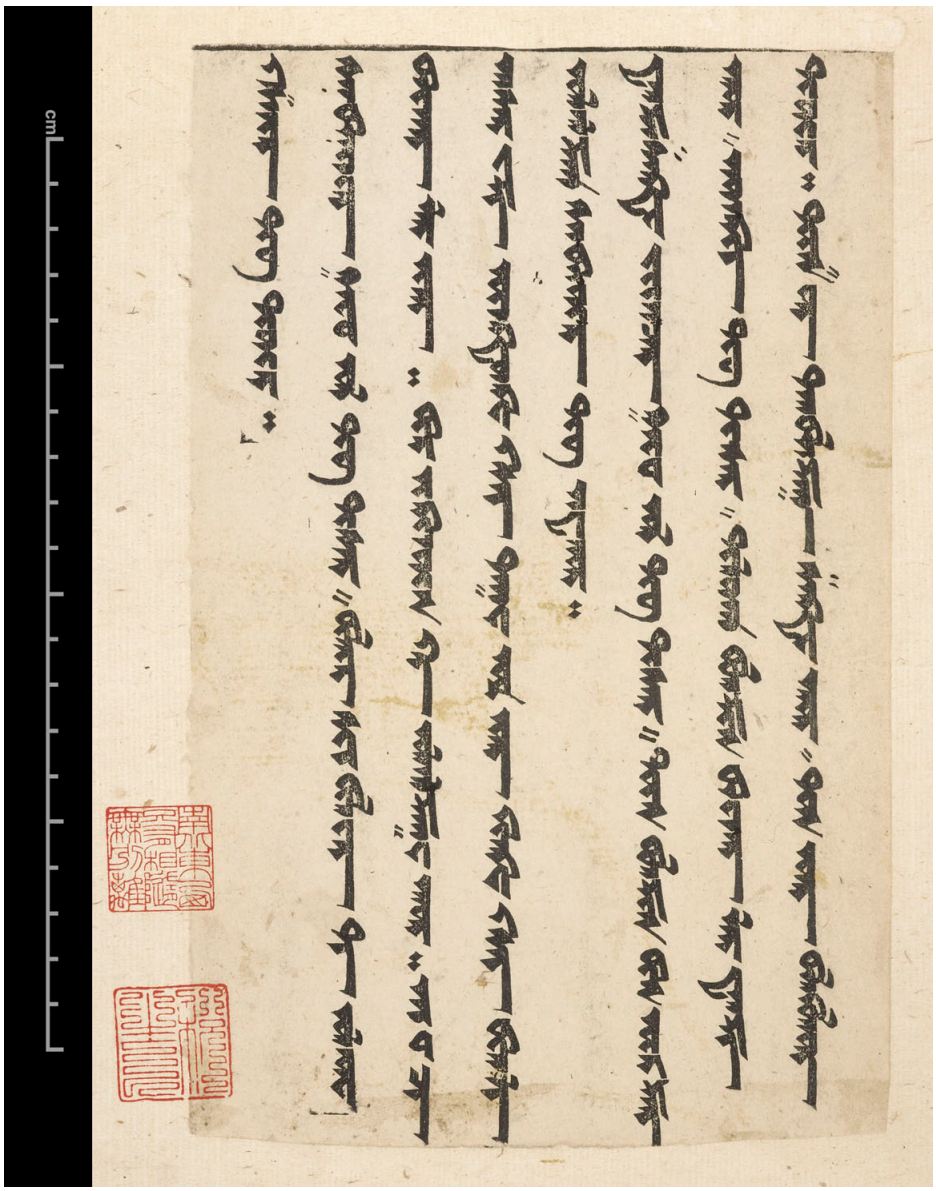


Figure 8: Princeton East Asian Library, PEALD 6a, recto. Block print.

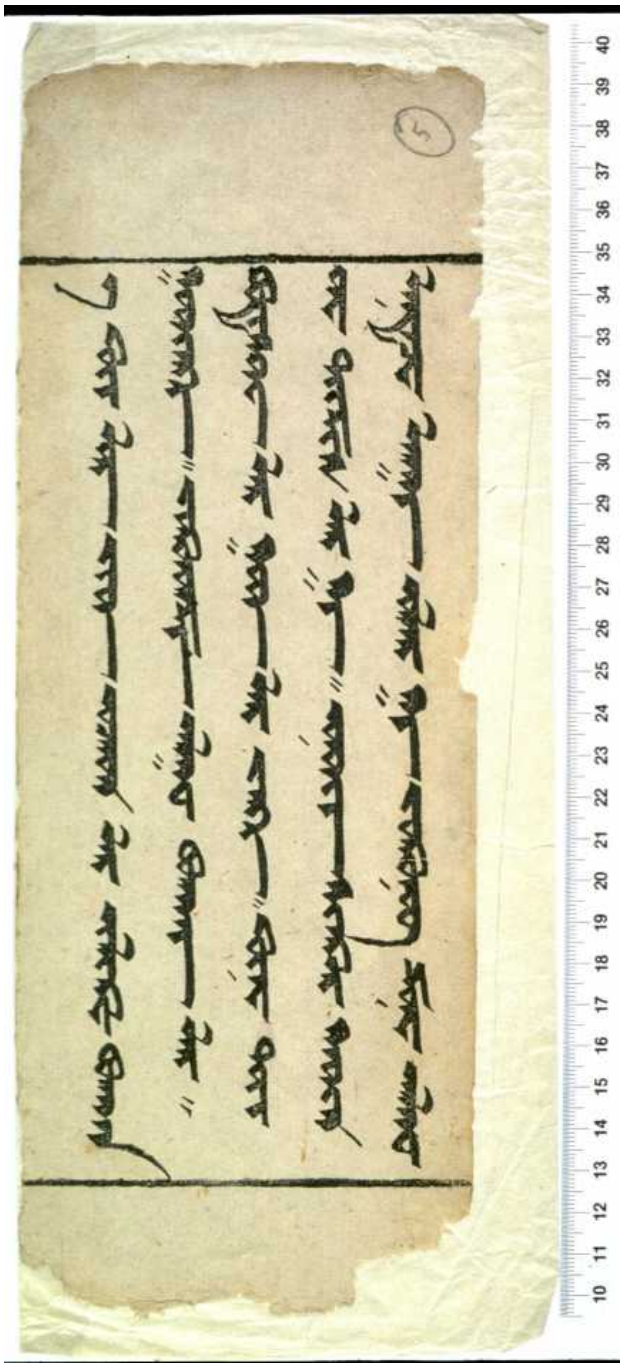


Figure 9: BBAW, U 0387, recto. Block print.

Figure 10: BBAW, U 4960, folio 1, recto. Block print. Seal in Han characters.

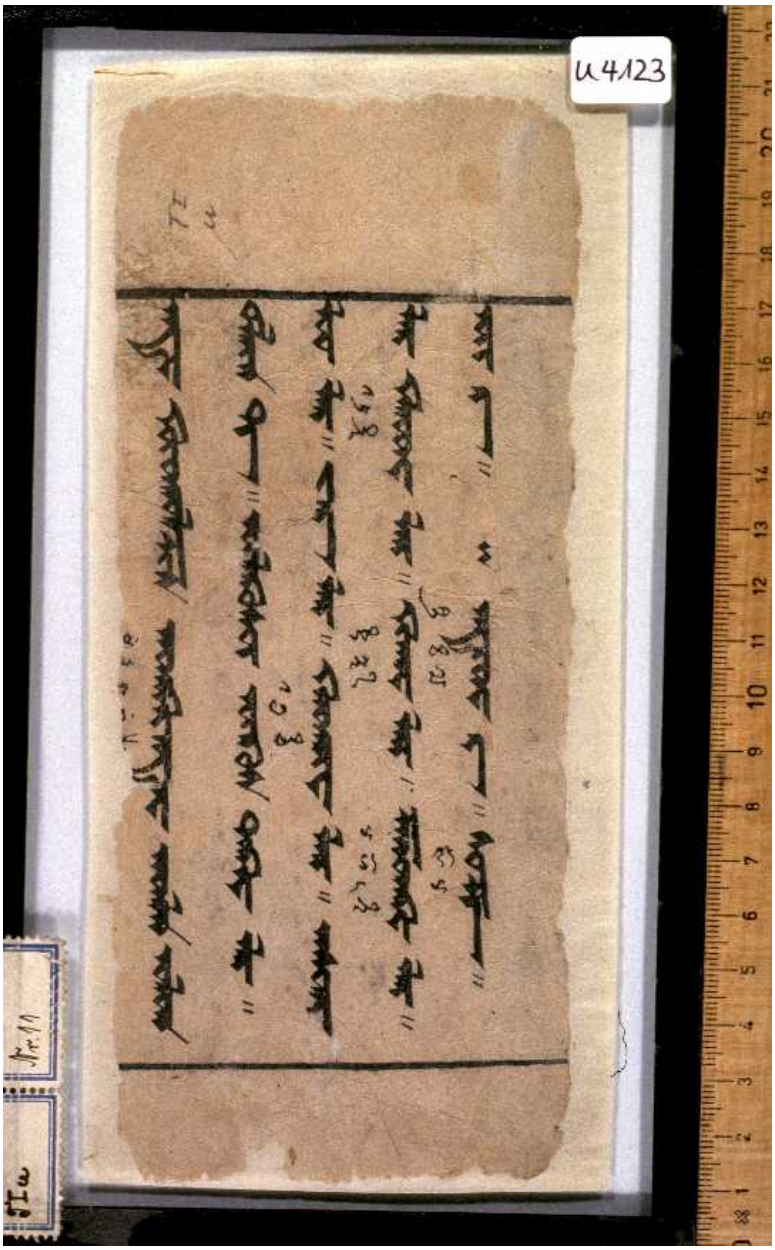


Figure 11: BBAW, U 4123. Block print.

