L2/19-016
2019-01-07

Revised proposal to encode Old Uyghur in Unicode

Anshuman Pandey
pandey@umich.edu

January 7, 2019

Document History

This proposal is a revision of the following:

- L2/18-126: “Preliminary proposal to encode Old Uyghur in Unicode”
- L2/18-333: “Proposal to encode Old Uyghur in Unicode”

It incorporates comments made by the UTC Script Ad Hoc Committee and other experts in:

- L2/18-168: “Recommendations to UTC #155 April-May 2018 on Script Proposals”
- L2/18-335: “Comments on the preliminary proposal to encode Old Uyghur in Unicode (L2/18-126)”

The major changes to L2/18-333 are as follows:

- Correction to glyphs for initial and medial *beth*, previously shown erroneously using forms for *yodh*
- Revision of glyphs for *aleph* and *nun* to reflect distinctive forms from the 9th century
- Revision of representative glyph for *zayin* for stylistic uniformity
- Unification of *gimel* and *heth* into a single letter, and addition of a letter for final *heth*
- Addition of an alternate form letter for both *aleph* and *nun*
- Expansion of description of the orientation of terminals for specific letters

A previous version of this proposal was reviewed by the following expert:

- Dai Matsui (Graduate School of Letters, Osaka University)

1 Introduction

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 is a cursive-joining alphabet with features of an *abjad*, and is characterized by its vertical orientation. The script flourished through the 15th century in Central Asia and parts of Iran, but it was eventually replaced by new orthographies based upon the Arabic script in the 16th century. Its usage was continued in Gansu through the 17th century.

Old Uyghur was the basis of vibrant multilingual, orthographic, scribal, and print cultures across Central Asia. It was used for recording religious, literary and administrative documents in Turkic languages, as well as Chinese, Mongolian, Sanskrit, Sogdian, and Tibetan. 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 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. 44–45). 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. 34–42). 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. 24–31 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. 46).

Old Uyghur is situated in the middle of a script continuum that originates from the Old Sogdian script of the 'Ancient Letters' and terminates at modern Mongolian. It developed from the 'cursive' style of the Sogdian script during the 8th–9th century (Kara 1996: 539). Just as speakers of Turkic languages adopted the Sogdian script, speakers of other languages in Central Asia turned to Old Uyghur to develop new orthographies. 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.

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. 47, 48). 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 modern 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).

2 Script identifier

The proposed Unicode identifier for the script is 'Old Uyghur', which is a scholarly designation. The name applies specifically to the script within the context of Unicode, and it does not apply to any language, culture, or community. The script is also known generically as 'Uyghur', without the descriptor 'Old'. The term 'Uyghur' has variant transliterations / transcriptions / spellings in English, such as 'Uighur', 'Uigur', 'Uygur', 'Uyğur', as well as 'Ouïgour' in French, 'Uigurisch' in German, etc.

To be sure, neither 'Uyghur' nor 'Old Uyghur' is an entirely accurate designation for the script. The renowned Turkolgist, Gerard 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 in Central Asia before the Uyghur language became prominent in the 8th century (1962: 43). However, Clauson concludes that "no useful purpose would be served by suggesting some other name" (1962: 100–101). This proposal abides by Clauson's conclusion. 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.

3 Encoding history

3.1 Previous proposals for Unicode

A proposal for Old Uyghur was previously submitted to the Unicode Technical Committee (UTC) by Omarjan Osman in 2013. Osman's "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. His proposed repertoire contains upright glyphs for the horizontal form 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 for purposes of character encoding. 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.

3.2 Existing standards

There are no existing formal standards for the Old Uyghur script. The closest related digital standard for the script is the Unicode encoding for Mongolian. Recently, the government of China published a standard

known as "GB/T 36331-2018 'Information technology – Uigur-Mongolian characters, presentation characters and use rules of controlling characters"'. According to Liang Hai, GB/T 36331-2018 is a subset of GB/T 26226-2010, which is China's standard for encoding Mongolian — based upon the complete Unicode encoding for the script — and equivalent to Mongolia's MNS 4932: 2000. Another subset of GB/T 26226-2010 is GB/T 25914-2010, which provides a standard for the modern writing system for the Mongolian language. Given the reference to "Uigur-Mongolian", it is apparent that the standard is intended for the representation of the early stages of the Mongolian script, using the phonemic model of the Unicode encoding and similar glyphs. However, it is not a character-encoding standard for Old Uyghur.

4 Script details

4.1 Structure

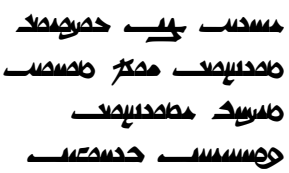
The Old Uyghur script is a cursive joining alphabet. It consists of 18 letters, which are derived from Sogdian, and ultimately from Imperial Aramaic. All vowel sounds are expressed using the letters *aleph*, *waw*, *yodh*; and the rich vowel repertoire of Turkic languages are represented using combinations of these letters in diagraphs and trigraphs (see § 8.1). Diacritics are used for diambiguating letters with similar appearances and for representing sounds for which distinctive letters do not exist (see § 8.2).

The cursive joining feature of the script is similar to that of Sogdian, with letters joined together at the baseline. Letters have an independent shape, which is used in isolation, and contextual shapes when they occur in initial, medial, or final position in a cursive string. All letters are dual joining; however, in some sources, the connection following *zayin* is suspended (see § 7).

4.2 Directionality

The conventional direction of writing for Old Uyghur is vertical, from top to bottom in columns that run from left to right. The vertical orientation is confirmed by biscriptal documents containing Han characters and Central Asian Brahmi. In some Iranian documents from the 14th century, the script is written horizontally. This may be influenced the Arabic script. When Old Uyghur texts were begun to be printed in the 20th century, publishers maintained fidelity to the standard vertical orientation (see fig. 47, 48). There are two appropriate orientations for Old Uyghur in digital representations:

Vertical *Horizontal*



- *Vertical* By default, the script should be oriented vertically, especially when an entire text block contains only Old Uyghur characters. A vertical orientation should also be used when Old Uyghur occurs with other scripts that can be rendered in the same direction.

- *Horizontal* In applications that do not support vertical layout or in contexts where the majority of surrounding text is non-vertical, Old Uyghur may be oriented horizontally and treated as a typical right-to-left script. In such instances, Old Uyghur character glyphs should be rotated 90 degrees clockwise with respect to their orientation in the code chart, and text should be set in horizontal lines that run from right to left, in successive lines from top to bottom. This orientation is identical to the conventional layout for scripts such as Sogdian and Arabic.

The horizontal, right-to-left orientation is used by modern scholars and publishers for short excerpts of Old Uyghur text because it is a convenient method to reference Old Uyghur words and phrases in multilingual contexts that also contain Arabic, Cyrillic, Devanagari, Tibetan, and other scripts (see fig. 54). 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, as in the regular display of Arabic, when it appears in horizonal environments.

Throughout this document, Old Uyghur characters are presented in their conventional vertical forms when they occur in examples, and in horizontal right-to-left orientation in Latin-script environments.

4.3 Repertoire

The traditional Old Uyghur alphabet consists of 18 letters. An inventory of the script appears in the margin of the manuscript U 40 (see fig. 38), which contains a Manichaean text and is dated to the 9th century. The excerpt with the inventory is given below, and shows 21 characters:

The first 17 are the basic letters of the script: *aleph*, *beth*, *gimel*, *waw*, *zayin*, *heth*, *yodh*, *kaph*, *lamedh*, *mem*, *nun*, *shin*, *pe*, *sadhe*, *resh*, *samekh*, *taw*. The names for letters follow the scholarly tradition of using the nomenclature for Aramaic names. The order of letters also follows the original Aramaic paradigm; however in this inventory, the positions of *samekh* and *shin* are swapped.

The four letters that follow are not clearly discernable due to blemishes and other damage to the manuscript. Clauson (1962: 107) suggests that they are 'hooked' *resh*, a final *samekh* (or *shin*), a final *mem*, and a two-dotted *heth*; however, he does not offer an explanation for this assignment of values for #18 and #19.[1]

[1] I should like to offer a comment on Clauson's identity of characters #18 and #19. Characters #20 and #21 are 𐽂 final *mem* and the 𐽃 two-dotted *heth*, respectively. The presence of final *mem* is understandable: as compared to other letters, the shape of *mem* varies considerably in its initial, medial, and final forms. The initial form 𐽄 of *mem* is given as #10, which is visibly different from its final form, 𐽂 (#20). I am less satisfied with Clauson's identification of #18 and #19. Clauson states that #18 is the 'hooked' *resh*. While, this letter follows *taw* in the natural alphabetic order, its shape in the manuscript resembles 𐽁, not the convention form 𐽆 of 'hooked' *resh*. This 𐽁 is an alternate form of final *aleph* / *nun*, which is a common form that the scribe would recognize as a 'special' character with a distinctive shape. Secondly, Clauson states that #19 is a 'final *samekh* (or shin)". However, *samekh* does not have a 'special' final shape that differs significantly from its 'regular' final form 𐽅 (or that of 𐽇 *shin*, for that matter). It appears to me that #19 is actually a poorly written 'hooked' *resh*, made evident by the semblance of a horizontal stroke at the end of the glyph. Accordingly, #18 is not the 'hooked' *resh*, but the downward turned final *aleph* / *nun*. Therefore, the values for #18 and #19 should be reconsidered.

In any case, the inventory is important in that it provides:

- attestation for the full repertoire and order of the alphabet. The scribe of the inventory in U 40 was quite cognizant of the Old Uyghur script. This is made evident by his distinctive depiction of the 17 core letters and the inclusion of the four 'special' characters.

- evidence for independent forms of letters.

- distinctive shapes for *aleph* #1 and *nun* (#11), in which the former is written with an initial horizontal stroke before the triangle, while the latter begins with a rounded stroke. These letters are also distinguished from *zayin* (#5) on account of their terminals.

- distinctive shapes for *beth* (#2) and *yodh* (#7), which are distinguished both by the curvature of their bodies and the length of their terminals.

- distinctive shapes for *gimel* (#3) and *heth* (#6), which are identified by their terminals.

- distinctive shapes for *samekh* (#12) and *shin* (#16), that being the presence of the elongated down- and rightward initial stroke in the latter, from which the second stroke merges at the midpoint.

- a sense that scribes were acutely aware of distinctions in contextual forms of letters, particularly those letters whose forms differed from the nominal forms. The final *mem* (#19) is likely included because its shape diverges from initial *mem* (#10).

- evidence for the usage of diacritics to expand the alphabet. The two-dotted *heth* is a common character used for representing /x/ or /q/.

The attestation of the complete repertoire is also palaeographically significant. After the 9th century, writing practices led to the merger of some letterforms, resulting in an abridged script in the 14th century. The following description of the changes to the Old Uyghur repertoire is based upon Clauson (1969: 109–110) and details provided by Dai Matsui (personal communication, August 2018–January 2019):

9th century

- All 18 letters are distinguishable in good manuscripts
- *samekh* and *shin* are distinctive
- initial and medial *aleph* and *nun* are distinguishable
- initial and medial *gimel* and *heth* are indistinguishable
- two dots above *heth* for representing /q/ or /x/

11th century

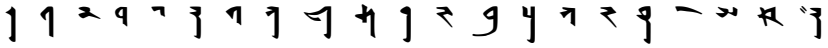
- *samekh* and *shin* become indistinguishable and represented using *shin*
- when necessary, two dots beneath *samekh* for representing *shin*
- *aleph* and *nun* become difficult to distinguish
- final *nun* indistinguishable from *zayin* without a dot over the former

14th century

- only *kaph*, *lamedh*, *mem*, *pe*, 'hooked' *resh* remain distinctive
- *beth* and *yodh* begin to merge and become indistinguishable
- in some instances *sadhe* is indistinguishable from *beth* / *yodh*
- *gimel* and *heth* are indistinguishable from consecutive *aleph* and/or *nun* without use of double dots
- medial and final *taw* indistinguishable from the sequence *waw-nun* unless the *nun* is dotted
- *samekh* / *shin* difficult to distinguish from *gimel* / *heth*
- *resh* begins to become indistinguishable from consecutive *aleph* and/or *nun*;

Identifiying the repertoire of 'the' Old Uyghur script requires understanding the periodization of the script's development. Given the significant deterioration of the script by the 14th century, the repertoire and letter-forms for the Unicode encoding for Old Uyghur should be based upon the earliest sources in order to enable the complete representation of texts in the script.

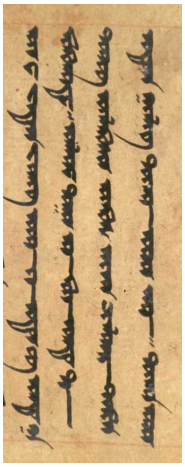
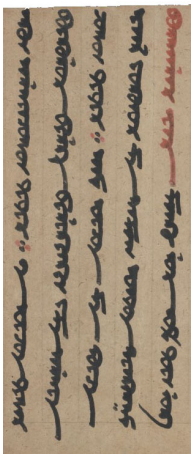
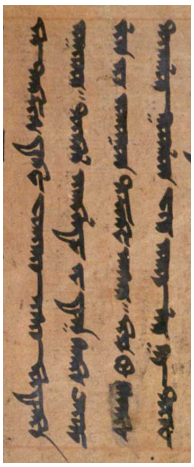
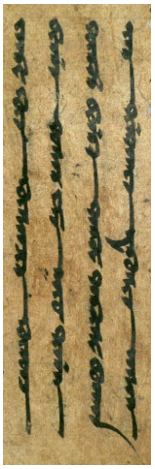
The inventory of the Old Uyghur script of the 9th century, as exhibited in U 40 — and as per my observation in the above footnote – would be displayed as follows when rendered in a basic digitized font, whose glyphs have been designed after analyzing distinctive letterforms across a variety of primary sources:



4.4 Styles of the script

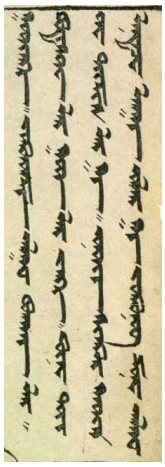
There are various styles of Old Uyghur, which may be classified into three broad categories: 'formal' (also: 'sūtra', 'semi-square', 'book'), 'block print', and 'cursive'. The 'formal' and 'cursive' styles are hand-written, while 'block print' styles were cut for printing using movable type.

Formal ('sūtra', 'semi-square', 'book') style

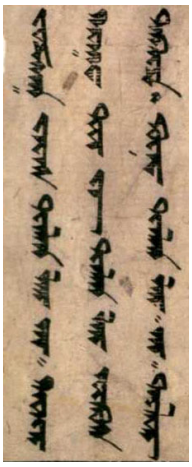
Mainz 119	Mainz 841	Pelliot Ouïgour 13	Mainz 819	Mainz 896	U 1071
					

Block-print style

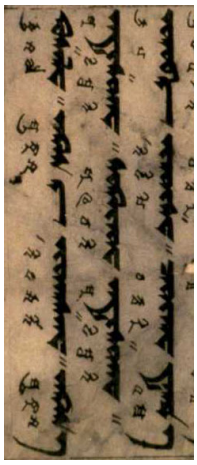
U 387



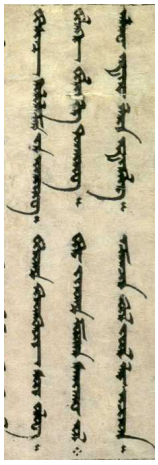
U 7008



Mainz 801

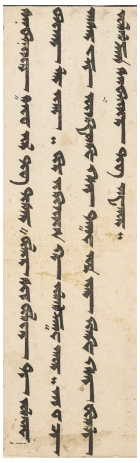


U 343



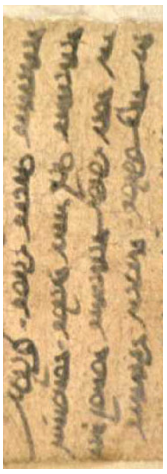
U 496

PEALD 6r

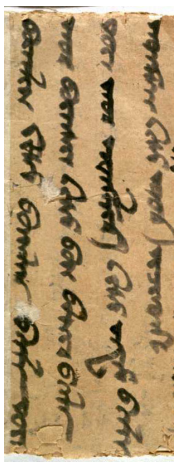


Cursive style

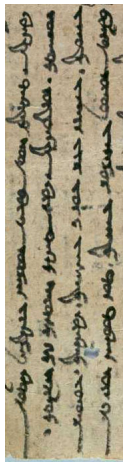
U 499



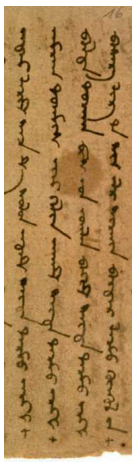
U 560



U 456



U 558



5 Approach to the Encoding

5.1 Scope

The proposed encoding for Old Uyghur provides a unified block for encoding texts written and printed in all styles of the script, across all periods.

5.2 Character repertoire

The proposed repertoire for Old Uyghur is based upon a fully attested set of 35 characters: 21 letters, 7 combining signs, 6 punctuation signs, 1 stem-extending sign, and 1 editorial sign. The code chart and names list follows p. 11.

The encoded set may contain characters that are not included in traditional and scholarly inventories of the script. Similarly, other characters may not be included, such as contextual forms of letters, etc. Such divergences naturally arise from the requirements of developing character-encoding standards and the distinctions between characters and glyphs.

The proposed repertoire is sufficient for representing the majority of Old Uyghur texts. There are other diacritics, punctuation, digits, and other symbols, as such those shown in Omarjan's proposal, that require additional research before a determination may be made regarding their suitability for encoding. These characters may be added to the repertoire in the future.

5.3 Character names

In the scholarly literature, Uyghur letters are known by the names of the original Aramaic letters from which they are derived. This nomenclature follows the scholarly convention for letters of the Sogdian scripts. The same is followed here. Throughout this proposal, italics are used for scholarly names for graphemes, while small capitals indicate Unicode character names, eg. 𐽀 is referred to as *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.

The descriptors 'right' and 'left' in the character names refer to the orientation of terminals or the placement of diacritics with respect to the base letter in the traditional vertical orientation of the script. When Old Uyghur occurs in non-vertical contexts, 'right' should be interpreted as 'down', and 'left' as "up'. For example, letters that possess a 'left' tail would be oriented such that the tail extends 'upwards', eg. 𐽁 ALEPH WITH LEFT TAIL would appear as 𐽀 in horizontal contexts. Similarly, 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 ◌𐽆, a 'below-base' sign.

5.4 Representative glyphs

The 'formal style' is most 'representative' of the Old Uyghur script because it expresses distinctive aspects of the characters. The block-print styles might have served as a very suitable basis for representative glyphs because the shapes of characters may be considered uniform and a print 'standard'. However, as the majority of block prints were produced after the 12th century, they do not reflect all distinctive aspects of the characters, especially those used in the 9th century, on account of the truncation of the alphabet. For this reason,

the representative glyphs are based upon normalizations of the 'formal' style, with designs influenced by the general aesthetics of the block-print style.

5.5 Justification for encoding

Although Old Uyghur is derived from Sogdian and is the ancestor of Mongolian, and shares similarities with both scripts, there is a requirement to represent Old Uyghur in plain text, particularly for distinguishing these scripts for the creation, processing, and digitization of text on the basis of character identity.

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 Mongolian encoding uses different names and ordering for letters, which reflect Mongolian preferences and pronunciations. Mongolian letter names do not correspond to Old Uyghur values. 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. The proposed encoding for Old Uyghur is based upon a graphetic model. The current model for Mongolian, which has a phonetic basis, presents several issues and is unsuitable for Old Uyghur. The proposed model for Uyghur offers a practical implementation for a vertical script that avoids the complications of the Mongolian model.

6 Proposed repertoire

6.1 Letters

The following letters are proposed for encoding:

Character name	Glyph	Joining	Latin
OLD UYGHUR LETTER ALEPH		dual	ʾ
OLD UYGHUR LETTER ALEPH WITH LEFT TAIL		right	-ʾ
OLD UYGHUR LETTER ALEPH-NUN WITH RIGHT TAIL		right	-ʾ, -n
OLD UYGHUR LETTER BETH		dual	β
OLD UYGHUR LETTER GIMEL-HETH		dual	γ, x, q
OLD UYGHUR LETTER WAW		dual	w
OLD UYGHUR LETTER ZAYIN		dual	z, ž
OLD UYGHUR LETTER FINAL HETH		right	-x, -q

OLD UYGHUR LETTER YODH	𐽰	dual	y
OLD UYGHUR LETTER KAPH	𐽱	dual	k
OLD UYGHUR LETTER LAMEDH	𐽲	dual	δ
OLD UYGHUR LETTER MEM	𐽳	dual	m
OLD UYGHUR LETTER NUN	𐽴	dual	n
OLD UYGHUR LETTER SAMEKH	𐽵	dual	s, š
OLD UYGHUR LETTER PE	𐽶	dual	p
OLD UYGHUR LETTER SADHE	𐽷	dual	c
OLD UYGHUR LETTER RESH	𐽸	dual	r
OLD UYGHUR LETTER SHIN	𐽹	dual	š
OLD UYGHUR LETTER TAW	𐽺	dual	t
OLD UYGHUR LETTER LESH	𐽻	dual	l

aleph The *aleph* is a palaeographically distinct letter in Old Uyghur. During the evolution of the script, forms of *aleph* and *nun* began to merge. The independent and initial forms of 𐽀 *aleph* are distinctive, however, its medial and final forms may resemble those of *nun*, compare initial 𐽀 *aleph* and 𐽴 *nun*, where the latter is rounder. In block prints and various manuscripts, they have the same final form 𐽀. Nonetheless, *aleph* has several features that define it as a distinctive letter separate from *nun*, such as alternate forms and shape variations (see fig. 11). Distinctive alternate forms attested in manuscripts and block prints are shown below. These have been included in the proposed repertoire as separate characters.

- 𐽁 In some manuscripts a word-final *aleph* is written independently, detached from the previous letter, regardless of the joining behavior of the latter (see fig. 10). This alternate 𐽁 is distinguished from the representative form 𐽀 by the vertical orientation of its terminal. In some cases, 𐽁 is used for distinguishing between final *a* (𐽀) and *ä*/*e* (𐽁), see fig. 2; also see forms used for *-a* in fig. 5. As both 𐽀 and 𐽁 may occur simultaneously, the latter is proposed for separate encoding as the right-joining character ALEPH WITH LEFT TAIL.