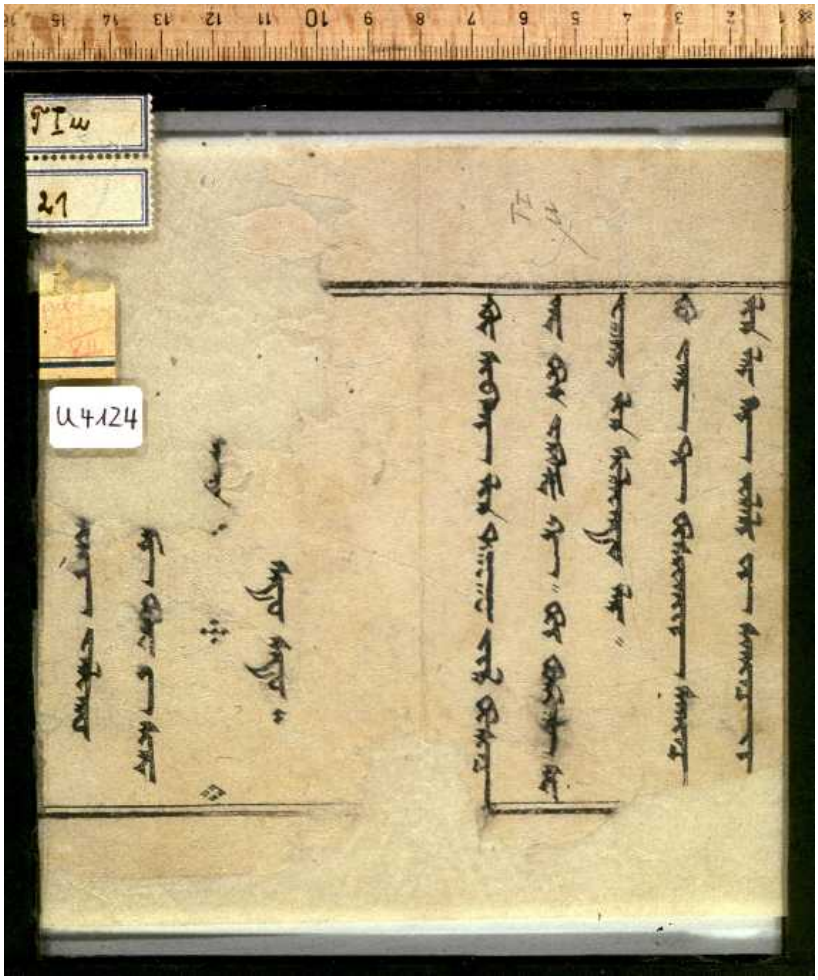


Figure 12: BBAW, U 4124. Block print.

Figure 13: BBAW, U 0343, folio 1, recto. Block print.

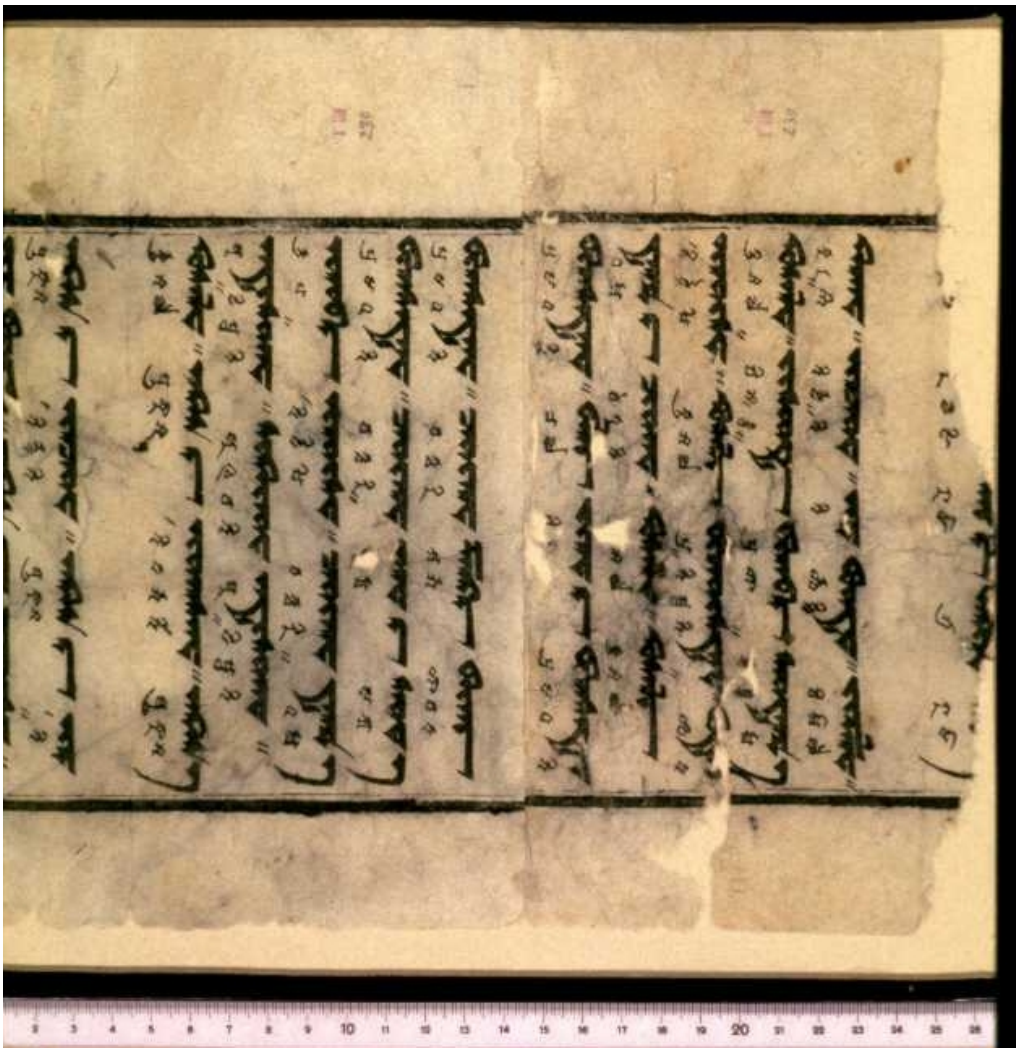


Figure 14: BBAW, Mainz 0801, middle portion. Block print. Annotations in Central Asian Brahmi.

Figure 15: BBAW, U 7008. Block print.

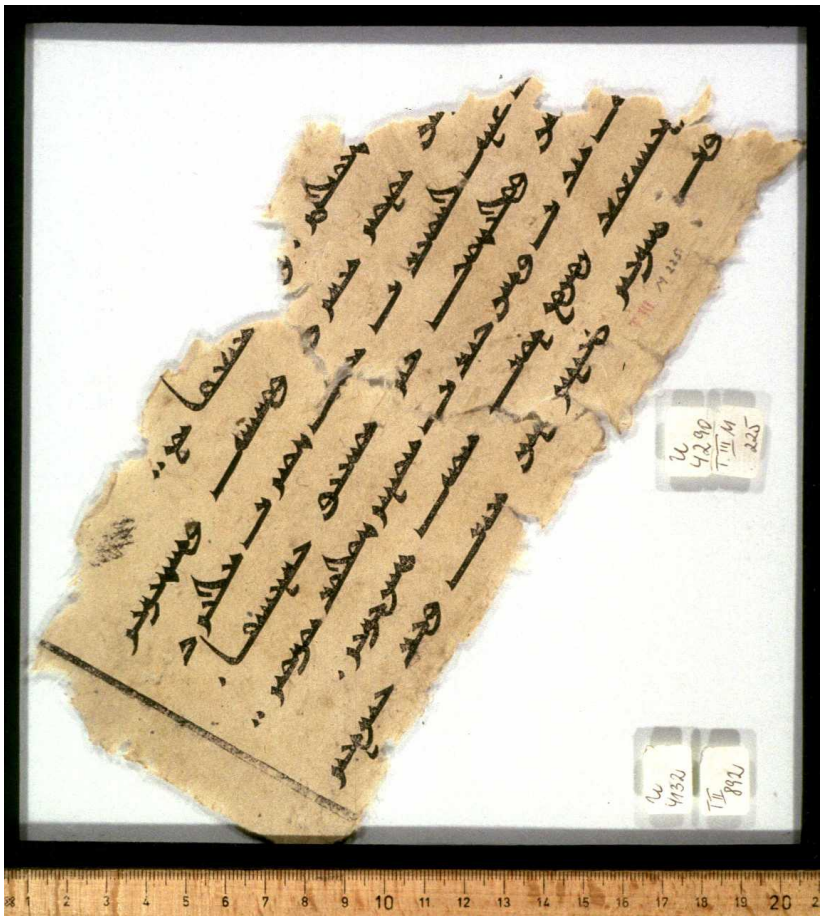


Figure 16: BBAW, U 4132 + 4290, folio 1. Block print.