- The *aleph* has an alternate final form , which occurs simultaneously with the regular final in several manuscripts. It occurs in the middle of words (see fig. 10), and is not merely a truncation of the regular horizontal terminal used at the end of line or at text margins. This is also used as an alternate form for final *nun*. For this reason, the character is proposed for encoding as the right-joining character ALEPH-NUN WITH RIGHT TAIL.

Also attested are the following stylistic variants. These are not proposed for encoding as separate character and should be handled using glyph substitutions in fonts.

- / The independent *aleph* is represented in some documents using the form (see fig 9). This stylistic variant resembles the letter *kaph*. The is produced by lifting the initial horizontal stroke of . When the variant is used, the ALEPH WITH LEFT TAIL takes a 'toothed' shape: . This form resembles the variant form of *kaph*. The 'toothed' variants / of independent *aleph* should be handled as a stylistic set when used as variants for / .

  This / 'toothed' form is also used in some sources as a contextual variant of final *aleph* after *kaph* and *pe*, eg. / *k* ʾ, / *p* ʾ (see fig. 12), compared to *k* ʾ, *p* ʾ. Such contextual glyph variation should be considered conventional behavior for relevant styles.

- In block-print documents, *aleph* takes the shape when it immediately follows *lamedh*, eg. (see fig. 12). This form is used in order to prevent clashing between the downward hook of *lamedh* and the body of *aleph*, ie. . The latter 'clashing' form occurs in hand-written manuscripts.

gimel **and** ***heth*** As evidenced by the inventory in U 40, these two letters are distinguished in independent and final positions using the glyphs and , respectively, but they have the same initial and medial forms (also see fig. 14). For this reason, the following model is proposed for representing these letters:

- The letters *gimel* and *heth* are unified as the dual-joining letter GIMEL-HETH, which is represented using the glyph . This letter is to be used for writing initial and medial *gimel* and *heth*, as well as for final *gimel*.

- The independent *heth* is encoded as the right-joining letter FINAL HETH. It is to be used for final *heth*.

- Dotted forms The diacritics ◌́ and ◌̋ may be placed above and for representing the sounds /q/ and /x/, eg. , , , (see § 8.2).

beth The regular final form of *beth* is , however, the final is also written as (see fig. 13). The leftward orientation of the tail is used likely for distinguishing *beth* from *yodh* when there is a limitation of space for extending the final stroke of the former. Such distinctions are necessary in block-print styles, where non-final forms of *beth* and *yodh* are highly identical. The is a glyphic variant and is not proposed for separate encoding. In digital plain-text representations, usage of for final *beth* is sufficient.

zayin The following points should be noted regarding this letter:

- Variant form The representative form of *zayin* is based upon the shape used in formal and block-print styles (see fig. 15). The glyphic variant 'sawtooth' form occurs in some documents (see fig. 8).

- Joining behavior In some sources ـز *zayin* does not connect to a following letter. It is defined as a dual-joining letter in order to enable joining on both the right and left, as needed. The control character ZWNJ ZWNJ may be placed after ZAYIN to prevent joining with the following letter.

- Dotted forms In some sources *zayin* is distinguished using the diacritics ◌́ and ◌̎, eg. ـزٕ and ـزٖ, in order to indicate /ž/ (see § 8.2).

kaph The regular final form of *kaph* is ـک, however, the final is also written as ـکا (see fig. 16). Both forms may occur simultaneously within a document. As there is no semantic difference between ـک and ـکا, the latter should be considered a stylistic variant. There is no need to encode it as a separate character. In digital plain-text representations, usage of ـک for final *kaph* is sufficient.

nun The ـن *nun* is represented using different shapes:

- Dotted form The diacritic ◌́ is commonly written above *nun* for distinguishing it from ـن *aleph*, eg. ـنَ (see § 8.2).

- Final forms The regular final form ـن of *nun* is similar to that of *aleph*. The final form is also written using ـ ALEPH-NUN WITH RIGHT TAIL.

samekh and ***shin*** As shown in U 40, the letters ـس *samekh* (/s/) and ـش *shin* (/š/) are palaeographically distinctive letters in the script. The two letters are distinguished by the fact that *shin* is written using two strokes (first with a right-sloping downward angle and the second as a leftward curve extending from the midpoint of the first), while *samekh* is a single stroke (see fig. 17). By the 11th century, however, they merged into a single letter, eg. ـس (see fig. 6). The regular form for *samekh* / *shin* in documents from this time is based upon the simpler ـش *shin* instead of ـس *samekh*. In such contexts, the diacritic ◌̎ is applied to ـش *shin* to express /š/, eg. ـشٖ, or 'marked' or 'dotted' *shin* (see § 8.2).

pe In various manuscripts and block prints, final ـپا *pe* is rendered as the ornamental form ـپـ (see fig. 18). The latter appears to occur at the end of line at the end of a section or a text. Although it is graphically distinct, it is to be considered a stylistic variant of the regular final *pe*. If there is a requirement to constrast both form of final *pe* in plain text, then the ornamental form may be added to the repertoire.

sadhe The regular final form of *sadhe* is ـص, however, the final is also written as ـڞ (see fig. 19). Both forms may occur simultaneously within a document. As there is no semantic difference between ـص and ـڞ, the latter should be considered a stylistic variant. There is no need to encode it as a separate character. In digital plain-text representations, usage of ـص for final *sadhe* is sufficient.

taw The regular final form of *taw* is ـت, however, the final is also written as ـټ (see fig. 20). Both forms may occur simultaneously. As there is no known semantic difference between the forms, the variant ـټ should be considered a stylistic variant. There is no need to encode it separately

lesh The letter ـ𐽔 represents the sound /l/. It is derived from 𐽔 U+10F44 SOGDIAN LETTER LESH, which is known as 'hooked *r*' (see Pandey 2016b for details). The Old Uyghur ـ𐽔 has been assigned the name 'LESH', following the name for the corresponding Sogdian letter. This is not a historical name, but one suggested by modern scholars as it aligns with the Aramaic name *resh*. The alias 'hooked *r*' has been specified in the names list.

6.1.1 Note on variation in terminal orientation

The following letters have attested variations in the orientation of their terminals:

	regular	alternate
aleph		
beth		
kaph		
nun		
nun		
sadhe		
taw		

There are various possible explanations for such variation:

- *Spacing adjustment* When letters with downward terminals occur at a margin with insufficient space to produce the regular elongated stroke, the terminal is curved to the left. In such cases, the direction of the tail has no semantic difference.

- *Stylistic preference* In some documents written in a highly cursive style, a scribe may have preferred to use rightward tails instead of downward terminals for all relevant letters, as a matter of preference. However, such an explanation may not bear relevance for early documents, where there is intentional alternation between convention and variant terminals.

- *Intentional alternation* A scribe or block-printer may have explicitly chosen to use a variant terminal instead of the conventional stroke. Such a conclusion may be drawn by the occurrence of both conventional and variant strokes in positions along a line other than at the end. Intentional alternation is also evident in cases where both the conventional and variant forms are used simultaenously in a document in independent contexts; this occurs frequently with *aleph*.

At present, only the  (ALEPH WITH LEFT TAIL) and (ALEPH-NUN WITH RIGHT TAIL) have been proposed for separate encoding. However, it may be practical to encode other alternate forms as separate letters. Encoding these provides a practical method for accurately representing Old Uyghur documents in plain text, as they appear on the page. This approach eliminates the need for switching fonts for a single character, and it avoid the usage of variation selectors. Moreover, experts, such as Dai Matsui, state that there is a practical need to distinguish between these variant forms for purposes of research: when studying Old Uyghur manuscripts, there is a requirement to reproduce the orientation of the tail in order to faithfully document fragmented and illegible letters for aiding future decipherment. At present, the variant forms should be considered stylistic variants and controlled through font selection

6.2 Combining signs

The following combining signs are used for disambiguation and representation of new sounds (see § 8.2):

Character name	Glyph
OLD UYGHUR COMBINING DOT RIGHT	◌𐽑
OLD UYGHUR COMBINING TWO DOTS RIGHT	◌𐽒
OLD UYGHUR COMBINING THREE DOTS RIGHT	◌𐽓
OLD UYGHUR COMBINING DOT LEFT	𐽐◌
OLD UYGHUR COMBINING TWO DOTS LEFT	𐽏◌
OLD UYGHUR COMBINING THREE DOTS LEFT	𐽎◌
OLD UYGHUR COMBINING HAMZA LEFT	𐽔◌

These signs are used as follows:

- The signs ◌́, ◌̤, etc. are analogous to Sogdian diacritics, eg. ◌̣ U+10F46 SOGDIAN COMBINING DOT BELOW and ◌̤ U+10F47 SOGDIAN COMBINING TWO DOTS BELOW. They are commonly used for differentiating between letters whose shapes are similar in particular styles of the script, and for indicating sounds for which distinctive letters do not exist in the script. These signs are commonly used with *nun*, *gimel*, *zayin*, *heth*, and *samekh*.

- The signs ◌̋, ◌̤, and ◌̃ were used in later documents of an administrative nature for representing non-Turkic sounds, especially those occuring in words of Arabic origin (see fig. 21). In such documents they occur with the letters *gimel*, *heth*, and *samekh*.

In Old Uyghur manuscripts, dot diacritics appear as elongated strokes, which are reflective of the scribal aesthetics of the script. In some manuscripts these diacritics are written as true dots or squared dots. Despite the variations in their shapes, these signs are palaeographically dots, and therefore, it is appropriate to refer to them as such in the names for the proposed character.

These signs function similarly to the *nuqṭa* diacritic, which is used in Brahmi-based scripts for representing sounds foreign to Indic languages, eg. ◌़ U+093C DEVANAGARI SIGN NUKTA. While it may be possible to encode combinations of base letter + combining sign as atomic letters, it is practical to avoid such an approach. Encoding such atomic letters is strongly not recommended as there are other combining signs used in Old Uyghur manuscripts, which have not been fully investigated for the present proposal. It is quite likely that additional combining signs will need to be encoded. As a result, it will be necessary to encode new sets of atomic letters for each every base letter + combining sign combination when a new combining sign is added to the repertoire. The proposed approach of using combining signs follows the model for Sogdian, from which Old Uyghur is derived.

There are other signs, such as ◌∘ ('ring right', as it would appear in a conventional vertical context), which are used in some documents for transcription. 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. These additional combining signs may be added to the proposed block in the future.

6.3 Punctuation signs

Spacing is used for separating words. The following signs are used for punctuation (see fig. 23 for examples):

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. 26, 31). The sign 𐽑 is also used as a general delimiter. When it occurs in documents where 𐽒 is used, it represents short segments of text and may function as a comma or semi-colon.

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 𐽕 is used as a general sign of punctuation and decoration, for example in fig. 28, 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 § 8.3 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	𐽑

6.5 Editoral signs

The following editorial signs are used in manuscripts:

Character name	Glyph
OLD UYGHUR DELETION MARK	◌𐽐

When written beneath a word or letter, this sign indicates that the respective text is an error and is to be omitted. The correct word is generally written after the mispelled word (see fig. 22).

6.6 Line-breaking

There are no formal rules for the breaking of Old Uyghur text at the 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.

6.7 Collation

The sort order for Old Uyghur follows the encoded order:

𐽀 ALEPH < 𐽁 ALEPH WITH LEFT TAIL < 𐽂 ALEPH-NUN WITH RIGHT TAIL < 𐽃 BETH <
𐽄 GIMEL-HETH < 𐽅 WAW < 𐽆 ZAYIN < 𐽇 FINAL HETH < 𐽈 YODH < 𐽉 KAPH <
𐽊 LAMEDH < 𐽋 MEM < 𐽌 NUN < 𐽍 SAMEKH < 𐽎 PE < 𐽏 SADHE < 𐽐 RESH <
𐽑 TAW < 𐽒 LESH

7 Joining behavior

The contextual forms of dual-joining letters are shown below:

Dual-joining letters

	iso	fin	med	init
ALEPH	𐽀	𐽀	𐽁	𐽁
BETH	𐽂	𐽂	𐽃	𐽃
GIMEL-HETH	𐽄	𐽄	𐽅	𐽅
WAW	𐽆	𐽆	𐽇	𐽇
ZAYIN	𐽈	𐽈	𐽉	𐽉
YODH	𐽊	𐽊	𐽋	𐽋
KAPH	𐽌	𐽌	𐽍	𐽍
LAMEDH	𐽎	𐽎	𐽏	𐽏
MEM	𐽐	𐽐	𐽑	𐽑
NUN	𐽒	𐽒	𐽓	𐽓
SAMEKH	𐽔	𐽔	𐽕	𐽕
PE	𐽖	𐽖	𐽗	𐽗
SADHE	𐽘	𐽘	𐽙	𐽙
RESH	𐽚	𐽚	𐽛	𐽛
SHIN	𐽜	𐽜	𐽝	𐽝
TAW	𐽞	𐽞	𐽟	𐽟
LESH	𐽠	𐽠	𐽡	𐽡

The contextual forms of right-joining letters are shown below:

Right-joining letters

	iso	fin
ALEPH WITH LEFT TAIL	𐽀	𐽀
ALEPH-NUN WITH RIGHT TAIL	𐽁	𐽁
FINAL HETH	𐽂	𐽂

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, 𐽆 TAW, 𐽅 WAW, 𐽃 NUN

bʾmsʾn
‘vamsan’

𐽅 BETH, 𐽀 ALEPH, 𐽉 MEM, 𐽑 SAMEKH, 𐽀 ALEPH, 𐽃 NUN

wyγwr
‘üigür’

𐽅 WAW, 𐽇 YODH, 𐽒 GIMEL, 𐽅 WAW, 𐽄 RESH

qwtlwγ
‘qutlug’

𐽂 HETH, ◌𐽙 COMBINING TWO DOTS LEFT, 𐽆 TAW, 𐽐 LESH, 𐽒 GIMEL

mncwšry
‘mancusari’

𐽉 MEM, 𐽃 NUN, 𐽋 SADHE, 𐽅 WAW, 𐽑 SAMEKH, ◌𐽘 COMBINING TWO DOTS RIGHT, 𐽄 RESH, 𐽇 YODH

swδwr
‘sutur’

𐽑 SAMEKH, 𐽅 WAW, 𐽁 LAMEDH, 𐽅 WAW, 𐽄 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

7.1 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 Unicode control character ZWNJ U+200C ZERO WIDTH NON-JOINER (abbreviated as 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.

ʾ*wyz*ʾ
‘üzä’

ALEPH, WAW, YODH, ZAYIN, ALEPH

ʾ*wyz*ʾ
‘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

7.2 Glyph interactions

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

7.2.1 *aleph*

When *aleph* occurs in final position after *kaph* and *pe*, it is rendered using a contextual variant. In block-print styles, when *aleph* follows *lamedh* it is written using a contextual variant. These are shown below:

Character sequence	Alternate	Regular
<KAPH, ALEPH>		
<LAMEDH, ALEPH>		
<PE, ALEPH>		

7.2.2 *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>		

7.2.3 *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>.

7.2.4 *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’		
kʾlmʾdwk ‘kälmädük’		

8 Encoded representations

8.1 Vowels

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

	Initial	Medial
ä	 ALEPH	 ALEPH
a, *e*	 ALEPH, ALEPH	 ALEPH
i, *ï*	 ALEPH, YODH	YODH
ī, *ī̈*	 ALEPH, YODH, YODH	 YODH, YODH

o, *u*	𐽰	𐽆 ALEPH, 𐽶 WAW	𐽱	𐽶 WAW
ö, *ü*	𐽲	𐽆 ALEPH, 𐽶 WAW, 𐽶 YODH	𐽱	𐽶 WAW
ö, *ü*	𐽳	𐽶 WAW, 𐽶 YODH	𐽳	𐽶 WAW, 𐽶 YODH
ō, *ō̈*, *ū*, *ṻ*	𐽴	𐽆 ALEPH, 𐽶 WAW, 𐽶 WAW	𐽳	𐽶 WAW, 𐽶 WAW

8.2 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:

		iso	fin	med	init	
dotted *gimel*	*γ*	𐽂◌̇	𐽂◌̇	𐽃◌̇	𐽃◌̇	𐽂 GIMEL-HETH, ◌̇ COMBINING DOT LEFT
two-dotted *gimel*	*γ*	𐽂◌̈	𐽂◌̈	𐽃◌̈	𐽃◌̈	𐽂 GIMEL-HETH, ◌̈ COMBINING TWO DOTS LEFT
dotted *zayin*	*ž*	𐽀◌̇	𐽀◌̇	𐽁◌̇	𐽁◌̇	𐽀 ZAYIN, ◌̇ COMBINING DOT RIGHT
two-dotted *zayin*	*ž*	𐽀◌̈	𐽀◌̈	𐽁◌̈	𐽁◌̈	𐽀 ZAYIN, ◌̈ COMBINING TWO DOTS RIGHT
dotted *heth*	*q*	𐽇◌̇	𐽇◌̇	—	—	𐽇 FINAL HETH, ◌̇ COMBINING DOT LEFT
two-dotted *heth*	*q*	𐽇◌̈	𐽇◌̈	𐽈◌̈	𐽈◌̈	𐽇 FINAL HETH, ◌̈ COMBINING TWO DOTS LEFT
dotted *nun*	*n*	𐽆◌̇	𐽆◌̇	𐽉◌̇	𐽉◌̇	𐽆 NUN, ◌̇ COMBINING DOT LEFT
two-dotted *shin*	*š*	𐽅◌̈	𐽅◌̈	𐽊◌̈	𐽊◌̈	𐽅 SHIN, ◌̈ COMBINING TWO DOTS RIGHT

8.3 Stem extension

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

tynly lr-r
'tinlag-lar-r'

𐽆 TAW, 𐽇 YODH, 𐽈 NUN, 𐽉 LESH, 𐽊 GIMEL, SP SPACE,
𐽉 LESH, 𐽇 RESH, SP SPACE,
𐽑 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.

9 Character Properties

9.1 Core data: `UnicodeData.txt`

```
10F70;OLD UYGHUR LETTER ALEPH;Lo;0;AL;;;;;N;;;;;
10F71;OLD UYGHUR LETTER ALEPH WITH LEFT TAIL;Lo;0;AL;;;;;N;;;;;
10F72;OLD UYGHUR LETTER ALEPH-NUN WITH RIGHT TAIL;Lo;0;AL;;;;;N;;;;;
10F73;OLD UYGHUR LETTER BETH;Lo;0;AL;;;;;N;;;;;
10F74;OLD UYGHUR LETTER GIMEL-HETH;Lo;0;AL;;;;;N;;;;;
10F75;OLD UYGHUR LETTER WAW;Lo;0;AL;;;;;N;;;;;
10F76;OLD UYGHUR LETTER ZAYIN;Lo;0;AL;;;;;N;;;;;
10F77;OLD UYGHUR LETTER FINAL HETH;Lo;0;AL;;;;;N;;;;;
10F78;OLD UYGHUR LETTER YODH;Lo;0;AL;;;;;N;;;;;
10F79;OLD UYGHUR LETTER KAPH;Lo;0;AL;;;;;N;;;;;
10F7A;OLD UYGHUR LETTER LAMEDH;Lo;0;AL;;;;;N;;;;;
10F7B;OLD UYGHUR LETTER MEM;Lo;0;AL;;;;;N;;;;;
10F7C;OLD UYGHUR LETTER NUN;Lo;0;AL;;;;;N;;;;;
10F7D;OLD UYGHUR LETTER SAMEKH;Lo;0;AL;;;;;N;;;;;
10F7E;OLD UYGHUR LETTER PE;Lo;0;AL;;;;;N;;;;;
10F7F;OLD UYGHUR LETTER SADHE;Lo;0;AL;;;;;N;;;;;
10F80;OLD UYGHUR LETTER RESH;Lo;0;AL;;;;;N;;;;;
10F81;OLD UYGHUR LETTER SHIN;Lo;0;AL;;;;;N;;;;;
10F82;OLD UYGHUR LETTER TAW;Lo;0;AL;;;;;N;;;;;
10F83;OLD UYGHUR LETTER LESH;Lo;0;AL;;;;;N;;;;;
10F84;OLD UYGHUR COMBINING DOT RIGHT;Mn;220;NSM;;;;;N;;;;;
10F85;OLD UYGHUR COMBINING TWO DOTS RIGHT;Mn;220;NSM;;;;;N;;;;;
10F86;OLD UYGHUR COMBINING THREE DOTS RIGHT;Mn;220;NSM;;;;;N;;;;;
10F87;OLD UYGHUR COMBINING DOT LEFT;Mn;230;NSM;;;;;N;;;;;
10F88;OLD UYGHUR COMBINING TWO DOTS LEFT;Mn;230;NSM;;;;;N;;;;;
10F89;OLD UYGHUR COMBINING THREE DOTS LEFT;Mn;230;NSM;;;;;N;;;;;
10F8A;OLD UYGHUR COMBINING HAMZA RIGHT;Mn;220;NSM;;;;;N;;;;;
10F8B;OLD UYGHUR PUNCTUATION BAR;Po;0;AL;;;;;N;;;;;
10F8C;OLD UYGHUR PUNCTUATION TWO BARS;Po;0;AL;;;;;N;;;;;
10F8D;OLD UYGHUR PUNCTUATION TWO DOTS;Po;0;AL;;;;;N;;;;;
10F8E;OLD UYGHUR PUNCTUATION FOUR DOTS;Po;0;AL;;;;;N;;;;;
10F8F;OLD UYGHUR PUNCTUATION FIVE DOTS;Po;0;AL;;;;;N;;;;;
10F90;OLD UYGHUR SECTION MARK;Po;0;AL;;;;;N;;;;;
10F91;OLD UYGHUR STEM EXTENDER;Po;0;AL;;;;;N;;;;;
10F92;OLD UYGHUR DELETION MARK;Mn;220;NSM;;;;;N;;;;;
```

9.2 Linebreak data: `LineBreak.txt`

```
10F70..10F83;AL # Lo [20] OLD UYGHUR LETTER ALEPH..OLD UYGHUR LETTER LESH
10F84..10F8A;CM # Mn  [7] OLD UYGHUR COMBINING DOT RIGHT..
                          OLD UYGHUR COMBINING HAMSA RIGHT
10F8B..10F90;AL # Po  [6] OLD UYGHUR PUNCTUATION BAR..OLD UYGHUR SECTION MARK
10F91;AL        # Po       OLD UYGHUR STEM EXTENDER
10F92;CM        # Mn       OLD UYGHUR DELETION MARK
```

9.3 Property list: `PropList.txt`

```
10F92           ; Extender # Po       OLD UYGHUR STEM EXTENDER
```

9.4 Shaping properties: `ArabicShaping.txt`

```
10F70; OLD UYGHUR ALEPH; D; No_Joining_Group
10F71; OLD UYGHUR ALEPH WITH LEFT TAIL; R; No_Joining_Group
10F72; OLD UYGHUR ALEPH-NUN WITH RIGHT TAIL; R; No_Joining_Group
10F73; OLD UYGHUR BETH; D; No_Joining_Group
10F74; OLD UYGHUR GIMEL-HETH; D; No_Joining_Group
10F75; OLD UYGHUR WAW; D; No_Joining_Group
10F76; OLD UYGHUR ZAYIN; D; No_Joining_Group
10F77; OLD UYGHUR FINAL HETH; R; No_Joining_Group
10F78; OLD UYGHUR YODH; D; No_Joining_Group
10F79; OLD UYGHUR KAPH; D; No_Joining_Group
10F7A; OLD UYGHUR LAMEDH; D; No_Joining_Group
10F7B; OLD UYGHUR MEM; D; No_Joining_Group
10F7C; OLD UYGHUR NUN; D; No_Joining_Group
10F7D; OLD UYGHUR SAMEKH; D; No_Joining_Group
10F7E; OLD UYGHUR PE; D; No_Joining_Group
10F7F; OLD UYGHUR SADHE; D; No_Joining_Group
10F80; OLD UYGHUR RESH; D; No_Joining_Group
10F81; OLD UYGHUR SHIN; D; No_Joining_Group
10F82; OLD UYGHUR TAW; D; No_Joining_Group
10F83; OLD UYGHUR LESH; D; No_Joining_Group
```

10 References

Anderson, Deborah, et. al. 2018. “Recommendations to UTC #155 April-May 2018 on Script Proposals” (L2/18-168). https://www.unicode.org/L2/L2018/18168-script-rec.pdf

China. 2018. “GB/T 36331-2018 ”Information technology – Uigur-Mongolian characters, presentation characters and use rules of controlling characters”. http://c.gb688.cn/bzgk/gb/showGb?type=online&hcno=DFE87CC79EA67F8BF8B37C9C41CF9348

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.

Hamilton James R. 1986. *Manuscrits ouïgours du IXè-Xè siècle de Touen-Houang*: textes établis, traduits, et commentés. Tome 1. Paris.

Israpil, Dilara and Yüsüp, Israpil. 2014. "Two Old Uighur Account Documents from Toqquzsaray Ruins in Maralbeši". *Studies on the Inner Asian Languages*, vol. 29, pp. 137–156.

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.

Matsui, Dai. 2017. "An Old Uigur Account Book for Manichaean and Buddhist Monasteries from Temple α in Qočo". *Zur lichten Heimat: Studien zu Manichäismus, Iranistik und Zentralasienkunde im Gedenken an Werner Sundermann*, Herausgegeben von einem Team „Turfanforschung", pp. 409–420. Wiesbaden: Harrassowitz Verlag.

———. 2018. "Comments on the preliminary proposal to encode Old Uyghur in Unicode (L2/18-126)". L2/18-335. https://www.unicode.org/L2/L2018/18335-old-uyghur-cmt.pdf

Moriyasu, Takao. 1989. "Notes on Uighur Documents (I)." *Studies on the Inner Asian Languages*, vol. 4, pp. 51–76.

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. 2016a. "Proposal to encode the Old Sogdian script in Unicode". L2/16-312R. http://www.unicode.org/L2/L2016/16312r-old-sogdian.pdf

———. 2016b. "Revised proposal to encode the Sogdian script in Unicode". L2/16-371R2. 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.

Shōgaito, Masahiro. 1988. "Passages from Abhidharma-nyāyānusāra-śāstra. Quoted in the Uighur Text Or. 8212-75B, British Library". *Studies on the Inner Asian Languages*, vol. 3, pp. 159–207.

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`

Wilkens, Jens. 2016. "Buddhism in the West Uyghur Kingdom and Beyond". *Transfer of Buddhism Across Central Asian Networks (7th to 13th Centuries)*, Carmen Meinert (ed.), pp. 191–249. Brill: Leiden and Boston.

Yakup, Abdurishid. 2011. "An Old Uyghur fragment of the Lotus Sūtra from the Krotkov collection in St. Petersburg." *Acta Orientalia Academiae Scientiarum Hungaricae*, vol. 64, no. 4 (December), pp. 411–426.

Zieme, Peter. 1975. "Zur Buddhistischen Stabreimdichtung der alten Uiguren". *Acta Orientalia Academiæ Scientiarum Hungaricæ*, vol. 29, no. 2, pp. 187–211.

———. 1991. *Die Stabreimtexte der Uiguren von Turfan und Dunhuang: Studien zur alttürkischen Dichtung*. Bibliotheca orientalis Hungarica, v. XXXIII. Budapest: Akadémiai Kiadó.

11 Acknowledgments

I express my gratitude to Dai Matsui (Graduate School of Letters, Osaka University) for sharing his wealth of knowledge about the Old Uyghur script, particularly its historical evolution, orthographic intracacies, and stylistic variations. I truly appreciate him for providing samples of manuscripts and for pointing me to specific examples of letterforms and orthographic issues.

I thank Liang Hai for providing extensive comments on all previous versions of this proposal. The present document owes its shape to his meticulous review and recommendations, especially regarding the representation of *gimel* and *heth*.

Roozbeh Pournader has provided invaluable comments and suggestions regarding the representation of right-to-left scripts. Not only this proposal, but every proposal I've submitted, and by extension every script I've encoded in Unicode, would not have been possible without his expertise and insight.

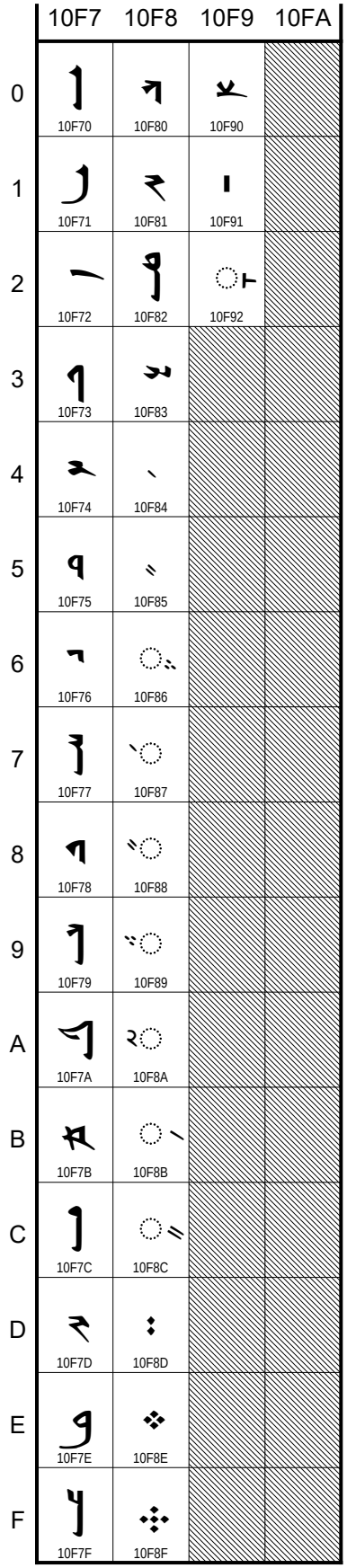
I also would like to thank Andrew West for providing reproductions and transcriptions of the Old Uyghur inscriptions on the walls of the Cloud Platform at Juyong Guan (fig. 46).

Digitized images of Old Uyghur documents from the Berlin Turfan Collection and the International Dunhuang Project have been used in accordance with academic fair-use conventions. I express my gratitude to Berlin-Brandenburgischen Akademie der Wissenchaften (Staatsbibliotek zu Berlin, Preussischer Kulturbesitz Orientabteilung) and to the British Library for making these images available. Documents from BBAW and BL are cited throughout this document with the shelfmark.

Deborah Anderson and the Script Ad Hoc Committee of the Unicode Technical Committee have patiently reviewed all version of this proposal and provided feedback necessary for developing a robust encoding for Old Uyghur. I thank Dr. Anderson for her diligence in collating and relaying comments and for providing me with the scholarly support necessary for detailed research of this script.

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. I thank Lisa Moore, president of the Consortium, and the board for their continued support of my work.

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 ALEPH WITH LEFT TAIL
10F72 OLD UYGHUR LETTER ALEPH-NUN WITH RIGHT TAIL
10F73 OLD UYGHUR LETTER BETH
10F74 OLD UYGHUR LETTER GIMEL-HETH
10F75 OLD UYGHUR LETTER WAW
10F76 OLD UYGHUR LETTER ZAYIN
10F77 OLD UYGHUR LETTER FINAL HETH
10F78 OLD UYGHUR LETTER YODH
10F79 OLD UYGHUR LETTER KAPH
10F7A OLD UYGHUR LETTER LAMEDH
10F7B OLD UYGHUR LETTER MEM
10F7C OLD UYGHUR LETTER NUN
10F7D OLD UYGHUR LETTER SAMEKH
10F7E OLD UYGHUR LETTER PE
10F7F OLD UYGHUR LETTER SADHE
10F80 OLD UYGHUR LETTER RESH
10F81 OLD UYGHUR LETTER SHIN
10F82 OLD UYGHUR LETTER TAW
10F83 OLD UYGHUR LETTER LESH
• hooked r

Combining signs

10F84 OLD UYGHUR COMBINING DOT RIGHT
10F85 OLD UYGHUR COMBINING TWO DOTS RIGHT
10F86 OLD UYGHUR COMBINING THREE DOTS RIGHT
10F87 OLD UYGHUR COMBINING DOT LEFT
10F88 OLD UYGHUR COMBINING TWO DOTS LEFT
10F89 OLD UYGHUR COMBINING THREE DOTS LEFT
10F8A OLD UYGHUR COMBINING HAMZA LEFT

Punctuation

10F8B OLD UYGHUR PUNCTUATION BAR
10F8C OLD UYGHUR PUNCTUATION TWO BARS
10F8D OLD UYGHUR PUNCTUATION TWO DOTS
10F8E OLD UYGHUR PUNCTUATION FOUR DOTS
10F8F OLD UYGHUR PUNCTUATION FIVE DOTS
10F90 OLD UYGHUR SECTION MARK

Stem extender

10F91 OLD UYGHUR STEM EXTENDER

Editorial mark

10F92 OLD UYGHUR DELETION MARK

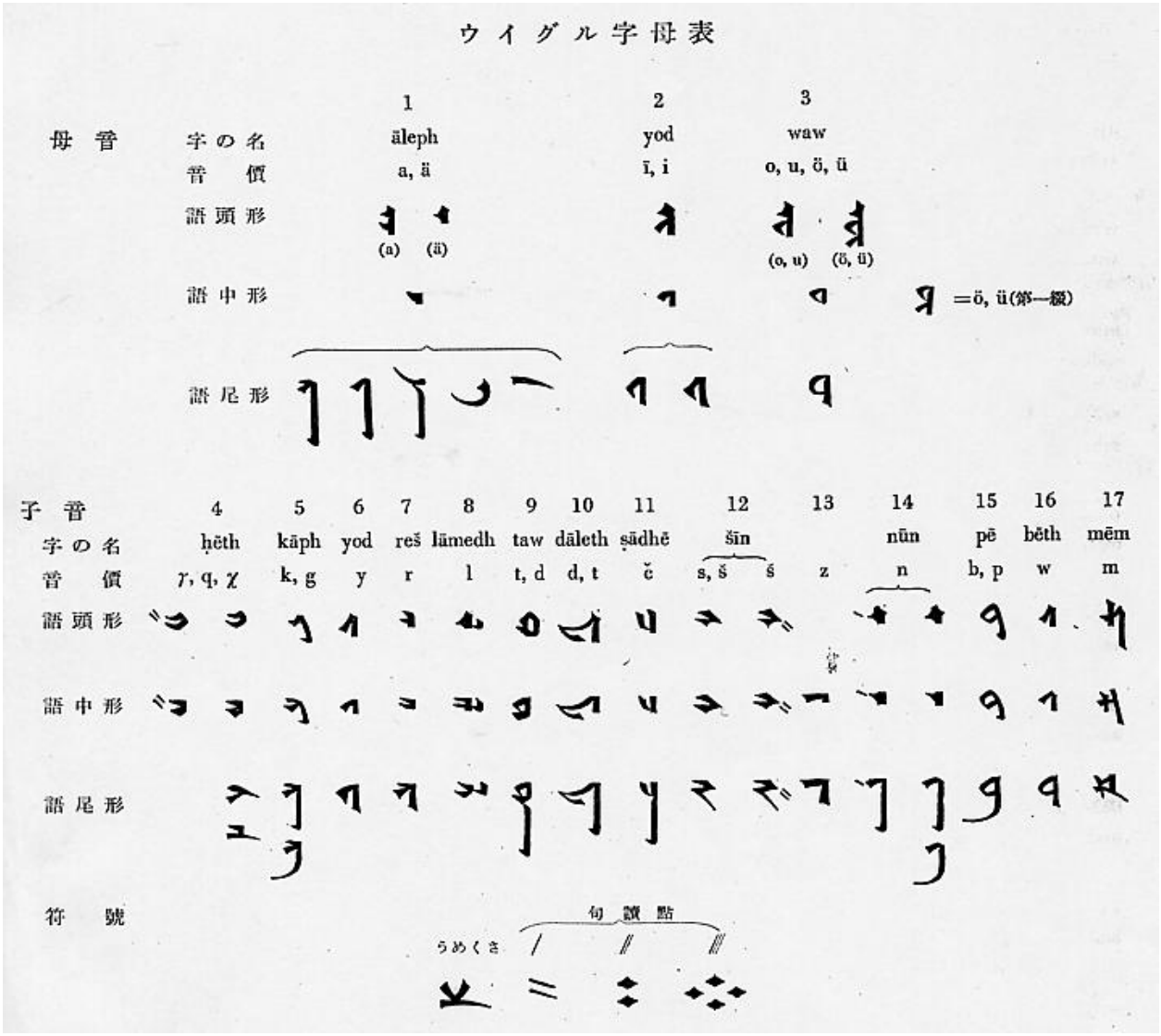


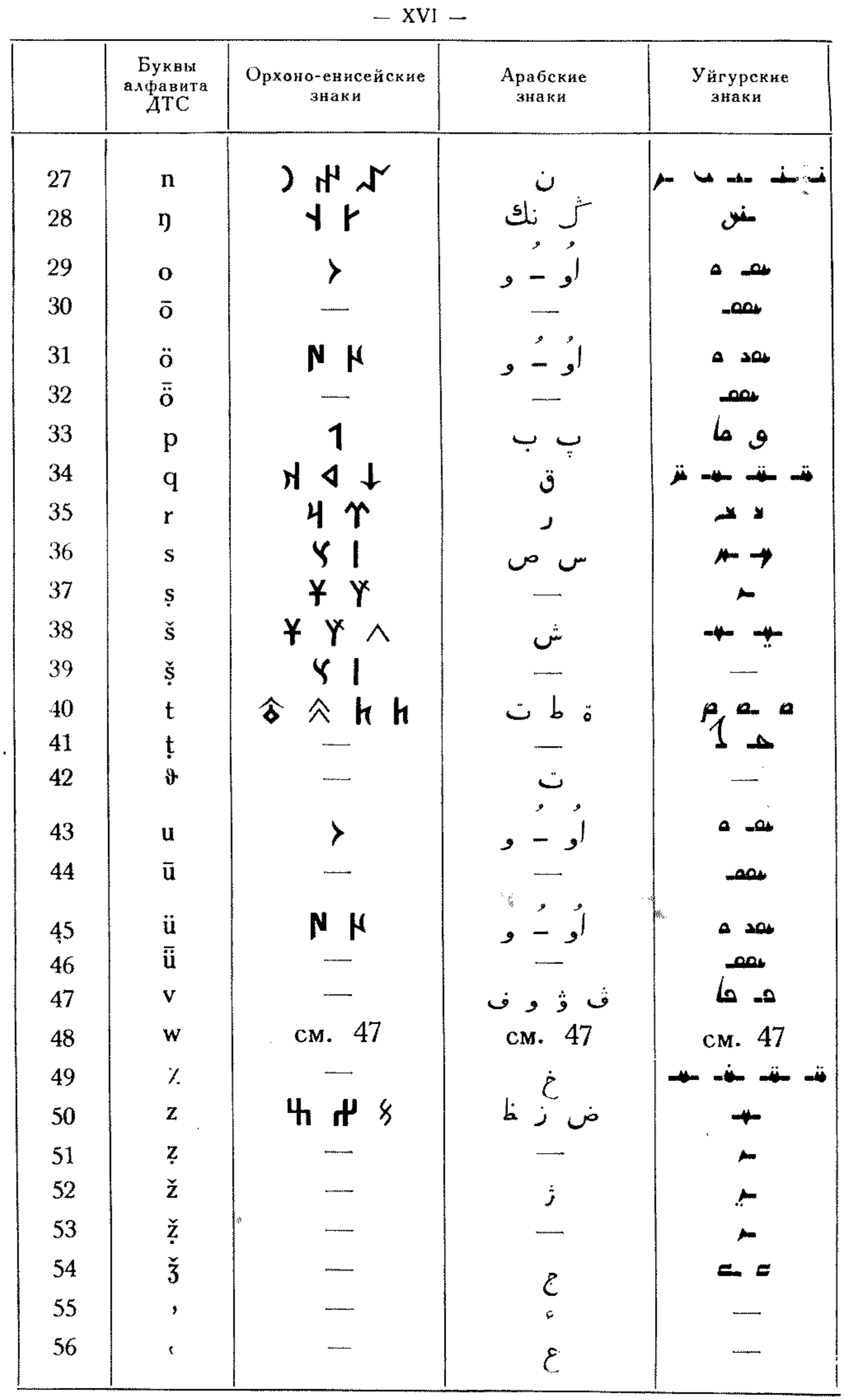
Figure 1: Table of Old Uyghur characters used in the Uyghur inscription in the multi-script Yuan dynasty inscriptions at Juyong Guan 居庸關 pass at the Great Wall northwest of Beijing (from Chü-Yung-Kuan 居庸關, "The Buddhist Arch of the Fourteenth Century A.D. at the Pass of the Great Wall Northwest of Peking", vol. 1, p. 165; reproduced from West 2006). See photograph containing an excerpt of the inscription in fig. 46.

Note: there are a few inaccurate assignment of names for graphemes based upon phonetic value. The glyphs shown for final *beth* (#16) is actually *waw*. The likely reason is that final /b/ does not occur in texts from this period and the original form became obsolete. #13 is unnamed, but it is clearly *zayin*. #10 is not *lamedh*, not *daleth*, which does not occur in Old Uyghur. #8 is the 'hooked' *resh* (LESH, not *lamedh*

— XV —

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

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		ن	
28	ŋ		ڭ نك	
29	o		او – و	
30	ō	—	—	
31	ö		او – و	
32	ȫ	—	—	
33	p		پ ب	
34	q		ق	
35	r		ر	
36	s		س ص	
37	ṣ		—	
38	š		ش	
39	ṣ̌		—	—
40	t		ة ط ت	
41	ṭ	—	—	
42	ϑ	—	ث	—
43	u		او – و	
44	ū	—	—	
45	ü		او – و	
46	ǖ	—	—	
47	v	—	ڤ ۋ و ف	
48	w	см. 47	см. 47	см. 47
49	γ	—	غ	
50	z		ض ز ظ	
51	ẓ	—	—	
52	ž	—	ژ	
53	ẓ̌	—	—	
54	ǯ	—	ج	
55	ʼ	—	ء	—
56	ʻ	—	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. 55.

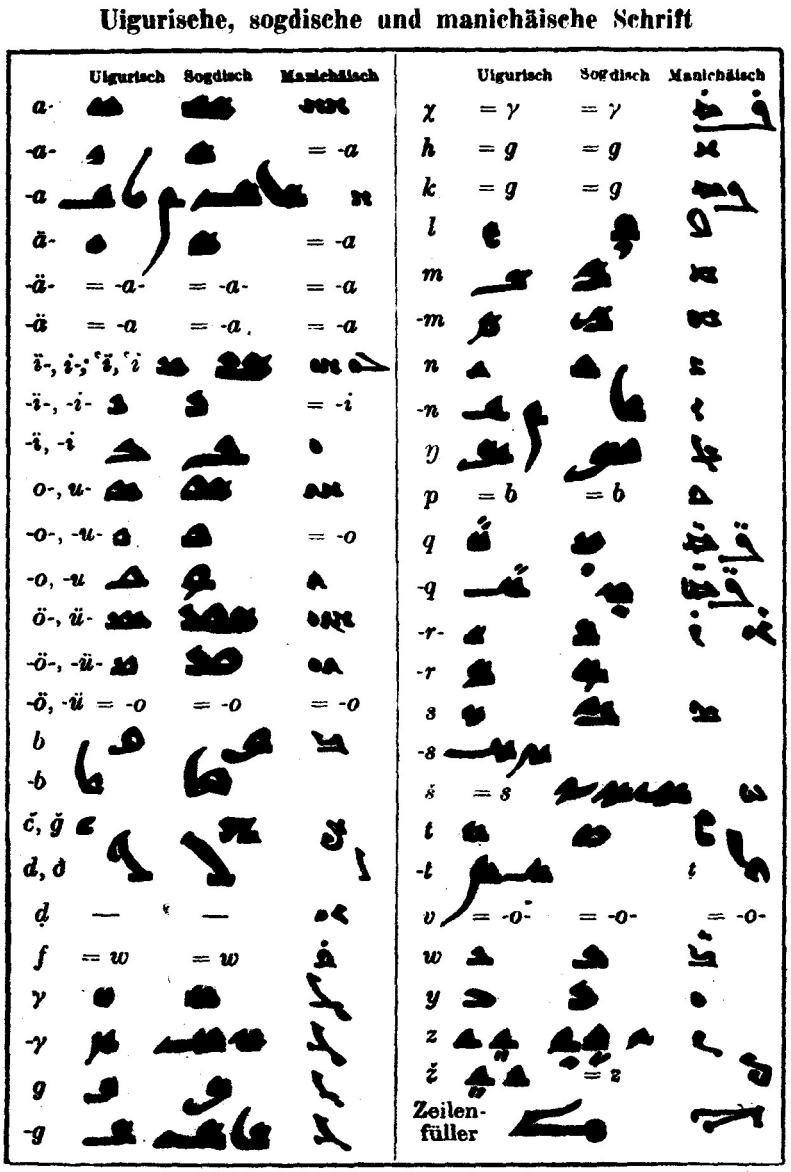


Figure 5: Comparison of Old Uyghur, Sogdian, and Manichaean letters (from von Gabain 1950: 17). For clearer examples of Old Uyghur letterforms referenced by von Gabain see the three Old Uyghur manuscripts, two in the formal script and the third in the cursive script, illustrated and transcribed in her work, reproduced here in fig. 49–53.