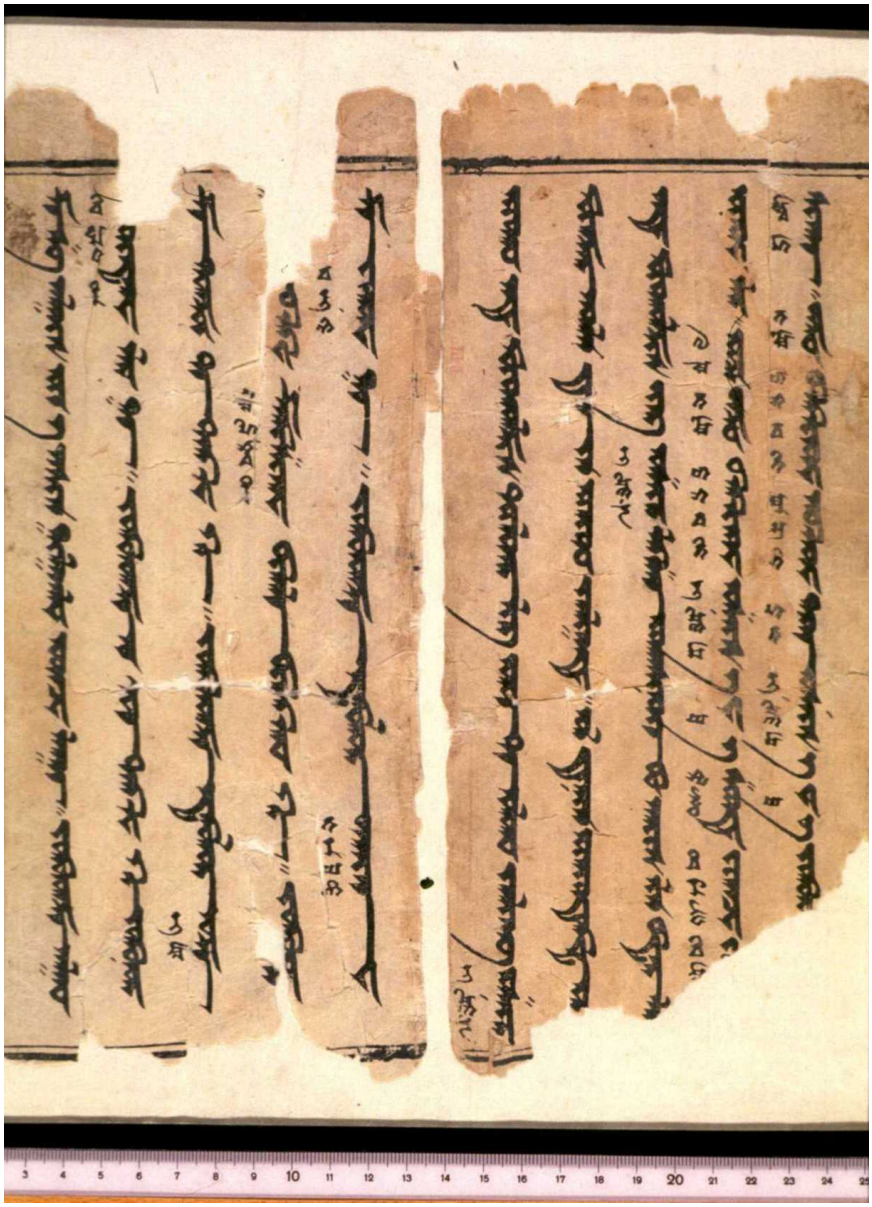


Figure 17: BBAW, Mainz 0764, middle. Formal script. Annotations in Central Asian Brahmi.

Figure 18: BBAW, U 3832, folio 1. Formal script.

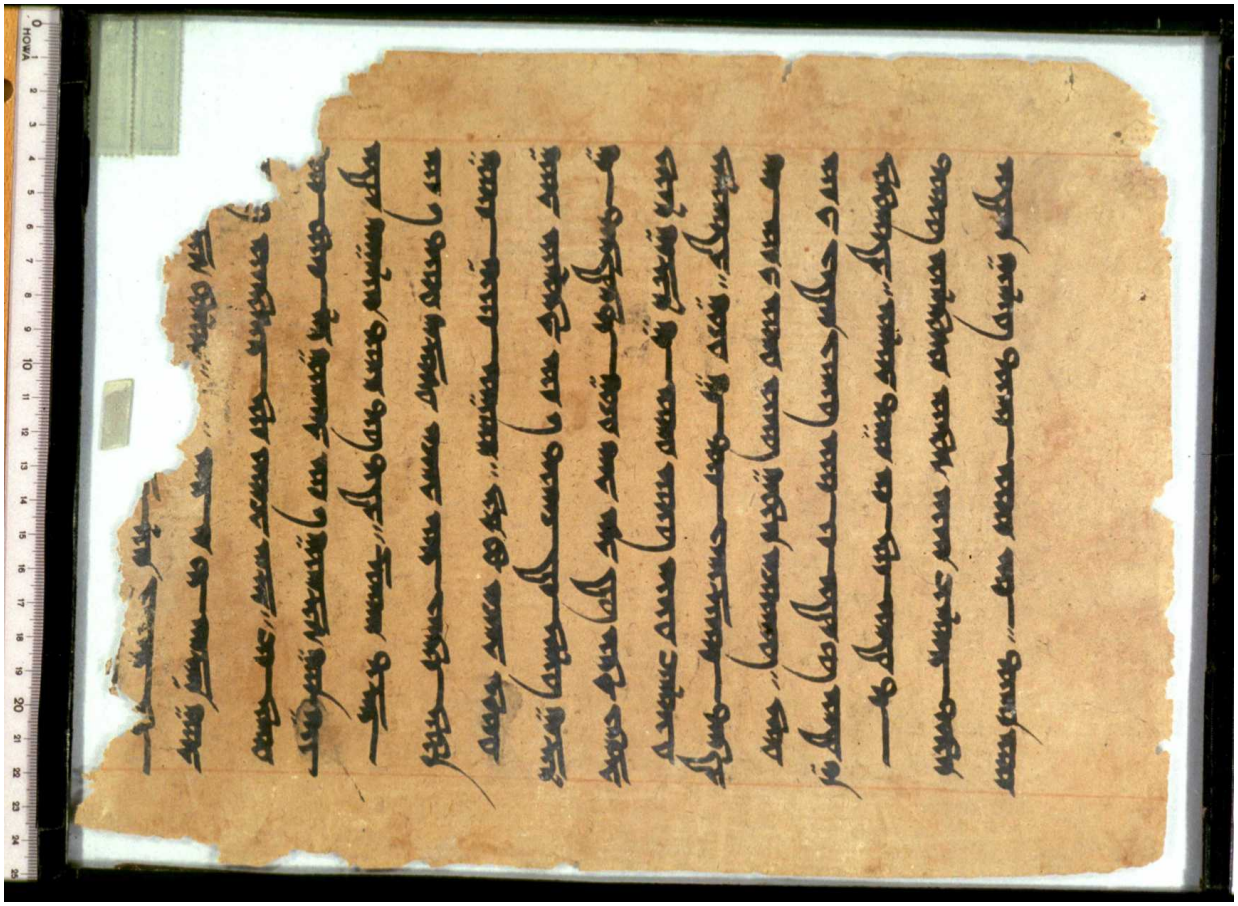


Figure 19: BBAW, Mainz 0841, folio 2.

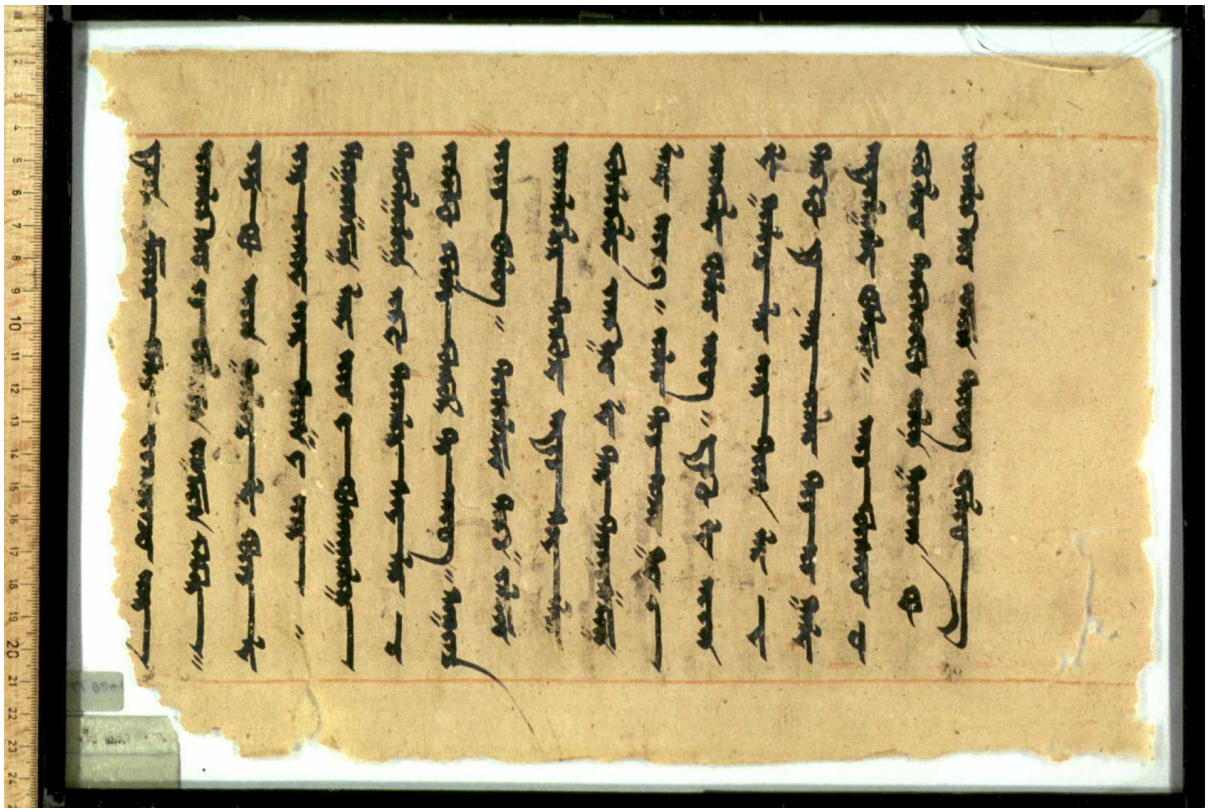


Figure 20: BBAW, U 0924, folio 2.