Schrifttabelle 349

Transliteration	1 *M* III Nr. 8 VII marg. (10. Jh. ?)	2 T IV Xusup (10. Jh. ?)	3 Kāšγārī Faksimile S. 6 (1072)	4 *ETṢ* Nr. 11 (Text 0) (13./14. Jh.)
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				

Figure 6: Chart showing development and variation in the Old Uyghur script from the 10th through 14th century (from Zieme 1991: 349).

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				

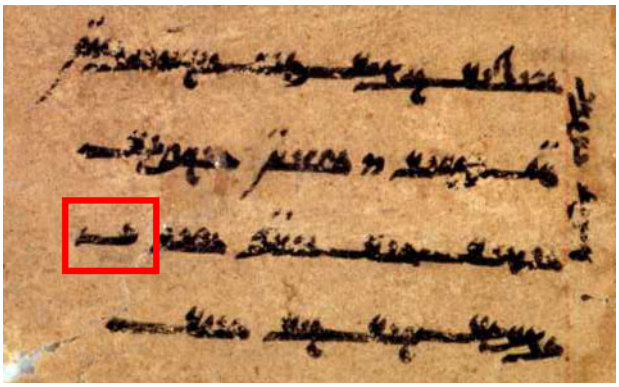
Table 2 Various forms of the Old Uighur alphabet from texts dating between the fourteenth and the tenth centuries BCE
Source: *adapted from Zieme 1991*

Figure 7: Comparison of Old Uyghur letterforms (from Coulmas 1996: 526). As stated by Coulmas, this chart is a copy of that shown in Zieme 1991 (shown here in fig. 6). Although it is an exact duplicate of Zieme’s chart, Coulmas’s chart is given here as an example of the inclusion of the Old Uyghur script in general reference handbooks on writing systems.

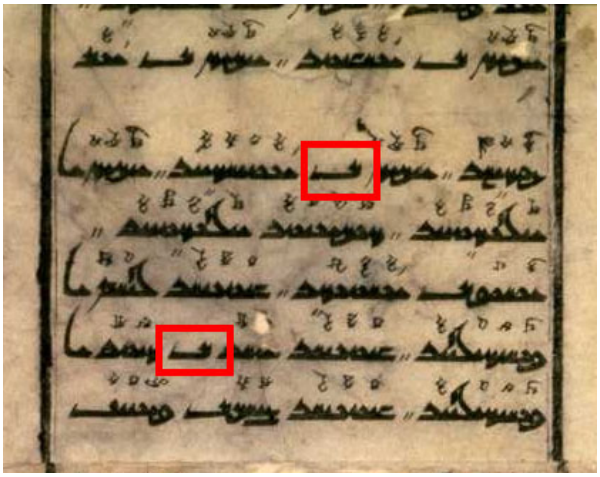
Compared transcription system for Old Uighur Alphabet

	Berliner Transkription system	Turkey	transcription at *Uigurisches Wörterbuch*	transliteration at *Uigurisches Wörterbuch*
𐽀	a, ạ	a, ạ	a	’’ / ’
𐽁	b	b	b	P
𐽂	č	ç	č	Č
𐽃	d, ṭ	d, ṭ	d, ḍ	D, T
𐽄	ä, ’ä	e, ’e	ä	’
𐽅	[e] i	ė / i	e	Y / ’Y
𐽆	g	g	g	K
𐽇	γ / γ́	g / ġ	g	Q, Q̈, Q̇
𐽈	h / χ, x, ẍ	h / ẖ, ḥ	h	H / X
𐽉	ï	ı	ı	Y, Y
𐽊	i	i	i	Y, ’Y
𐽋	ǰ̣, ž	j	ǰ̣, ž	J̣̌, Ẓ̌, Ẓ
𐽌	k	k	k	K
𐽍	[k] q, q̈, q̇	k / ḳ	k	K / Q, Q̈, Q̇
𐽎	l	l	l	L
𐽏	m	m	m	M
𐽐	n, ṅ	n, ṅ	n	N, Ṅ
𐽑	ng, ñ, ŋ	n͡g, ng, ñ	ŋ	NK
𐽒	o	o	o	W / ’W
𐽓	ö, o̱	ö, o̱	ö	W / WY / ’WY
𐽔	p	p	p	P
𐽕	r	r	r	R
𐽖	s, ẓ	s, ẓ	s, ṣ	S, Z
𐽗	š	ṣ	š	Ṣ, Ṣ
𐽘	t, ḍ	t, ḍ	t, ṭ	T, D
𐽙	u	u	u	W / ’W
𐽚	ü, u̱	ü, u̱	ü	W / WY / ’WY
[𐽛] 𐽜	v	v	v	V
𐽝	y	y	y	Y
𐽞	z, ṣ	z, ṣ	z, ẓ	Z, S

Figure 8: Comparison of transliteration schemes for Old Uyghur (from Olmez 2016).

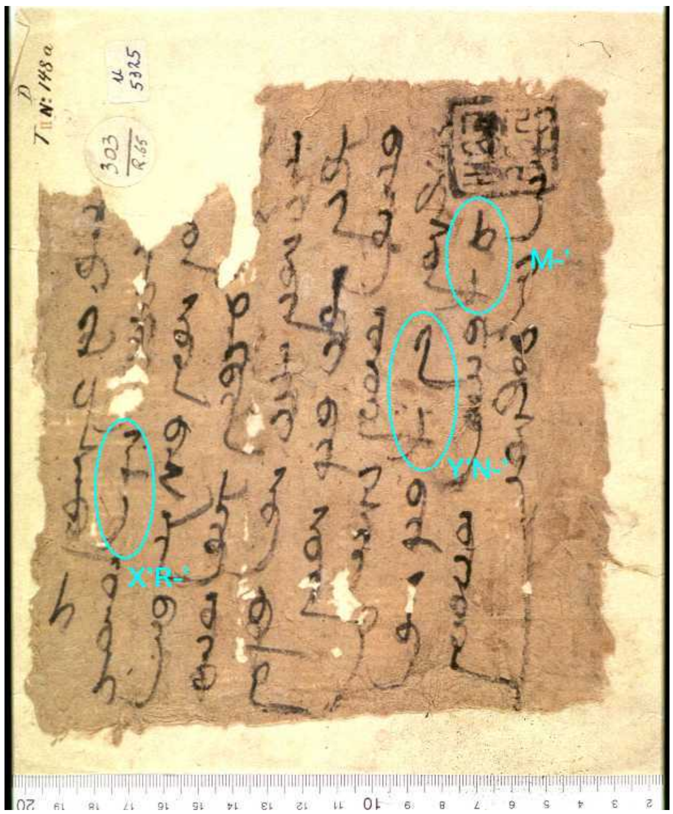


Example of the regular 𐽀 independent *aleph* (excerpt from U 2215).



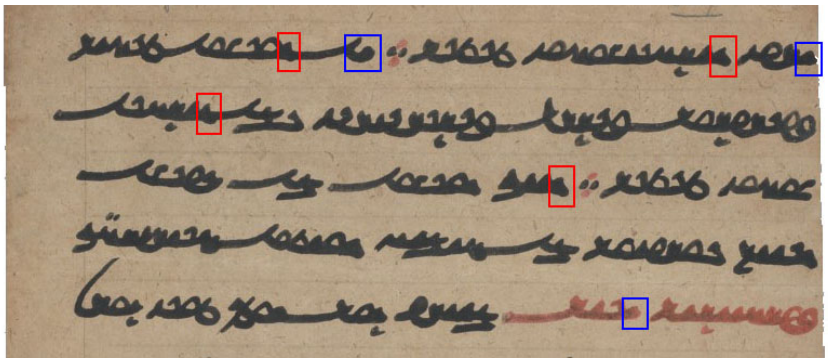
Examples of the alternate form 𐽁 of independent *aleph* (excerpt from Mainz 801).

Figure 9: Forms of independent *aleph*. Images have been rotated 90 degrees counter-clockwise for layout purposes.

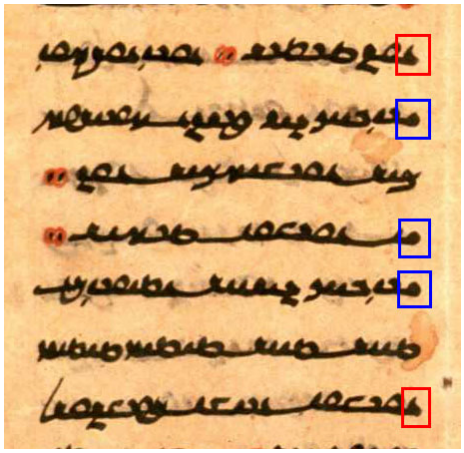


Usage of ALEPH-NUN WITH RIGHT TAIL and ALEPH WITH LEFT TAIL in U 5325. Annotations produced by Dai Matsui, November 2018.

Figure 10: Alternate forms of final *aleph*.

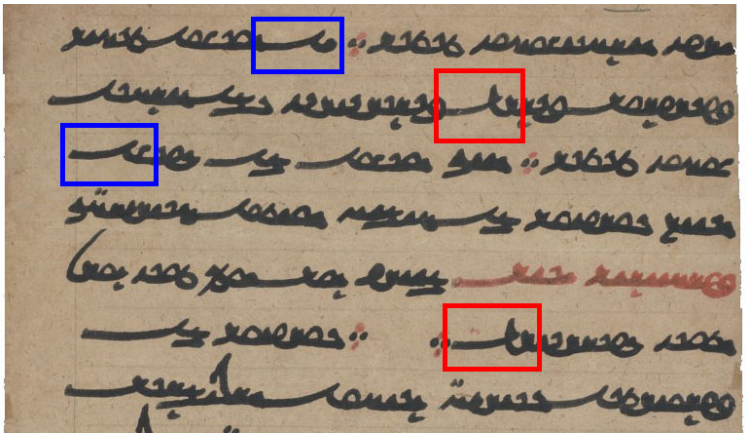


Excerpt from Pelliot Ouïgour 13 showing initial forms of ▲ *aleph* (red) and ◓ *nun* (blue).

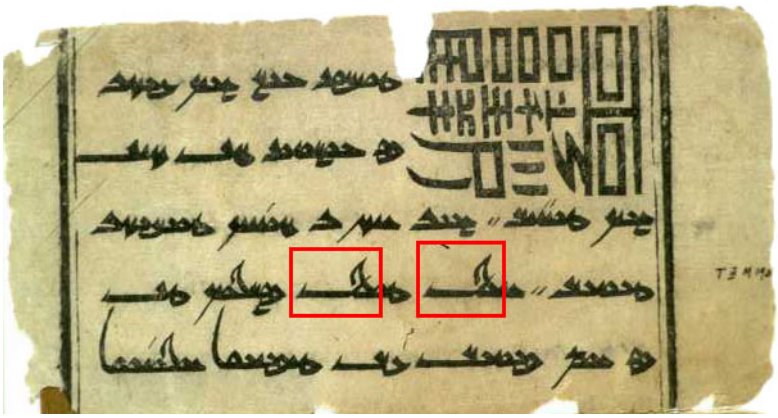


Excerpt from Mainz 126 showing initial forms of ▲ *aleph* (red) and ◓ *nun* (blue).

Figure 11: Examples of *aleph* and *nun*. Images have been rotated 90 degrees counter-clockwise for layout purposes.

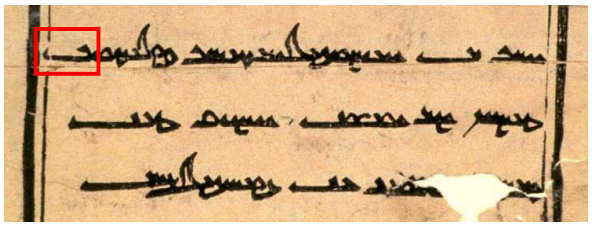


Excerpt from Pelliot Ouïgour 13 showing the final form of *aleph* used with *kaph* (red), compared with the regular form (blue).

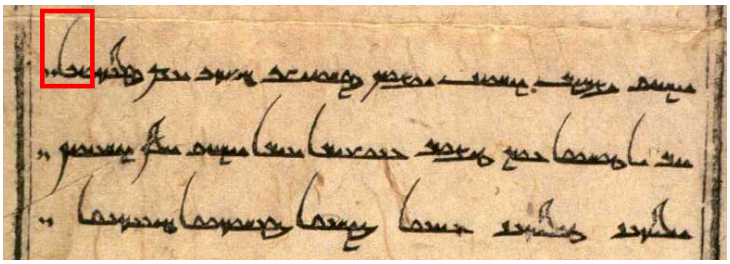


Folio from U 4960 showing the contextual form of *aleph* used with *lamedh*.

Figure 12: Examples of contextual variants of *aleph*. Images have been rotated 90 degrees counter-clockwise for layout purposes.

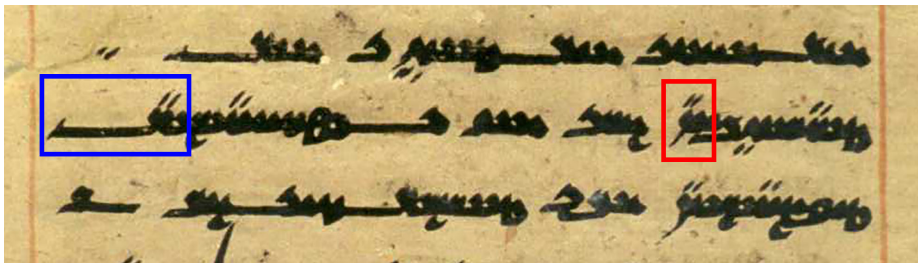


Excerpt from U 4708 showing final 𐽆 in *bodistb* ‘bodhisattva’.



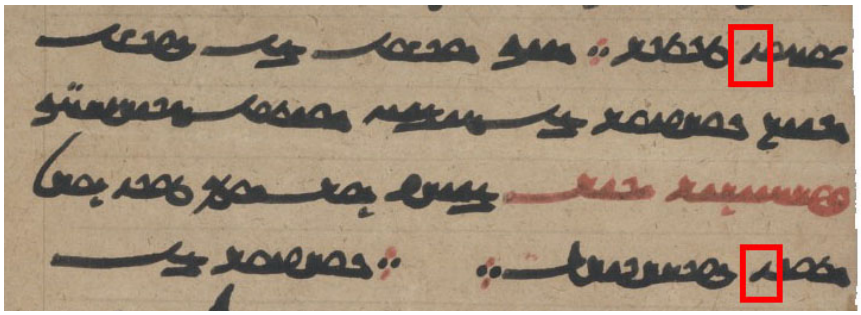
Excerpt from U 4707 showing final *beth* written using a variant form 𐽆 with left-ward tail in *bodistb* ‘bodhisattva’.

Figure 13: Examples of *beth*. Images have been rotated 90 degrees counter-clockwise for layout purposes.

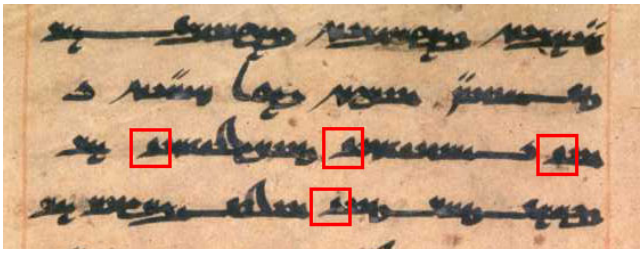


Excerpt from U 924 showing final forms of *gimel* (red) and *heth* (blue).

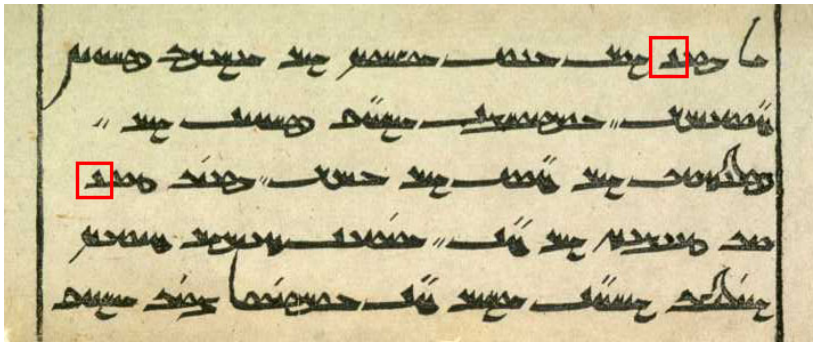
Figure 14: Examples of *gimel* and *heth*. Images have been rotated 90 degrees counter-clockwise for layout purposes.



Excerpt from Pelliot Ouïgour 13 showing a hand-written form of ◌ *zayin*.

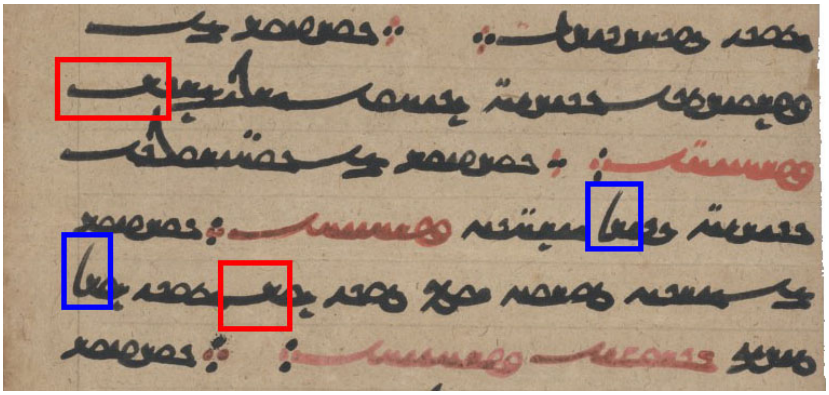


Excerpt from Mainz 119 showing a hand-written form of ◌ *zayin*.

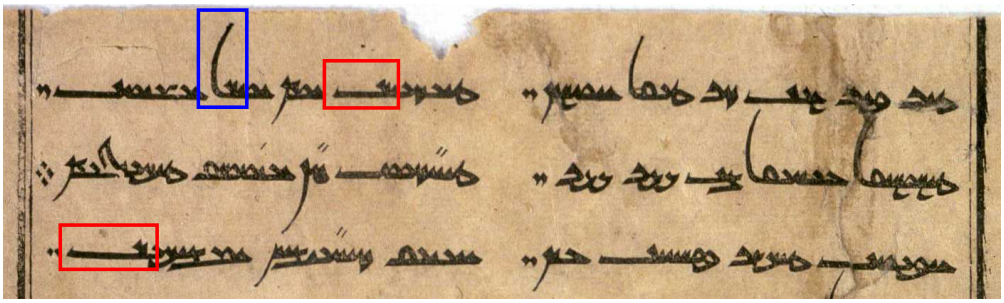


Excerpt from U 387 showing a block-print form of ◌ *zayin*.

Figure 15: Examples of *zayin*. Images have been rotated 90 degrees counter-clockwise for layout purposes.



Usage of the regular final form ـک of *kaph* (red) and the alternate final form ـڪ (blue) in a manuscript (excerpt from Pelliot Ouïgour 13)



Usage of the regular final form ـک of *kaph* (red) and the alternate final form ـڪ (blue) in a block print (excerpt from U 4301)

Figure 16: Examples of *kaph*. Images have been rotated 90 degrees counter-clockwise for layout purposes.

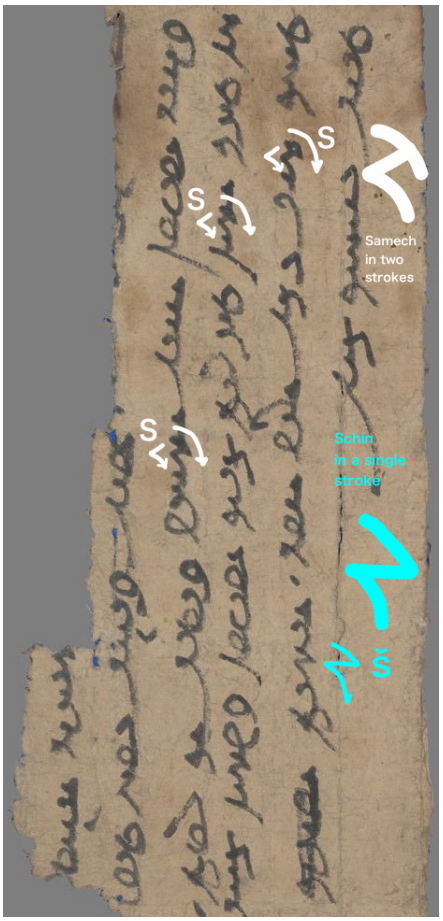
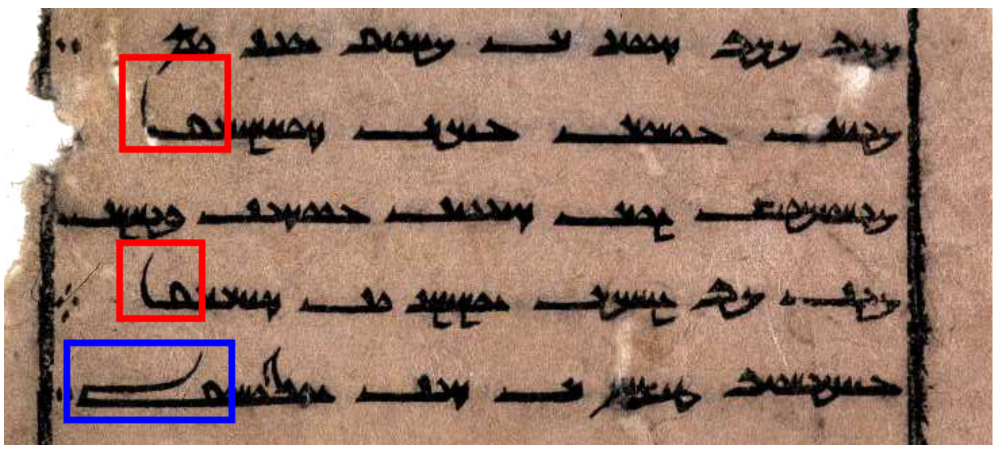
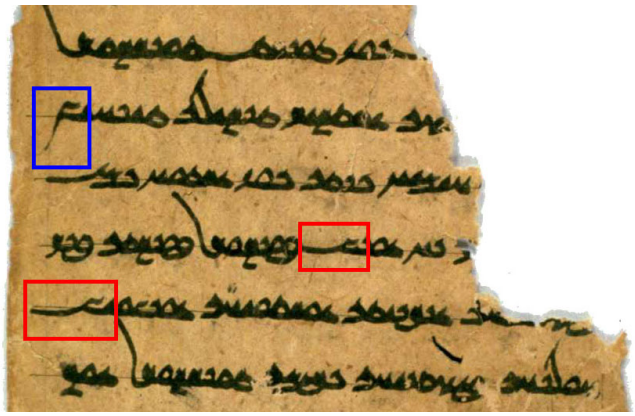


Figure 17: Excerpt from Pelliot Ouïgour 5 (9th–10th c.) showing the distinction between 𐽎 SAMEKH and 𐽔 SHIN. Annotations produced by Dai Matsui, August 2018.