Figure 21: PEALD 6r, recto.

Figure 22: BBAW, U 0040, recto. Note the inventory of Old Uyghur letters at the bottom of the folio. The letters 𐽀 *aleph* through 𐽔 *taw* are legible. The four letters following *taw* are unclear.

Figure 23: BBAW, U 0320, folio 1.

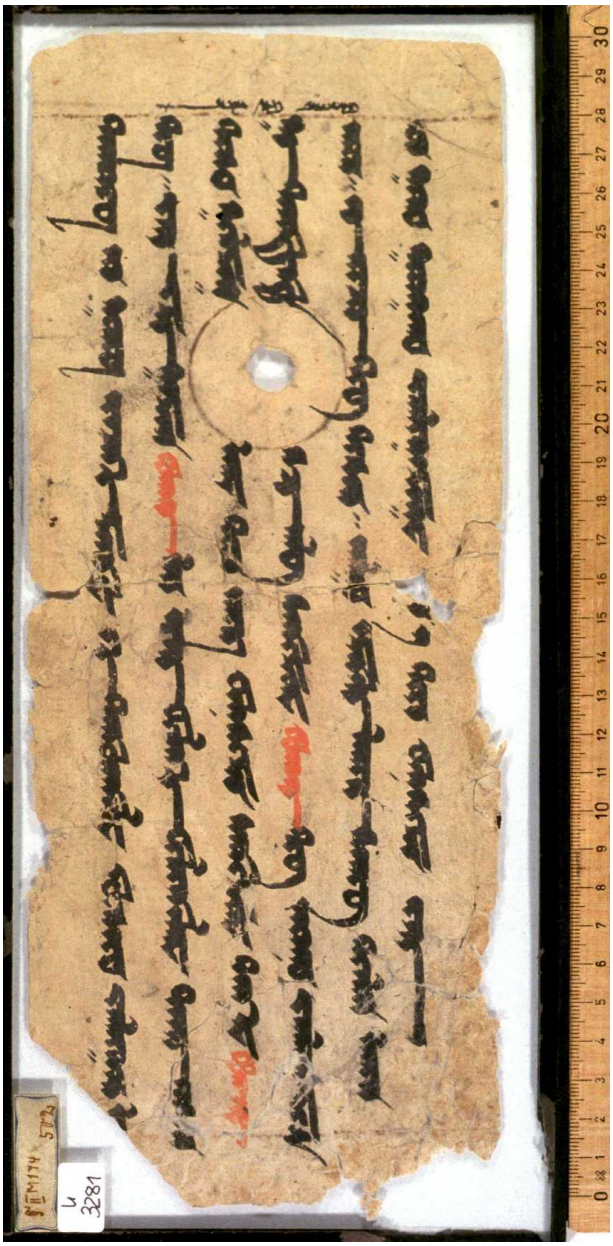


Figure 24: BBAW, U 3281, folio 1.

Figure 25: BBAW, Mainz 0843, folio 2.

Figure 26: BBAW, U 7123, recto.

Figure 27: BL / BBAW, Ch 5555, recto. *Ekottaragamasutra* / 增一阿含經 *Zeng yi e han jing*.

Figure 28: BBAW, U 0456, folio 1.

Figure 29: BBAW, Ch/U 7730, verso.

Figure 30: Detail of the Old Uyghur text of the multi-script Yuan dynasty Buddhist inscriptions on the west wall of the Cloud Platform at Juyong Guan 居庸關 pass at the Great Wall northwest of Beijing. Photograph by Andrew West, 2011.

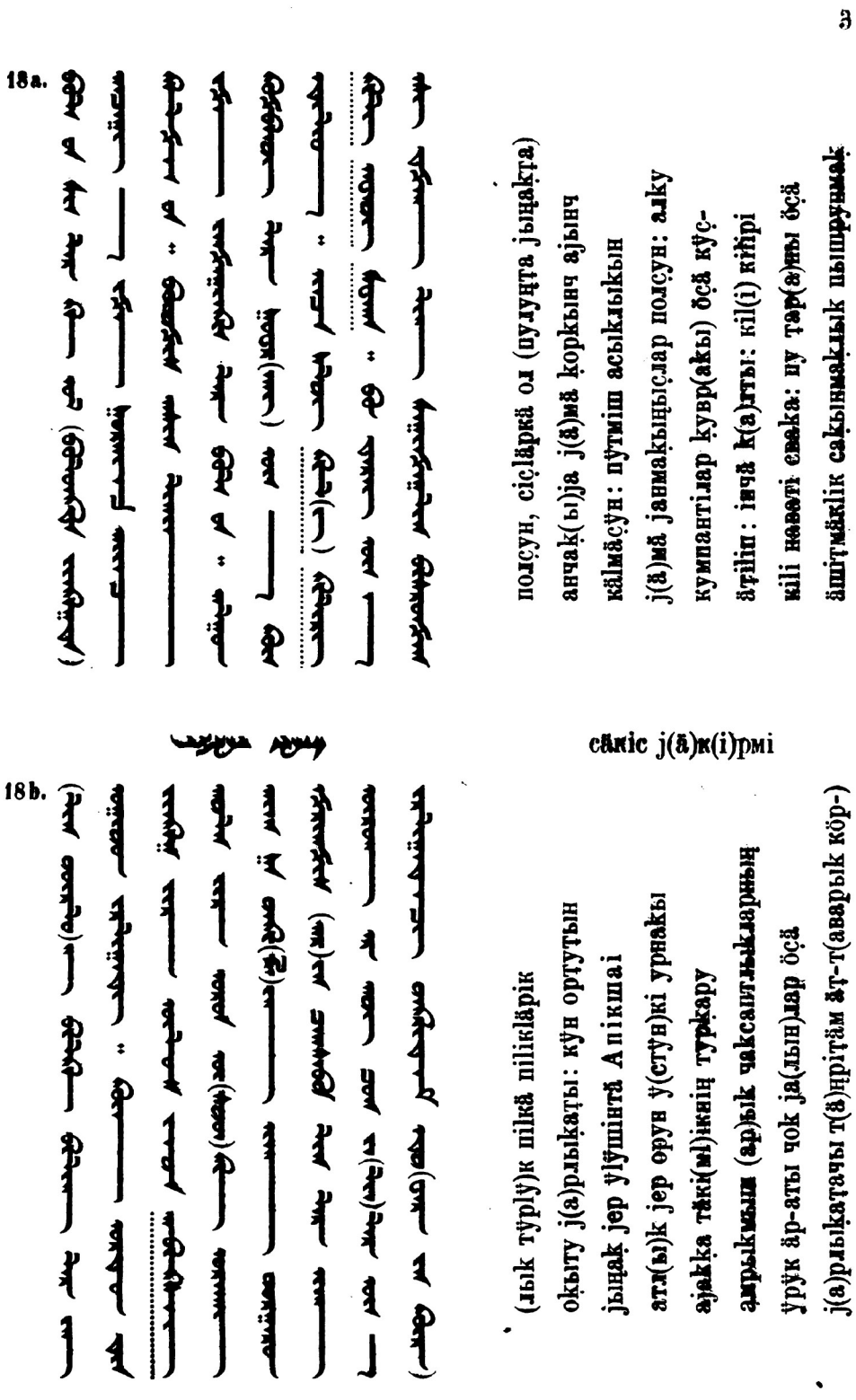


Figure 31: Printed edition of *Ṭišastvustik* in the Old Uyghur script (from Radloff 1910: 3).