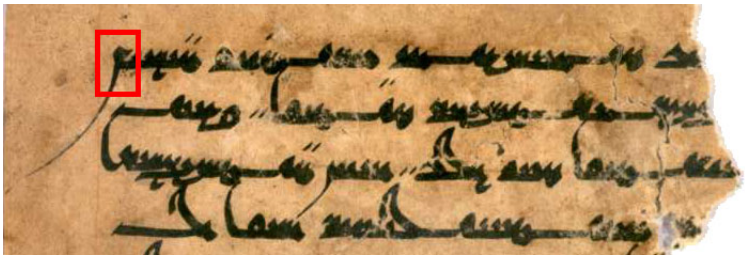


Regular 𐾆 (red) and ornamental 𐾆 (blue) forms forms of final *pe* in a block print (excerpt from U 4750)

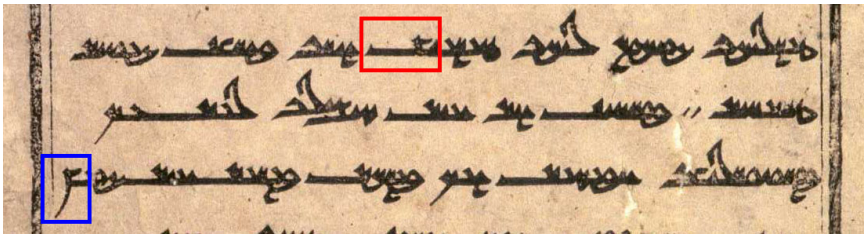
Figure 18: Examples of *pe*. Images have been rotated 90 degrees counter-clockwise for layout purposes.



Usage of the alternate 𐽆 (blue) and regular final 𐽇 (red) of *sadhe* in a manuscript (excerpt from Mainz 302)

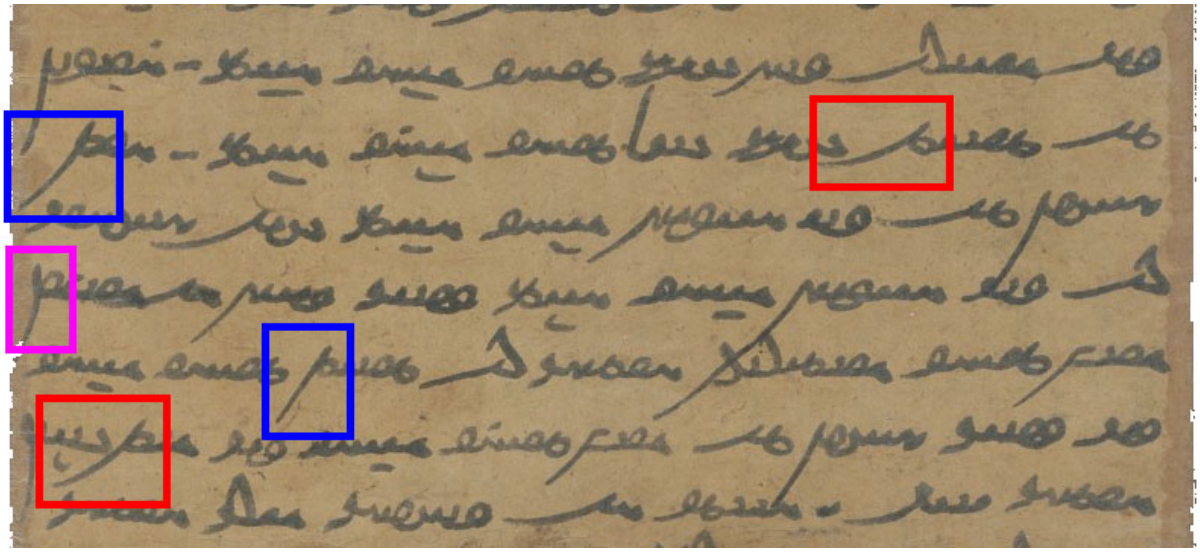


Usage of the alternate 𐽆 *sadhe* in a manuscript (excerpt from Mainz 393)



Usage of the alternate 𐽆 (blue) and regular final 𐽇 (red) of *sadhe* in a block print (excerpt from U 4680)

Figure 19: Examples of final *sadhe*. Images have been rotated 90 degrees counter-clockwise for layout purposes.



Comparison of the regular final form ـه (red) of *taw* with the alternate ـم (blue) and the sequence *waw*+*nun* (*magenta*) in a manuscript (excerpt from Pelliot Chinois 3046).

Figure 20: Examples of *taw*. Images have been rotated 90 degrees counter-clockwise for layout purposes.

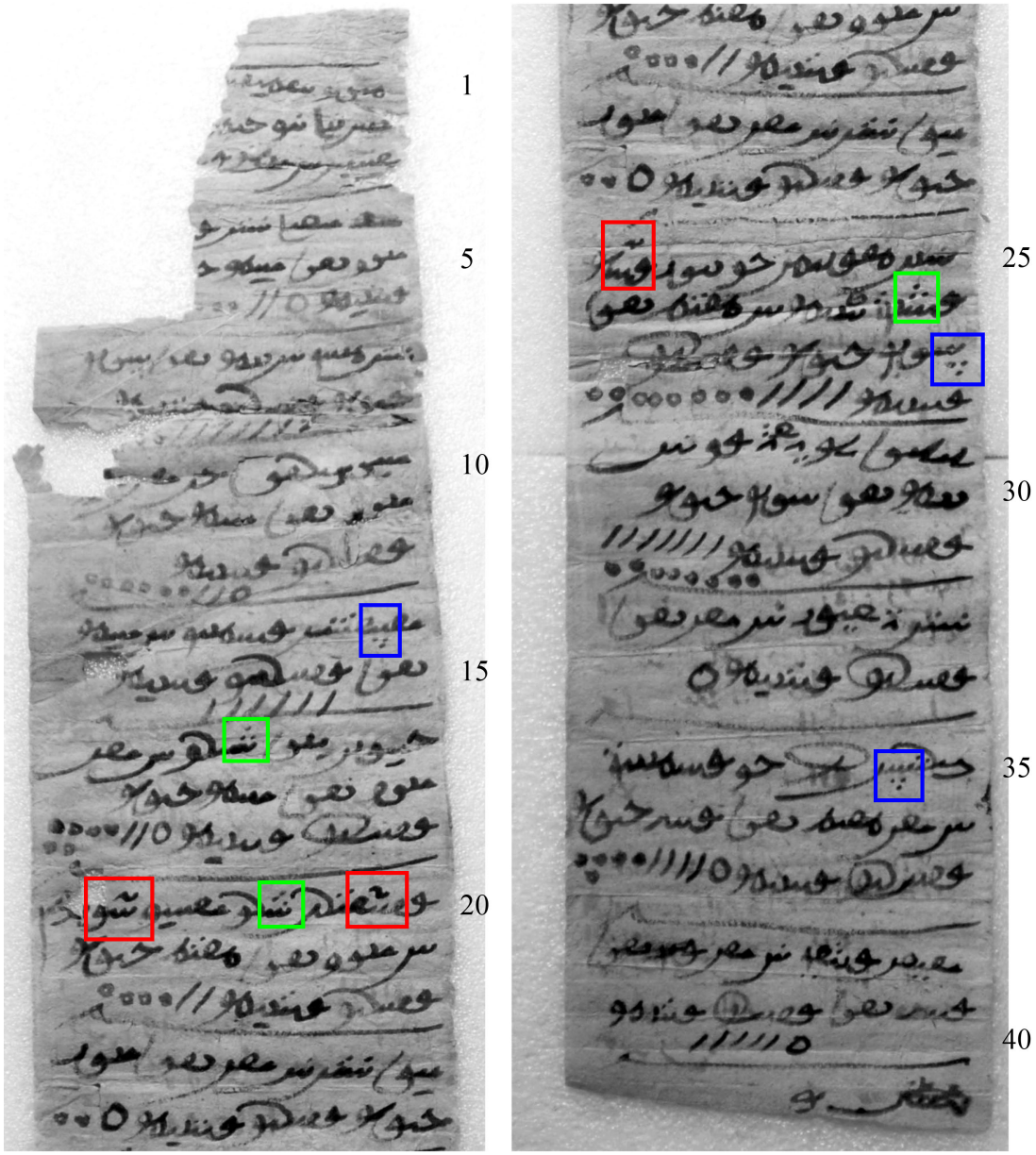
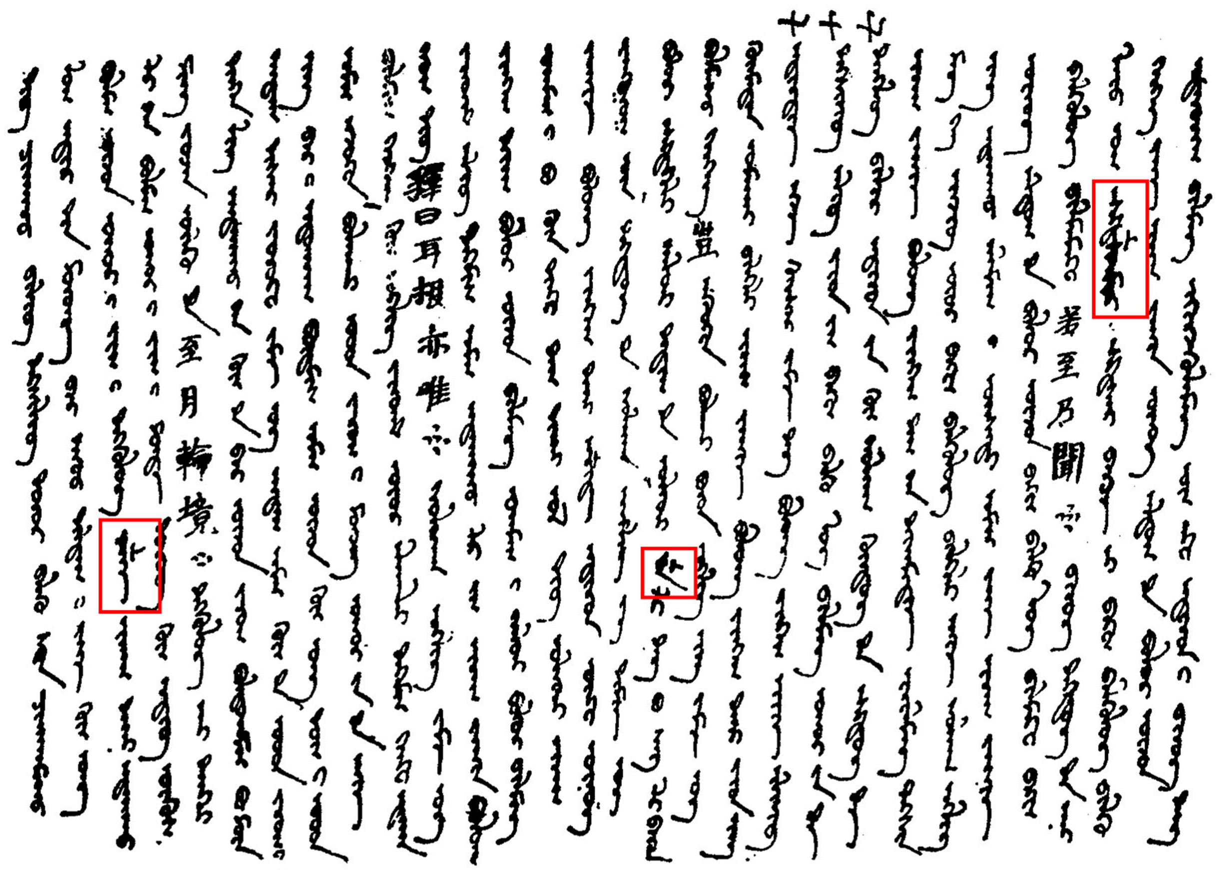
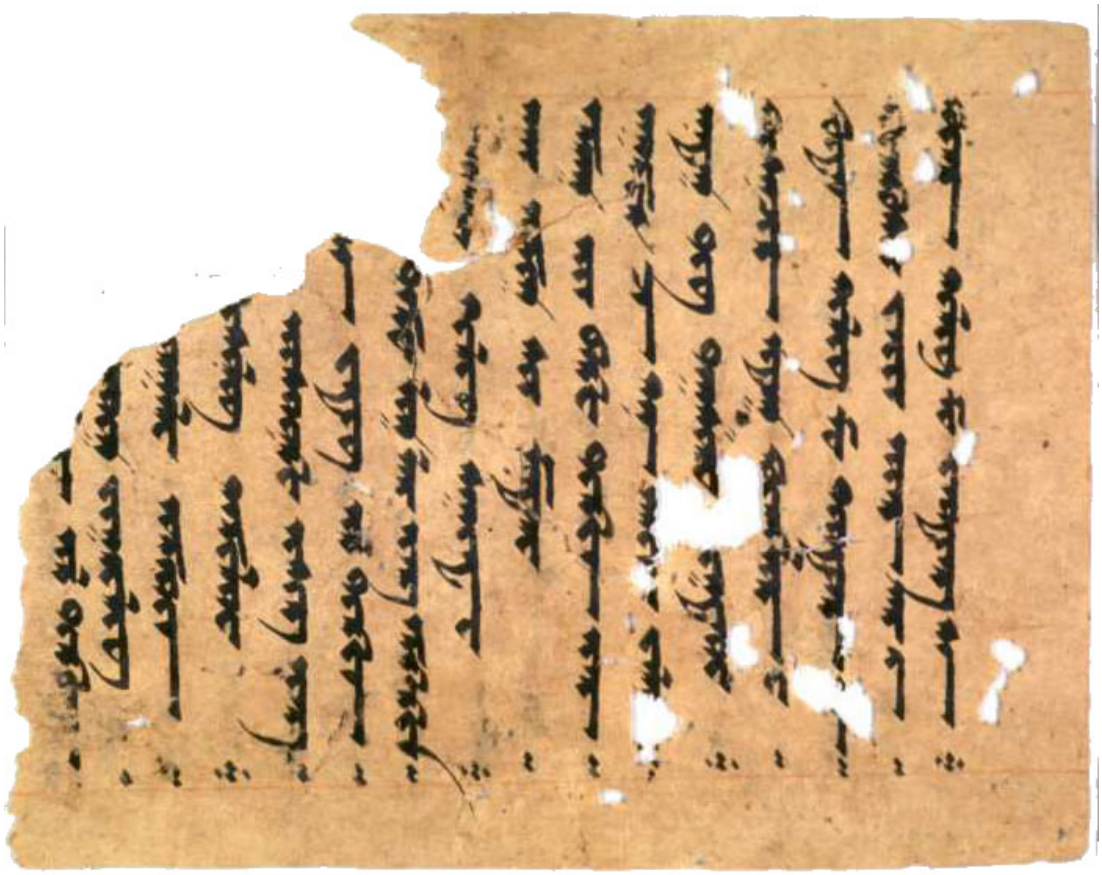


Figure 21: Usage of ◌ (blue), ◌ (green), and ◌ (red) for transcribing Arabic in a Old Uyghur administrative document (from Israpil 2014: plate I).



大英図書館所蔵 Or. 8212-75B 頁76B-77A（本文テクスト219—239）

Figure 22: Usage of the ◌⊢ deletion mark for indicating error correction in Or. 8212/75, an Old Uyghur manuscript containing passages of the of the Buddhist text *Abhidharma-nyāyānusāra-śāstra* (from Shōgaito 1988: 207). Note the intralinear text in Han characters.



The punctuation signs ⁌ TWO DOTS and ❖ FOUR DOTS at the bottom margin (Mainz 36).

Figure 23: Examples of punctuation signs

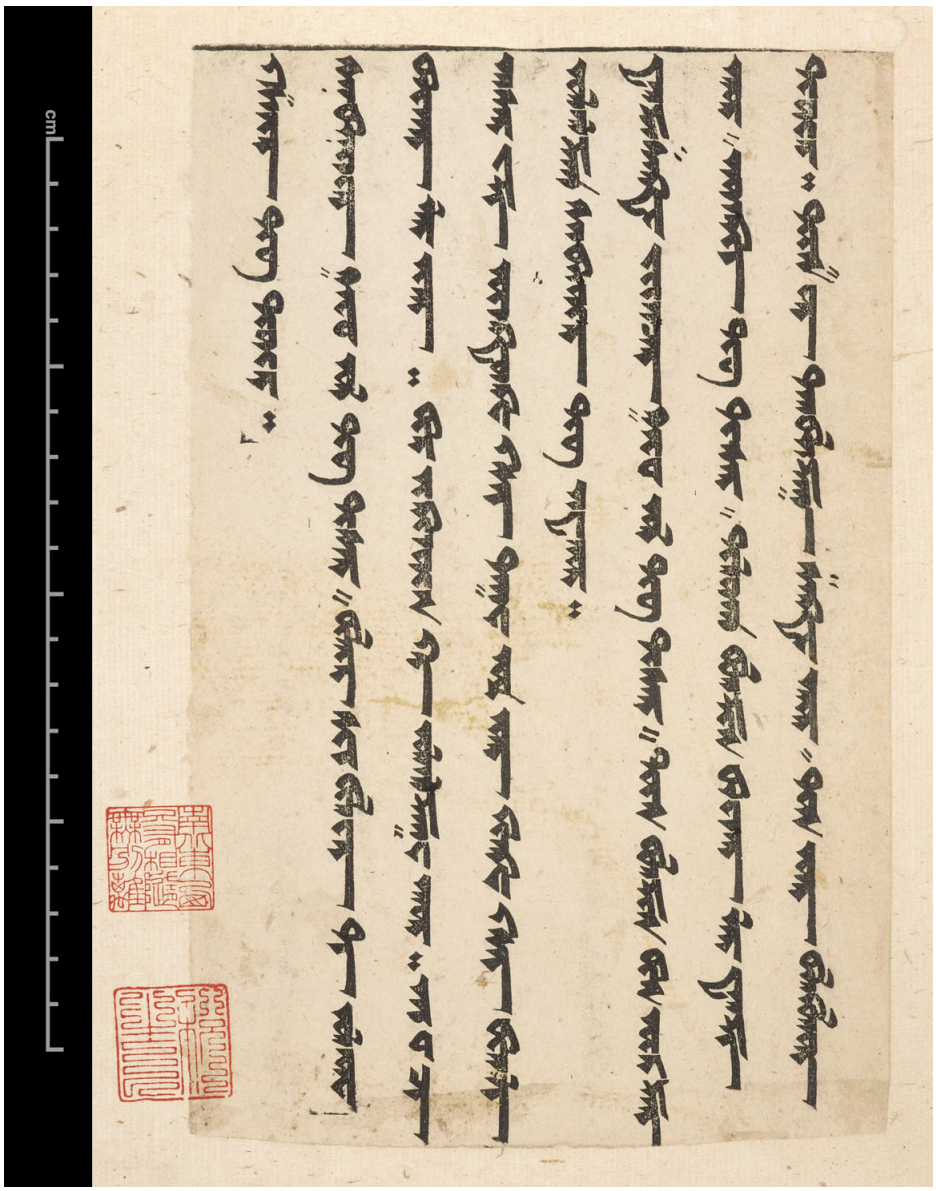


Figure 24: Princeton East Asian Library, PEALD 6a, recto. Block print.

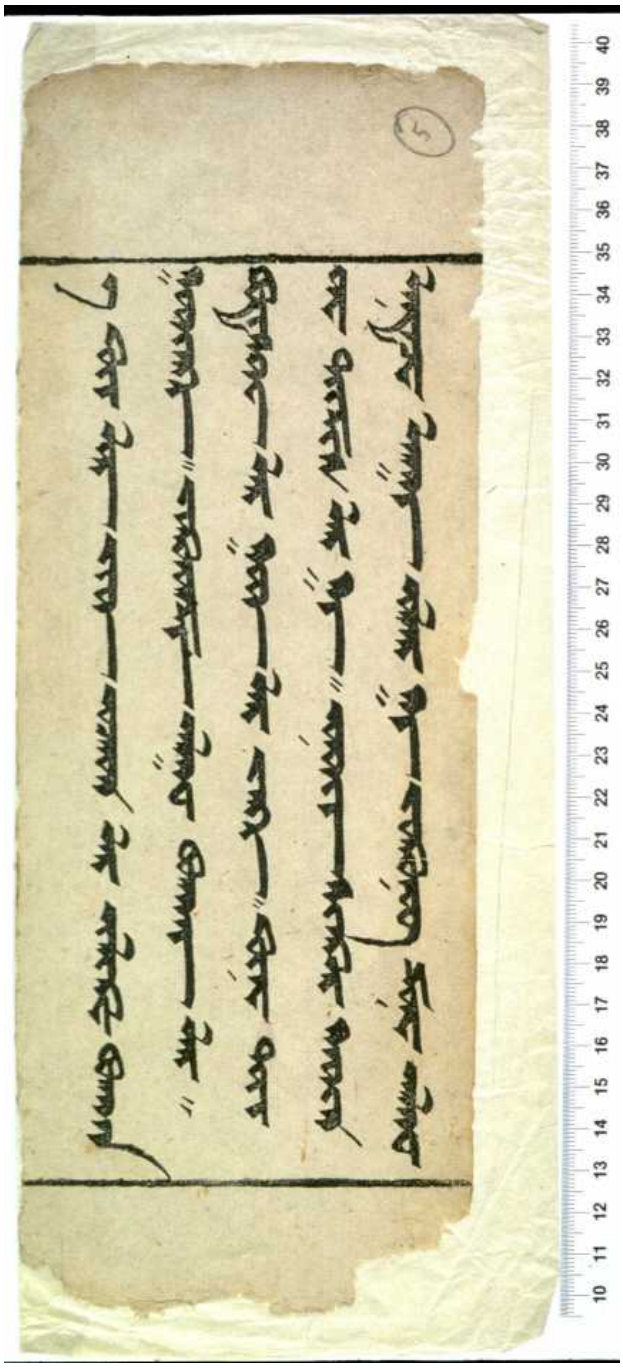


Figure 25: BBAW, U 387, recto. Block print.

Figure 26: BBAW, U 4960, folio 1, recto. Block print. Seal in Han characters.

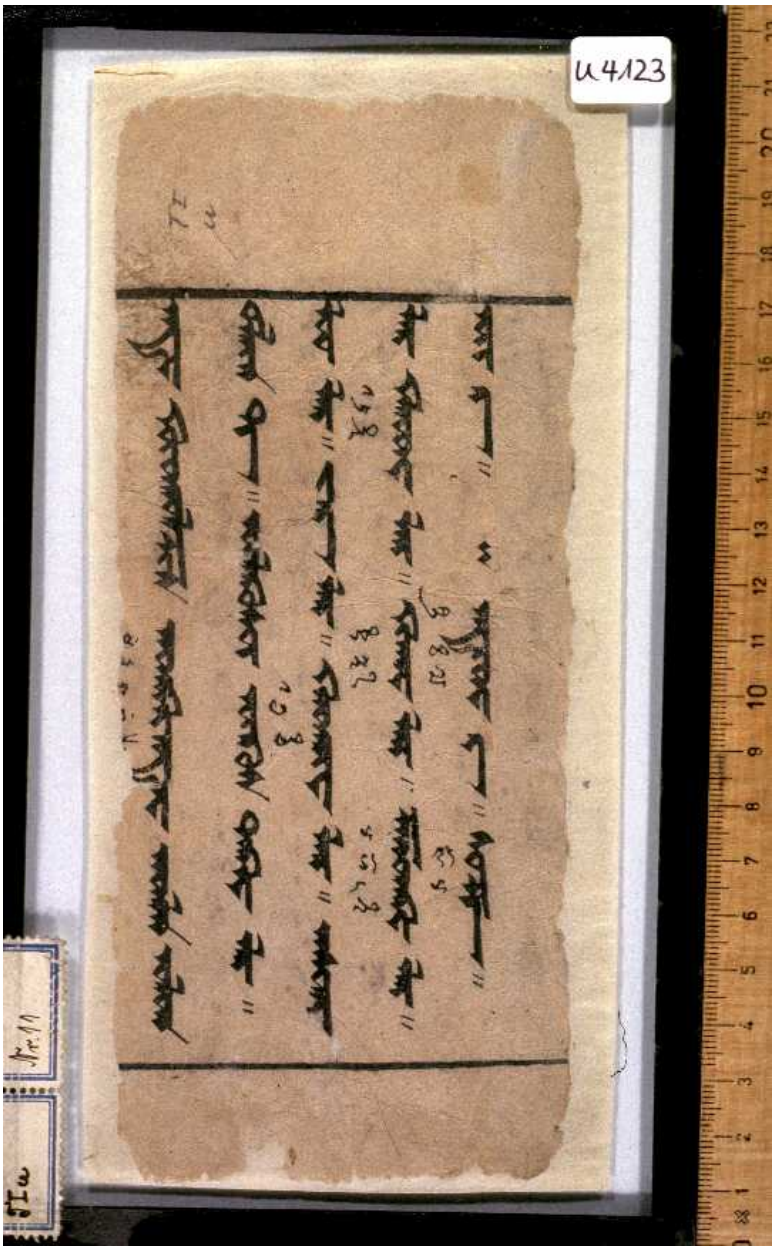


Figure 27: BBAW, U 4123. Block print.

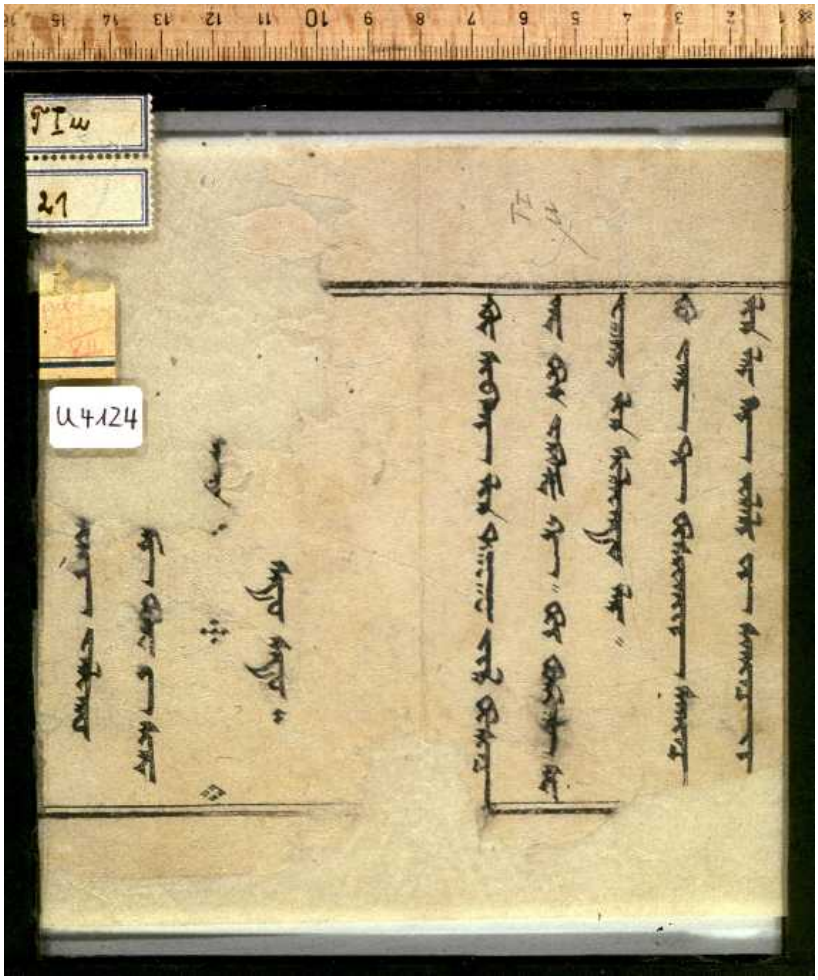


Figure 28: BBAW, U 4124. Block print.

Figure 29: BBAW, U 343, folio 1, recto. Block print.

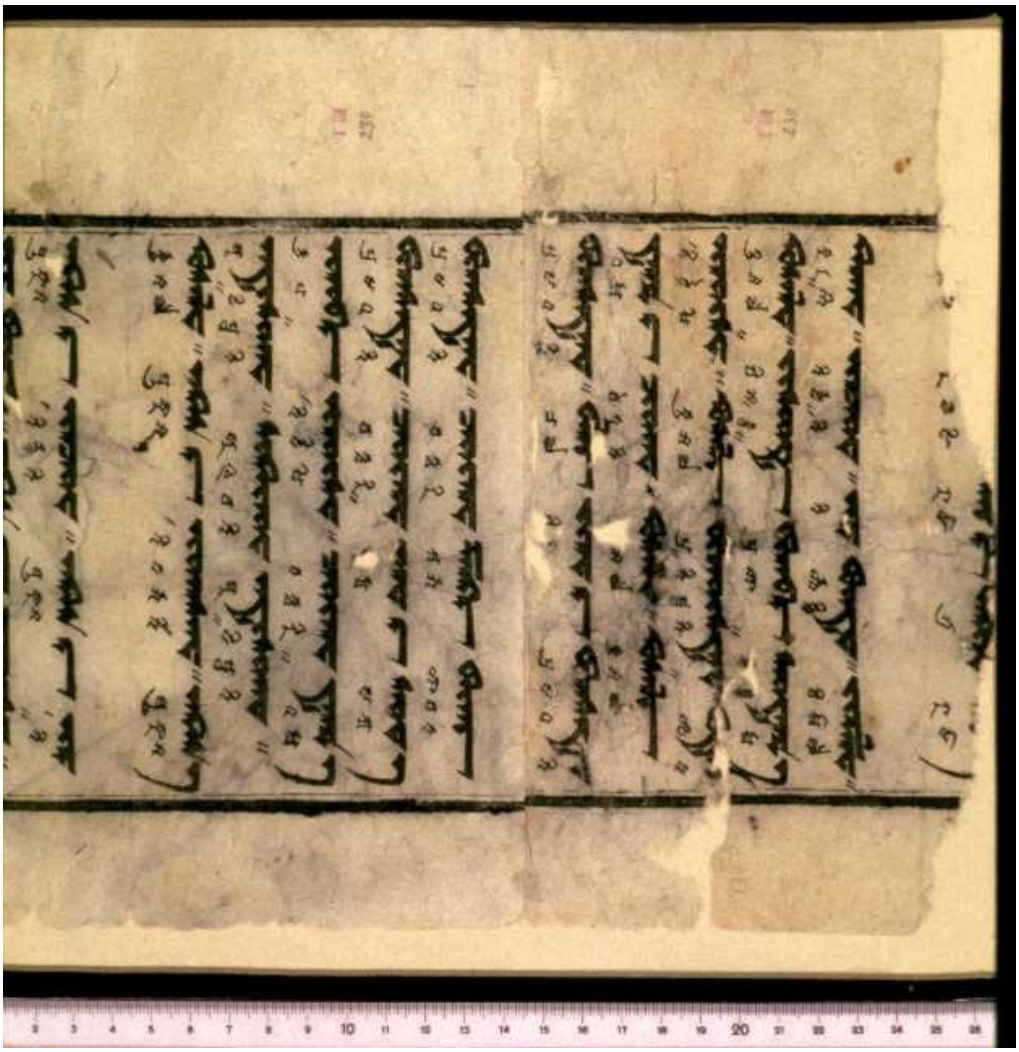


Figure 30: BBAW, Mainz 801, middle portion. Block print. Annotations in Central Asian Brahmi.

Figure 31: BBAW, U 7008. Block print.

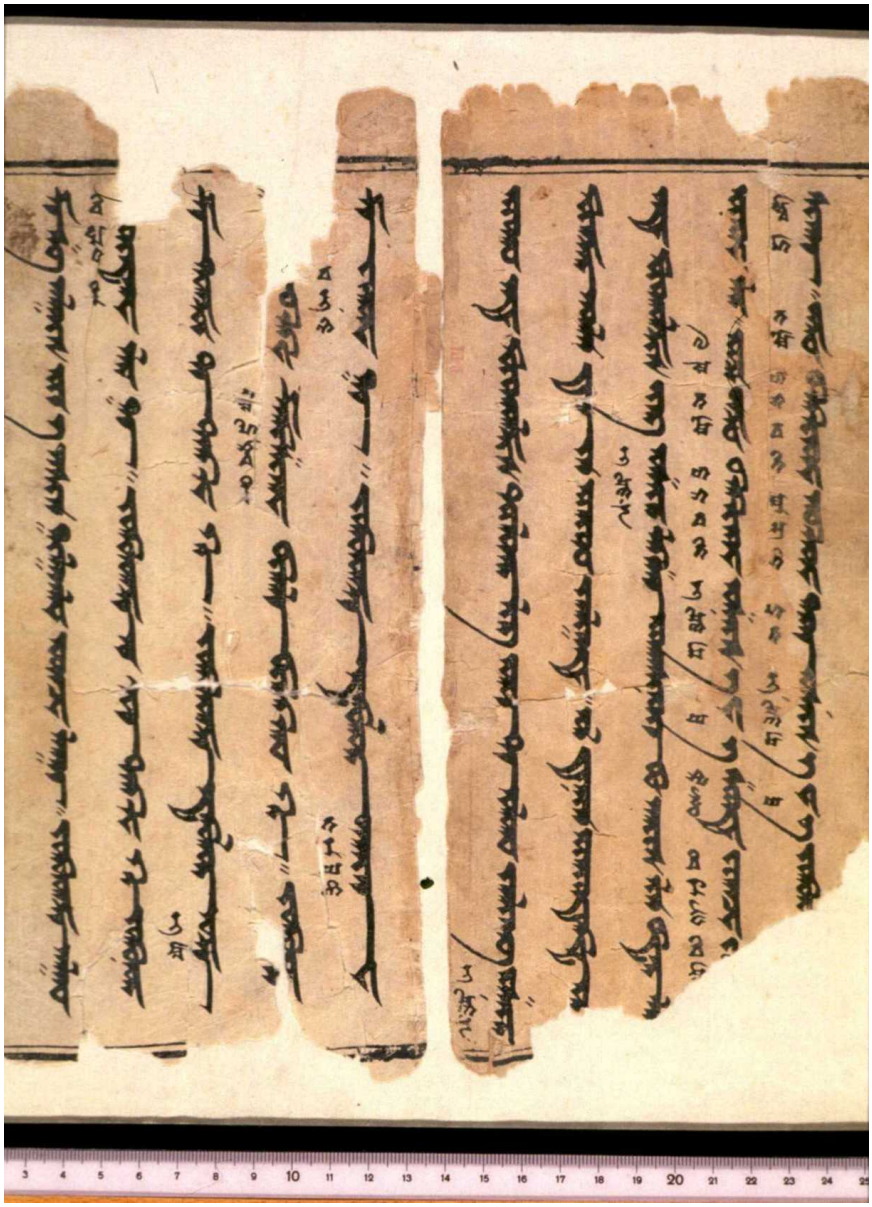


Figure 32: BBAW, Mainz 764, middle. Formal script. Annotations in Central Asian Brahmi.

Figure 33: BBAW, U 3832, folio 1. Formal script.

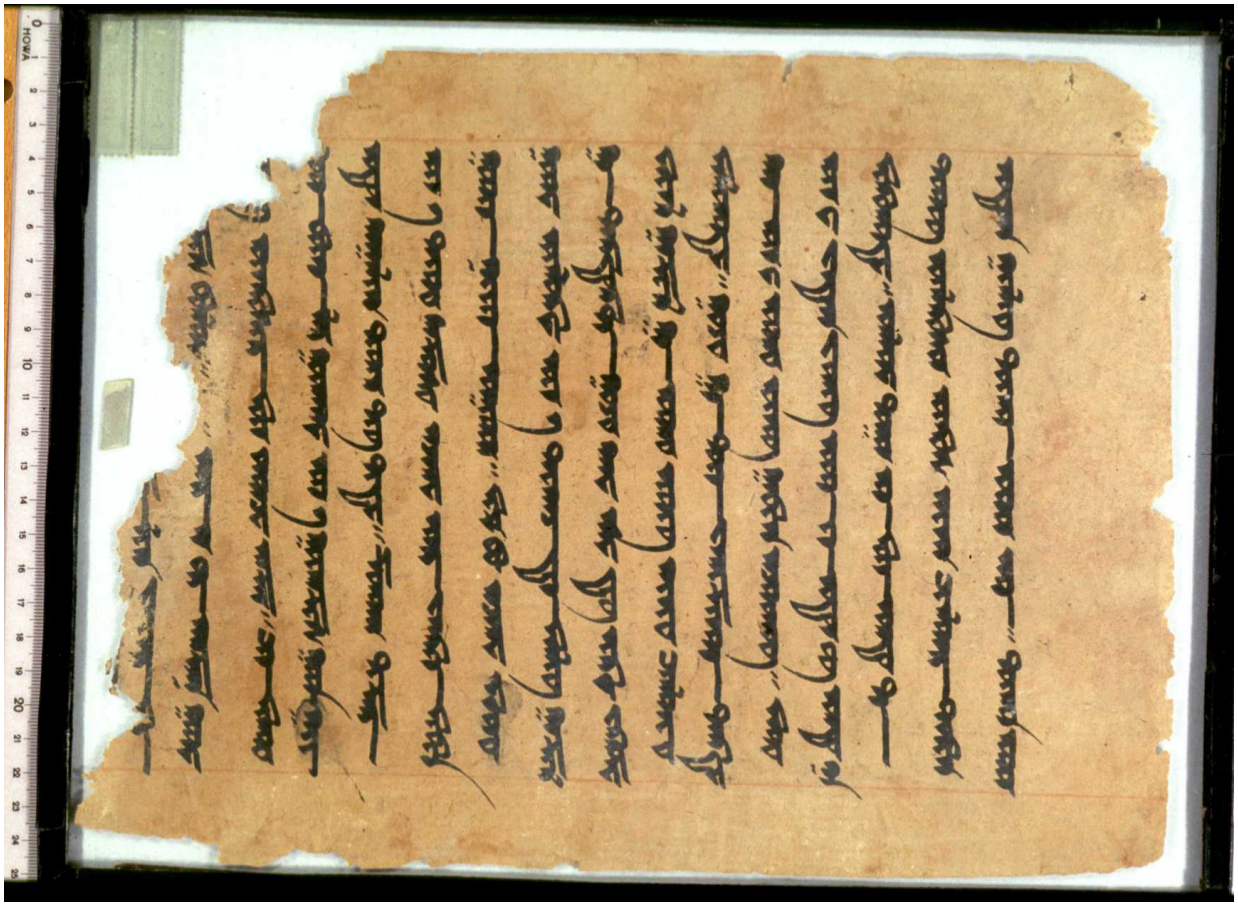


Figure 34: BBAW, Mainz 841, folio 2.

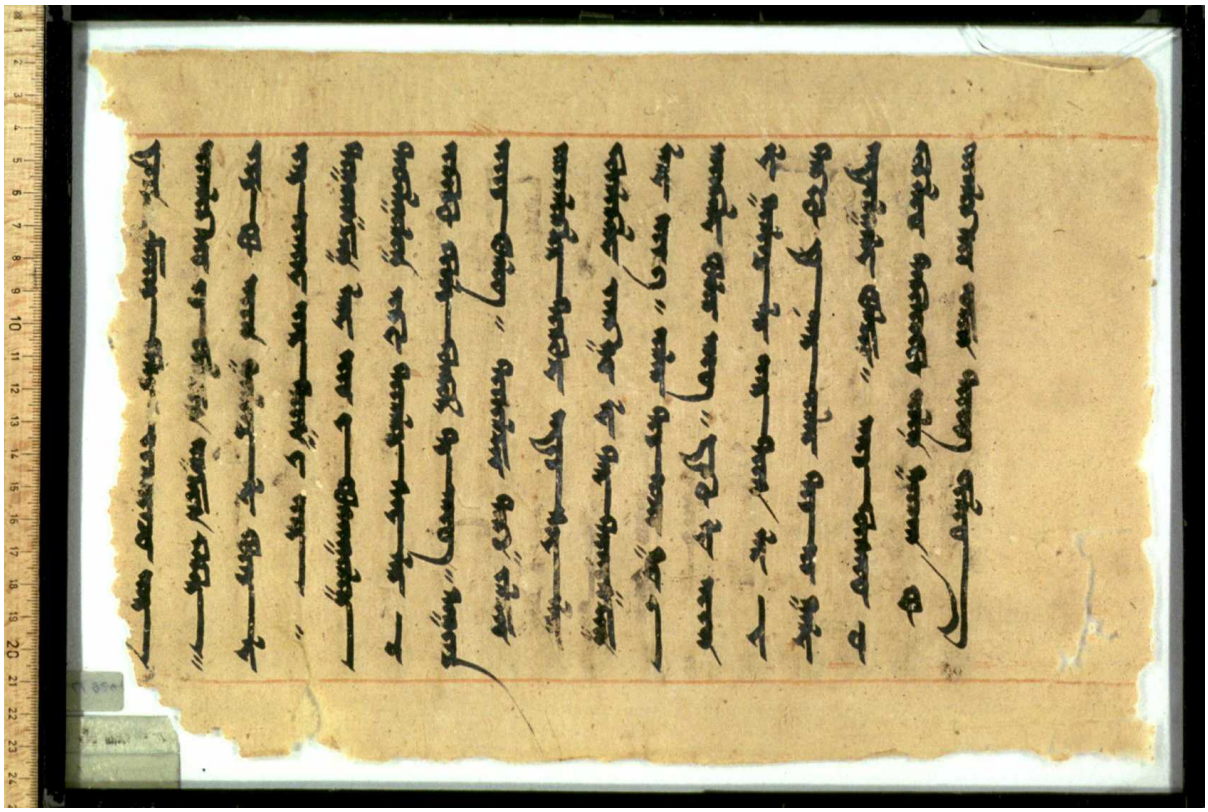


Figure 35: BBAW, U 924, folio 2.

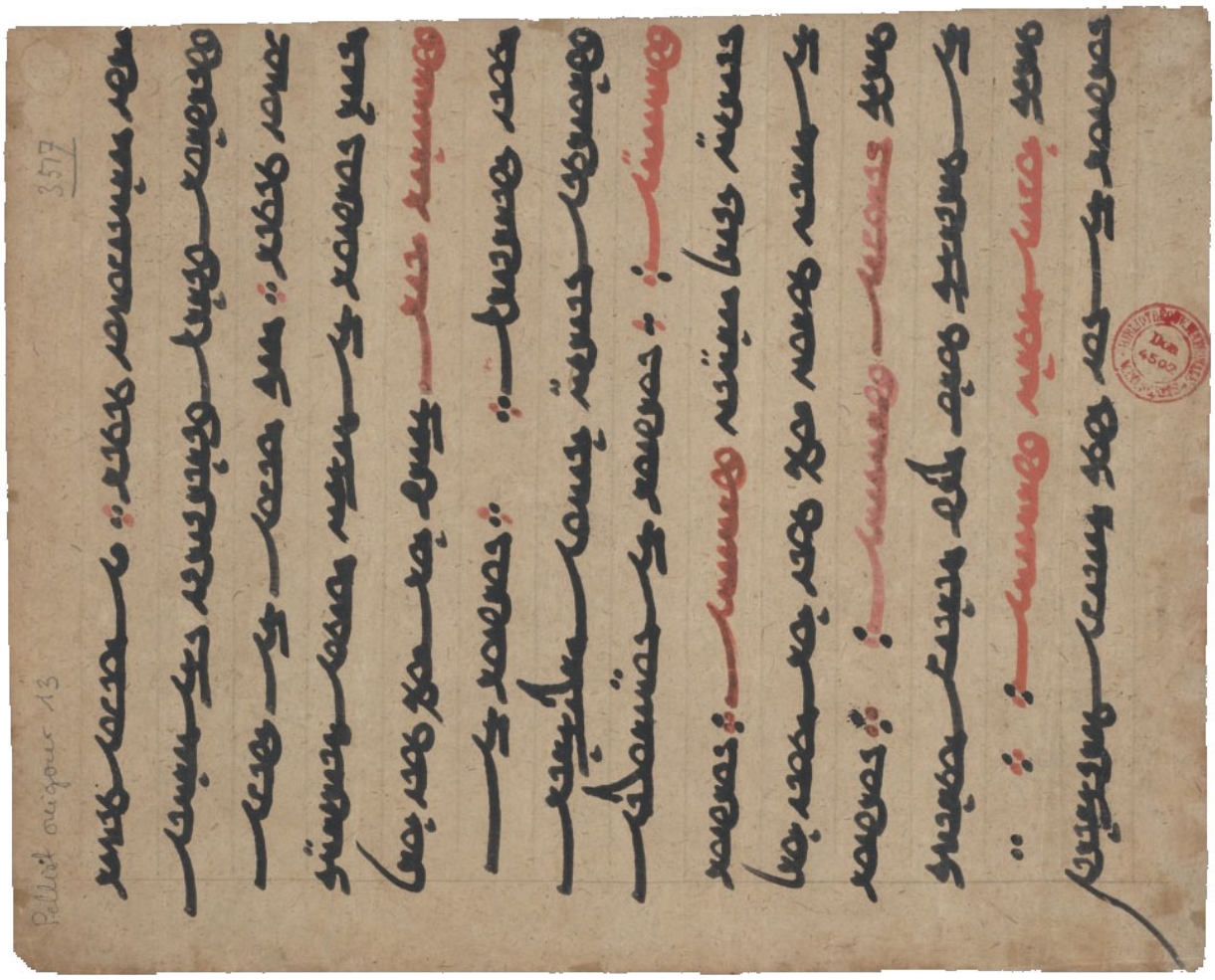


Figure 36: Pelliot Ouïgour 13.