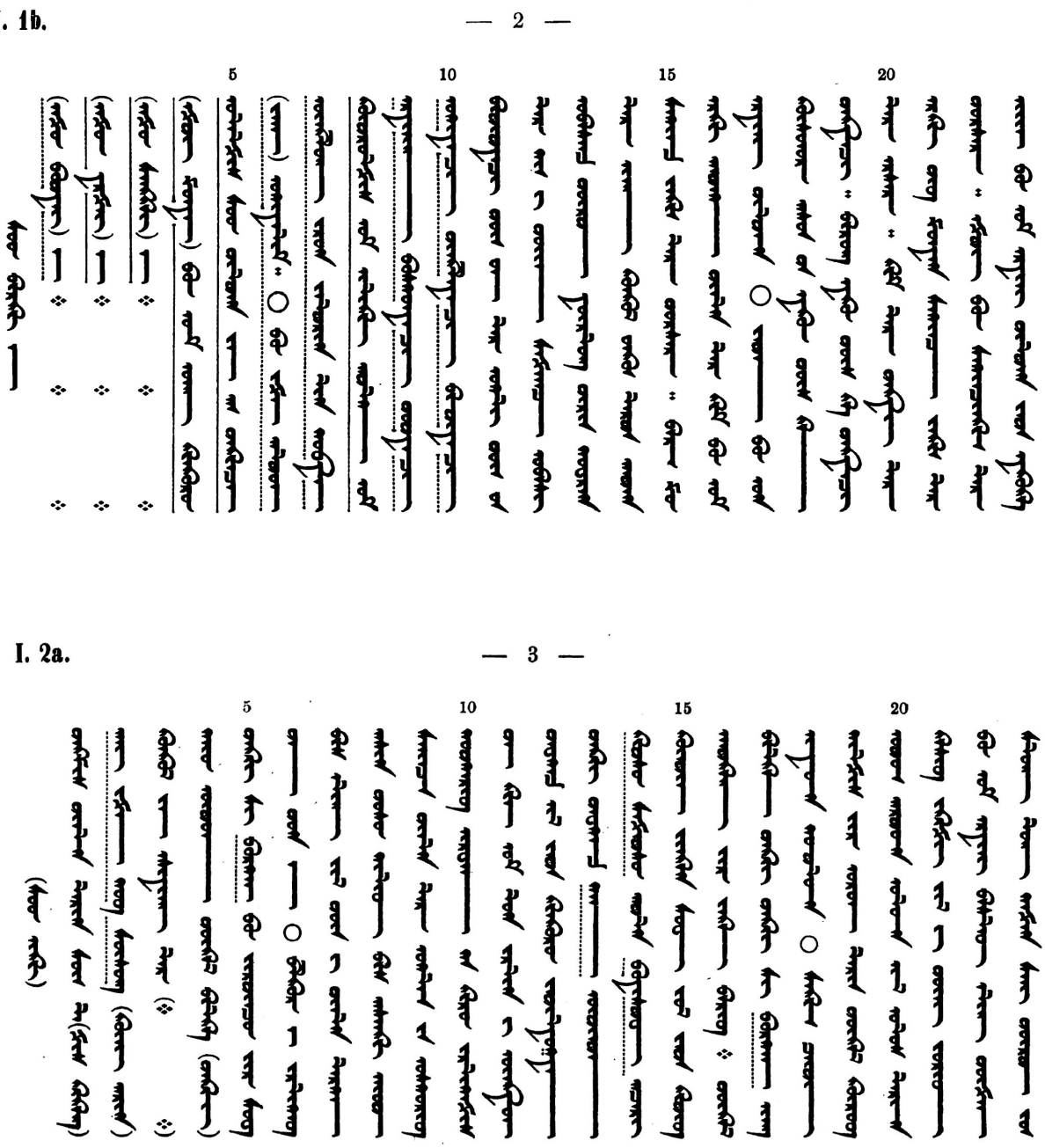
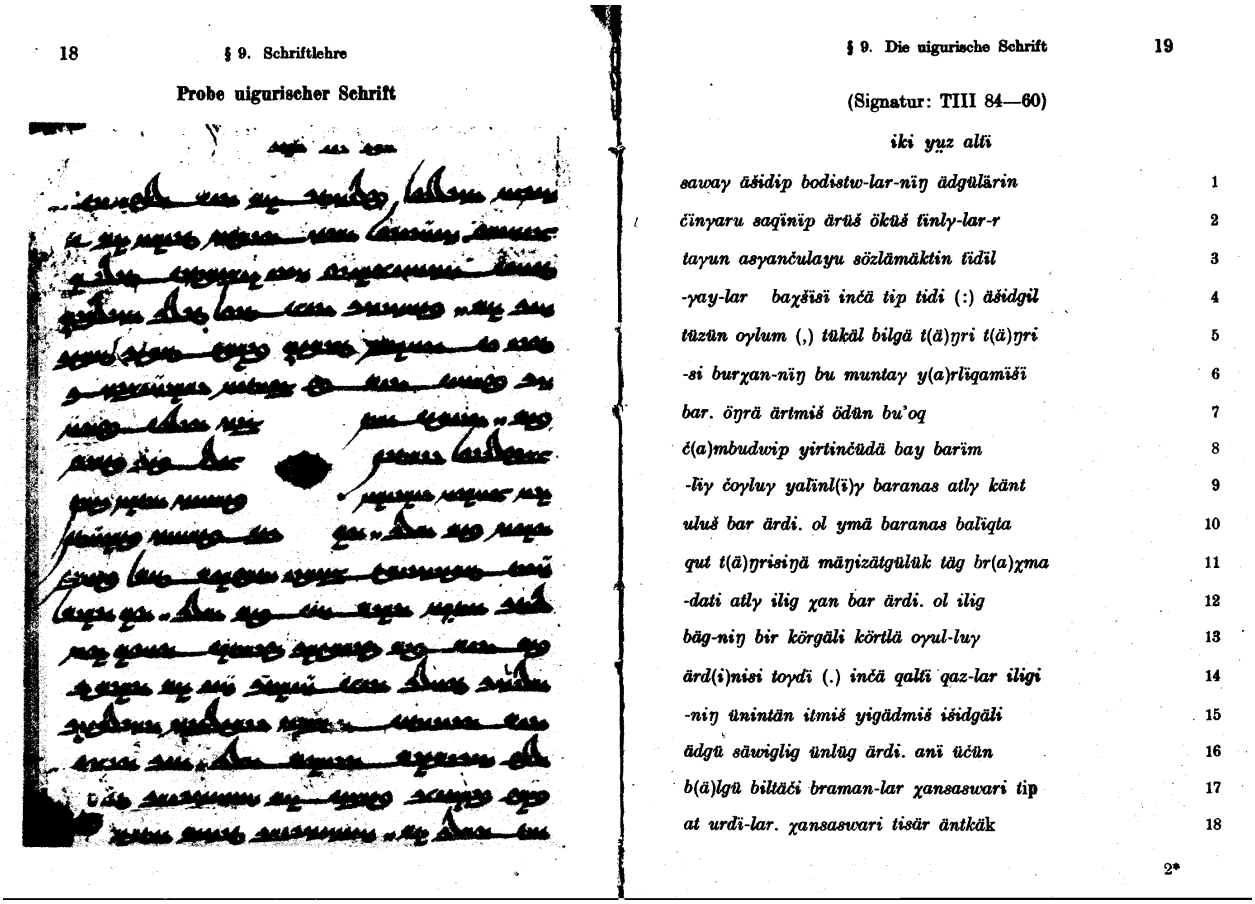


Figure 32: Printed edition of *Suvarṇaprabhāsa*, a Mahayana Buddhist text, in the Old Uyghur script (from Radlov and Malov 1913: 2–3).



18 § 9. Schriftlehre

Probe uigurischer Schrift

§ 9. Die uigurische Schrift 19

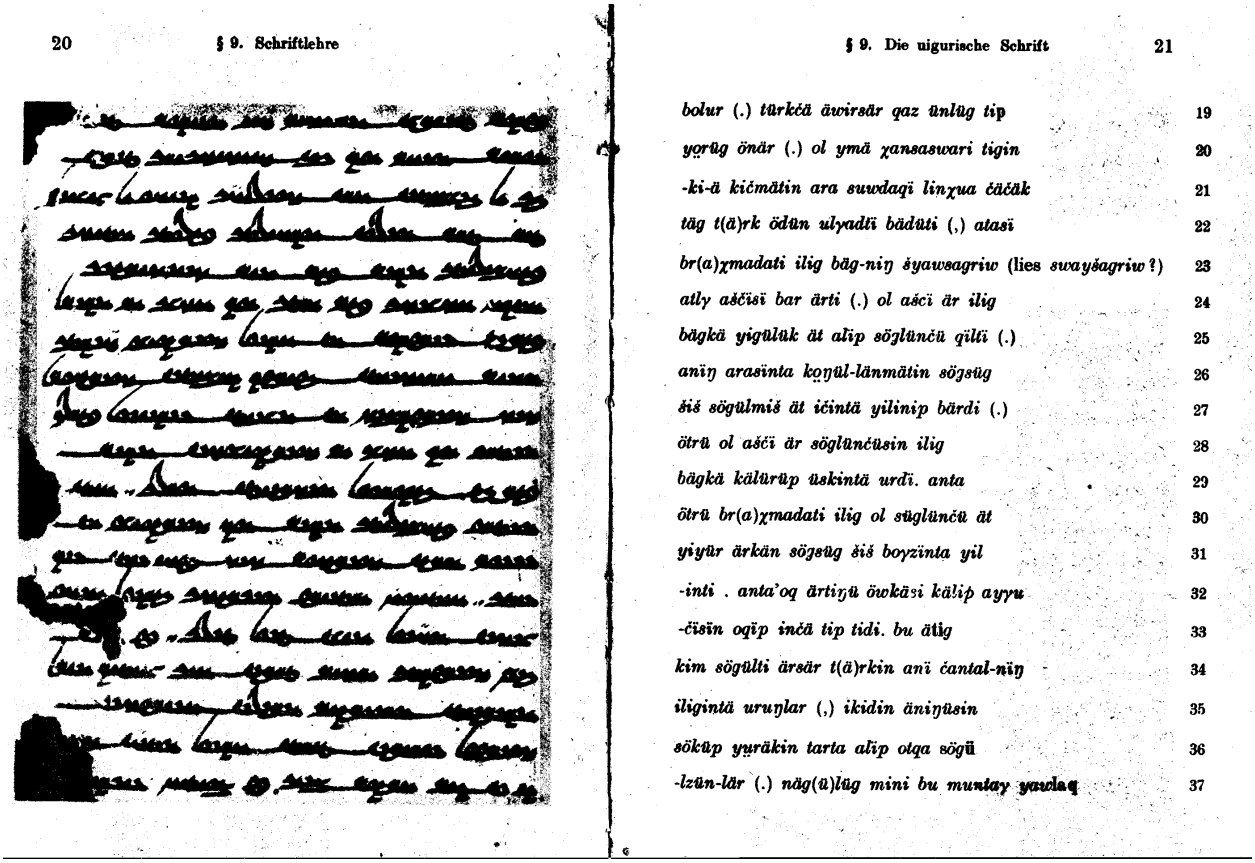
(Signatur: TIII 84—60)

iki yuz altï

sawaγ äšidip bodistw-lar-nïŋ ädgülärin	1
čïnγaru saqïnïp ärüš öküš tïnlγ-lar-r	2
tayun asγančulayu sözlämäktin tïdïl	3
-γay-lar baχšïsï inčä tip tidi (:) äšidgil	4
tüzün oylum (,) tükäl bilgä t(ä)ŋri t(ä)ŋri	5
-si burχan-nïŋ bu muntaγ γ(a)rlïqamïšï	6
bar. öŋrä ärtmiš ödün bu'oq	7
č(a)mbudwip yirtinčüdä bay barïm	8
-lïγ čoγluγ yalïnl(ï)γ baranas atlγ känt	9
uluš bar ärdi. ol ymä baranas balïqta	10
qut t(ä)ŋrisiŋä mäŋizätgülük täg br(a)χma	11
-dati atlγ ilig χan bar ärdi. ol ilig	12
bäg-niŋ bir körgäli körtlä oγul-luγ	13
ärd(i)nisi toγdï (.) inčä qaltï qaz-lar iligi	14
-niŋ ünintän itmiš yigädmiš išidgäli	15
ädgü säwiglig ünlüg ärdi. anï üčün	16
b(ä)lgü biltäči braman-lar χansaswari tip	17
at urdï-lar. χansaswari tisär äntkäk	18

2*

Figure 33: Transcription of an Old Uyghur manuscript (from von Gabain 1950: 18–19). Continued in fig. 34.

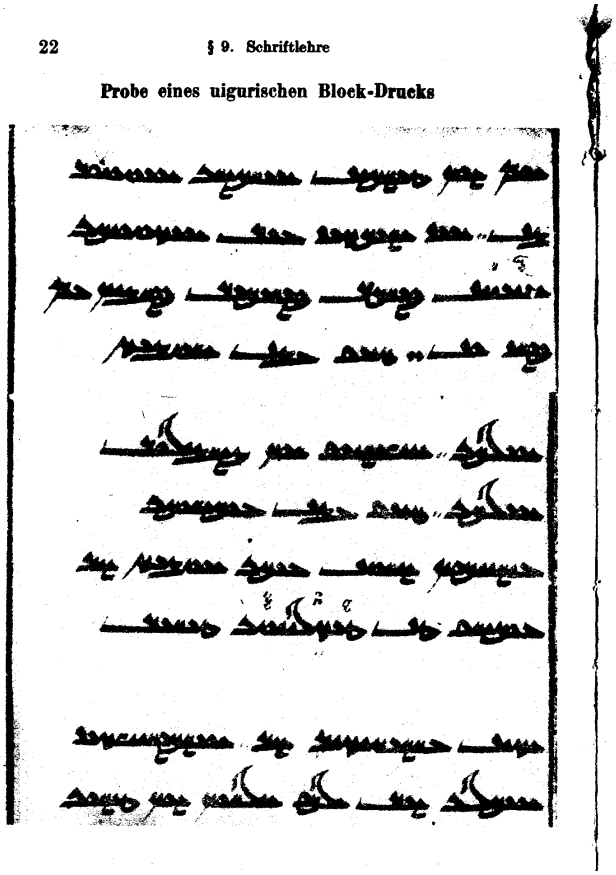


20 § 9. Schriftlehre

§ 9. Die uigurische Schrift 21

bolur (.) *türkčä äwirsär qaz ünlüg tip* 19
yo̤rüg önär (.) *ol ymä χansaswari tigin* 20
-ki-ä kičmätin ara suwdaqï linχua čäčäk 21
täg t(ä)rk ödün ulyadtï bädüti (,) *atasï* 22
br(a)χmadati ilig bäg-niŋ šyawsagriw (lies *swayšagriw*?) 23
atly aščïsï bar ärti (.) *ol ašcï är ilig* 24
bägkä yigülük ät alïp söglünčü qïltï (.) 25
anïŋ arasïnta ko̤ŋül-länmätin sögsüg 26
šiš sögülmiš ät ičintä yilinip bärdi (.) 27
ötrü ol ašči är söglünčüsin ilig 28
bägkä külürüp üskintä urdï. anta 29
ötrü br(a)χmadati ilig ol süglünčü ät 30
yiyür ärkän sögsüg šiš boyzïnta yil 31
-inti . anta'oq ärtiŋü öwkäsi kälip ayyu 32
-čïsïn oqïp inčä tip tidi. bu ätig 33
kim sögülti ärsär t(ä)rkin anï čantal-niŋ 34
iligintä uruŋlar (,) *ikidin äniŋüsin* 35
söküp yṳräkin tarta alïp olqa sögü 36
-lzün-lär (.) *näg(ü)lüg mini bu muntay yavlaq* 37

Figure 34: Transcription of an Old Uyghur manuscript in a grammar of Old Turkic (from von Gabain 1950: 20–21). Continued from fig. 33.



22 § 9. Schriftlehre

Probe eines uigurischen Block-Drucks

§ 9. Die uigurische Schrift 23

(Signatur: TIIIM 195)

nom-luγ tilgän äwirgäli ötünür 1

m(ä)n. üzäliksiz yig üstünki 2

inyana [darüber in *brāhmī*: *jñāna*] ***bilgä biligig bulmaqïm*** 3

bolzun. qayu ymä ärtmiš 4

ödki. ančulayu oq kälmädük 5

ödki. qayu ymä yuγärüki 6

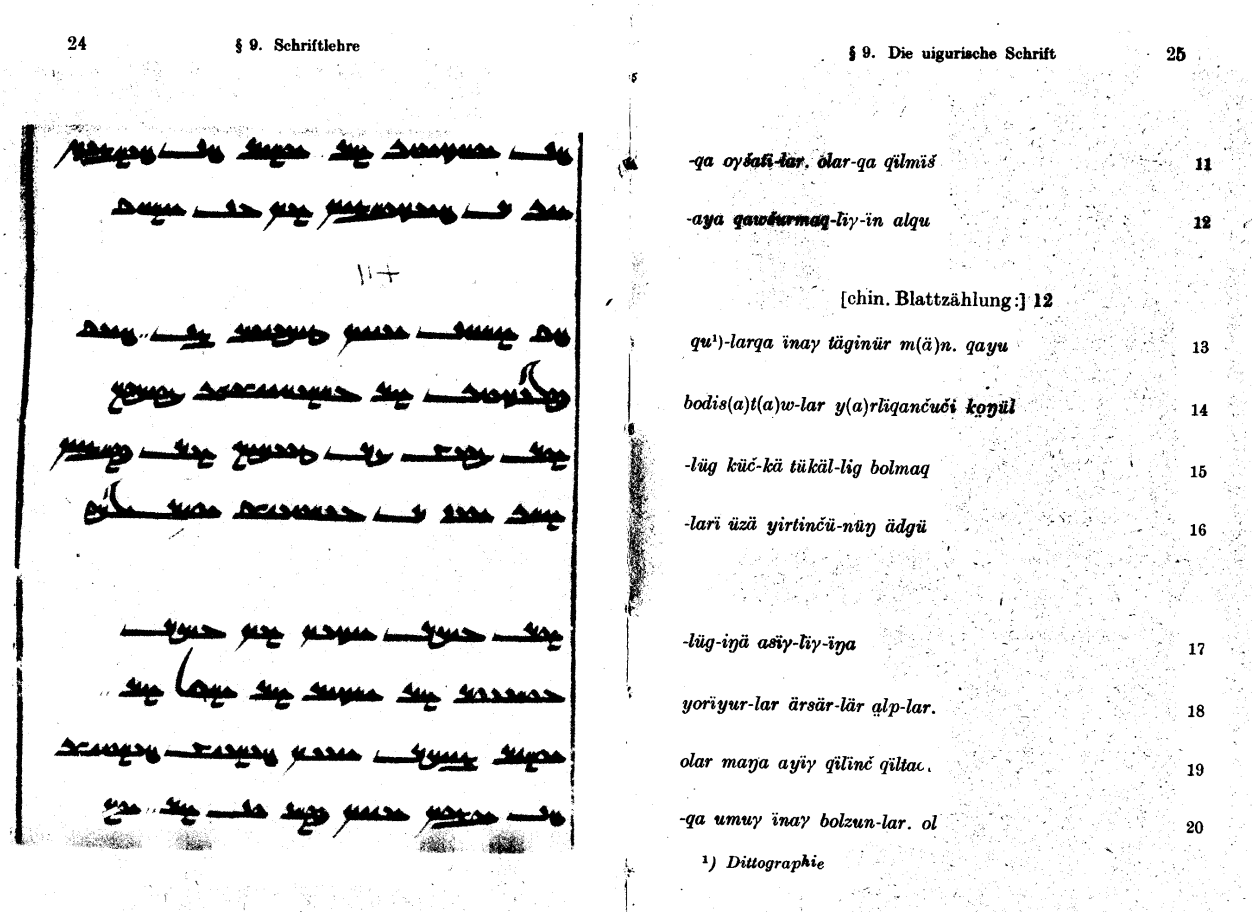
yalaγuq-larta yigi utmiš-lar 7

yuγärü-tä tiṯdanti [darüber in *brāhmī*: *dhiṣṭhanti*] *tirig* 8

äsän y(a)rlïqasar-lar. ülgüläńčsiz 9

ögdi-lig ädgü adruq-luγ taluγ 10

Figure 35: Transcription of an Old Uyghur manuscript in a grammar of Old Turkic (from von Gabain 1950: 22–23). Continued in fig. 36.



24 § 9. Schriftlehre

§ 9. Die uigurische Schrift 25

-qa oγšatï-lar. olar-qa qïlmïš 11

-aya qawšurmaq-lïγ-ïn alqu 12

[chin. Blattzählung:] 12

qu[1]*-larqa ïnaγ täginür m(ä)n. qayu* 13

bodis(a)t(a)w-lar y(a)rlïqančuči kǫŋül 14

-lüg küč-kä tükäl-lig bolmaq 15

-larï üzä yirtinčü-nüŋ ädgü 16

-lüg-iŋä asïγ-lïγ-ïŋa 17

yoriyur-lar ärsär-lär alp-lar. 18

olar maŋa ayïγ qïlïnč qïltac. 19

-qa umuγ ïnaγ bolzun-lar. ol 20

[1]) *Dittographie*

Figure 36: Transcription of an Old Uyghur manuscript in a grammar of Old Turkic (from von Gabain 1950: 24–25). Continued from fig. 35.

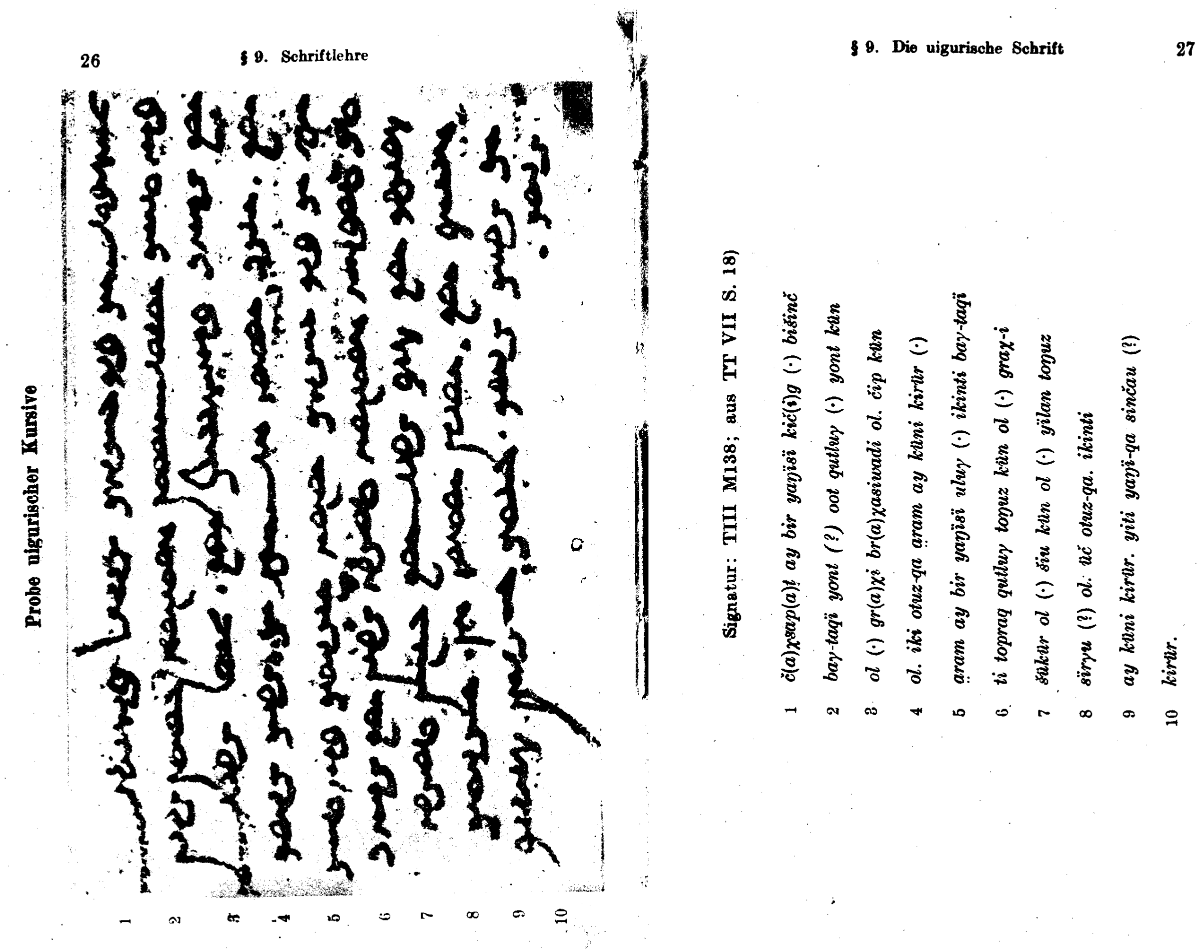
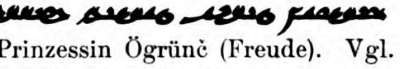


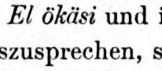
Figure 37: Transcription of an Old Uyghur manuscript in a grammar of Old Turkic (from von Gabain 1950: 26–27).

S. 46. *čaidan* stammt vielleicht aus dem chinesischen 齋壇 *čai-t'an* (alte, aus der Intonation zu erschließende Form: *čai-dan*), wörtlich »Fasten-Platz« oder »Fasten-Halle« (Giles, Lex., gibt die Bedeutung »altars of abstinence«, — »Taoist temples or halls«).

S. 48. Zu dem Ausdruck *ymki* »sitzen« (*olur-*) sind die chinesisch-buddhistischen, mit 坐 *tso* »sitzen« zusammengesetzten Ausdrücke zu vergleichen:

坐臘 to retreat during the twelfth moon, W. Williams, Dict;
打坐參禪 to meditate in a retreat, ebenda;
坐安居 rester en retraite (St.-Julien, Ex. prat. S. 169) sc. retraite religieuse, ebenda;
坐夏 être sédentaire dans la retraite d'été, ebenda S. 191.

Ebenda. *tngrim* war tatsächlich eine Titulatur, denn unter den Fresken der Turfanexpedition II (A. von Le Coq) befindet sich die Abbildung einer uigurischen Prinzessin mit der Beischrift  *ögrünč tigin tngrim körki* = das Bild der Prinzessin Ögrünč (Freude). Vgl. auch den Titel *tngrilär* im Bekenntnis der Üträt, S. 80 Z. 64.

Ebenda.  [nach Radloff *El ökäsi* und ihm zufolge »Volksmutter« zu übersetzen] ist nicht *Il ögäsi* auszusprechen, sondern *Il ügäsi*, wie die chinesische Umschreibung beweist. Auf einem Fragmente des Kara Balgassun-Denkmals findet sich nämlich der Titel

內宰相頡于伽思,

aus dem Schlegel (Chinesische Inschrift auf dem uigurischen Denkmal in Kara Balgassun S. 11) einen »inneren Minister, Kit-kan ka-su« oder »Kirkhan-kaš« (ebenda S. 11, 12) herausliest. Schlegel hat eigenmächtig 于 (*ü*) in 干 (*kan*) verändert, da nach seiner Meinung die Bücher der T'ang-Dynastie maßgebend seien, nicht die Steininschriften! Umgekehrt vielmehr sind die durch Büchertradition überlieferten Titel

大相 頡于迦斯 und 內宰相 頡于伽思
in 大相 頡于迦斯 und 內宰相 頡于伽思
der Premierminister Il ü- gä- si der innere Minister Il ü- gä- si

zu restituieren. *Il ügäsi* »Ruhm des Reichs« (ungefähr فخر الملك) wird (wie نظام الملك) ein Titel gewesen sein, nicht ein Name. Damit entfällt auch die sachliche Schwierigkeit, den Il ügäsi, der schon a. 781 erster Minister war, noch 60 Jahre später, a. 841, fast am Ende der Glanzzeit des

Figure 38: Excerpt from Müller's *Uigurica* showing Old Uyghur text in a horizontal layout (1910: 93). Note the orientation of the glyphs, turned 90 degrees clockwise in relation to their appearance in the code chart.

Uighur		Mongolian	
𐽰	*aleph*	𐼲	U+1820 MONGOLIAN LETTER A
		𐽷	U+1887 MONGOLIAN LETTER ALI GALI A
𐽱	*beth*	𐼱	U+1838 MONGOLIAN LETTER WA
𐽲	*gimel*	𐼬	U+182C MONGOLIAN LETTER QA
		𐼭	U+182D MONGOLIAN LETTER GA
𐽳	*waw*	𐼱	U+1820 MONGOLIAN LETTER WA
		𐽦	U+18A6 MONGOLIAN LETTER ALI GALI HALF U
𐽴	*zayin*	𐼽	U+183D MONGOLIAN LETTER ZA
𐽵	*heth*	𐼬	U+182C MONGOLIAN LETTER QA
𐽶	*yodh*	𐼵	U+1835 MONGOLIAN LETTER JA
		𐼶	U+1836 MONGOLIAN LETTER YA
𐽷	*kaph*	𐼺	U+183A MONGOLIAN LETTER KA
𐽸	*lamedh*	𐽡	U+1891 MONGOLIAN LETTER ALI GALI DA
		𐼳	U+1833 MONGOLIAN LETTER DA
𐽹	*mem*	𐼸	U+1828 MONGOLIAN LETTER MA
𐽺	*nun*	𐼸	U+1828 MONGOLIAN LETTER NA
𐽻	*samekh*	𐼰	U+1830 MONGOLIAN LETTER SA
		𐼯	U+1831 MONGOLIAN LETTER SHA
𐽼	*pe*	𐼩	U+182A MONGOLIAN LETTER BA
		𐼪	U+182B MONGOLIAN LETTER PA
𐽽	*sadhe*	𐼨	U+182A MONGOLIAN LETTER CHA
𐽾	*resh*	𐼻	U+1837 MONGOLIAN LETTER RA
𐽿	*taw*	𐼮	U+1832 MONGOLIAN LETTER TA
𐾀	*lesh*	𐼼	U+182F MONGOLIAN LETTER LA

Figure 39: Comparison of Old Uyghur letters with Unicode Mongolian letters.

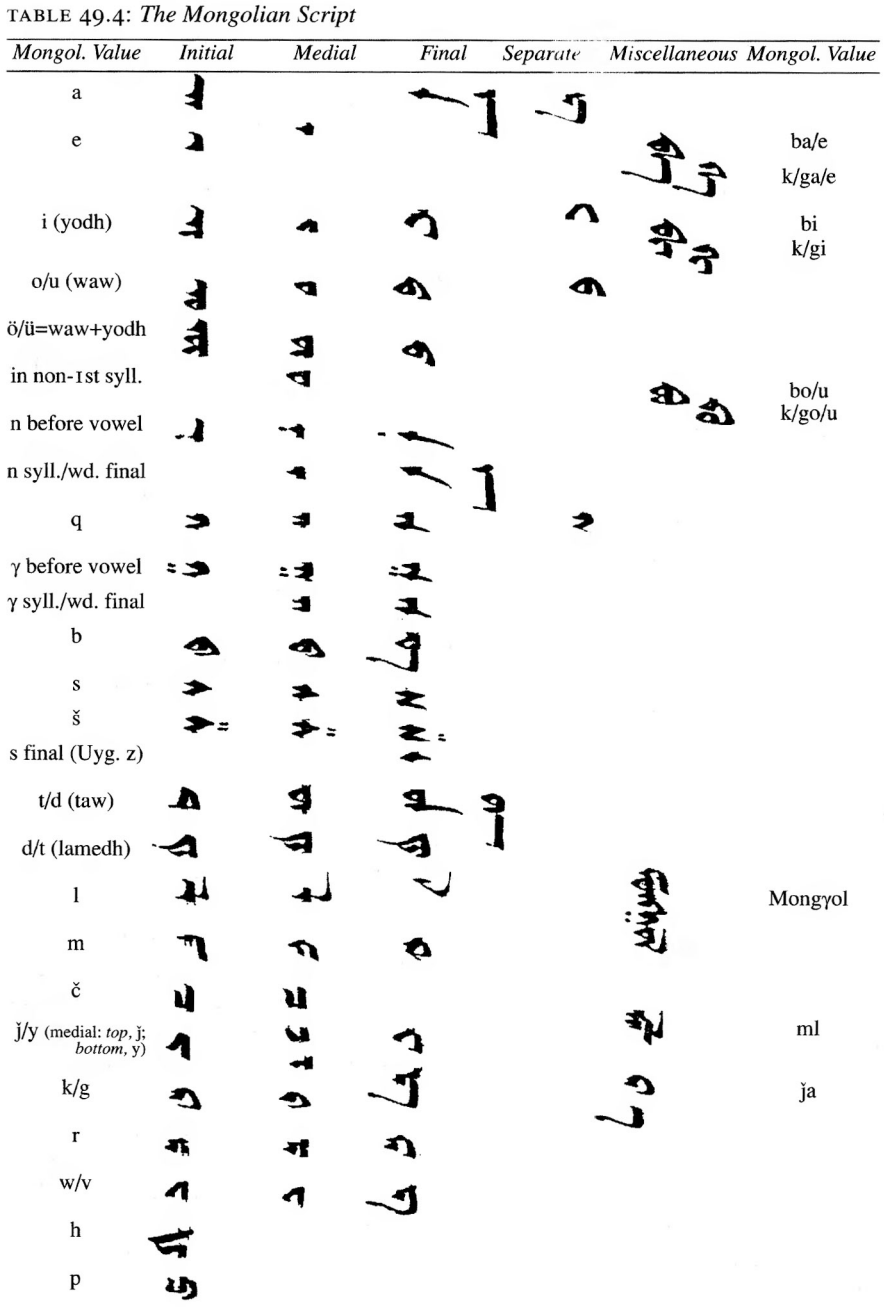
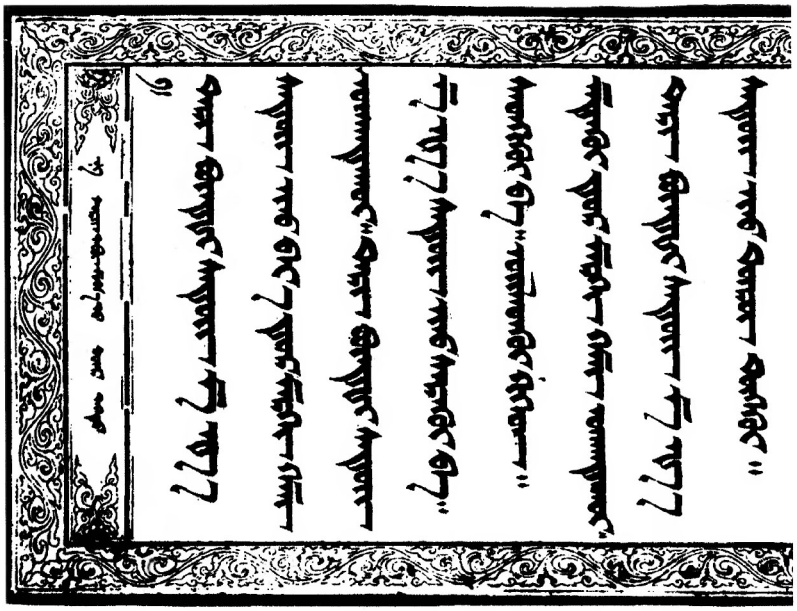


Figure 40: Table showing letters of the Mongolian script (from Kara 1996: 545). See table of Old Uyghur letters from the same source in fig. 4.

SAMPLE OF MONGOLIAN



1. *Transliteration:*	tʾrʾ	pwβʾδhy	sʾδwβʾ	mʾ hʾ ʾ /sʾδwβʾ	ʾynw	pʾy ʾ δwr	mʾrkʾn
2. *Normalization:*	tere	bôdhi-saduva		ma-hā-saduva	inu	bey-e-dür	mergen
3. *Gloss:*	that	bodhisattva		mahāsattva	3POSS	body-DAT	wise

1. kʾmʾn / ʾwqʾqδʾqwy :	tʾrʾ	pwβʾδhy	sʾδwβʾ	mʾ hʾ ʾ / sʾδwβʾ	ʾynw	serekü
2. kemen / uqaγdaqui	tere	bôdhi-saduva		ma-hā-saduva	inu	sereküi
3. saying should.know	that	bodhisattva		mahāsattva	3POSS	waking

1. ba : / sʾtkykwy	pʾ :	ʾwylʾtkwy	kykʾt : /	mʾδʾkwy	dwr	mʾrkʾn	kʾmʾn
2. ba : / sedkiküi	ba	üiledküi	kiged	medeküi-dür		mergen	kemen
3. and thinking	and	acting	as.well	knowing-DAT		wise	saying

1. ʾwqʾqδʾqwy : /	tʾrʾ	pwβʾδhy	sʾδwβʾ	mʾ hʾ ʾ/sʾδwβʾ	ʾynw	twyrwn	twykʾkwy :
2. uqaγdaqui	tere	bôdhi-saduva		ma-hā-saduva	inu	törön	tügeküi
3. should.know	that	bodhisattva		mahāsattva	3POSS	born	spreading

'You should know: that bodhisattva and mahasattva is wise in (the knowledge of) body. You should know: that bodhisattva and mahasattva is wise in watchfulness, thinking, acting as well as perceiving. That bodhisattiva (is wise in the knowledge of) the sense organs and sense objects (lit. what is being generated and what is spreading.'

—*From the printed Mongol Kanjur, vol. 49, folio 2A. Text without diacritics. Early 18th century blockprint.)*

Figure 41: Sample Mongolian text (from Kara 1996: 546). Compare the Mongolian block print with the Old Uyghur block print in fig. 8.