Figure 37: PEALD 6r, recto.

Figure 38: BBAW, U 40, recto. Note the inventory of Old Uyghur letters at the bottom of the folio (see § 4.3 for additional details).

Figure 39: BBAW, U 320, folio 1.

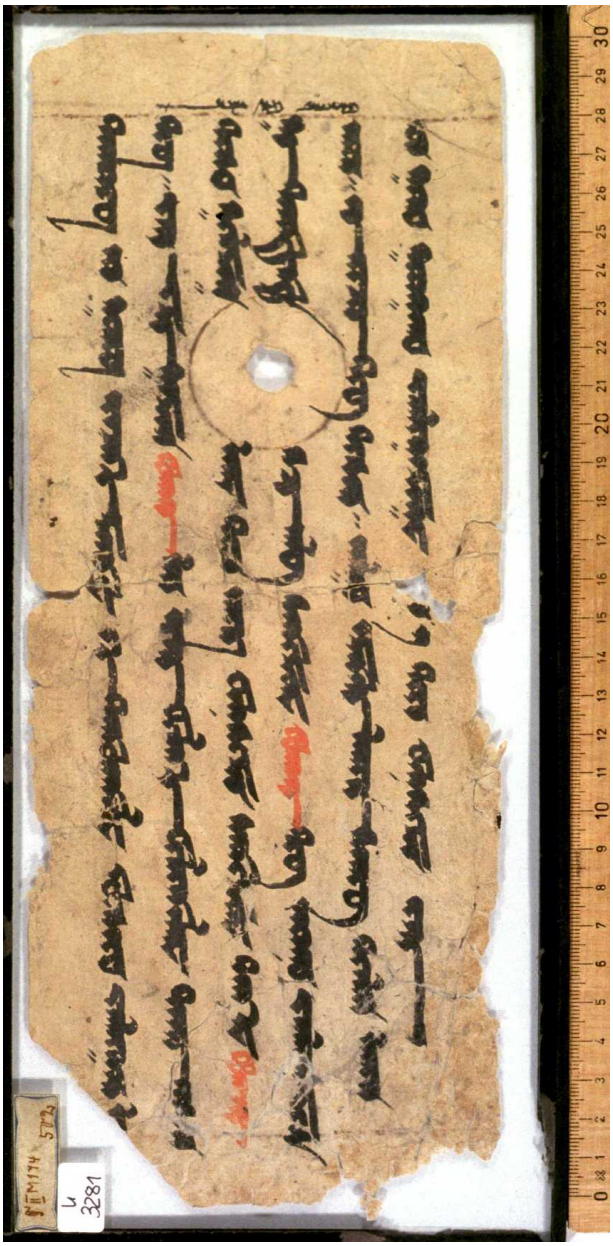


Figure 40: BBAW, U 3281, folio 1.

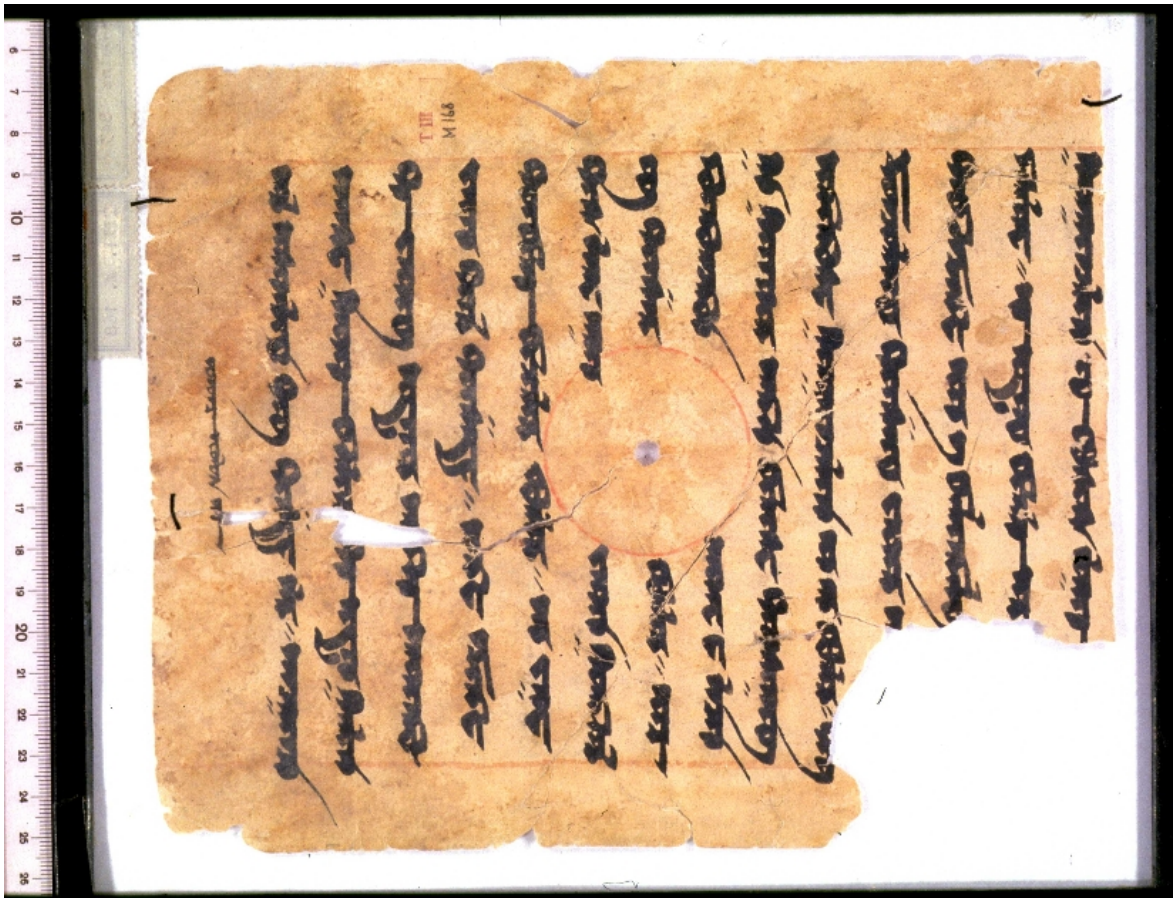


Figure 41: BBAW, Mainz 843, folio 2.

Figure 42: BBAW, U 7123, recto.

Figure 43: BL / BBAW, Ch 5555, recto. *Ekottaragamasutra* / 增一阿含經 *Zeng yi e han jing*.

Figure 44: BBAW, U 456, folio 1.

Figure 45: BBAW, Ch/U 7730, verso.

Figure 46: 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.

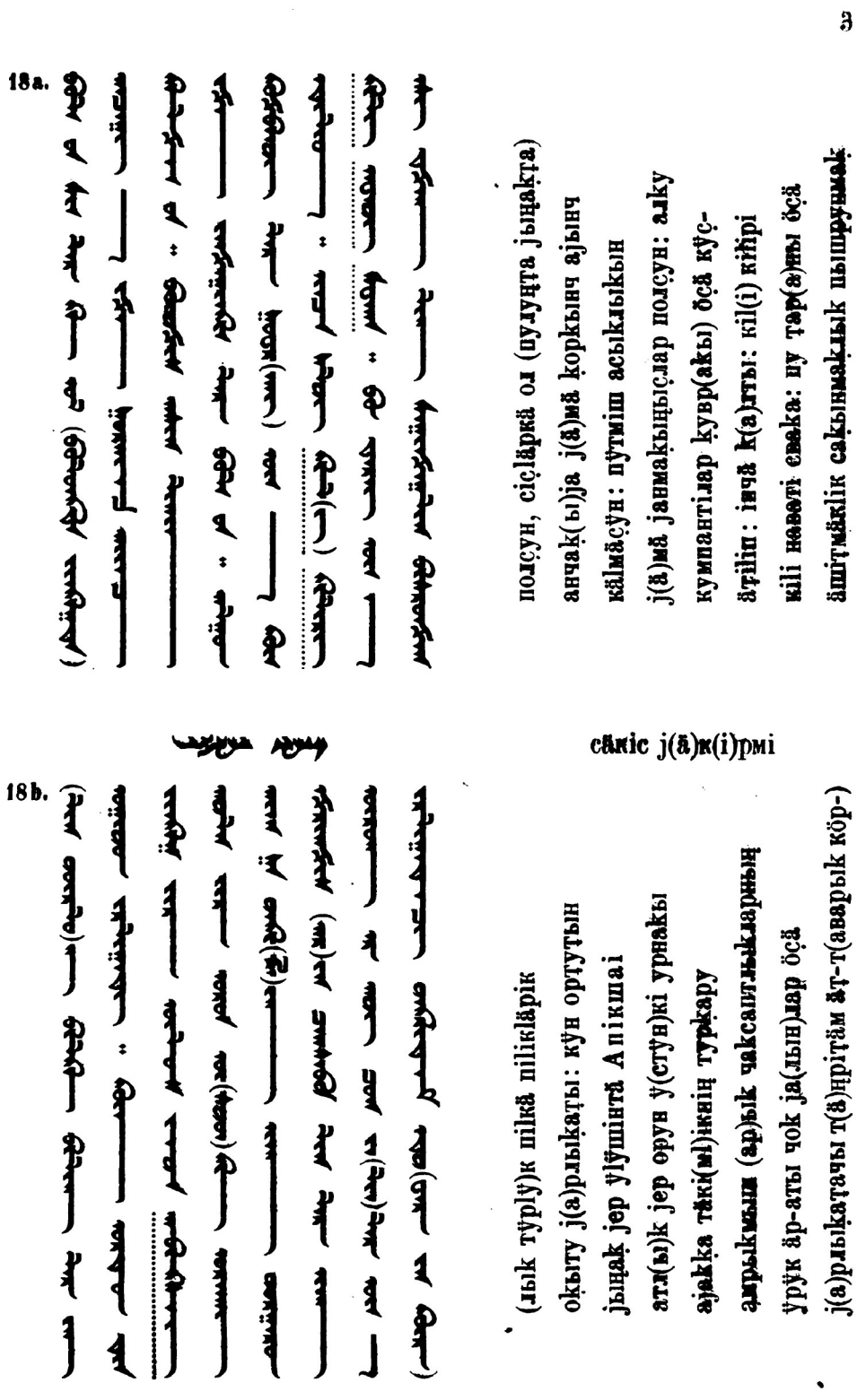


Figure 47: Printed edition of *Ṭišastvustik* in the Old Uyghur script (from Radloff 1910: 3).

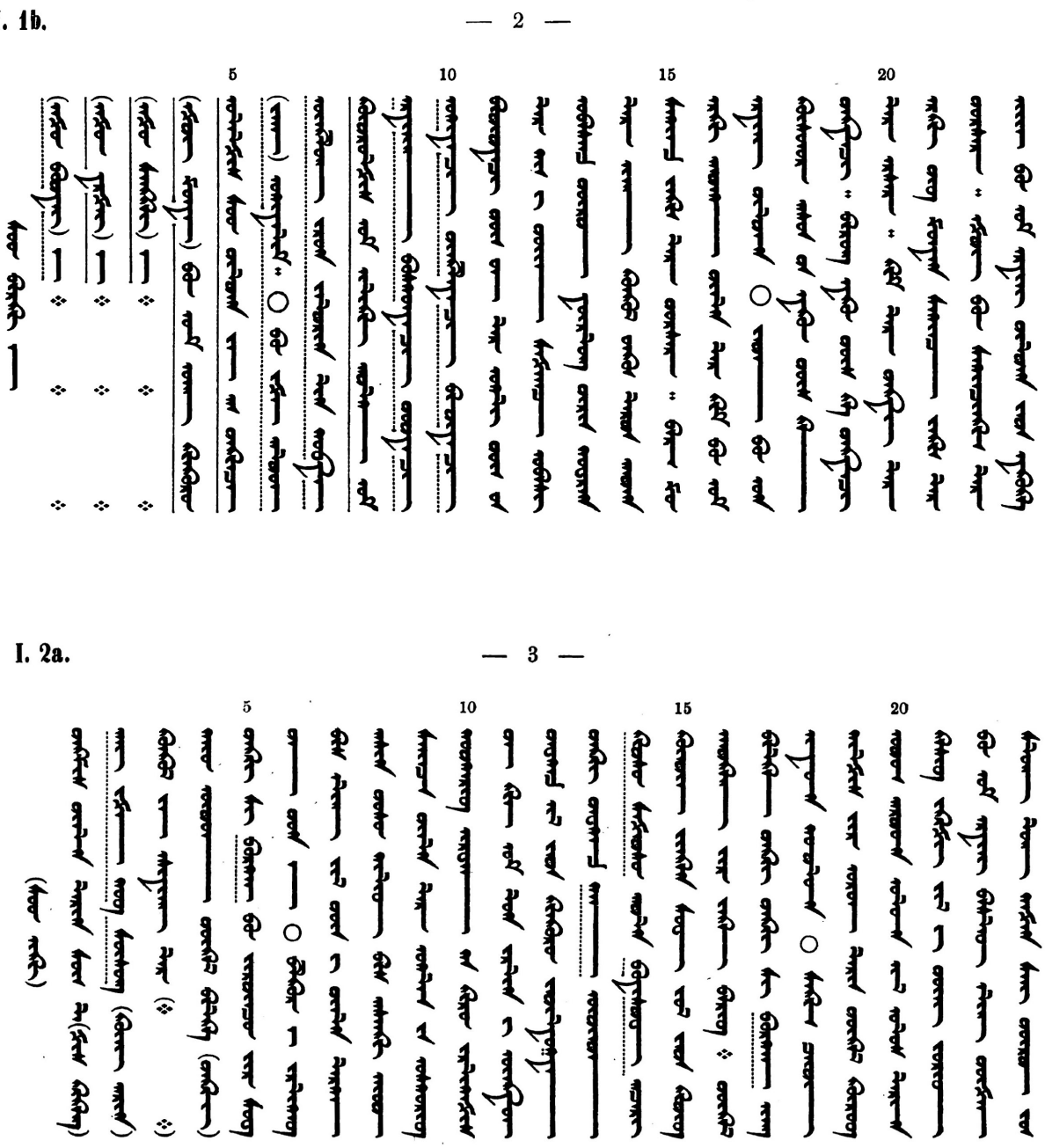
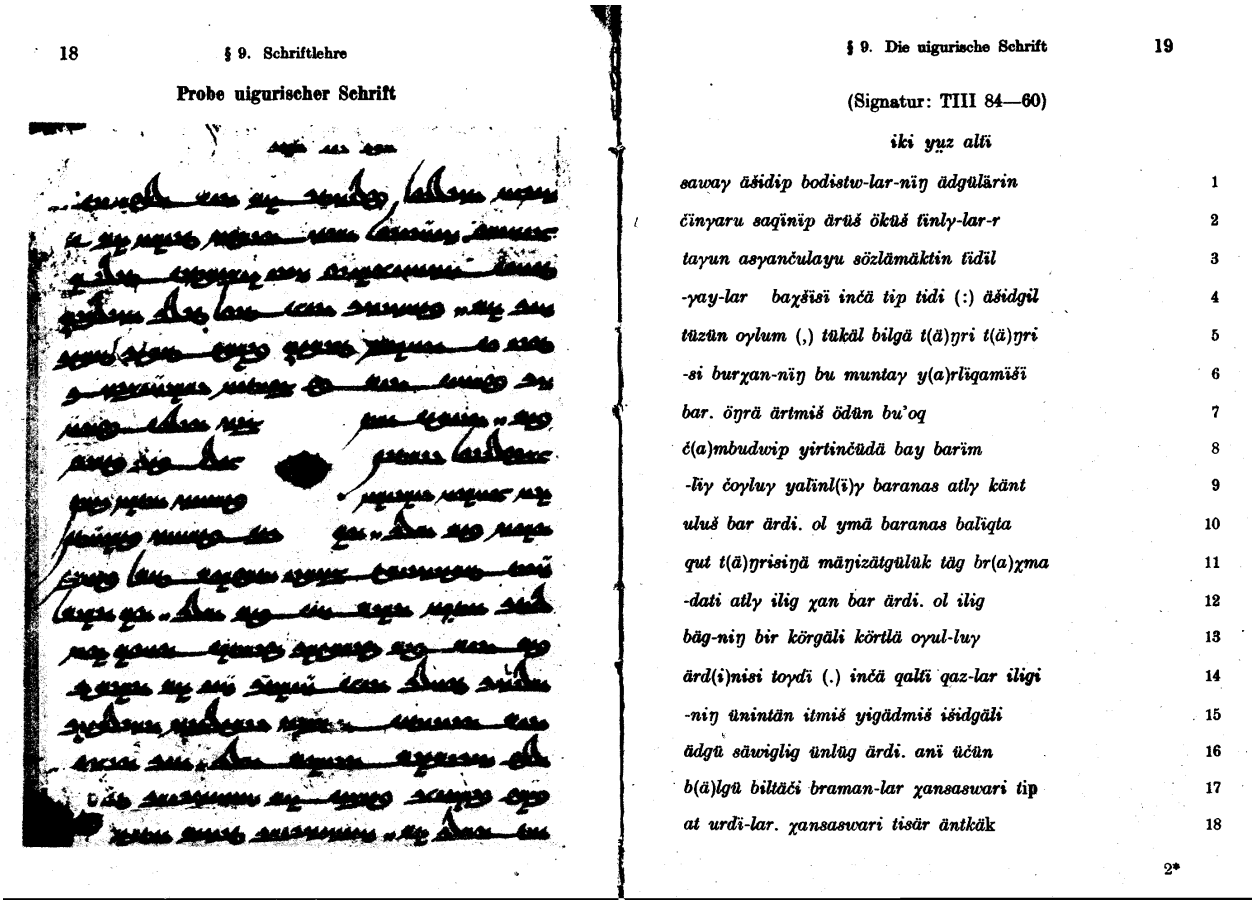


Figure 48: Printed edition of *Suvarṇaprabhāsa*, a Mahayana Buddhist text, in the Old Uyghur script (from Radlov and Malov 1913: 2–3).



18 § 9. Schriftlehre

Probe uigurischer Schrift

§ 9. Die uigurische Schrift 19

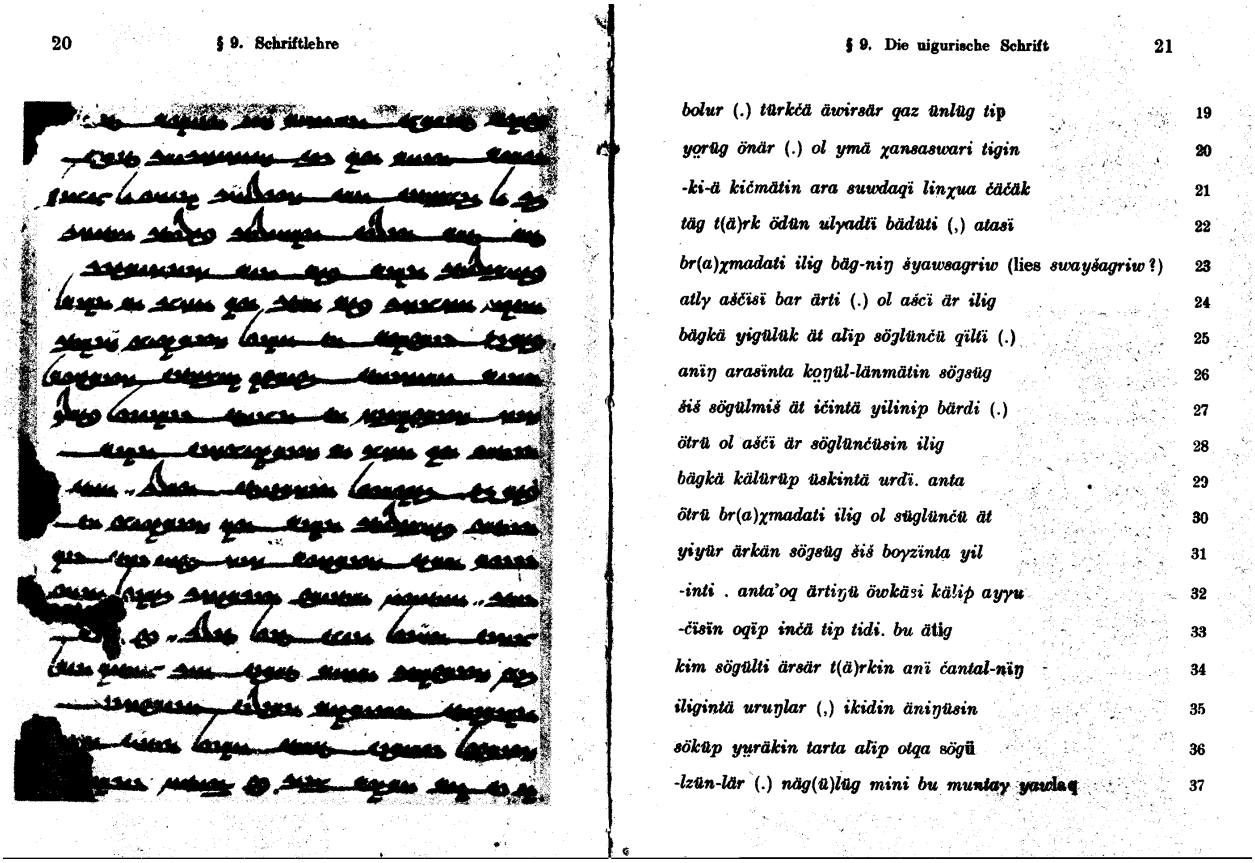
(Signatur: TIII 84—60)

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 oγlum (,) 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 49: Transcription of an Old Uyghur manuscript (from von Gabain 1950: 18–19). Continued in fig. 50.

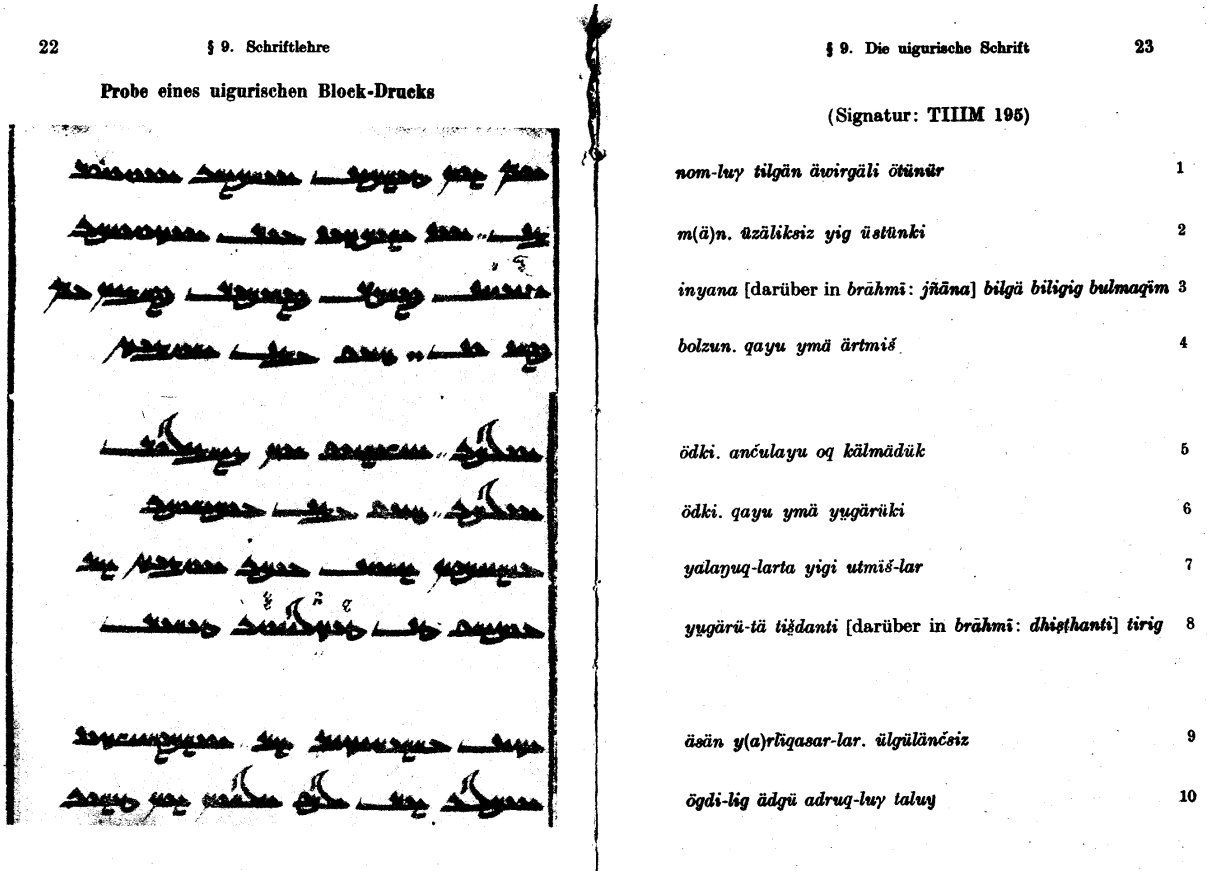


20 § 9. Schriftlehre

§ 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 50: Transcription of an Old Uyghur manuscript in a grammar of Old Turkic (from von Gabain 1950: 20–21). Continued from fig. 49.



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. äzäliksiz 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

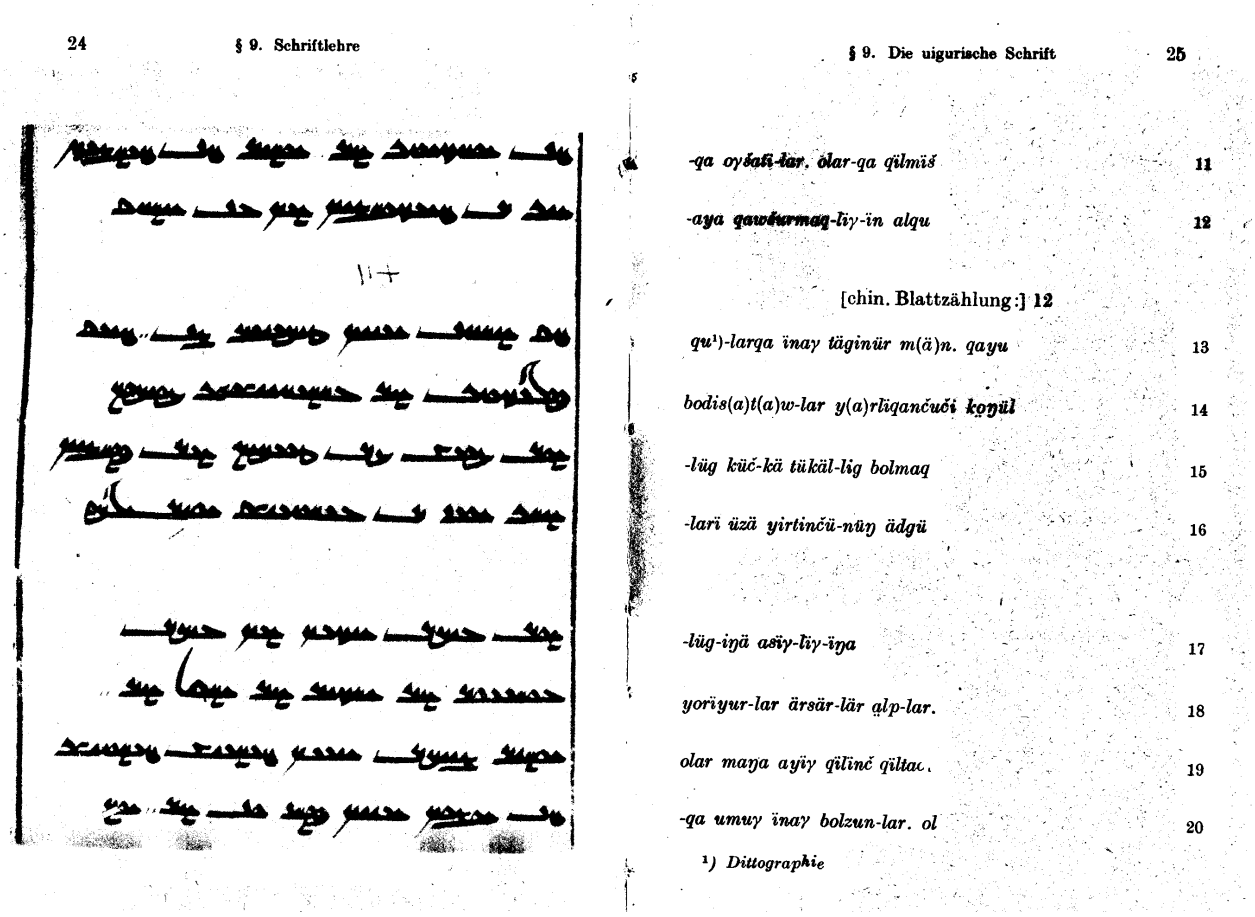
yalaŋuq-larta yigi utmïš-lar 7

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

äsän y(a)rlïqasar-lar. ülgüläńčsiz 9

ögdi-lig ädgü adruq-luγ taluγ 10

Figure 51: Transcription of an Old Uyghur manuscript in a grammar of Old Turkic (from von Gabain 1950: 22–23). Continued in fig. 52.



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

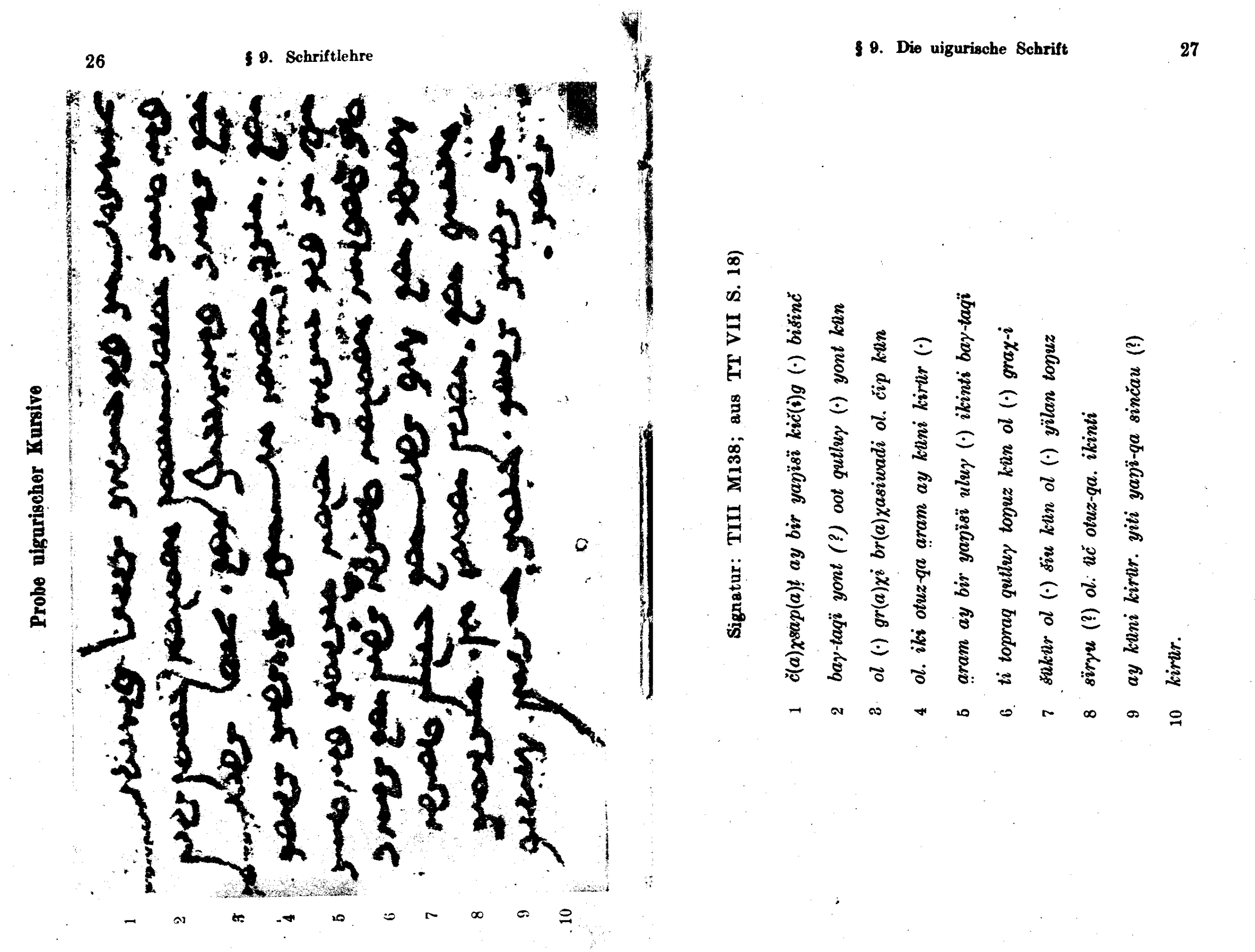
yoriyur-lar ärsär-lär alp-lar. 18

olar maŋa ayïγ qïlïnč qïltac. 19

-qa umuγ ïnaγ bolzun-lar. ol 20

[1]) *Dittographie*

Figure 52: Transcription of an Old Uyghur manuscript in a grammar of Old Turkic (from von Gabain 1950: 24–25). Continued from fig. 51.



26 § 9. Schriftlehre

Probe uigurischer Kursive

§ 9. Die uigurische Schrift 27

Signatur: TIII M138; aus TT VII S. 18)

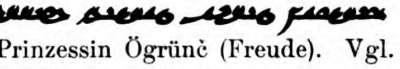
1 *č(a)χšap(a)ṭ ay bir yaŋïsï kič(i)g (·) bišinč*
2 *bay-taqï yont (?) oot qutluγ (·) yont kün*
3 *ol (·) gr(a)χi br(a)χasiwadi ol. čip kün*
4 *ol. iki otuz-qa aram ay küni kirür (·)*
5 *aram ay bir yaŋïsï uluγ (·) ikinti bay-taqï*
6 *ti topraq qutluγ toŋuz kün ol (·) graχ-i*
7 *šükür ol (·) šiu kün ol (·) yïlan toŋuz*
8 *siryu (?) ol. üč otuz-qa. ikinti*
9 *ay küni kirür. yiti yaŋï-qa sinčau (?)*
10 *kirür.*

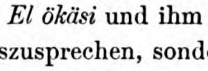
Figure 53: Transcription of an Old Uyghur manuscript in a grammar of Old Turkic (from von Gabain 1950: 26–27).

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 54: 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.

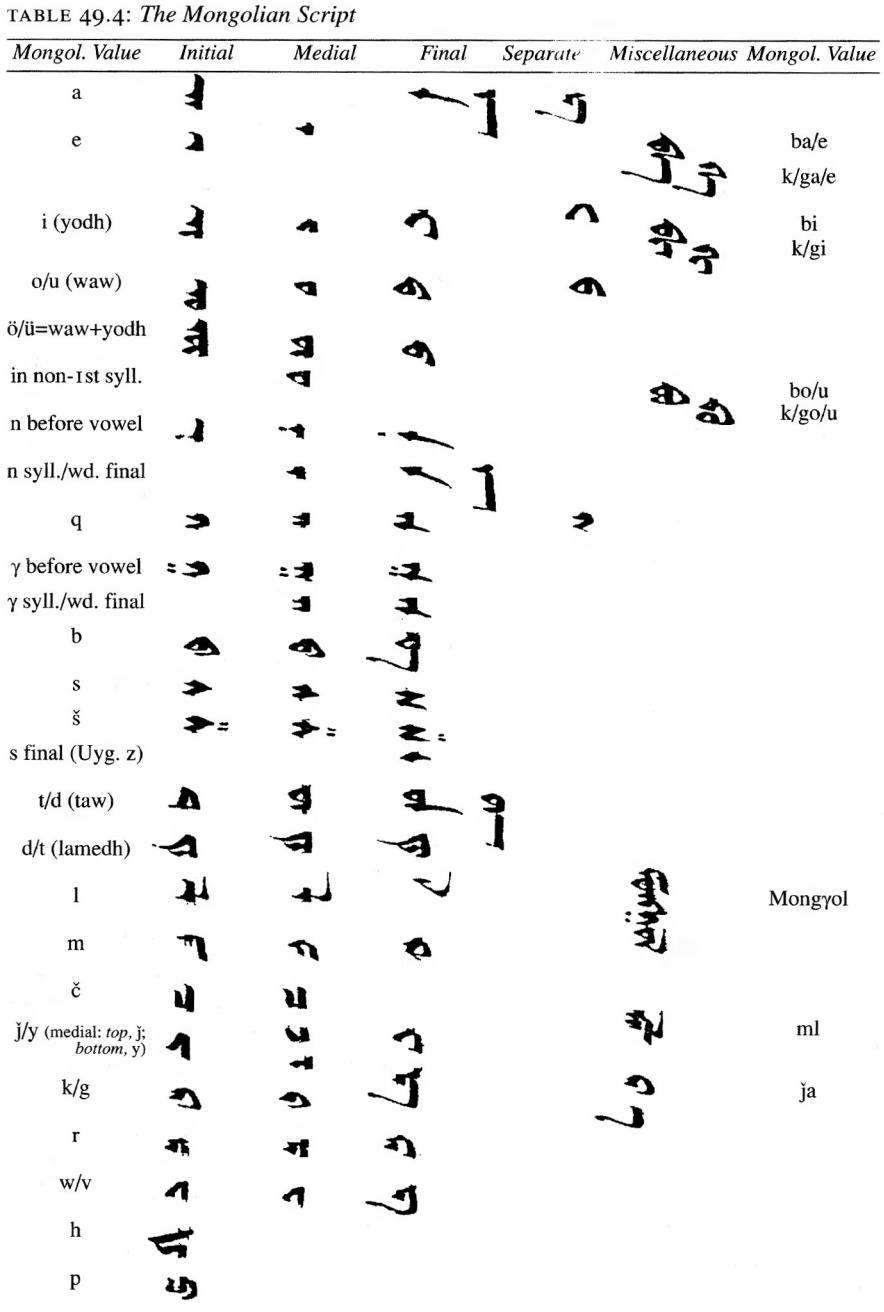
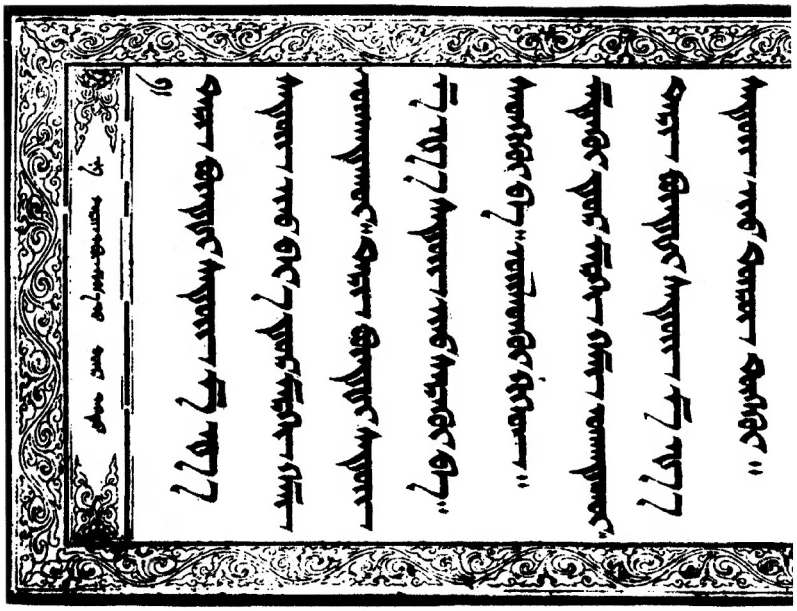


Figure 55: Table showing letters of the Mongolian script (from Kara 1996: 545). See table of Old Uyghur letters from the same source in fig. 4.

SAMPLE OF MONGOLIAN



1. Transliteration:	t'r'	pwβ'δhy	s'δwβ'	m' h' '	/s'δwβ'	'ynw	p'y ' δwr	m'rk'n
2. Normalization:	tere	bôdhi-saduva		ma-hā-saduva		inu	bey-e-dür	mergen
3. Gloss:	that	bodhisattva		mahāsattva		3POSS	body-DAT	wise

1. k'm'n /	'wq'qδ'qwy :	t'r'	pwβ'δhy	s'δwβ'	m' h' '	/ s'δwβ'	'ynw	serekü
2. kemen /	uqaγdaqui	tere	bôdhi-saduva		ma-hā-saduva		inu	sereküi
3. saying	should.know	that	bodhisattva		mahāsattva		3POSS	waking

1. ba : /	s'tkykwy	p' :	'wyl'tkwy	kyk't : /	m'δ'kwy	dwr	m'rk'n	k'm'n
2. ba : /	sedkiküi	ba	üiledküi	kiged	medeküi-dür		mergen	kemen
3. and	thinking	and	acting	as.well	knowing-DAT		wise	saying

1. 'wq'qδ'qwy : /	t'r'	pwβ'δhy	s'δwβ'	m' h' '/	s'δwβ'	'ynw	twyrwn	twyk'kwy :
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 56: Sample Mongolian text (from Kara 1996: 546). Compare the Mongolian block print with the Old Uyghur block print in fig. 24.

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

A. Administrative

1. **Title:** *Revised proposal to encode Old Uyghur in Unicode*
2. Requester's name: *Anshuman Pandey <pandey@umich.edu>*
3. Requester type (Member body/Liaison/Individual contribution): *Expert contribution*
4. Submission date: *2019-01-07*
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: *Old Uyghur*
   b. The proposal is for addition of character(s) to an existing block:
      Name of the existing block:
2. Number of characters in proposal: *35*
3. Proposed category (select one from below - see section 2.2 of P&P document):
   A-Contemporary ( ) B.1-Specialized (small collection) ( ) B.2-Specialized (large collection) ( )
   C-Major extinct (X) 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
6. References:
   a. Are references (to other character sets, dictionaries, descriptive texts etc.) provided? *Yes*
   b. Are published examples of use (such as samples from newspapers, magazines, or other sources) of proposed characters attached? *Yes*
7. Special encoding issues:
   Does the proposal address other aspects of character data processing (if applicable) such as input, presentation, sorting, searching, indexing, transliteration etc. (if yes please enclose information)? *Yes*

8. Additional Information:

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

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

C. Technical - Justification

1. Has this proposal for addition of character(s) been submitted before? *No*
   If YES explain
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? *Dai Matsui <dmatsui@let.osaka-u.ac.jp>*
   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: *See text of proposal*
4. The context of use for the proposed characters (type of use; common or rare) *Common*
   Reference: *See text of proposal*
5. Are the proposed characters in current use by the user community? *Yes;*
   If YES, where? Reference: *Currently used by scholars of Turkic and Central Asian studies*
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 characters for diacritics*
    Is a list of composite sequences and their corresponding glyph images (graphic symbols) provided? *N/A*
    If YES, reference:
12. Does the proposal contain characters with any special properties such as control function or similar semantics? *No*
    If YES, describe in detail (include attachment if necessary)
13. Does the proposal contain any Ideographic compatibility characters? *No*
    If YES, are the equivalent corresponding unified ideographic characters identified?
    If YES, reference: