L2/20-003R
2020-02-16

Revised proposal to encode Old Uyghur in Unicode

Anshuman Pandey

pandey@umich.edu
pandey.github.io/unicode

February 16, 2020

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”
- L2/19-016: “Revised 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)”
- L2/19-047: “Recommendations to UTC #158 January 2019 on Script Proposals”
- L2/20-046: “Recommendations to UTC #162 January 2020 on Script Proposals”

The major changes to L2/19-016 are as follows:

- Addition of letters for generic *aleph-nun* (§ 7.1.1), *beth-yodh* (§ 7.1.2) for handling ambiguous readings
- Inclusion of a baseline modifier for producing an ornamental terminal (§ 7.5)
- List of characters not proposed for encoding (§ 5.2)
- Tables showing comparisons of letterforms from various sources (tables 2–4)

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

- Yukiyo Kasai (Centrum für Religionswissenschaftliche Studien, Ruhr-Universität Bochum)
- Dai Matsui (Graduate School of Letters, Osaka University)
- Mehmet Ölmez (Department of Modern Turkic Languages and Literatures, Istanbul University)

1 Introduction

The 'Uyghur' or 'Old Uyghur' script was used between the 8th and 17th centuries across Central Asia for recording religious, literary, and administrative documents in Turkic languages, as well as Chinese, Mongolian, Sogdian, and Tibetan. There are two major styles of the script: square and cursive. The script was developed further through the usage of block printing in the 14th century.

Derived from the 'cursive' style of the Sogdian script during the 8th–9th century, Uyghur is situated in the middle of a script continuum that originates from the Sogdian script of the 'Ancient Letters' and terminates at modern Mongolian. Just as speakers of Turkic languages adopted the Sogdian script, speakers of other languages in Central Asia turned to the Uyghur script to develop new orthographies. A popular narrative states that in the 13th century, the scholar Tata Tonga, who was also a chancellor of the Naiman Khanate, developed an orthography for writing the Mongolian language using the Uyghur script during the reign of Genghis Khan. The Uyghur-based Mongolian script developed into a distinctive script with its own scribal and print culture, and itself generated a few offshoots.

The Uyghur script was basis for vibrant textual cultures across Central Asia. It was used in multilingual documents alongside major Asian scripts. There are documents containing Uyghur script with intralinear Han characters; Manichaean script with Uyghur on the reverse; Chinese manuscripts with Turkic translations in Uyghur script; and texts written in Uyghur with interlinear Sanskrit annotations in 'Turkestani' or Central Asian styles of Brahmi. The 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 Uyghur and the Arabic scripts are also extant. The script was also used in parts of Iran. By the 16th century the Uyghur script was replaced by new orthographies for Turkic languages based upon the Arabic script; although its usage in Gansu is attested through the 17th century.

There has been active scholarship on the Uyghur script and its written record since the early 20th century. It was during this time that European expeditions to Turfan unearthed vast amounts of materials in Uyghur and other scripts. German and Russian scholars adapted the Uyghur script for modern typesetting. Texts in the Uyghur script were edited and published by F. W. Max Müller, V. V. Radlov, and others (see fig. 42–44). At least two styles of metal types were produced for printing these editions, based upon the square style used in manuscripts and the style used in block prints. Over the past century, interest in the Uyghur script has continued to grow, especially within studies of the 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 name 'Uyghur' (/ʊjɣʊr/) has various transliterations, transcriptions, and spellings in European languages. In English it is spelled 'Uighur', 'Uigur', 'Uygur'; 'Ouïgour' in French; 'Uigurisch' in German. In English scholarly nomenclature, the script is referred to as both 'Uyghur' and 'Old Uyghur'. The usage of 'old' arises from the fact that the 'Uyghur script' may refer to both the historical Sogdian-based script and the later Arabic-based orthography used for the modern Uyghur language, which is not directly related to the Uyghur language of the 8th century, for which the original Uyghur script was used for writing. In order to differentiate between the two scripts, the descriptor 'Old Uyghur' is used for referring to the historical script. To be sure, neither 'Uyghur' nor 'Old Uyghur' is an entirely accurate designation for the script. The renowned Turkologist, 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 follows Clauson's conclusion. The proposed Unicode identifier for the script is 'Old Uyghur'. The name pertains specifically to the script within the context of Unicode, and it does not refer to any other language, culture, or community.

3 Encoding history

3.1 Justification for encoding

Although the Old Uyghur script is derived from Sogdian and is the ancestor of Mongolian, and shares similarities with both scripts, is has requirements that justify an independent encoding for it in Unicode:

- Distinguishing Old Uyghur from related scripts in plain text and preserving its glyphic and stylistic distinctive in multilingual contexts. Sogdian and Mongolian glyphs do not adequately transmit the aesthetic and orthographic features of formal and block-printed Uyghur letters.

- The encoded repertoire contains characters that are specific to Old Uyghur, such as generic letters to be used for representing textual ambiguities resulting from usage of a single character for two letters.

- A simpler encoding model than is used for Mongolian in Unicode. The proposed Old Uyghur repertoire is based upon a palaeographic and graphetic model, while the Unicode encoding for Mongolian is based upon a phonetic model.

- A separate encoding preserves character identity and semantics. The Mongolian encoding uses different names and ordering for letters, which reflect Mongolian preferences and pronunciations. Mongolian letter names do not correspond directly to the values of Old Uyghur letters.

- The proposed model for Old Uyghur offers a practical implementation for a vertical script that avoids the complications of the Mongolian model, and has a default horizontal representation that differs from Mongolian.

3.2 Previous Unicode proposals

Proposals to encode Old Uyghur were previously submitted to the Unicode Technical Committee (UTC) by Omarjan Osman: "Proposal for encoding the Uygur script in the SMP" (L2/12-066) and "Proposal to Encode the Uyghur Script in ISO/IEC 10646". These proposals provide 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 through character encodinng. Thus, 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.

3.3 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. The structure is similar to that of Sogdian, with letters joined together at the baseline. The basic letters have an independent shape and contextual forms when they occur in initial, medial, or final positions. All letters are dual joining, except for *zayin*, which does not join to the left. Diacritics are used for diambiguating letters with similar appearances and for indicating phonetic distinctions between such letters (see § 8.3).

Word boundaries are generally demarcated in manuscripts and block prints using spaces. However, in numerous manuscripts, letters with elongated terminals may be written as a swash, such that the final stroke touches the initial letter of the following word. This is a calligraphic or space-filling calligraphic technique; there is no joining behavior between such a final letter and the following initial letter. In plain encoded text, a space would be expected after the final letter in such cases.

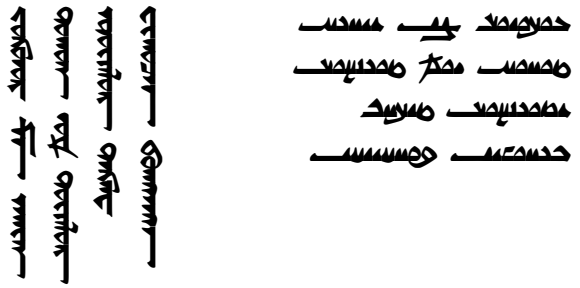
In addition to the elongation of terminals, other space-filling techniques are observed in manuscripts. In some documents, at the margin of the page the final letter of the last word may be written as a separate, independent letter. If the letter contains a terminal, that stroke may be elongated; otherwise the letter may be preceded by a baseline extender of variable width.

There are no formal rules for indicating the breaking of words at the end of line, or usage of hyphens or other continuation devices. In the majority of texts, words are not split at line boundaries; however, in very few texts a word is split at the end of line by continuing it on the next line. In digital layouts, line-breaks should occur after words.

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 by the Arabic script. When Old Uyghur texts were begun to be printed in the 20th century, some publishers maintained fidelity to the standard vertical orientation (see fig. 43, 44). 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 scholars and publishers for short excerpts of Old Uyghur text because it is a convenient method to print Uyghur words and phrases in multilingual contexts that also contain Arabic, Cyrillic, Devanagari, Tibetan, and other scripts (see fig. 50). Given the global range of scholars of 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. The code chart, following p. 12, shows the default orientation of Old Uyghur characters.

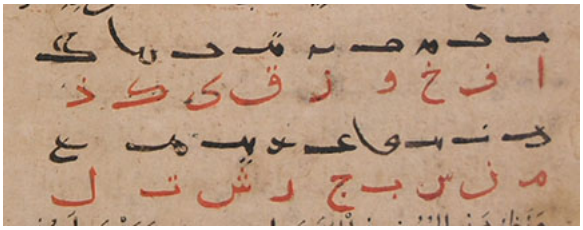
4.3 Repertoire

The traditional Old Uyghur alphabet consists of 18 letters, which are derived from Sogdian. The letters represent consonantal sounds. The letters *aleph*, *waw*, *yodh* are used for expressing vowels, following the Semitic convention inherited from Sogdian. The vocalic repertoire of Turkic languages is represented using combinations of these letters in digraphs and trigraphs (see § 8.2).

The historical repertoire is attested in the manuscript U 40 (see fig. 1), dated to the 9th century:

The inventory contains 21 characters (as read from left to right). The first 17 are the basic letters of the script. Following the scholarly nomenclature, these are *aleph*, *beth*, *gimel*, *waw*, *zayin*, *heth*, *yodh*, *kaph*, *lamedh*, *mem*, *nun*, *samekh*, *pe*, *sadhe*, *resh*, *shin*, *taw*. The four letters that follow are not clear due to blemishes in the manuscript. Clauson (1962: 107) suggests that they are 'hooked *resh*', a final *samekh* (or *shin*), a final *mem*, and a two-dotted *heth*.[1] The inventory is important in that it provides attestation for the full repertoire and order of the alphabet, and evidence for the independent forms of letters, and special forms, eg. final *mem*, two-dotted *heth*. It also provides evidence for the usage of diacritics to expand the alphabet and specify phonetic distinctions, eg. two-dotted *heth* represents /x/ or /q/.

A repertoire of the Uyghur script of the 11th century is attested in the ديوان لغات الترك *Dīwān lughāt al-turk*, a dictionary of Turkic languages compiled by the Kara-Khanid scholar Mahmud Kashgari (see fig. 2–3). An excerpt from the text shows Old Uyghur letters (black ink) with their Arabic analogues (red ink):



The repertoire is *aleph*, *beth*, *gimel*, *waw*, *zayin*, two-dotted *heth*, *yodh*, *kaph*, *lamedh*, *mem*, dotted *nun*, *shin*, *pe*, *sadhe*, *resh*, two-dotted *shin*, *taw*, 'hooked r'. The inventory indicates the loss of distinction of some traditional letters. A single shape is used for *samekh* and *shin*; *samekh* is written using the palaeographical *shin*, *shin*is written using diacritics. The shape of *nun* differs from *aleph*, but it is denoted using a diacritic. Similarly, *gimel* and *heth* are distinctively, but the latter is written using diacritics. Apart from illustrating the dynamic orthography of the script, the attestation is noteworthy because the Arabic transliteration provides a sense of the phonetic values of Uyghur letters during this time period in the Kara-Khanid Khanate. It also indicates that the Uyghur script may have been written horizontally in some contexts during this period.

The attested repertoires in U 40 and Kashgari are significant for palaeographical reasons. After the 9th century, different documents show various transformations of the script, such as the coalescing of letters with similar graphical structures. Based upon Clauson (1969: 109–110)[2] and details provided by Dai Matsui (per-

[1] The final *mem* is likely included because it differs in shape from the independent form; the dotted *heth* has a high frequency of usage. I am not satisfied with Clauson's identification of letters #18 and #19. He states that #18 is the 'hooked' *resh*. While, this letter follows *taw* in the alphabetic order, its shape here resembles 𐽆 — an alternate final form of *aleph* and *nun* that differs from the regular finals — not the 𐽙 'hooked' *resh*. Secondly, he states that #19 is a "final *samekh* (or shin)"; however, these letters do not have a 'special' final shape that differs greatly from their regular finals. I propose that #19 is actually a poorly written 'hooked' *resh*, as supported by the below-base horizontal stroke in the letter.

[2] Clauson writes: "In good early manuscripts it is reasonably easy to tell all the eighteen letters apart. Samech and schin have slightly different outlines; initial, and even medial, aleph and nun are just distinguishable, and gimel-cheth, although the two letters themselves are indistinguishable, is identified by two superscribed dots when it represents velar k (or x?)."

sonal communication, August 2018–January 2019), the major orthographic practices observed in documents are as follows:

Documents from the 9th century indicate that:

- palaeographic shapes of all 18 letters are distinguishable in good manuscripts
- final *aleph* and *nun* may be written similarly
- initial and medial *gimel* and *heth* are indistinguishable
- two dots above *heth* for representing /q/ or /x/

By the 11th century, the following are observed in some documents:

- *samekh* and *shin* are written in some documents using a single form, resembling *shin*
- two dots beneath *samekh* or *shin* for distinguishing /š/ and /s/
- medial and final *aleph* and *nun* become difficult to distinguish
- in less carefully written documents, final *nun* and *zayin* without a dot over the former

Additional observation in documents of the 14th century include:

- only *kaph*, *lamedh*, *mem*, *pe*, 'hooked' *resh* remain clearly distinctive
- *beth* and *yodh* may not be clearly differentiated or are written using a similar form
- *sadhe* may not be clearly differentiated from *beth* / *yodh*
- *gimel* / *heth* may be indistinct from consecutive *aleph* and/or *nun* without usage of diacritics
- medial and final *taw* indistinguishable from the sequence *waw-nun* unless the *nun* is dotted
- *samekh* / *shin* difficult to distinguish from *gimel* / *heth* without dots
- *resh* may be written similarily to consecutive *aleph* and/or *nun*

The above phenomena do not suggest a linear or systematic evolution of the script from the 9th to 14th century, and the observations are not uniform across all documents within a period. Rather, variations in orthography may be related to regional scribal practices; the language used by scribes; familiarity of the scribe with the source text being copied, and the accuracy of the source; the type of document being written; and the style of script and degree of careful writing. As shown in Hamilton (2005), there are variable degrees of fidelity to letterforms in documents from the same century. It is difficult to know if the writing of two letters with similar graphical structures using a single ambiguous sign is due to rapid writing or to an actual merger of the two letters in the scribe's understanding of the script or some broader orthographic trend. Such cases could result from simplification of the repertoire due to assimilation of sounds, for example the usage of *shin* for both *samekh* and *shin* due to loss of sibilants, and the displacement of one or the other letter.

Various charts of the script have been published in scholarly materials. Of these, Zieme's chart shows an overview of the representations of letters in different periods (see fig. 9). Other charts, unfortunately, do not provide a full repertoire of attested letters, but appear to be snapshots of the script from a particular document or a period. For instance, von Gabain's chart shows letters that are typical of the square style (see fig. 7), while Kara's chart shows letters that resemble those used in block prints (see fig. 6). However, neither of these charts depict all palaeographically attested letters.

4.4 Styles of the script

Old Uyghur documents may be classified by style into two major categories: 'square' and 'cursive' (Moriyasu 2004). The 'square' style is the 'formal' or 'book' style used for religious and literary manuscripts from the 9th through 12th century. A variant known as 'semi-square' is used for a less formal style of the 'square' script. The 'cursive' style occurs in numerous civil and administrative documents from the 12th through 15th centuries. The term 'semi-cursive' is used for documents written in a style that is less loose than the 'cursive'.

In the 14th century, the Old Uyghur script was adapted for block-printing. This advancement established a style that may be considered a 'print standard'. Block prints resemble the 'square' style, but given the period in which they were produced, the character repertoire is abridged and does not contain distinctive forms for palaeographic letters. Numerous folios and fragments of block-printed books have been preserved. This 'standard' block-print style is similar to the inscriptional type, which appears on the stone walls of the Cloud Platform at Juyong Guan, Beijing, erected in the 14th century (see fig. 41).

5 Character repertoire

The proposed encoding for Old Uyghur provides characters that support the encoding of typical documents in the script, from the earliest period to the latest. Its basis is the inventory of the alphabet attested in the 10th century manuscript U 40, which lists the independent forms of all letters. It is reinforced by the inventory provided in the 11th century treatise *Dīwān lughāt al-turk* by Kashgari. The proposed repertoire is further supported by attestations for the full range of letters used in facsimilies of Uyghur manuscripts printed by Müller in 1908 (see table 4). Numerous manuscripts written in various styles have been analyzed to identify diacritics and other characters for inclusion in the repertoire. The proposed repertoire provides for encoding of the following:

- All palaeographically distinct letters as attested in extant Old Uyghur manuscripts, block prints, and printed facsimiles.

- Generic representation of a pair of high-frequency letters that are written in some documents using a single grapheme. Letters are included for handling generic or merged forms of *aleph-nun* and *beth-yodh*. These letters are to be used in contexts where the pair of letters are written using the same shape, eg. medial and final forms in numerous documents, and the actual underlying letter is graphically unidentifiable, but may be discerned morphologically. These characters enable non-specialist users to encode texts where the true reading of a letter is unknown, and without the need for understanding Turkic morphology.

- Alternate form of *aleph* to be used where a semantic distinction may be conveyed. Typically, the Unicode character-glyph model would consider the orientation of a terminal to the west instead of the typical southward direction as a glyphic variant of the normative glyph. But, for *aleph*, such a curved variant has been encoded as an atomic character on account of its semantic function. This approach eliminates the need for using variation selectors or font changes for representing contrastive usage.

- Dotted forms of letters used for disambiguation and extension of the alphabet. The script is known for its usage of diacritics, namely dots, to distinguish between different phonetic values expressed by a single letter. As there is no standard for usage for these diacritics, they are encoded as combining signs. These signs may be applied to any letter to represent the diacritic form. This approach reduces the need to encode precomposed letters consisting of the base letter + diacritic.

The encoding does not aim to, nor could it be expected to, provide a means for handling all ambiguous readings that result from cursive or careless writing. A Unicode encoding cannot attempt to account for all idiosyncratic scribal practices that result in ambiguous readings. Indecipherability of a piece of text is not so much a problem of what is written — the underlying text was likely written to convey a single value by the writer and may have been comprehensible to a reader familiar with the styles and nuances of writing during that time – but it is a problem for a modern reader who is unfamiliar with the script or underlying language. It is unreasonable to expect that a Unicode encoding would fully enable a user attempting to transcribe a piece of text without knowing how to distinguish one letter from the other without knowing the underlying language or rules of the script, whether it is Old Uyghur or Latin, Cyrillic, Kaithi or any other script with quite dynamic cursive traditions. The natural ambiguity of a cursive text that could be read in multiple way by a person unfamiliar with the language, might probably be deciphered quickly by someone familiar with the language through morphological and syntactic or other linguistic contexts. Yet, a Unicode encoding should also enable any user — be they a scholar of Turkology, a general cataloguer, or someone entirely unfamiliar with the script — to represent a text in front of them using a combination of ingenuity and a repertoire of characters that match the graphemes on the page. To that extant, if there are pairs of letters in addition to *aleph-nun* and *beth-yodh* that are consistently confounded, then they may be encoded in the future.

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 repertoire is sufficient for representing the majority of Old Uyghur texts. There are other diacritics, punctuation, digits, and other symbols, that require additional research before being proposed for encoding in the future.

The representative glyphs are based upon the independent shapes of letters showing in U 40 and Kashgari. Such depictions suggest a tradition of representing the script using these forms. These representative glyphs differ from the practice of basing independent forms upon the final forms of letter for most cursive joining scripts in Unicode. Contextual forms of the letters are based upon normalizations of forms attested across the available sources, and verified by the forms used in the printed facsimilies by (see table 4) The representative glyphs have been designed to reflect the general aesthetics of the block-print style.

The names of Old Uyghur letters are based upon scholarly names for the original Sogdian letters, which in turn reflect the ancestral Aramaic names. Throughout this proposal, italics are used for scholarly names for graphemes, while small capitals indicate Unicode character names, eg. 𐽀 is referred to as the grapheme *aleph* and the Unicode character OLD UYGHUR LETTER ALEPH. For brevity, in references to the Unicode character, 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. In horizontal 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, eg. in horizontal layout the ◌𐽆 COMBINING DOT RIGHT would appear as ◌́, a 'dot below' sign.

5.1 Characters proposed for encoding

The repertoire contains 35 characters: 19 basic letters, 2 generic letters, 7 combining signs, 6 punctuation signs, 2 baseline modifiers, and 1 editorial sign. The code chart and names list follows p. 12.

Basic letters

Character name	Glyph	Joining	Latin
OLD UYGHUR LETTER ALEPH		dual	ʾ
OLD UYGHUR LETTER ALEPH WITH LEFT TAIL		right	-ʾ
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		dual	y
OLD UYGHUR LETTER KAPH		dual	k
OLD UYGHUR LETTER LAMEDH		dual	δ
OLD UYGHUR LETTER MEM		dual	m
OLD UYGHUR LETTER NUN		dual	n
OLD UYGHUR LETTER SAMEKH		dual	s
OLD UYGHUR LETTER PE		dual	p
OLD UYGHUR LETTER SADHE		dual	c
OLD UYGHUR LETTER RESH		dual	r

OLD UYGHUR LETTER SHIN	𐽂	dual	š
OLD UYGHUR LETTER TAW	𐽆	dual	t
OLD UYGHUR LETTER LESH	𐽊	dual	l

Letters for representing ambiguous forms

Character name	Glyph	Joining	Latin
OLD UYGHUR LETTER GENERIC ALEPH-NUN	𐽐	dual	ʾ, n
OLD UYGHUR LETTER GENERIC BETH-YODH	𐽑	dual	β, y

Combining signs

Character name	Glyph
OLD UYGHUR COMBINING DOT RIGHT	◌𐽘
OLD UYGHUR COMBINING TWO DOTS RIGHT	◌𐽙
OLD UYGHUR COMBINING THREE DOTS RIGHT	◌𐽚
OLD UYGHUR COMBINING DOT LEFT	𐽘◌
OLD UYGHUR COMBINING TWO DOTS LEFT	𐽙◌
OLD UYGHUR COMBINING THREE DOTS LEFT	𐽚◌
OLD UYGHUR COMBINING HAMZA LEFT	𐽛◌

Punctuation signs

Character name	Glyph
OLD UYGHUR PUNCTUATION BAR	
OLD UYGHUR PUNCTUATION TWO BARS	
OLD UYGHUR PUNCTUATION TWO DOTS	
OLD UYGHUR PUNCTUATION FOUR DOTS	
OLD UYGHUR PUNCTUATION FIVE DOTS	
OLD UYGHUR SECTION MARK	

Editorial sign

Character name	Glyph
OLD UYGHUR DELETION MARK	

Baseline modifiers

Character name	Glyph	Joining
OLD UYGHUR STEM EXTENDER		dual
OLD UYGHUR ORNAMENTAL TERMINAL		right

5.2 Characters not proposed for encoding

The following characters are not proposed for encoding at this time. They may be considered glyphic variants of regular letters (see description for each respective letter in § 7.1). If there is a need for representing these characters in plain text, then a proposal for adding them to the repertoire may be submitted in the future.

Characters not proposed for encoding

Name	Glyph
‘toothed’ *aleph*	
‘toothed’ *aleph* with curved terminal	
aleph / *nun* with rightward terminal	
beth with curved terminal	
kaph with curved terminal	
sadhe with rightward terminal	
taw with rightward terminal	

6 Shaping behavior

The model is based upon that used for cursive joining scripts in Unicode. Each basic letter of the script is included in the encoded repertoire, with representative glyphs based upon the independent shape. The contextual forms of each letter are produced using a shaping engine, which substitutes the atomic letter with the appropriate positional glyph.

Dual-joining characters

	independent	final	medial	inital
ALEPH				
GENERIC ALEPH-NUN				
BETH				
GENERIC BETH-YODH				
GIMEL-HETH				
WAW				

YODH	𐽁	𐽁	𐽁	𐽁
KAPH	𐽂	𐽂	𐽂	𐽂
LAMEDH	𐽃	𐽃	𐽃	𐽃
MEM	𐽄	𐽄	𐽄	𐽄
NUN	𐽅	𐽅	𐽅	𐽅
SAMEKH	𐽆	𐽆	𐽆	𐽆
PE	𐽇	𐽇	𐽇	𐽇
SADHE	𐽈	𐽈	𐽈	𐽈
RESH	𐽉	𐽉	𐽉	𐽉
SHIN	𐽊	𐽊	𐽊	𐽊
TAW	𐽋	𐽋	𐽋	𐽋
LESH	𐽌	𐽌	𐽌	𐽌
STEM EXTENDER	𐽐	𐽐	𐽐	𐽐

Right-joining characters

	independent	final
ALEPH WITH LEFT TAIL	𐽀	𐽀
FINAL HETH	𐽍	𐽍
ZAYIN	𐽎	𐽎
ORNAMENTAL TERMINAL	𐽏	𐽏

6.1 Collation

The sort order for Old Uyghur letters follows the encoded order:

𐽀 ALEPH < 𐽁 GENERIC ALEPH-NUN < 𐽂 ALEPH WITH LEFT TAIL < 𐽃 BETH <

𐽄 GENERIC BETH-YODH < 𐽅 GIMEL-HETH < 𐽆 WAW < 𐽇 ZAYIN < 𐽈 FINAL HETH <

𐽉 YODH < 𐽊 KAPH < 𐽋 LAMEDH < 𐽌 MEM < 𐽍 NUN < 𐽎 SAMEKH < 𐽏 PE <

𐽐 SADHE < 𐽑 RESH < 𐽒 SHIN < 𐽓 TAW < 𐽔 LESH

7 Description of proposed characters

7.1 Letters

7.1.1 *aleph* and *nun*

The 𐽀 *aleph* and 𐽁 *nun* are distinctive letters of the script. They are derived, respectively, from Sogdian 𐼰 *aleph* and 𐼱 *nun*. Palaeographically, the body of the Uyghur *aleph* is triangular and has a sharp point at the top left; while the Uyghur *nun* is rounded. These two letters present some challenges for character encoding. In some texts their shapes are contrasted in all positions; in others, the distinctions between them are less evident in some positions. It is significant to note that the contrast between *aleph* and *nun* is maintained in the printed reproductions of Uyghur manuscripts in Müller's *Uigurica*, published in 1908 (see fig. 42). A description of the letters in various positions is given below:

- *Independent* The independent 𐽀 *aleph* and 𐽁 *nun* are attested in U 40 and by Kashgari. The independent *aleph* appears commonly (see fig. 13), and has the following alternate forms:

  – A word-final *aleph* may be written independently, detached from the previous letter, regardless of the joining behavior of the latter (see fig. 16). It is written using the regular independent form 𐽀 or the alternate 𐽂 with a vertical terminal. In some cases, the two are used concurrently for distinguishing between final *a* (𐽀) and *ä* or *e* (𐽂), see fig. 4; also see forms used for *-a* in fig. 7. The 𐽂 is not used for *nun*.

  – The independent 𐽀 *aleph* is represented in some documents using the 'toothed' form 𐽕 (see fig. 13). This stylistic variant resembles the letter 𐽖 *kaph*. When the 𐽕 variant is used, the 𐽂 takes also has a 'toothed' shape 𐽗. The 'toothed' variants 𐽕 / 𐽗 should be handled as a stylistic set when used in place of 𐽀 / 𐽂. These 'toothed' forms are not used for *nun*.

- *Initial* The initial 𐽘 *aleph* and 𐽙 *nun* are preserved in carefully written texts, such as Mainz 126 and Pelliot Ouïgour 13 (see fig. 14), and printed facsimiles (see fig. 12). In other documents where contrast between the letters is not well maintained, the initial form of *aleph* may resemble that of *nun*; or initial *nun* may resemble *aleph*; or the two may be written using a generic shape that approximates their structures, such as 𐽚.

- *Medial* In Müller (1908), there is a clear distinction between the medial ◮ *aleph* and medial ◮ *nun*, where the former is more hooked and shorter than the latter (see fig. 12). However, in the majority of documents the medial forms are not contrasted consistently, or at all. Some perceived lack of contrast may be ascribed to the thick strokes that are characteristic of some scribal practice. Some actual lack of contraction may be due to the ambiguities inherent in cursive or rapid writing where there is less consideration for producing letters carefully. Some In such cases the medial form of both letters is written using a shape resembling that of *aleph* or *nun*, or a generic shape such as ◮.

- *Final* In Müller (1908), there is a clear distinction between the final ـــ *aleph* and final ـــ *nun*, where the former is more hooked and shorter than the latter, and the latter is characterized by a slightly curved terminal (see fig. 12). However, in several manuscripts and block prints, the final forms of both letters are written using a single form:

  – *Manuscripts* ـــ or ـــ or a swash variant in which the body of the letter forms a curved stroke with the terminal (see fig. 15).

  – *Block prints* in some block prints it appears that the finals are differentiated: the body of final ـــ *aleph* is triangular with points at east and south (from a vertical perspective), while the final ـــ *nun* is slanted eastward. The difference is supported by the chart in fig. 8 showing the forms of Uyghur letters used in the inscription at Juyong Guan pass. Whether or not ـــ is in fact distinct from ـــ, the highly similar structures of these glyphs lend themselves to being interpreted as the same letter, and in several documents, that is the case.

  – The *aleph* and *nun* is also written as ـ when final. This form occurs concurrently with the regular final *aleph* (ـــ, ـــ) in several manuscripts. This form has both semantic and stylistic functions. It is used in the middle of words as a morphological separator (Matsui, personal correspondence, November 2018; see also fig. 16). Also, it is used at the end of a line or at a text margin when there is limited space for the horizonal terminal of the *aleph* or *nun*

  – Nonetheless, there are exceptions for final *aleph*, which are as follows:

    * Following *kaph* or *pe* Whether or not *aleph* is written distinctively from *nun*, when it follows *kaph* or *pe*, *aleph* is written using the palaeographical *aleph*, eg. ـــ *k ›*, as is the convention for *kaph* or *pe* before *aleph* (see § 9.0.1). Even when *aleph* is not distinguished from *nun* in medial or final position, when it follows *kaph* or *pe*, it is written distinctively. This final form occurs concurrently with the regular final *aleph*, and is attested in manuscripts and block prints (see fig. 17). In documents where the ـــ / ـا 'toothed' form of independent *aleph*, described above, is used instead of the regular independent form ـــ, it is used with penultimate *kaph* and *pe* as well: ـــ / ـا *k ›*, ـــ / ـا *p ›* (see fig. 17), compare to ـــ *k ›*, ـــ *p ›*. Such contextual glyphic variation should be considered conventional behavior.

- *Disambiguation* Due to the ambiguity of these two letters in some documents, the ◌́ is written above *nun* in order to distinguish it from *aleph* when the two letters are indistinct, compare ـــ vs ـــ for /n/ and /a/, respectively (see § 8.3).

The various forms of *aleph* and *nun* are summarized in the table below:

		X_n	X_f	X_m	X_i
aleph	regular				
	alternate			—	—
	variants			—	—
nun	regular				
	variant	—		—	—
merged	regular				
	variant	—		—	—

The ambiguity posed by the loss of contrast between *aleph* and *nun* in medial and final positions in various sources adds complexity for uniquely encoding characters that have distinct shapes in some contexts, but that have similar or identical shapes in others. Despite the fact that the rendering of *aleph* and *nun* using a single glyph in various contexts is an inherent aspect of some styles of the writing system, the encoding model should enable a means for uniquely encoding a string containing *aleph* and *nun* such that there is a one-to-one correspondence between a glyph and the identity of the underlying character. The encoding model for *aleph* and *nun* should enable representation of the following in plain text:

- the distinctive independent  *aleph* and *nun*, and the merged
- the distinctive initial forms *aleph* and *nun*, and the merged initial form
- the distinctive medial forms *aleph* and *nun*, and the merged medial form
- the distinctive final forms *aleph* and *nun*, and the merged final form
- the alternate independent form of *aleph*

Given the above, the following model is practical for encoding *aleph* and *nun*:

		X_n	X_f	X_m	X_i
ALEPH	dual				
ALEPH WITH LEFT TAIL	right			—	—
NUN	dual				
GENERIC ALEPH-NUN	dual				

- This approach follows the typical model for cursive joining scripts and can fully represent all occurrences of *aleph* and *nun*.
- It encodes the palaeographical forms *aleph* and *nun* as separate characters.
- It also encodes a generic unified *aleph-nun* to be used in cases where the forms of the two letters are not contrasted.
- Alternate forms are represented as atomic letters, without need for variation selection or font switching.
- The contextual substitution for ـــه *aleph* following penultimate *kaph* and *pe* is to be handled by the font as part of the regular shaping behavior for the script.
- The 'toothed' forms ـــه / ـا of *aleph* are to be treated as stylistic variants of ـــه / ـا.
- The alternate final form ـ of *aleph* and *nun* is to be treated as a stylistic variant.
- The final forms ـــه *aleph* and ـــه *nun* used in block prints are to be treated as stylistic variants.

Use the regular letters ALEPH and NUN when letterforms can be distinguished, and the unified character ALEPH-NUN when the letterform is ambiguous.

7.1.2 *beth* and *yodh*

The letters ـــه *beth* and ـه *yodh* are palaeographically distinctive letters in the script. They are distinguished in all positions (see fig. 18, 19, 20).

In some less carefully written documents, they are written using an ambiguous form ـه that approximates the general outline of the two letters (see fig. 21). In such documents, this generic form may be used concurrently with distinctive forms of the letter, which further contributes to the internal ambiguous identity of the letters. A native reader of the script would have recognized the intended value of the ambigious sign through morphological contexts: *beth* occurs much less frequently than *yodh*, and typically in transcriptions of non-Turkic words. Although § 4.3 warns that the proposed encoding cannot be expected to accommodate ambiguities presented by cursive writing, the case of a generic *beth-yodh* is warranted given the available evidence. The ambiguous representations of *beth* in a form that resembles *sadhe*, however, require additional research for determining the need for encoding a generic character; if necessary, such a generic character may be proposed for encoding in the future. Given the above, the following model is practical for encoding complete representation of *beth* and *yodh*:

		X_n	X_f	X_m	X_i
BETH	dual	ـ	ـ	ـ	ـ
YODH	dual	ـ	ـ	ـ	ـ
GENERIC BETH-YODH	dual	ـ	ـ	ـ	ـ

The regular final form of *beth* is ـڢ, however, the final is also written as ـڤ (see fig. 20). The left-ward orientation of the tail is used likely for distinguishing ـڢ *beth* from ـۑ *yodh* when there is a limitation of space for extending the final stroke of the former. This curved form is to be treated as a stylistic variant.

Use the regular letters BETH and YODH when letterforms can be distinguished; use the character that best resembles the glyph used for both letters in the given source.

7.1.3 *gimel* and *heth*

As evidenced by the inventories in U 40 and Kashgari, these two letters are distinguished in independent positions using the glyphs ـۿ and ـۼ, respectively. However, they have the same ۿ initial and ۾ medial forms (also see fig. 22). Final *heth* occurs as both ـۿ and ـۼ, but the latter is not used for *gimel*. The ـۼ is also used for distinguish *heth* from *gimel* in final position. To enable the complete representation of these two letters, the following model is proposed:

		X_n	X_f	X_m	X_i
GIMEL-HETH	dual	ۿ	ۿ	۾	۾
FINAL HETH	right	ۼ	ۼ	—	—

For representing inventories of the script, GIMEL-HETH is to be used for *gimel* and FINAL-HETH for *heth*. Use the unified letter GIMEL-HETH, but use FINAL HETH for representing the alternate form as necessary. For distinguishing between different phonetic contexts of *gimel* and *heth*, the diacritics ◌ٙ and ◌ٚ may be used with the letters, eg. ـۿٙ, ـۿٚ, ـۼٙ, ـۼٚ (see § 8.3).

There is not sufficient evidence available at present to determine the need for encoding a generic character to handle the ambigious representation of *gimel-heth* and *shin* using a single character. If a requirement to encode such a character is made clear in the future, then a proposal may be submitted at that time. As mentioned in § 4.3, the encoding cannot be expected to accommodate all ambiguities that occur in written texts.

7.1.4 *waw*

The letter ـۄ *waw* is consistently represented in Old Uyghur documents.

7.1.5 *zayin*

The representative form ـۯ of *zayin* is based upon the shape used in block-print styles (see fig. 23). The glyphic variant ـۮ 'sawtooth' form occurs in manuscripts (see fig. 11). In some sources *zayin* is distinguished using the diacritics ◌ٕ and ◌ٖ, eg. ـۯٕ and ـۯٖ, in order to indicate /ž/ (see § 8.3).

7.1.6 *kaph* and *resh*

The letters ـڬ *kaph* and ـڭ *resh* have a similar body, but are distinguished by their contexual forms.

The regular final form of *kaph* is 𐽔, however, the final is also written as 𐽕 (see fig. 24). The left-ward orientation of the tail is used to accommodate space constraints on a line. It is to be treated as a stylistic variant.

7.1.7 *lamedh*

The letter 𐽖 *lamedh* is consistently represented in Old Uyghur documents.

7.1.8 *mem*

As attested in the inventory in U 40, the *mem* has two distinctive graphemes: 𐽙 and 𐽚. These are the independent and final forms, respectively. Following the cursive joining model, the final form would be rendered when *mem* occurs in final position in a string.

7.1.9 *samekh* and *shin*

As shown in U 40, the letters 𐽛 *samekh* (/s/) and 𐽜 *shin* (/š/) are palaeographically distinctive letters in the script. The two letters are distinguished by the fact that *samekh* is written using two strokes (the first with a right-sloping downward angle and the second as a leftward curve extending from the midpoint of the first), while *shin* is a single stroke (right-sloping downward angle with a sharp pivot to the left). These distinctions are observed in several texts in initial, medial and final positions (see fig. 25, 26, 27).

By the 11th century, both letters were written using a similar glyph (see fig. 9). The form for *samekh* / *shin* in documents from this time is based upon the simpler 𐽜 *shin* instead of 𐽛 *samekh*. In such contexts, the diacritic ◌̈ is applied to 𐽜 *shin* to express /š/, eg. 𐽜̈, or 'marked' or 'dotted' *shin* (see § 8.3).

Use the regular letters SAMEKH and SHIN when letterforms can be distinguished. In texts where /s/ and /š/ are written using a single letter, depending upon the shape of the letter, either SAMEKH or SHIN should be used. Use the character that best resembles the glyph used for both letters in the given source.

7.1.10 *pe*

In various manuscripts and block prints, final 𐽗 *pe* is rendered as the ornamental form 𐽗𐽲 (see fig. 28). 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 may be considered a stylistic variant of the regular final *pe*. However, there is some evidence that the ornamental terminal may in fact be a separate grapheme 𐽲, which would mean that 𐽗𐽲 is actually a sequence of non-final *pe* and the space-filling sign 𐽲. This sign has been encoded as a combining sign (see § 7.5).

7.1.11 *sadhe*

The regular final form of *sadhe* is 𐽝, however, the final is also written as 𐽞 (see fig. 29). The left-ward orientation of the tail is used to accommodate space constraints on a line. It is to be treated as a stylistic variant.

7.1.12 *taw*

The body of the initial form 𐽅 of *taw* sits below the baseline, as compared to its medial 𐽅 and final 𐽆 forms. This practice is exhibited in manuscripts and block prints, and may be accepted as normative behavior. The depth of the body of the initial form differs by source. In some cases, the final stroke of the loop meets the stroke of the next letter at the baseline. In other sources, where the terminal looped stroke of *taw* connects with the initial vertical that produces the spine of the letter, the following letter connects to the initial *taw* where the spine of the *taw* meets the baseline.

The regular final form of *taw* is 𐽆, however, the final is also written as 𐽇 (see fig. 30). The left-ward orientation of the tail is used to accommodate space constraints on a line. It is to be treated as a stylistic variant.

7.1.13 *lesh*

The letter 𐽄 represents the sound /l/. It is derived from 𐾴 U+10F44 SOGDIAN LETTER LESH, which is known as 'hooked *r*' (see Pandey 2016b for details). The Uyghur 𐽄 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*, from which it is ultimately derived. The alias 'hooked *r*' has been specified in the names list.

7.1.14 Note on variation in terminal orientation

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

	regular	alternate
aleph	𐽀	𐽀, 𐽁
beth	𐽁	𐽂
kaph	𐽃	𐽄
nun	𐽉	𐽊
pe	𐽋	𐽌
sadhe	𐽍	𐽎
taw	𐽅	𐽇

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 alternate forms of *aleph* are proposed for encoding as separate characters. The alternate *pe* may be represented using a sequence of the letter and a combining sign for the ornamental terminal. The other alternate final forms are to be treated as glyphic variants. If a semantic difference between a variant and regular form is identified, then the variant form may be considered for encoding at that time.

7.2 Combining signs

The following combining signs are used for disambiguation and representation of new sounds (see § 8.3):

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. 31). In such documents they occur with the letters *gimel*, *heth*, and *samekh*.

In 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 ◌𐽑 ('ring right', as it would appear in a conventional vertical context, or ◌˚ in a horizontal 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.

7.3 Punctuation signs

The following signs are used for punctuation (see fig. 33 for examples):

- The signs 𐽙, 𐽚 are common forms of punctuation (see Knüppel 2002). They are used for delimiting text segments of various lengths, such as sentences. When these two signs are used together, 𐽚 indicates smaller segments, while 𐽛 closes longer sections (see fig. 38, 40). The sign 𐽙 is also used as a general delimiter. When it occurs in documents where 𐽚 is used, it represents short segments of text and may function as a comma or semi-colon.

- The signs 𐽛 and ❖ are used for indicating the end of larger portions of text. In some documents, 𐽛 is used in place of 𐽚, especially in cases of minimal punctuation. The sign ❖ generally indicates the end of a section or the completion of a text. While this sign is similar to the generic ⁖ U+2058 FOUR DOT PUNCTUATION already encoded in Unicode, the Old Uyghur ❖ 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. 39. It seems to have been borrowed from Sogdian scribal traditions; however, it is encoded as a script-specific sign on account of directional considerations.

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

7.4 Stem extender

The 𐽐 OLD UYGHUR STEM EXTENDER is used for extending the baseline (see § 8.4 for details). It is used as a typographic filler and also for indicating a suffix that is separated from the stem. The stem-extending sign is defined as a left-joining character.

7.5 Ornamental terminal

The 𐽑 OLD UYGHUR ORNAMENTAL TERMINAL is used for representing an ornamental terminal, eg. the final 𐽑𐽆 *pe* written at the end of a line (see fig. 28). Von Gabain shows the sign in her chart of the script as a "Zeilenfüller" (German "row-filler"; see fig. 7). The proposed ornamental terminal is defined as a left-joining character.

7.6 Editoral sign

When written beneath a word or letter, ◌𐽆 OLD UYGHUR DELETION MARK indicates that the respective text is an error and is to be omitted. In authentic representations of manuscripts, it is to be placed after the letter that carries the mark. The correct word is generally written after the mispelled word (see fig. 32).

8 Encoded representations

8.1 Examples

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:

ʾrslan
‘arslan’

𐽀 ALEPH, 𐽀 ALEPH, 𐽂 LESH, 𐽃 TAW, 𐽁 WAW, 𐽀 NUN

ʾʾltwn
‘altun’

𐽀 ALEPH, 𐽀 ALEPH, 𐽂 LESH, 𐽃 TAW, 𐽁 WAW, 𐽀 NUN

bʾmsʾn
‘vamsan’

𐽆 BETH, 𐽀 ALEPH, 𐽄 MEM, 𐽅 SAMEKH, 𐽀 ALEPH, 𐽀 NUN

wyγwr
‘üigür’

𐽁 WAW, 𐽉 YODH, 𐽇 GIMEL-HETH, 𐽁 WAW, 𐽈 RESH

qwtlwγ
‘qutlug’

𐽇 GIMEL-HETH, ◌𐽍 COMBINING TWO DOTS LEFT, 𐽃 TAW, 𐽂 LESH, 𐽁 WAW, 𐽇 GIMEL-HETH

mncwšry
‘mancusari’

𐽄 MEM, 𐽀 NUN, 𐽋 SADHE, 𐽁 WAW, 𐽅 SAMEKH, ◌𐽎 COMBINING TWO DOTS RIGHT, 𐽈 RESH, 𐽉 YODH

swδwr
‘sutur’

𐽅 SAMEKH, 𐽁 WAW, 𐽊 LAMEDH, 𐽁 WAW, 𐽈 RESH

pwδystb
'bodisatav'

𐽆 PE, 𐽼 WAW, 𐽶 LAMEDH, 𐽹 YODH, 𐽴 SAMEKH, 𐽺 TAW 𐽸 BETH,

pwrγʾn
'burxan'

𐽆 PE, 𐽼 WAW, 𐽷 RESH, 𐽲 GIMEL-HETH, 𐽸 ALEPH, 𐽸 NUN

pwlz-wn
'bolz-un'

𐽆 PE, 𐽼 WAW, 𐽵 LESH, 𐽳 ZAYIN, 𐽼 WAW, 𐽸 NUN

pylyk
'bilig'

𐽆 PE, 𐽹 YODH, 𐽵 LESH, 𐽹 YODH, 𐽻 KAPH

twykʾl
'tükäl'

𐽺 TAW, 𐽼 WAW, 𐽹 YODH, 𐽻 KAPH, 𐽸 ALEPH, 𐽵 LESH

tnkry
'tängri'

𐽺 TAW, 𐽸 NUN, 𐽻 KAPH, 𐽷 RESH, 𐽹 YODH

8.2 Vowels

In general, all vowels are indicated in Old Uyghur. There is an exception to the rule for writing out all vowels: in some words, the short /a/ (= /ä/) is not expressed, eg. *tängri* is written as , without an explicit *aleph* for /ä/.

The representation of vowels follows the basic '*matres lectionis*' pattern for Semitic scripts, in which 𐽸 *aleph*, 𐽼 *waw*, and 𐽹 *yodh* are used for indicating vowels. These letters are combined in digraphs and trigraphs in order to express the rich vowel repertoire of Turkic languages, as shown below.

	Initial	Medial
ä	𐽸 ALEPH	𐽸 ALEPH
a, *e*	𐽸 ALEPH, 𐽸 ALEPH	𐽸 ALEPH
i, *ï*	𐽸 ALEPH, 𐽹 YODH	𐽹 YODH

ī, *ī̄*	𐽼	𐽀 ALEPH, 𐽆 YODH, 𐽆 YODH	𐽼	𐽆 YODH, 𐽆 YODH
o, *u*	𐽻	𐽀 ALEPH, 𐽇 WAW	𐽻	𐽇 WAW
ö, *ü*	𐽽	𐽀 ALEPH, 𐽇 WAW, 𐽆 YODH	𐽻	𐽇 WAW
ö, *ü*	𐽽	𐽇 WAW, 𐽆 YODH	𐽽	𐽇 WAW, 𐽆 YODH
ō, *ō̄*, *ū*, *ū̄*	𐽾	𐽀 ALEPH, 𐽇 WAW, 𐽇 WAW	𐽾	𐽇 WAW, 𐽇 WAW

The final forms of all vowels are represented using the final form of ALEPH, WAW, or YODH, respectively.

8.3 Disambiguation and extension of letters

The combining signs enumerated in § 7.2 are written with letters to diambiguate consonants or to represent consonants for which distinctive letters do not exist. The following forms are attested. Combining signs are placed after a letter in encoded text:

		X_n	X_f	X_m	X_i	
dotted *gimel, heth*	*γ*	𐽂◌̇	𐽂◌̇	𐽂◌̇	𐽂◌̇	𐽂 GIMEL-HETH, ◌̇ COMBINING DOT LEFT
two-dotted *gimel, heth*	*γ*	𐽂◌̈	𐽂◌̈	𐽂◌̈	𐽂◌̈	𐽂 GIMEL-HETH, ◌̈ COMBINING TWO DOTS LEFT
dotted *zayin*	*ž*	𐽅◌̇	𐽅◌̇	–	–	𐽅 ZAYIN, ◌̇ COMBINING DOT RIGHT
two-dotted *zayin*	*ž*	𐽅◌̈	𐽅◌̈	—	—	𐽅 ZAYIN, ◌̈ COMBINING TWO DOTS RIGHT
dotted *heth*	*q*	𐽃◌̇	𐽃◌̇	—	—	𐽃 FINAL HETH, ◌̇ COMBINING DOT LEFT
two-dotted *heth*	*q*	𐽃◌̈	𐽃◌̈	—	—	𐽃 FINAL HETH, ◌̈ COMBINING TWO DOTS LEFT
dotted *nun*	*n*	𐽀◌̇	𐽀◌̇	𐽀◌̇	𐽀◌̇	𐽀 NUN, ◌̇ COMBINING DOT LEFT
two-dotted *shin*	*š*	𐽁◌̈	𐽁◌̈	𐽁◌̈	𐽁◌̈	𐽁 SHIN, ◌̈ COMBINING TWO DOTS RIGHT

8.4 Stem extension

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

tynly lr-r
'tinlag-lar-r'

𐽅 TAW, 𐽁 YODH, 𐽂 NUN, 𐽃 LESH, 𐽄 GIMEL, SP SPACE,
𐽃 LESH, 𐽁 RESH, SP SPACE,
𐽆 STEM EXTENDER, 𐽁 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 Glyph interactions

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

9.0.1 *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>	𐽋	𐽋
<PE, ALEPH>	𐽍	𐽍

9.0.2 *waw*

In initial and medial position, the tails of *kaph* and *pe* attach below the baseline of the following letter, eg. 𐽋𐽂 <KAPH, NUN> and 𐽍𐽂 <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>	𐽍	𐽍

9.0.3 *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. 𐽀𐽆𐽏 <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. 𐽀𐽀𐽆𐽏 <ALEPH, ALEPH, MEM, WAW>.

9.0.4 *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 may attach to the terminal of the latter for aesthetic considerations:

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

10 Character Properties

10.1 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;;;;;
10F73;OLD UYGHUR LETTER ALEPH-NUN;Lo;0;AL;;;;;N;;;;;
10F74;OLD UYGHUR LETTER BETH;Lo;0;AL;;;;;N;;;;;
10F75;OLD UYGHUR LETTER GIMEL-HETH;Lo;0;AL;;;;;N;;;;;
10F76;OLD UYGHUR LETTER WAW;Lo;0;AL;;;;;N;;;;;
10F77;OLD UYGHUR LETTER ZAYIN;Lo;0;AL;;;;;N;;;;;
10F78;OLD UYGHUR LETTER FINAL HETH;Lo;0;AL;;;;;N;;;;;
10F79;OLD UYGHUR LETTER YODH;Lo;0;AL;;;;;N;;;;;
10F7A;OLD UYGHUR LETTER KAPH;Lo;0;AL;;;;;N;;;;;
10F7B;OLD UYGHUR LETTER LAMEDH;Lo;0;AL;;;;;N;;;;;
10F7C;OLD UYGHUR LETTER MEM;Lo;0;AL;;;;;N;;;;;
10F7D;OLD UYGHUR LETTER NUN;Lo;0;AL;;;;;N;;;;;
10F7E;OLD UYGHUR LETTER SAMEKH;Lo;0;AL;;;;;N;;;;;
```

```
10F7F;OLD UYGHUR LETTER PE;Lo;0;AL;;;;;N;;;;;
10F80;OLD UYGHUR LETTER SADHE;Lo;0;AL;;;;;N;;;;;
10F81;OLD UYGHUR LETTER RESH;Lo;0;AL;;;;;N;;;;;
10F82;OLD UYGHUR LETTER SHIN;Lo;0;AL;;;;;N;;;;;
10F83;OLD UYGHUR LETTER TAW;Lo;0;AL;;;;;N;;;;;
10F84;OLD UYGHUR LETTER LESH;Lo;0;AL;;;;;N;;;;;
10F85;OLD UYGHUR COMBINING DOT RIGHT;Mn;220;NSM;;;;;N;;;;;
10F86;OLD UYGHUR COMBINING TWO DOTS RIGHT;Mn;220;NSM;;;;;N;;;;;
10F87;OLD UYGHUR COMBINING THREE DOTS RIGHT;Mn;220;NSM;;;;;N;;;;;
10F88;OLD UYGHUR COMBINING DOT LEFT;Mn;230;NSM;;;;;N;;;;;
10F89;OLD UYGHUR COMBINING TWO DOTS LEFT;Mn;230;NSM;;;;;N;;;;;
10F8A;OLD UYGHUR COMBINING THREE DOTS LEFT;Mn;230;NSM;;;;;N;;;;;
10F8B;OLD UYGHUR COMBINING HAMZA RIGHT;Mn;220;NSM;;;;;N;;;;;
10F8C;OLD UYGHUR PUNCTUATION BAR;Po;0;AL;;;;;N;;;;;
10F8D;OLD UYGHUR PUNCTUATION TWO BARS;Po;0;AL;;;;;N;;;;;
10F8E;OLD UYGHUR PUNCTUATION TWO DOTS;Po;0;AL;;;;;N;;;;;
10F8F;OLD UYGHUR PUNCTUATION FOUR DOTS;Po;0;AL;;;;;N;;;;;
10F90;OLD UYGHUR PUNCTUATION FIVE DOTS;Po;0;AL;;;;;N;;;;;
10F91;OLD UYGHUR SECTION MARK;Po;0;AL;;;;;N;;;;;
10F92;OLD UYGHUR STEM EXTENDER;Po;0;AL;;;;;N;;;;;
10F93;OLD UYGHUR ORNAMENTAL TERMINAL;Lo;0;AL;;;;;N;;;;;
10F94;OLD UYGHUR DELETION MARK;Mn;220;NSM;;;;;N;;;;;
```

10.2 Linebreak data: `LineBreak.txt`

```
10F70..10F84;AL # Lo [21] OLD UYGHUR LETTER ALEPH..OLD UYGHUR LETTER LESH
10F85..10F8B;CM # Mn  [7] OLD UYGHUR COMBINING DOT RIGHT..
                          OLD UYGHUR COMBINING HAMSA RIGHT
10F8C..10F91;AL # Po  [6] OLD UYGHUR PUNCTUATION BAR..OLD UYGHUR SECTION MARK
10F92;AL        # Po      OLD UYGHUR STEM EXTENDER
10F93;AL        # Po      OLD UYGHUR ORNAMENTAL TERMINAL
10F94;CM        # Mn      OLD UYGHUR DELETION MARK
```

10.3 Property list: `PropList.txt`

```
10F94           ; Extender # Po       OLD UYGHUR STEM EXTENDER
```

10.4 Shaping properties: `ArabicShaping.txt`

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

```
10F84; OLD UYGHUR LESH; D; No_Joining_Group
```

11 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

———. 2019. "Recommendations to UTC #158 January 2019 on Script Proposals" (L2/19-047). https://www.unicode.org/L2/L2019/19047-script-adhoc-recs.pdf

———. 2020. "Recommendations to UTC #162 January 2020 on Script Proposals" (L2/20-046). https://www.unicode.org/L2/L2020/20046-script-adhoc-rept.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; 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.

Klaproth, Heinrich Julius. 1812. *Abhandlung über die Sprache und Schrift der Uiguren*. Berlin.

Knüppel, Michael. 2002. "Zu den „Daṇḍas" und „Doppel-Daṇḍas" in der uigurischen Schrift". *Acta Orientalia Academiae Scientiarum Hungaricae*, vol. 55, no. 4, pp. 339–343.

Le Coq, Albert von. 1919. "Kurze Einführung in die uigurische Schriftkunde". *Mitteilungen des Seminars für Orientalische Sprachen an der Friedrich-Wilhelms-Universität zu Berlin*, vol. 22, pt. 2, pp. 93–109.

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.

———. 2004. “From Silk, Cotton and Copper Coin to Silver. Transition of the Currency Used by the Uighurs during the Period from the 8th to the 14th Centuries”. *Turfan Revisited: The First Century of Research into the Arts and Cultures of the Silk Road*, Desmond Durkin-Meisterernst, et al. (eds.), pp. 228–239. Berlin: Dietrich Reimer Verlag.

Müller, Friedrich Wilhelm Karl. 1908. *Uigurica*. Abhandlungen der Königlich Preußischen Akademie der Wissenschaften, Philosophisch-historische Klasse 2. Berlin: Verlag der Königlichen Akademie der Wissenschaften.

———. 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.

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

Osman, Omarjan. 2012. “Proposal for encoding the Uygur script in the SMP”. L2/12-066. https://www.unicode.org/L2/L2012/12066-uygur.pdf

———. 2013. “Proposal to Encode the Uyghur Script in ISO/IEC 10646”. L2/13-071. http://www.unicode.org/L2/L2013/13071-uyghur.pdf

Pandey, Anshuman. 2016. “Revised proposal to encode the Sogdian script in Unicode”. L2/16-371R2. http://www.unicode.org/L2/L2016/16371r2-sogdian.pdf

Radloff, W [Radlov, Vasiliĭ Vasil′evich]. 1908. “Die vorislamitisehen Schriftarten der Türken und ihr Verhältniss zu der Sprache derselben”. *Bulletin de l’Académie Impériale des Sciences de St.-Pétersbourg*, pp. 835–856.

———. 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.

———. 1911. *Kuan-ši-im Pusar: Eine türkische Übersetzung des XXV. Kapitels der chinesischen Ausgabe des Saddharmapuṇḍarīka*. Bibliotheca Buddhica XIV. St.-Pétersbourg: Commissionnaires de l’Académie Impériale des Sciences.

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

Ščerbak, A. M. 1982. “De l’alphabet ouigour”. *Acta Orientalia Academiæ Scientiarum Hungaricæ*, vol. 36, no. 1–3, pp. 469–474.

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. Mit 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 Academiæ Scientiarum Hungaricæ*, 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.

12 Acknowledgments

I express my gratitude to Dr. 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 am very grateful to Dr. Mehmet Ölmez (Department of Modern Turkic Languages and Literatures, Istanbul University) for sharing excerpts of his book with me, as well as images of the *Dīwān lughāt al-turk* of Kashgari, and several other publications.

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.

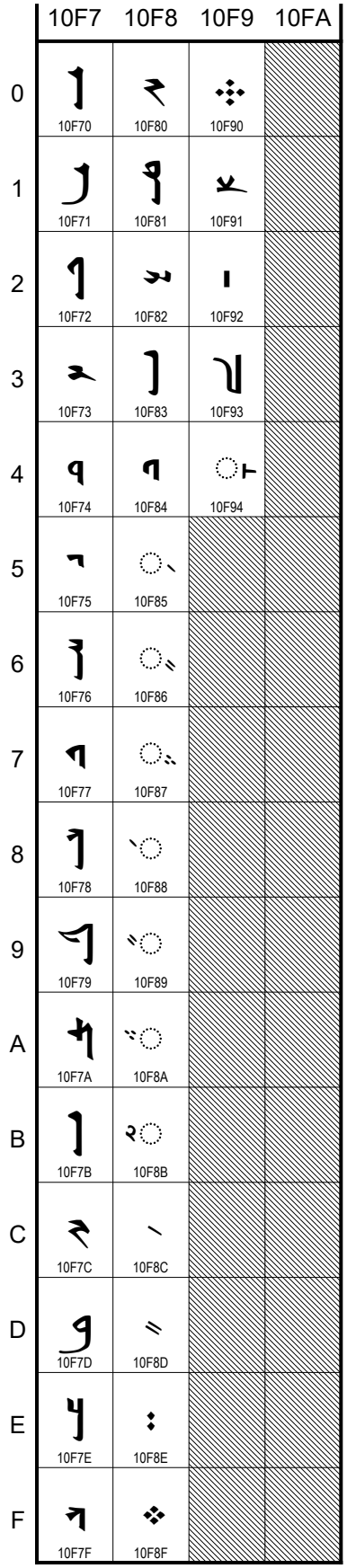
I also would like to thank Andrew West for providing reproductions and transcriptions of the Old Uyghur inscriptions on the walls of the Cloud Platform at Juyong Guan.

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

Letters for ambiguous readings

10F83 OLD UYGHUR LETTER GENERIC ALEPH-NUN
10F84 OLD UYGHUR LETTER GENERIC BETH-YODH

Combining signs

10F85 OLD UYGHUR COMBINING DOT RIGHT
10F86 OLD UYGHUR COMBINING TWO DOTS RIGHT
10F87 OLD UYGHUR COMBINING THREE DOTS RIGHT
10F88 OLD UYGHUR COMBINING DOT LEFT
10F89 OLD UYGHUR COMBINING TWO DOTS LEFT
10F8A OLD UYGHUR COMBINING THREE DOTS LEFT
10F8B OLD UYGHUR COMBINING HAMZA LEFT

Punctuation

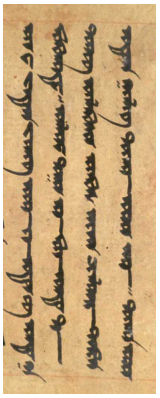
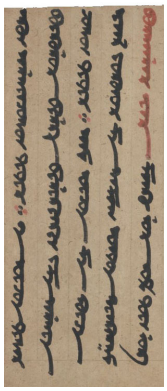
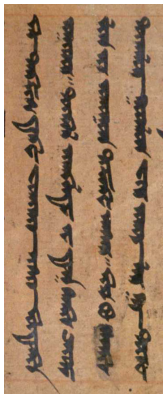
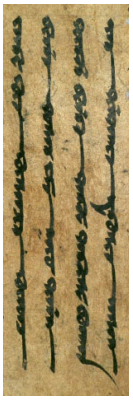
10F8C OLD UYGHUR PUNCTUATION BAR
10F8D OLD UYGHUR PUNCTUATION TWO BARS
10F8E OLD UYGHUR PUNCTUATION TWO DOTS
10F8F OLD UYGHUR PUNCTUATION FOUR DOTS
10F90 OLD UYGHUR PUNCTUATION FIVE DOTS
10F91 OLD UYGHUR SECTION MARK

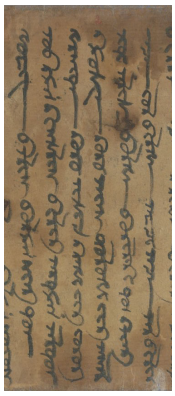
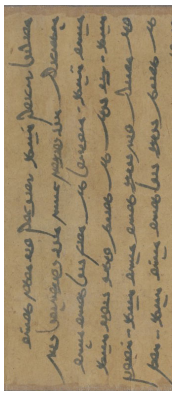
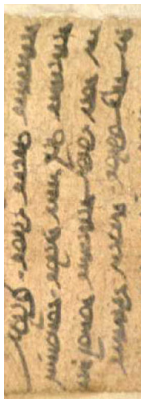
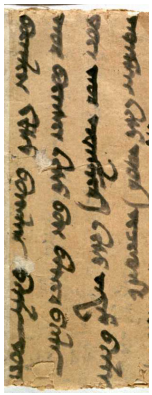
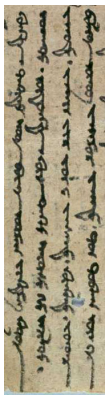
Baseline modifiers

10F92 OLD UYGHUR STEM EXTENDER
10F93 OLD UYGHUR ORNAMENTAL TERMINAL

Editorial signs

10F94 OLD UYGHUR DELETION MARK

Uyghur ‘square’ style					
Mainz 119	Mainz 841	Pelliot ouïgour 13	Mainz 819	Mainz 896	U 1071
					

Uyghur ‘semi-square’ style				
Pelliot chinois 2998	Pelliot chinois 3046	U 499	U 560	U 456
				

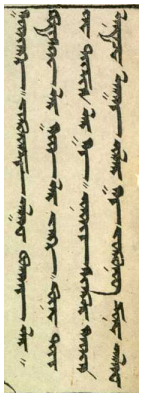
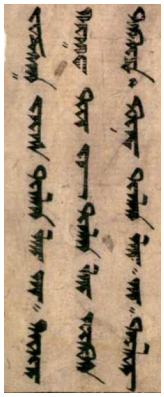
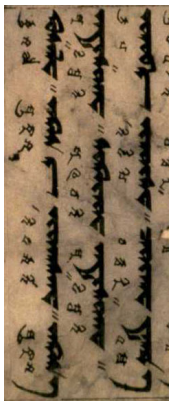
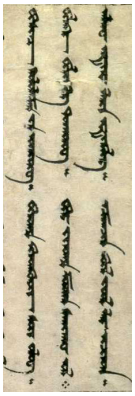
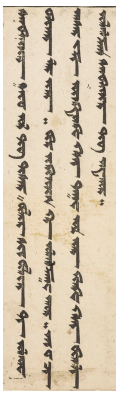
Uyghur block-print styles					
U 387	U 7008	Mainz 801	U 343	U 496	PEALD 6r
					

Table 1: Specimens of Old Uyghur ‘square’ styles

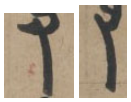
	PO 13 (fig. 34)			Mainz 126 (fig. 35)		
aleph						
beth	—	—		—	—	—
gimel / heth						
waw			—			
zayin		—	—		—	—
yodh						
kaph						
lamedh	—		—			—
mem						
nun						
samekh	—	—	—	—	—	—
pe	—	—			—	
sadhe						
resh			—			—
shin	—					
taw	—					
lesh						

Table 2: Specimens of letters from manuscripts

	U 387 – U 390 (fig. 37)		
aleph			
beth		—	—
gimel / heth			
waw			
zayin		—	—
yodh			
kaph			
lamedh	—		—
mem			
nun	—		
samekh	—	—	—
pe		—	
sadhe	—		
resh			—
shin			
taw			
lesh			

Table 3: Specimens of letters from block prints

	Muller (1908: 36–44)			
aleph				
beth	—			
gimel / heth				
waw				
zayin		—	—	
yodh				
kaph				
lamedh				
mem				
nun				
samekh			—	
pe				
sadhe				
resh				
shin				
taw				
lesh				

Table 4: Specimens of letters from printed facsimilies

Figure 1: BBAW, U 40, recto. Note the inventory of Old Uyghur letters at the bottom of the folio (see § 4.3 for additional details).

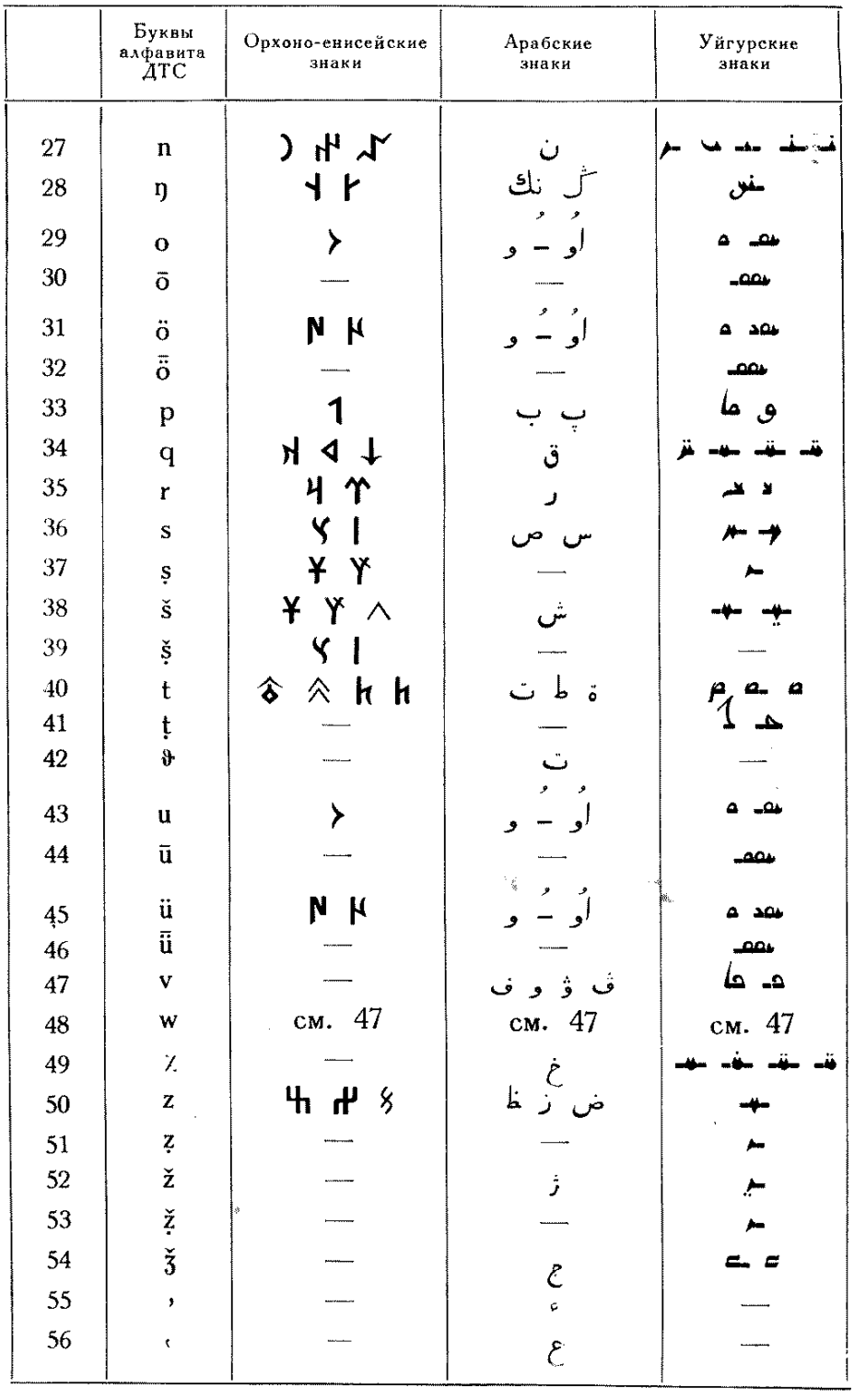
Figure 2: A folio from the *Dīwān lughāt al-turk*, written in the 11th century by Mahmud Kashgari. Note the Old Uyghur repertoire (black ink) with Arabic analogues (red ink). See fig. 3 for a mnemonic device containing the letters. Image courtesy of Mehmet Ölmez.

Figure 3: A folio from the *Dīwān lughāt al-turk*, written in the 11th century by Mahmud Kashgari. Note the Old Uyghur phrase at the top in red ink, which is a mnemonic device containing all letters of the script. See fig. 2 for a repertoire of the script. Image courtesy of Mehmet Ölmez.

— XV —

	Буквы алфавита ДТС	Орхоно-енисейские знаки	Арабские знаки	Уйгурские знаки
1	a	𐰀 𐰁	آ اَ ـَ	ـا ـا ـا ـا
2	ā	—	اَا	—
3	ä	𐰀 𐰁	اَ ـَ	ـا ـا ـا ـا
4	ǟ	—	—	—
5	b	𐰉 𐰋 𐰌	ب	ـو ـو
6	č	𐰍 𐰎	ج ح	ـج ـج ـج
7	d	𐰏 𐰑 𐰓	د (ض)	ـد ـد
8	ḍ	—	—	ـض ـض
9	ð	—	ذ	—
10	e	𐰀 𐰃	اِ ـِ ـَ	ـا ـا
11	ę	𐰇 𐰁 𐰀	ـَ اَ	ـا ـا
12	ē	—	اِ اَا	—
13	f	—	ف	ـف ـف
14	g	𐰍	ك گ	ـگ ـگ
15	ɣ	𐰴 𐰺 𐰺	غ	ـغ ـغ ـغ
16	h	—	ه	—
17	ḥ	—	ح	—
18	i	𐰀 𐰃	اِ ـِ ـَ ی	ـا ـا ـا
19	ī	—	اِ ـِ ـَ	ـا ـا
20	ï	𐰀 𐰃	اِ ـِ ـَ ی	ـا ـا ـا
21	ï̄	—	اِ ـِ ـَ	ـا ـا
22	j	𐰚 𐰜	ی	ـج ـج ـج
23	ǰ	𐰛 𐰝	—	—
24	k	𐰴 𐰺 𐰼 𐰽	ك	ـگ ـگ
25	l	𐰸 𐰺	ل	ـل ـل
26	m	𐰢 𐰣	م	ـم ـم ـم

Figure 4: Representation of Old Turkic sounds in the Orkhon, Arabic, and Old Uyghur scripts (from Nadeliaev, et al. 1969: xv). Continued in fig. 5.



— XVI —

	Буквы алфавита ДТС	Орхоно-енисейские знаки	Арабские знаки	Уйгурские знаки
27	n	𐰽 𐰢 𐰴	ن	ـنـ ـن ـنـ ـن
28	ŋ	𐰴 𐰺	ڭ نك	ـڭـ
29	o	𐰾	اُو ـُـو	ـوـ ـو
30	ō	—	—	ـوـ
31	ö	𐰆 𐰋	اُو ـُـو	ـوـ ـو
32	ȫ	—	—	ـوـ
33	p	𐰃	پ ب	ـپـ ـپ
34	q	𐰋 𐰇 𐰉	ق	ـقـ ـقـ ـق
35	r	𐰼 𐰚	ر	ـر ـر
36	s	𐰴 𐰃	س ص	ـسـ ـس
37	ṣ	𐰲 𐰳	—	ـصـ
38	š	𐰲 𐰳 𐰀	ش	ـشـ ـشـ
39	ṣ̌	𐰴 𐰃	—	—
40	t	𐰓 𐰔 𐰏 𐰏	ة ط ت	ـطـ ـطـ ـط
41	ṭ	—	—	ـطـ
42	ⱦ	—	ت	—
43	u	𐰾	اُو ـُـو	ـوـ ـو
44	ū	—	—	ـوـ
45	ü	𐰆 𐰋	اُو ـُـو	ـوـ ـو
46	ȳ	—	—	ـوـ
47	v	—	ف ۋ و ف	ـفـ ـف
48	w	см. 47	см. 47	см. 47
49	ɣ	—	غ	ـغـ ـغـ ـغـ ـغ
50	z	𐰏 𐰢 𐰽	ض ز ظ	ـضـ
51	ẓ	—	—	ـز
52	ž	—	ژ	ـژ
53	ẓ̌	—	—	ـز
54	ǯ	—	ع	ـعـ ـع
55	ʾ	—	ء	—
56	ʿ	—	ع	—

Figure 5: Representation of Old Turkic sounds in the Orkhon, Arabic, and Old Uyghur scripts (from Nadeliaev, et al. 1969: xvi). Continued from fig. 4.

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 6: Table showing letters of the Old Uyghur script (from Kara 1996: 540). See table of Mongolian letters from the same source in fig. 51.

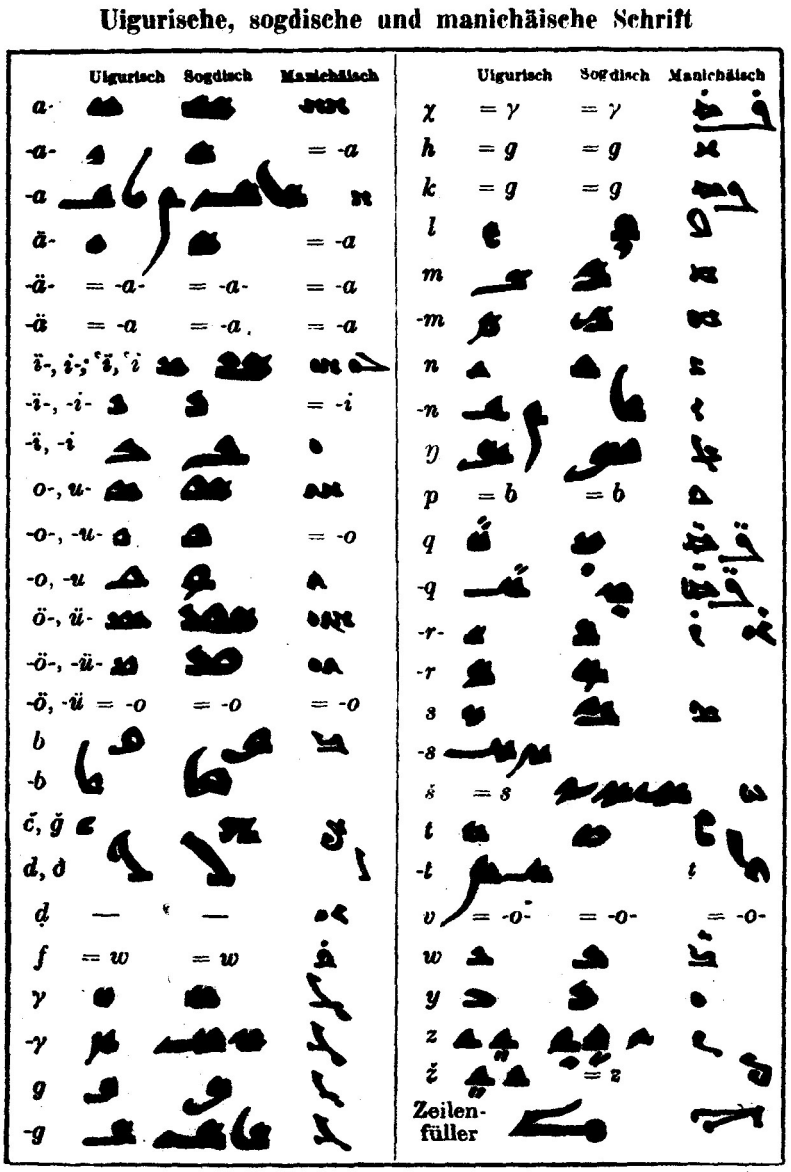


Figure 7: Comparison of Old Uyghur, Sogdian, and Manichaean letters (from von Gabain 1950: 17). For clearer examples of Old Uyghur letterforms referenced by von Gabain see the three Old Uyghur manuscripts, two in the formal script and the third in the cursive script, illustrated and transcribed in her work, reproduced here in fig. 45–49.

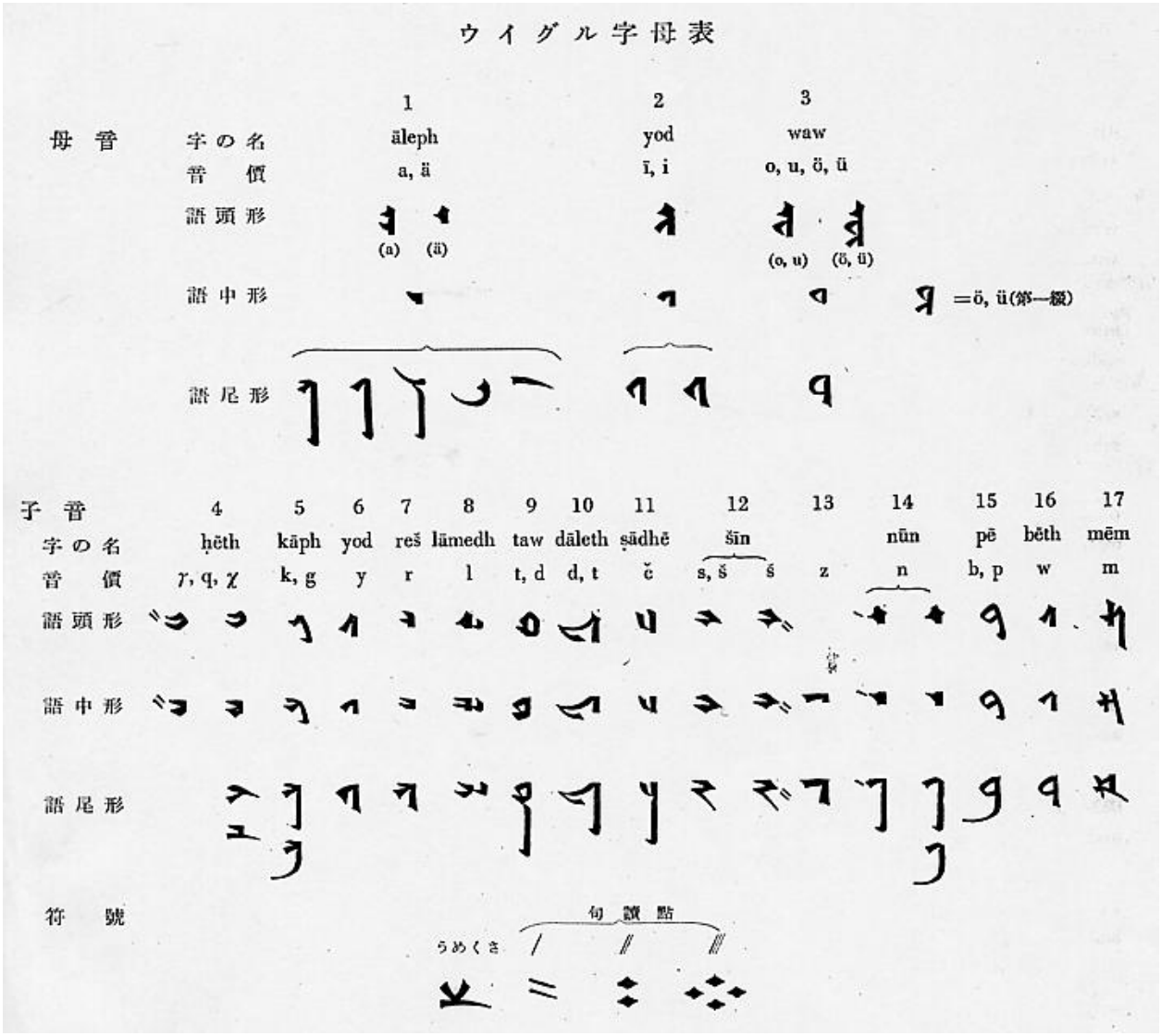


Figure 8: Table of Old Uyghur characters used in the Uyghur inscription in the multi-script Yuan dynasty inscriptions at Juyong Guan 居庸關 pass at the Great Wall northwest of Beijing (from Chü-Yung-Kuan 居庸關, "The Buddhist Arch of the Fourteenth Century A.D. at the Pass of the Great Wall Northwest of Peking", vol. 1, p. 165; reproduced from West 2006). See photograph containing an excerpt of the inscription in fig. 41.

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*

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 9: 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				

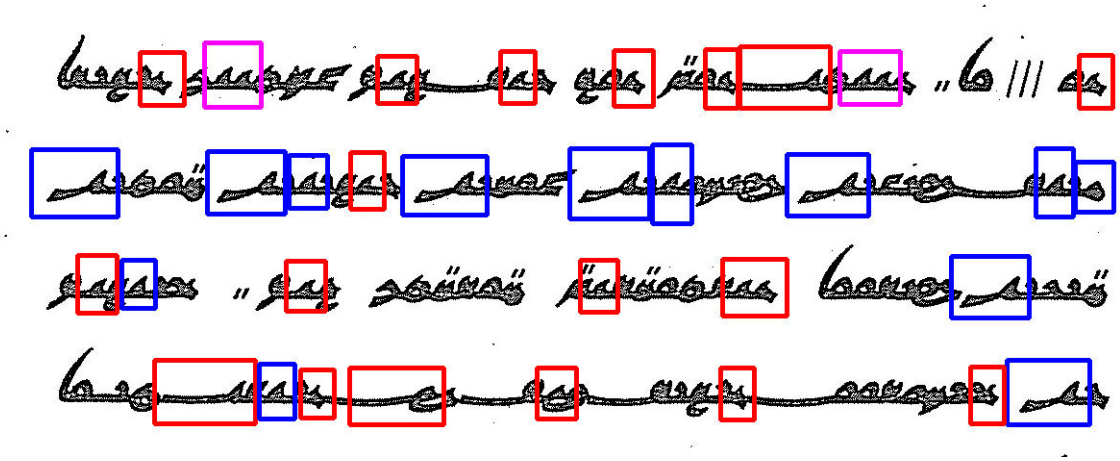
Table 2 Various forms of the Old Uighur alphabet from texts dating between the fourteenth and the tenth centuries BCE
Source: *adapted from Zieme 1991*

Figure 10: 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. 9). 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

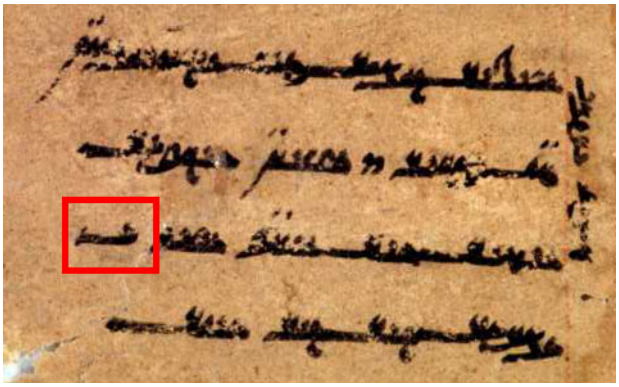
	Berliner Transkription system	Turkey	transcription at *Uigurisches Wörterbuch*	transliteration at *Uigurisches Wörterbuch*
𐽀	a, ạ	a, ạ	a	’’ / ’
𐽁	b	b	b	P
𐽂	č	ç	č	Č
𐽃	d, ṭ	d, ṭ	d, ḍ	D, T
𐽄	ä, ’ä	e, ’e	ä	’
𐽅	[e] i	ė / i	e	Y / ’Y
𐽆	g	g	g	K
𐽇 𐽈 𐽉	γ / γ́	g / ġ	g	Q, Q̈, Q̇
𐽊 𐽋 𐽌	h / χ, x, ẍ	h / ẖ, ḥ	h	H / X
𐽍	ï	ı	ı	Y, Y
𐽎	i	i	i	Y, ’Y
𐽏 𐽐	ǰ̣, ž	j	ǰ̣, ž	J̌, J̌, J̣
𐽑	k	k	k	K
𐽒	[k] q, q̈, q̇	k / ḳ	k	K / Q, Q̈, Q̇
𐽓	l	l	l	L
𐽔	m	m	m	M
𐽕	n, ṅ	n, ṅ	n	N, Ṅ
𐽖	ng, ñ, ŋ	n͡g, ng, ñ	ŋ	NK
𐽗 𐽘	o	o	o	W / ’W
𐽙 𐽚	ö, o̱	ö, o̱	ö	W / WY / ’WY
𐽛	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 11: Comparison of transliteration schemes for Old Uyghur (from Ölmez 2016).

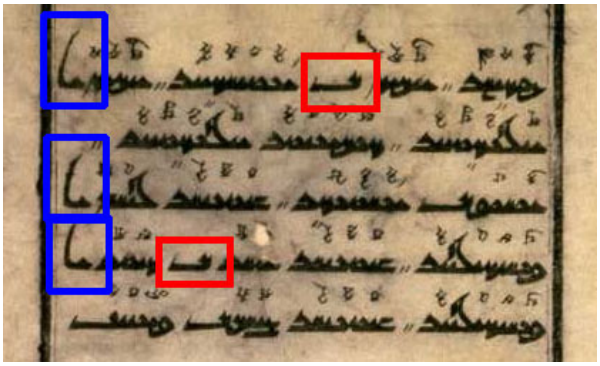


u . . . p ○○ anta oq̈ ol yäk-lär čštani ilig
? als nun jene Dämonen des Königs Tschastani
ning küčin küsünin čoγïn yaltï[nïn] . . q̈utïn
Kraft, Macht und Majestät
q̈ïvïn körüp artuq̈raq̈ q̈orq̈tï-lar ○○ ünlär-
sahen, noch mehr gerieten sie in Furcht. Ihre Stimme
in ösürüp ilig bäg-kä inčä tip
anschwellen lassend, zu dem Könige folgendermaßen

Figure 12: Attestations of contrastive forms of *aleph* (red) and *nun* (blue) in all positions in a printed facsimile of an Old Uyghur text (from Müller 1908: 42, 43). Sequences of *aleph* and *nun* are highlighted (magenta), as these clearly show the differences in the medial forms of the letters.

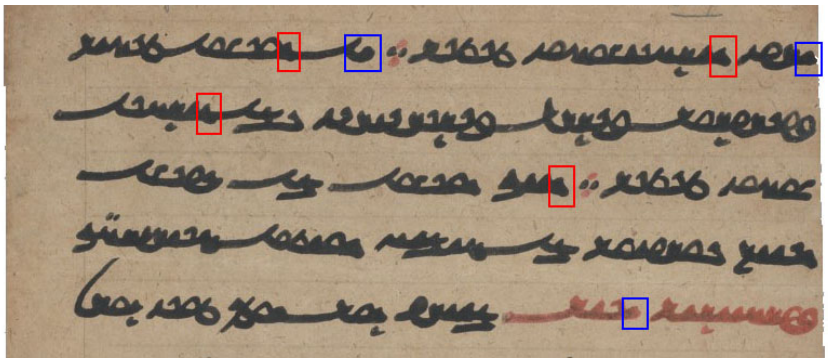


Example of the independent form ـــ of *aleph* (from U 2215).

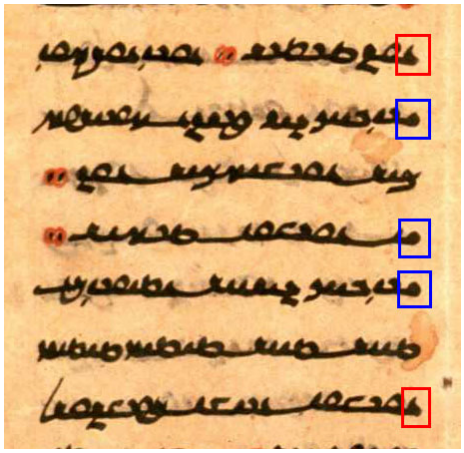


Examples of the 'toothed' variant ـــ and the curved variant ـــ of independent *aleph* (from Mainz 801).

Figure 13: Forms of independent *aleph*. Images have been rotated 90 degrees counter-clockwise for layout purposes.

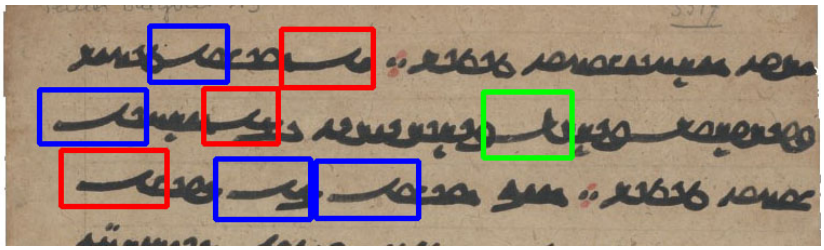


Excerpt from Pelliot Ouïgour 13 showing initial forms of 𐽀 *aleph* (red) and 𐽁 *nun* (blue).



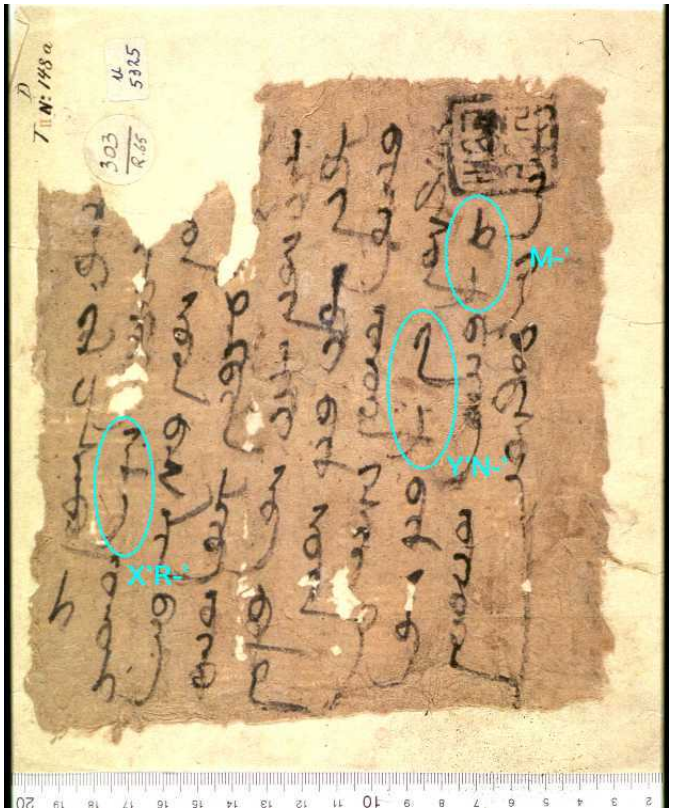
Excerpt from Mainz 126 showing initial forms of 𐽀 *aleph* (red) and 𐽁 *nun* (blue).

Figure 14: Contrastive representation of inital *aleph* and initial *nun*. Images have been rotated 90 degrees counter-clockwise for layout purposes.



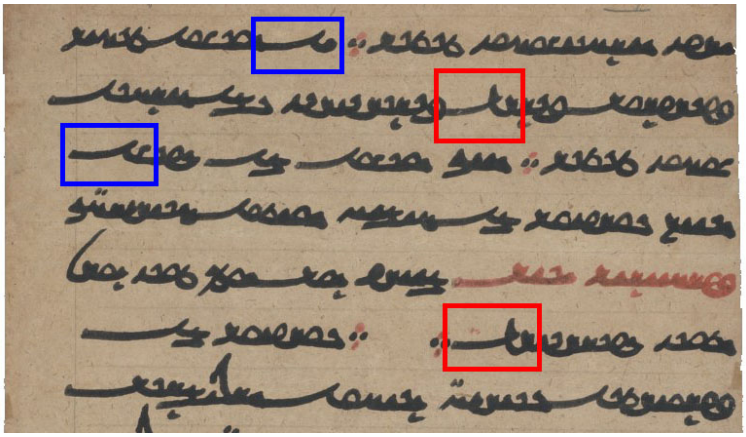
A single form used for final *aleph* (red) and *nun* (blue), and the distinctive final *aleph* (green) used after penultimate *kaph* (from Pelliot Ouïgour 13)

Figure 15: Examples of the generic form used for final *aleph* and *nun*. Images have been rotated 90 degrees counter-clockwise for layout purposes.

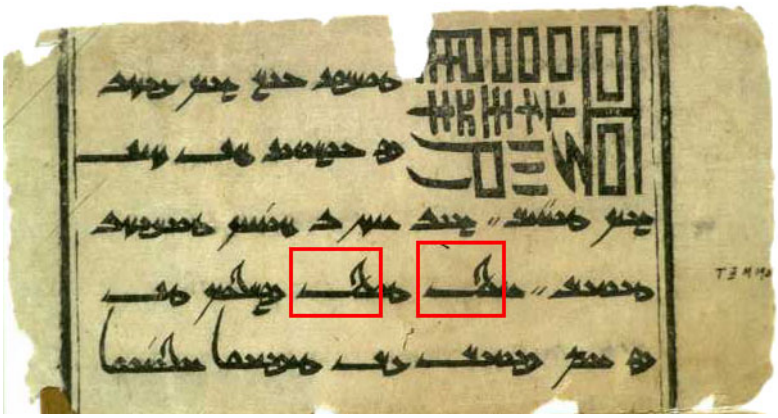


Usage of ﻟ ALEPH WITH RIGHT TAIL and ﺎ ALEPH WITH LEFT TAIL in U 5325. Annotations produced by Dai Matsui, November 2018.

Figure 16: Alternate forms of final *aleph*.

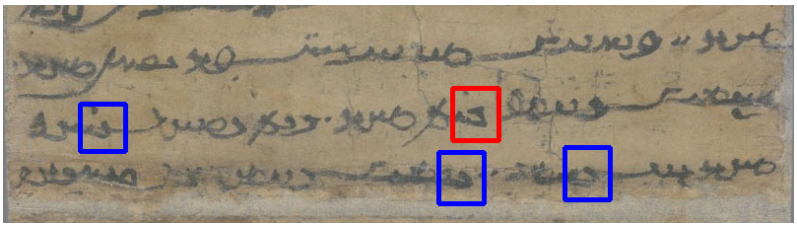


Excerpt from Pelliot Ouïgour 13 showing the final form of *aleph* used with *kaph* (red), compared with the regular form (blue).

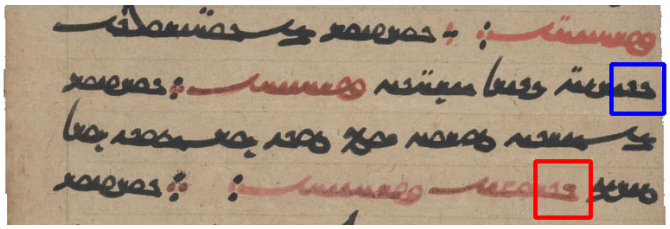


Folio from U 4960 showing the contextual form of *aleph* used with *lamedh*.

Figure 17: Examples of contextual variants of *aleph*. Images have been rotated 90 degrees counter-clockwise for layout purposes.

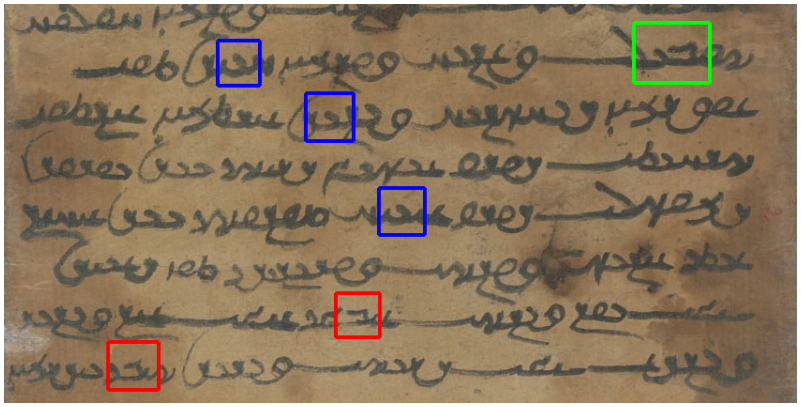


Contrast between initial *beth* and initial *yodh* in a cursive text (from Pelliot chinois 3049). The medial ◌ *beth* has a more angular stroke than the medial ◌ *yodh*.

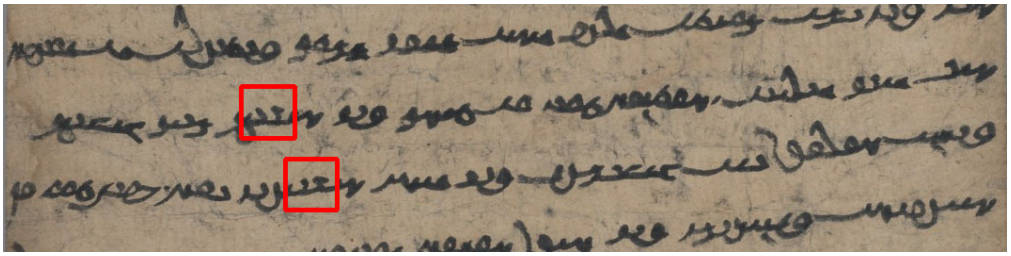


Contrastive representation of initial *beth* and initial *yodh* in the sequence *beth*, *yodh* (red) and *yodh*, *yodh* (blue) in a cursive text (from Pelliot ouïgour 13).

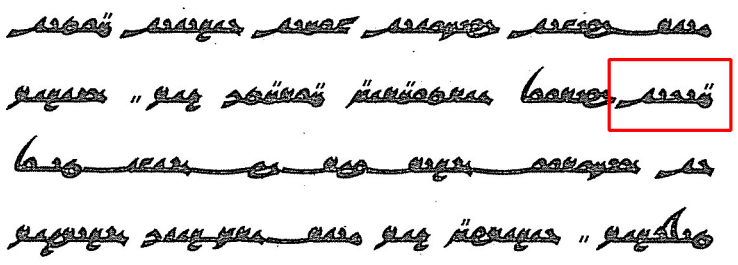
Figure 18: Contrastive representation of *beth* and *yodh* in initial position



Contrast between medial *beth* (red) and medial *yodh* (blue), and a sequence of *beth*, *yodh* (green) (from Pelliot chinois 2998).

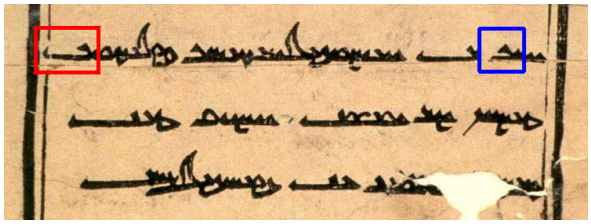


Contrast between medial *beth* and medial *yodh* in sequence in a cursive text (from Pelliot ouïgour 2). The medial *beth* has a notch at the head, while the medial *yodh* is a simple stroke.

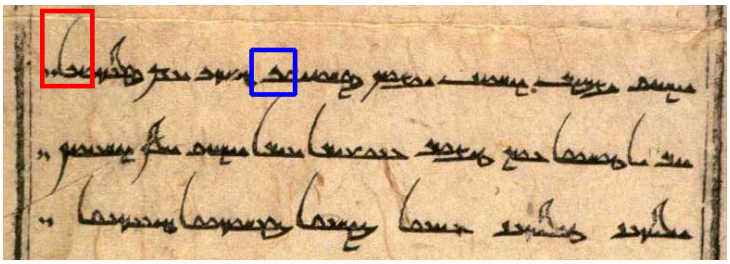


Contrastive representation of a sequence of medial *yodh*, *beth*, *yodh* in a printed facsimilie (from Müller 1908: 42, 43). The medial 𐽆 *beth* has a more angular stroke than the medial 𐽅 *yodh*.

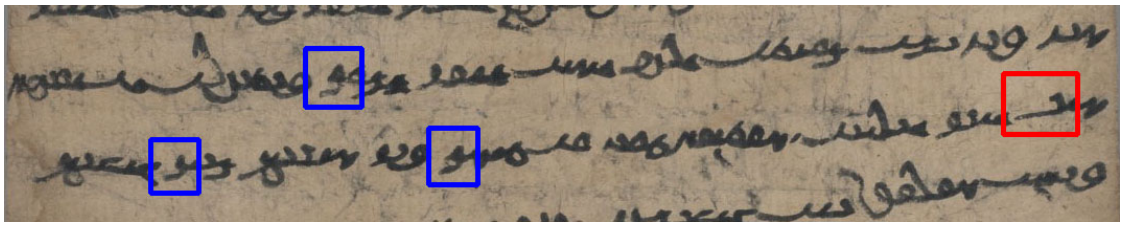
Figure 19: Contrastive representation of *beth* and *yodh* in medial position.



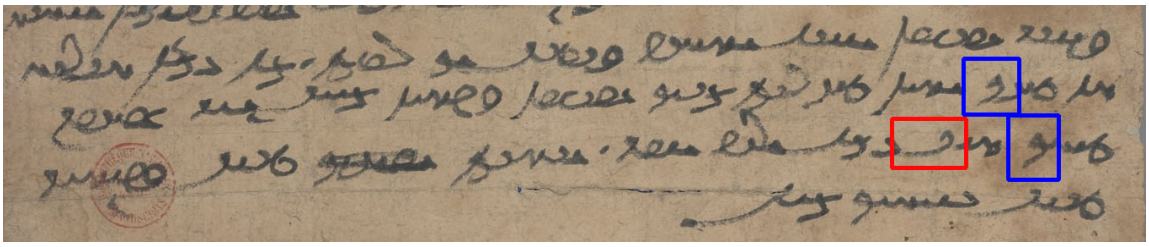
Contrast between final 𐽂 *beth* (red) and final 𐽃 *yodh* (blue) in a block print (from U 4708).



The variant final 𐽁 form of 𐽂 *beth* with a left-ward tail, contrasted with final 𐽃 *yodh* in a block print (U 4704).

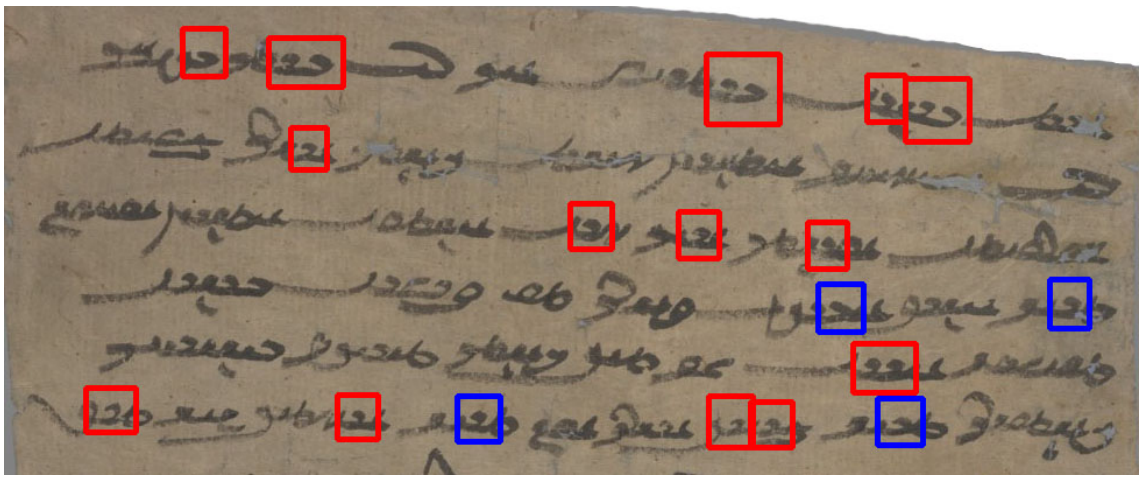


Contrast between final 𐽂 *beth* (red) and final 𐽃 *yodh* (blue) in cursive script (from from Pelliot ouïgour 3).



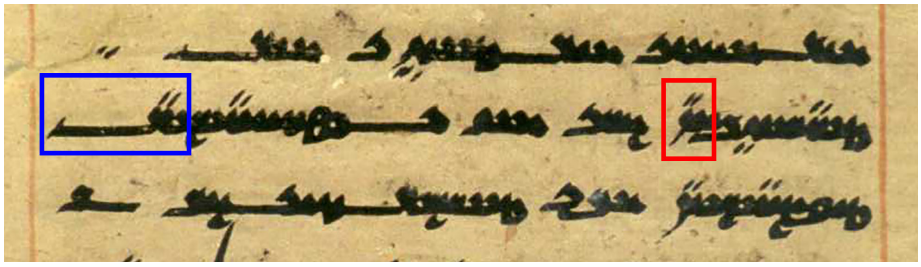
Contrastive representation of final 𐽂 *beth* (red) and final 𐽃 *yodh* (blue) in cursive script (from from Pelliot ouïgour 5). The *beth* is characterized by the length of its terminal, while *yodh* is characterized by both the shape of the body and its short terminal.

Figure 20: Contrastive representation of *beth* and *yodh* in final position.

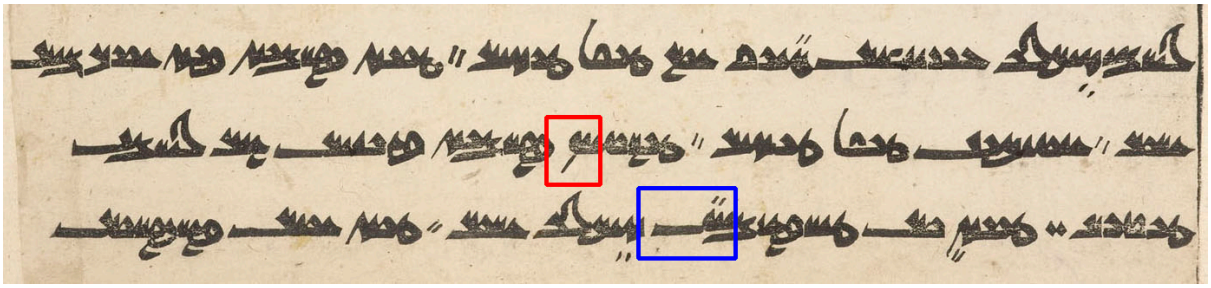


Ambigious representation of *beth* and *yodh* using a sign (from Pelliot ouïgour 3). There is a lack of consistency in distinguishing *beth* (red) from *yodh* (blue).

Figure 21: Ambiguous representation of *beth* and *yodh*. Images have been rotated 90 degrees counter-clockwise for layout purposes.

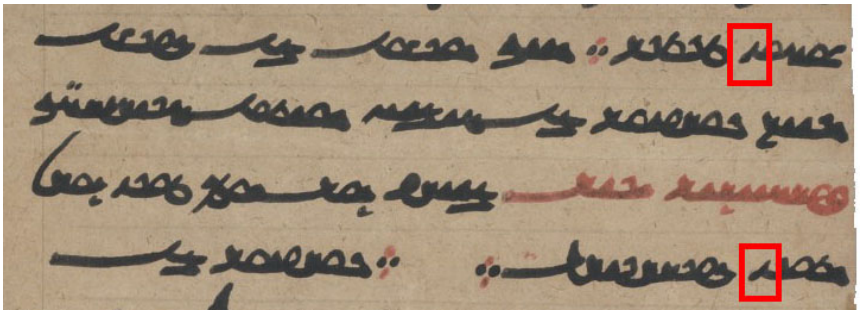


Excerpt from U 924 showing distinctive usage final forms of 𐽂 *gimel* (red) and 𐽆 *heth* (blue).

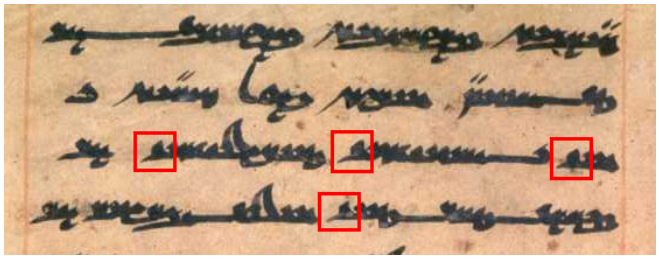


Excerpt from PEALD 6a showing final forms of 𐽂 *gimel* (red) and 𐽆 *heth* (blue).

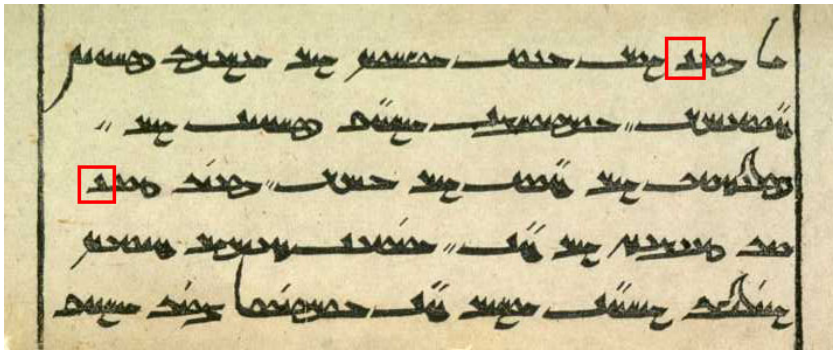
Figure 22: Examples of *gimel* and *heth*. Images have been rotated 90 degrees counter-clockwise for layout purposes.



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

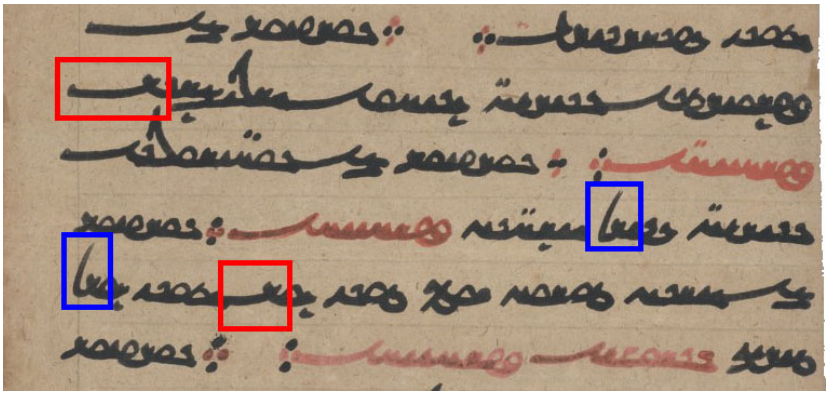


Excerpt from Mainz 119 showing a hand-written form of ـز *zayin*.

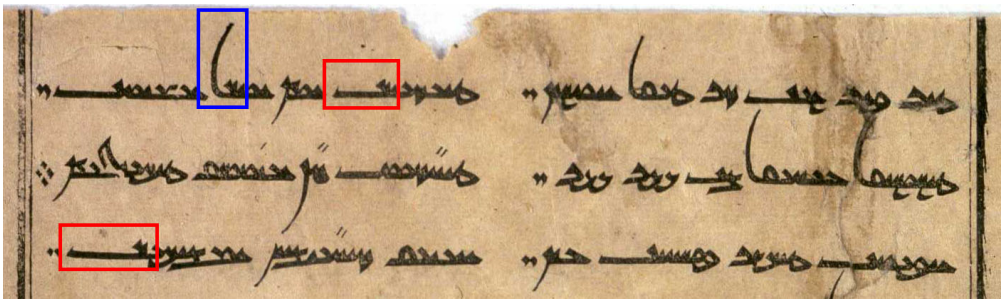


Excerpt from U 387 showing a block-print form of ـز *zayin*.

Figure 23: Examples of *zayin*. Images have been rotated 90 degrees counter-clockwise for layout purposes.

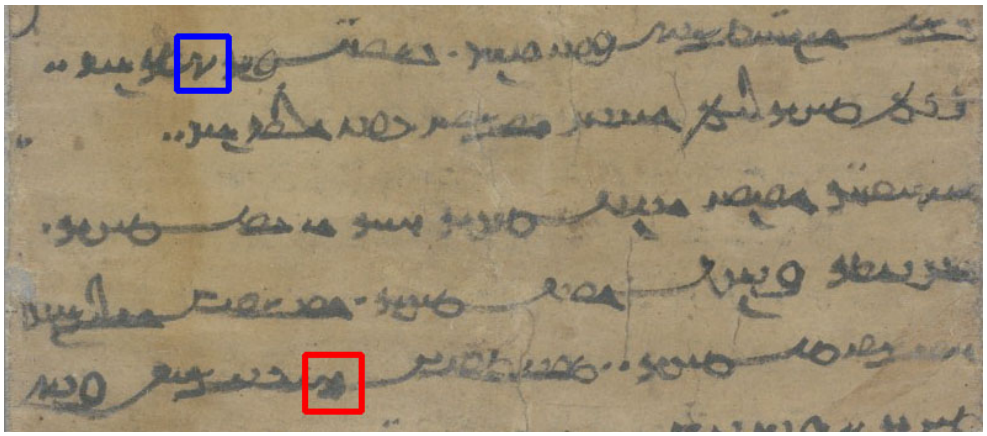


Usage of the regular final form ـك of *kaph* (red) and the alternate final form ـکـ (blue) in a manuscript (excerpt from Pelliot Ouïgour 13)

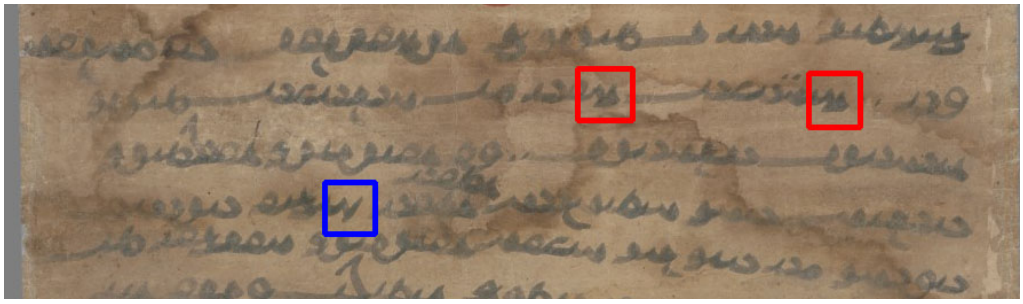


Usage of the regular final form ـك of *kaph* (red) and the alternate final form ـکـ (blue) in a block print (excerpt from U 4301)

Figure 24: Examples of *kaph*. Images have been rotated 90 degrees counter-clockwise for layout purposes.

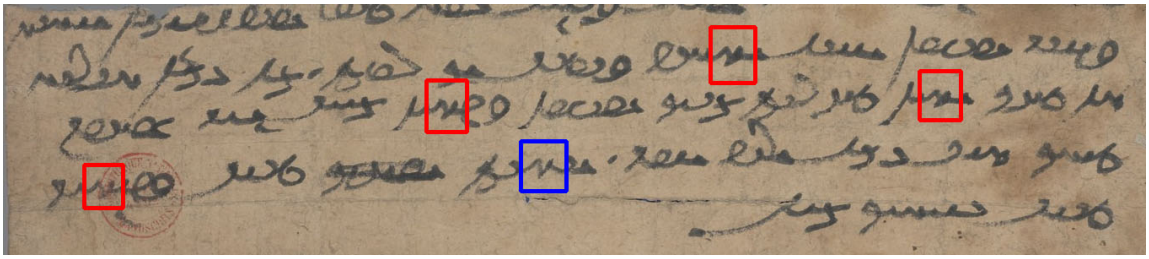


Distinctive forms of initial 𐽆 *samekh* (red) and medial 𐽆 *shin* (blue) from a cursive document (Pelliot chinoise 3049).

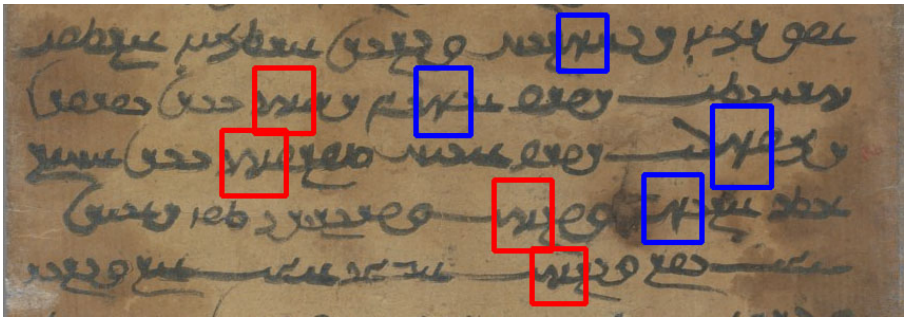


Distinctive forms of initial 𐽆 *samekh* (red) and medial 𐽆 *shin* (blue) from a cursive document (Pelliot chinoise 3072).

Figure 25: Contrastive representations of initial *samekh* and *shin*.



Distinctive forms of medial 𐽆 *samekh* (red) and medial 𐽇 *shin* (blue) from a cursive document (Pelliot ouïgour 5).



Distinctive forms of medial 𐽆 *samekh* (red) and medial 𐽇 *shin* (blue) from a cursive document (Pelliot chinois 2998).

Figure 26: Contrastive representations of *samekh* and *shin*. Images have been rotated 90 degrees counter-clockwise for layout purposes.

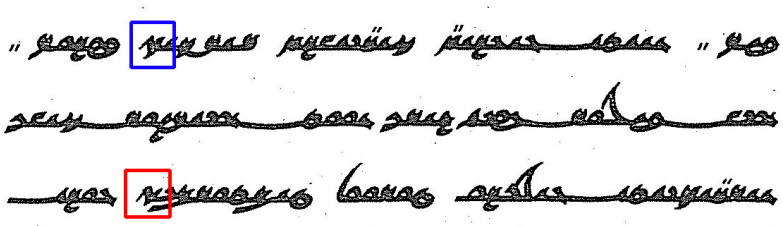
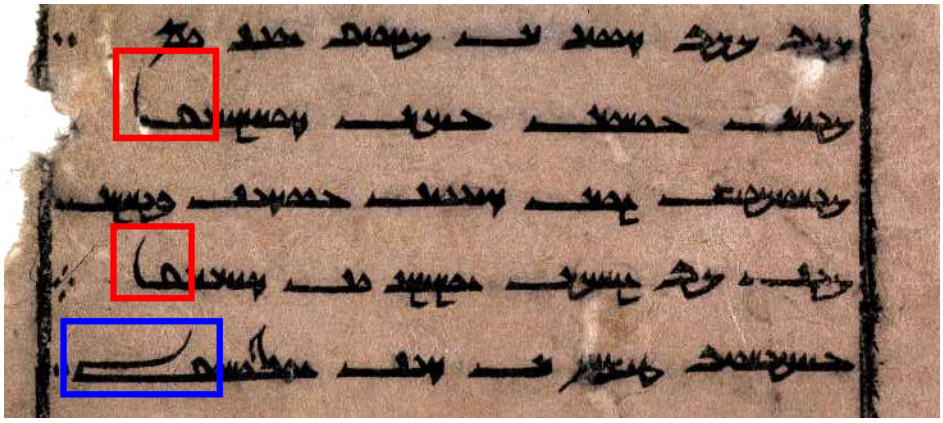
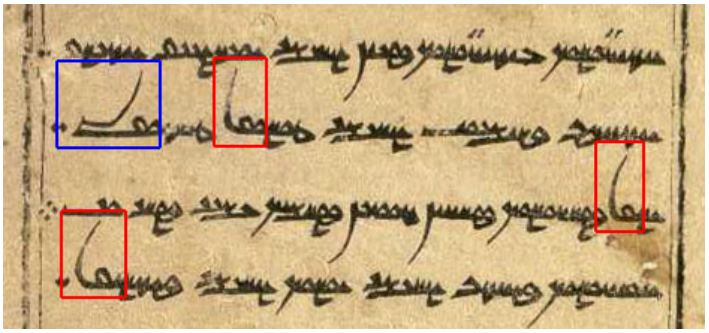


Figure 27: Distinctive final forms of 𐽆 *samekh* (blue) and 𐽈 *shin* (blue) (from Müller 1908: 44).



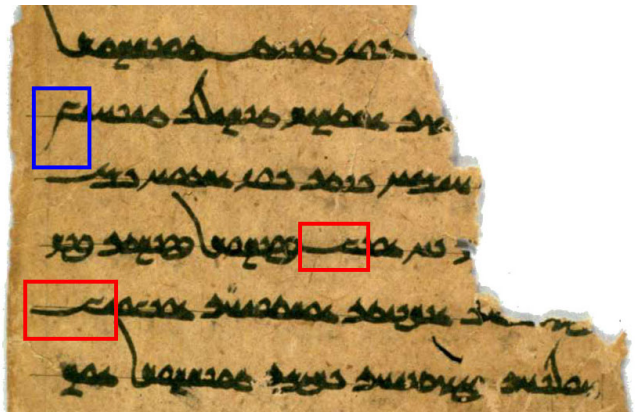
Regular ﻫﺎ (red) and ornamental ﻫﮯ (blue) forms of final *pe* (excerpt from U 4750)



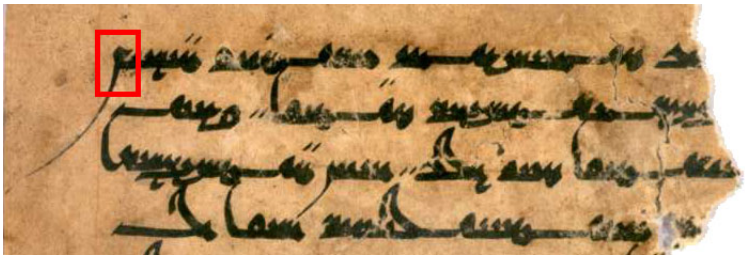
Regular ﻫﺎ (red) and ornamental ﻫﮯ (blue) forms of final *pe* (excerpt from U 4162)

Regular ﻫﺎ (red) and ornamental ﻫﮯ (blue) forms of final *pe* (excerpt from Mainz 34)

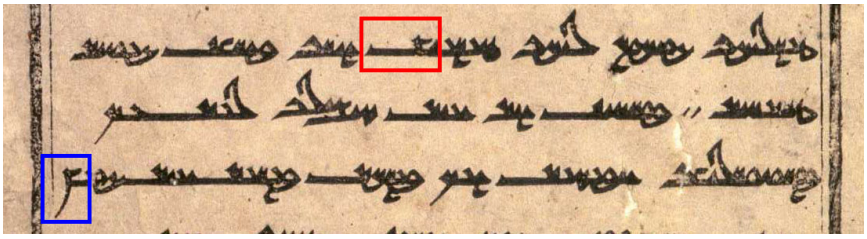
Figure 28: Examples of *pe*. Images have been rotated 90 degrees counter-clockwise for layout purposes.



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

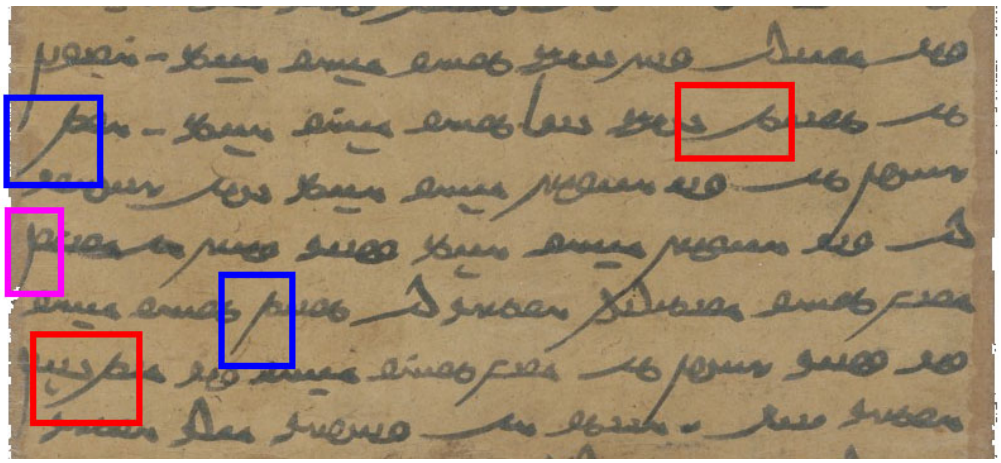


Usage of the alternate 𐽆 *sadhe* in a manuscript (excerpt from Mainz 393)



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

Figure 29: Examples of final *sadhe*. Images have been rotated 90 degrees counter-clockwise for layout purposes.



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

Figure 30: Examples of *taw*. Images have been rotated 90 degrees counter-clockwise for layout purposes.

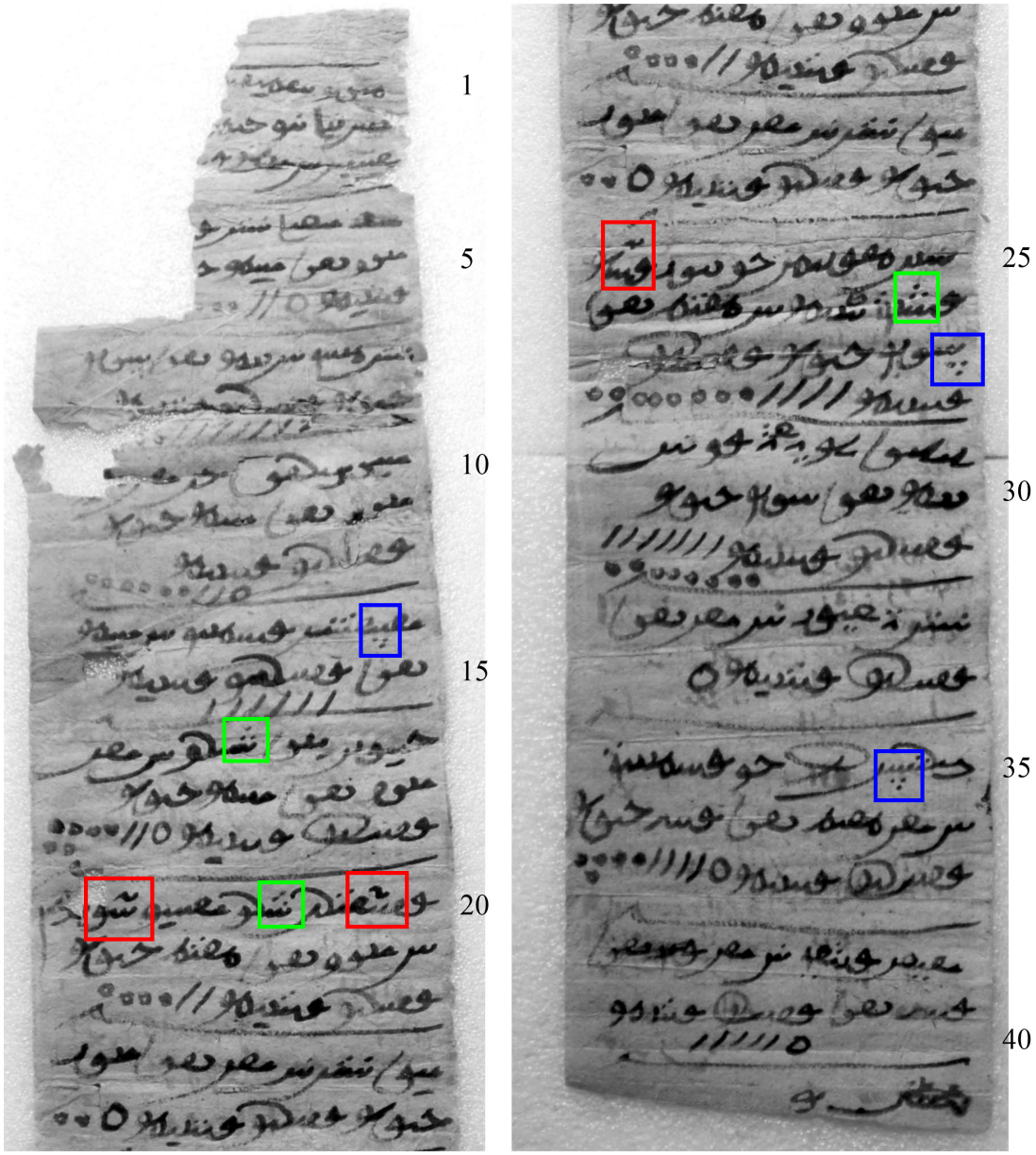
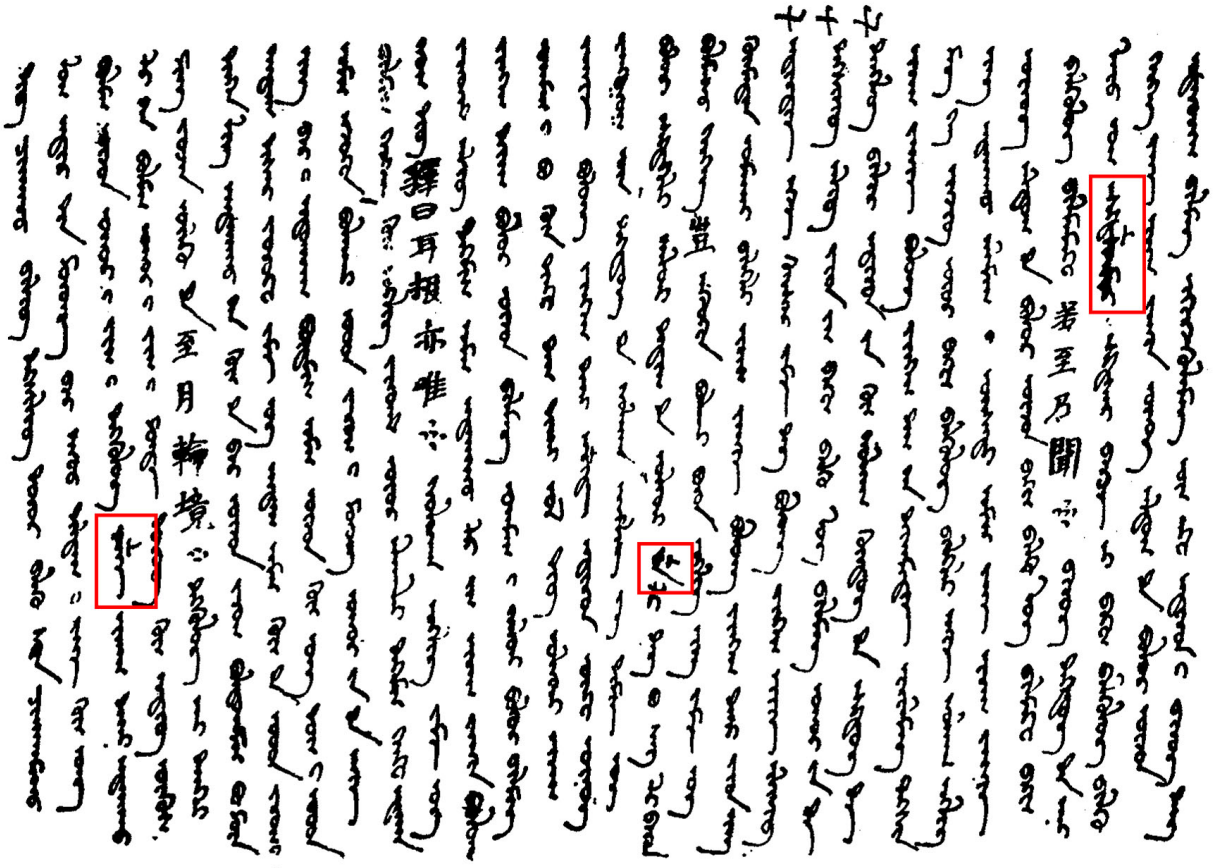
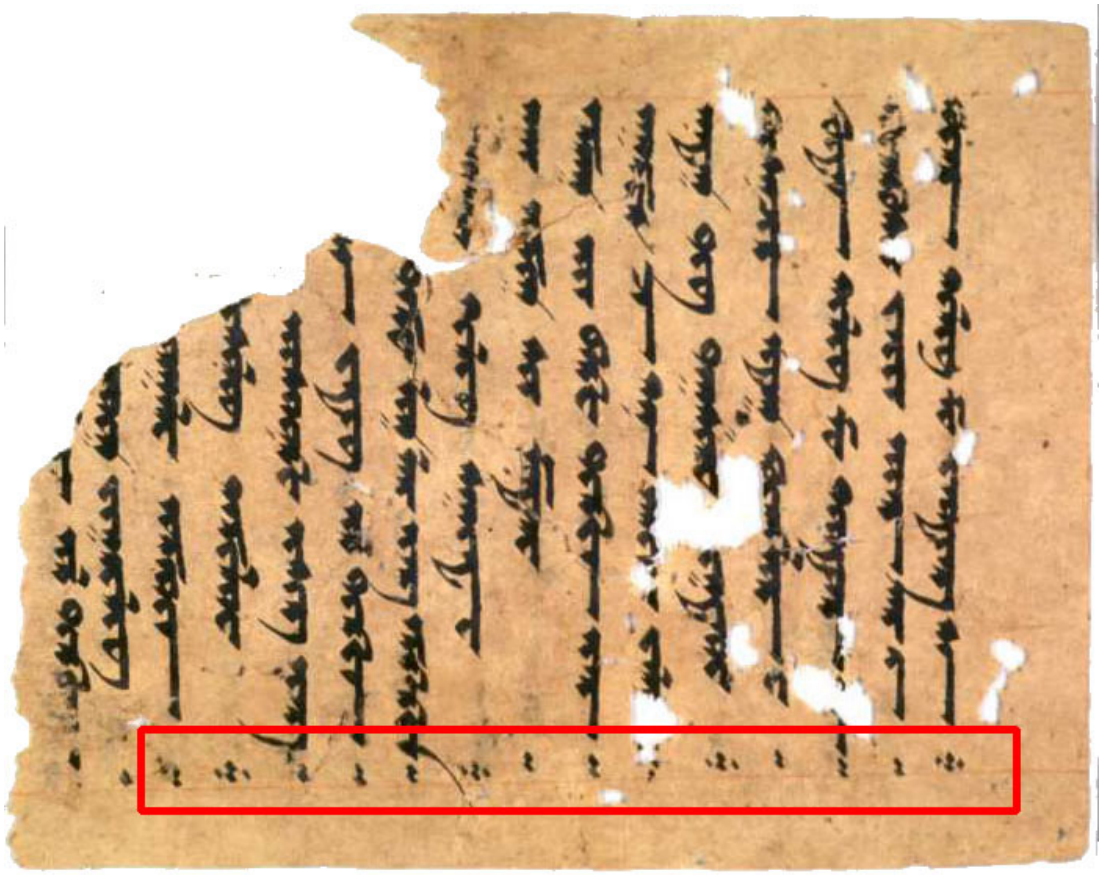


Figure 31: Usage of ◌ (blue), ◌ (green), and ◌ (red) for transcribing Arabic in a Old Uyghur administrative document (from Israpil 2014: plate I).



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

Figure 32: Usage of the ◌ꞈ 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 33: Examples of punctuation signs

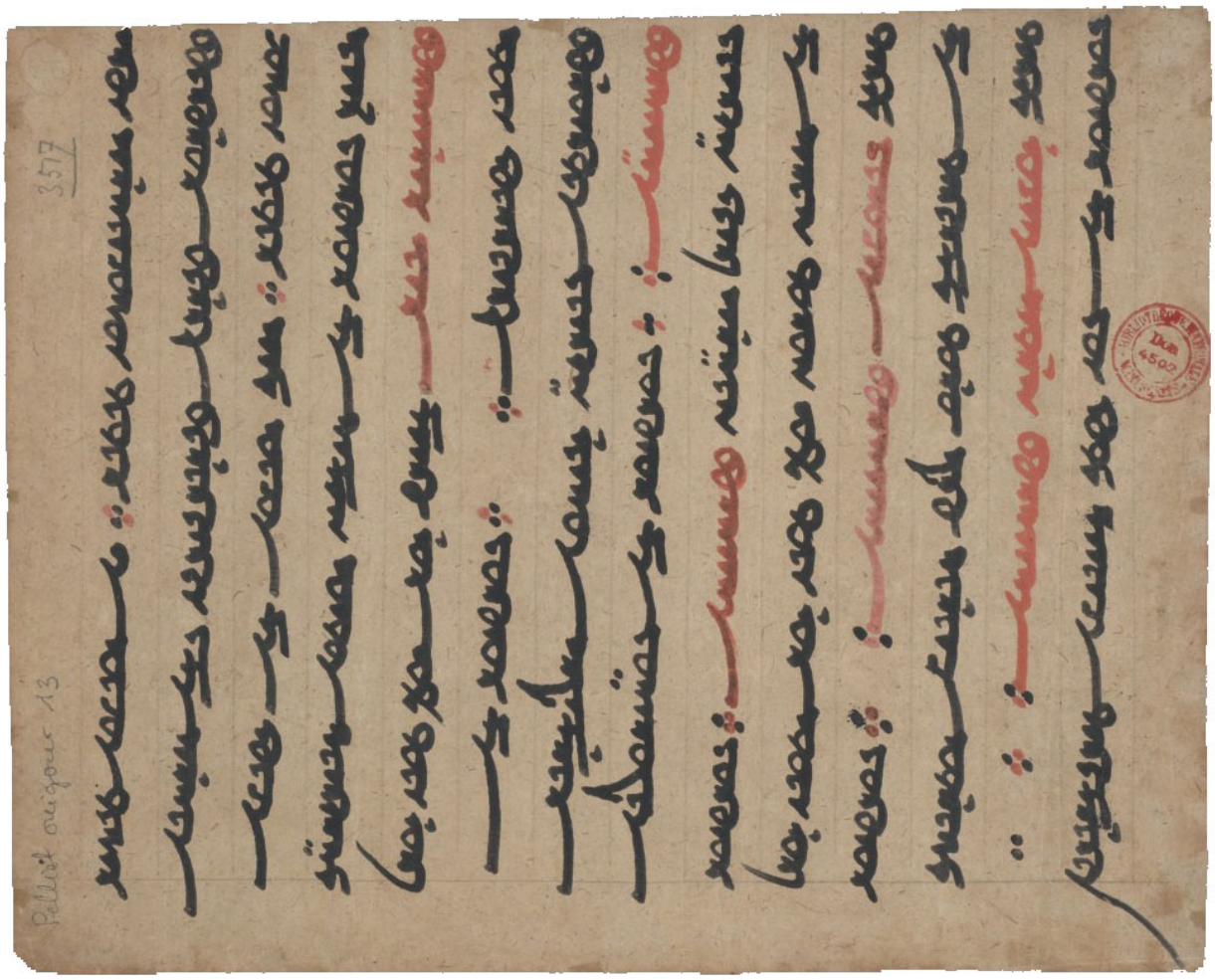


Figure 34: Pelliot Ouïgour 13.

Figure 35: Mainz 126.

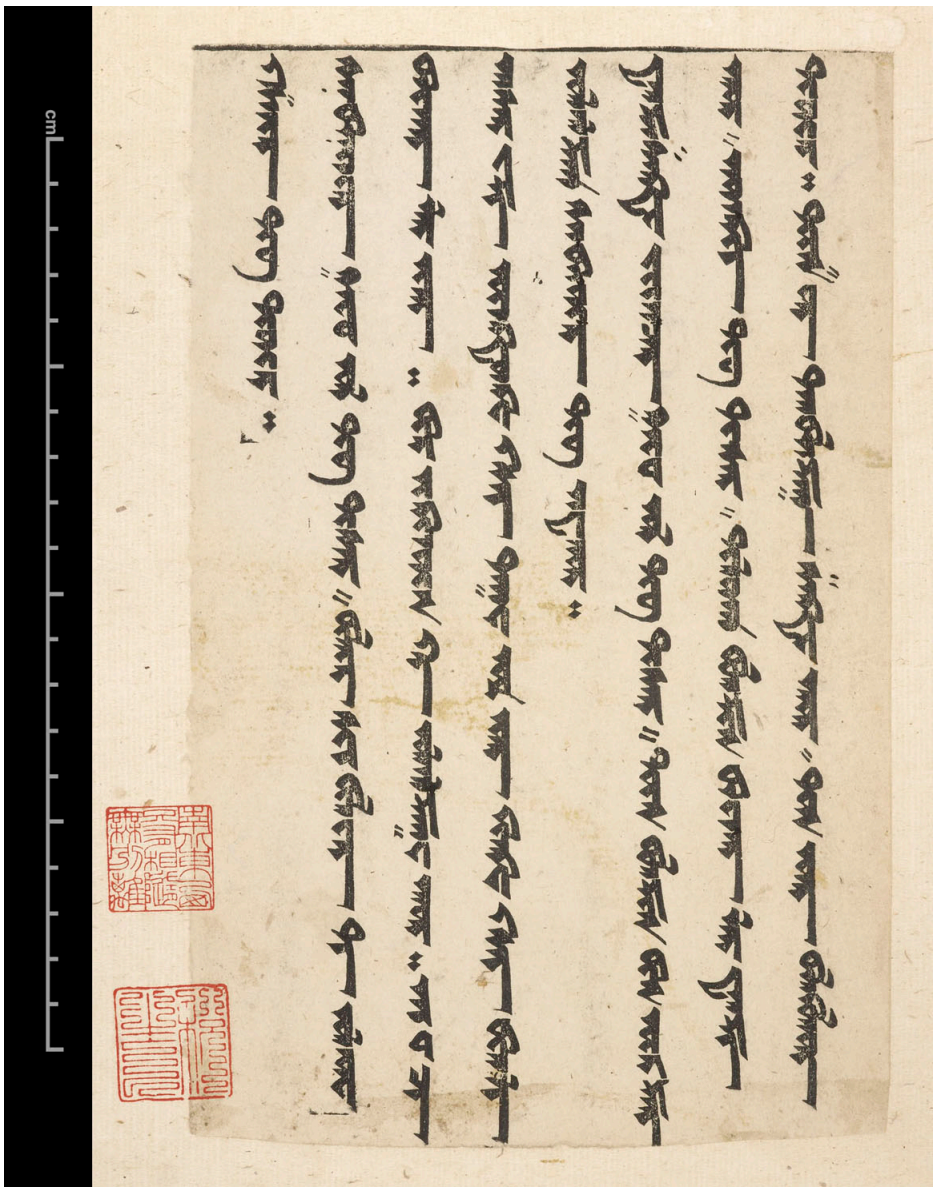


Figure 36: Princeton East Asian Library, PEALD 6a, recto. Block print.

Figure 37: BBAW, U 387 & U 388, recto. Block print.

Figure 38: BBAW, U 4960, folio 1, recto. Block print. Seal in Han characters.

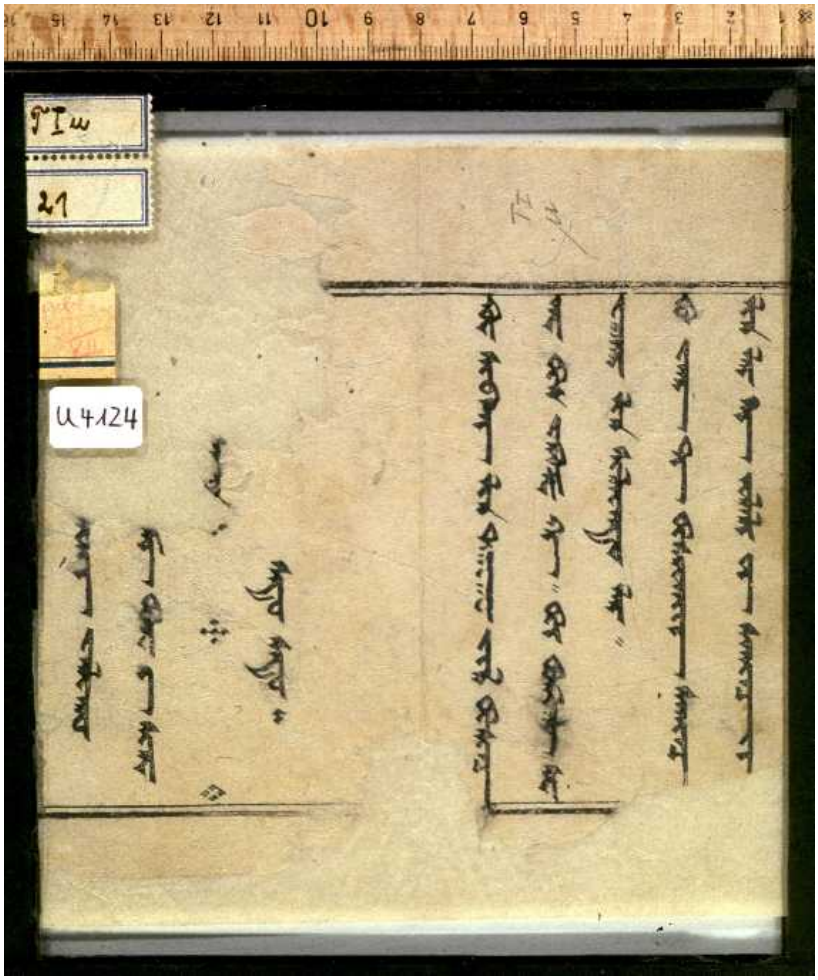


Figure 39: BBAW, U 4124. Block print.

Figure 40: BBAW, U 343, folio 1, recto. Block print.

Figure 41: 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.

Proben der nicht zum *Altun y(a)ruq̈* gehörigen Texte[1].

T III 84, 13.

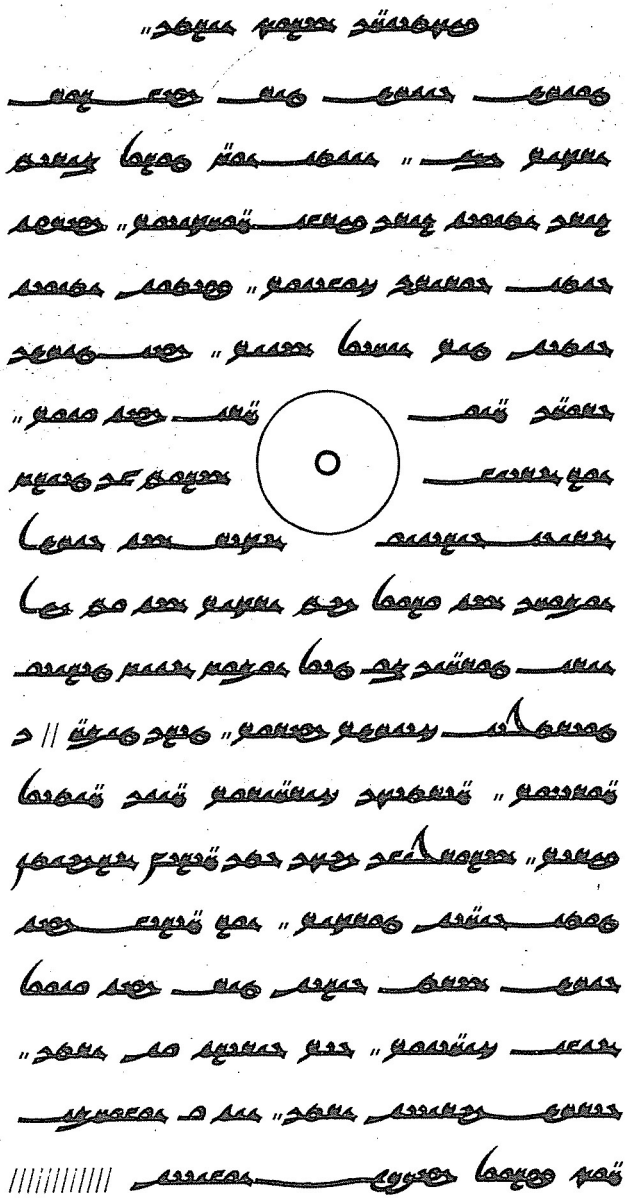


Figure 42: Excerpt from a printed edition of *Altun Yaruq* in the Old Uyghur script (from Müller 1908: 36).

Figure 43: Excerpt from a printed edition of *Ṭišastvustik* in the Old Uyghur script (from Radloff 1910: 3).

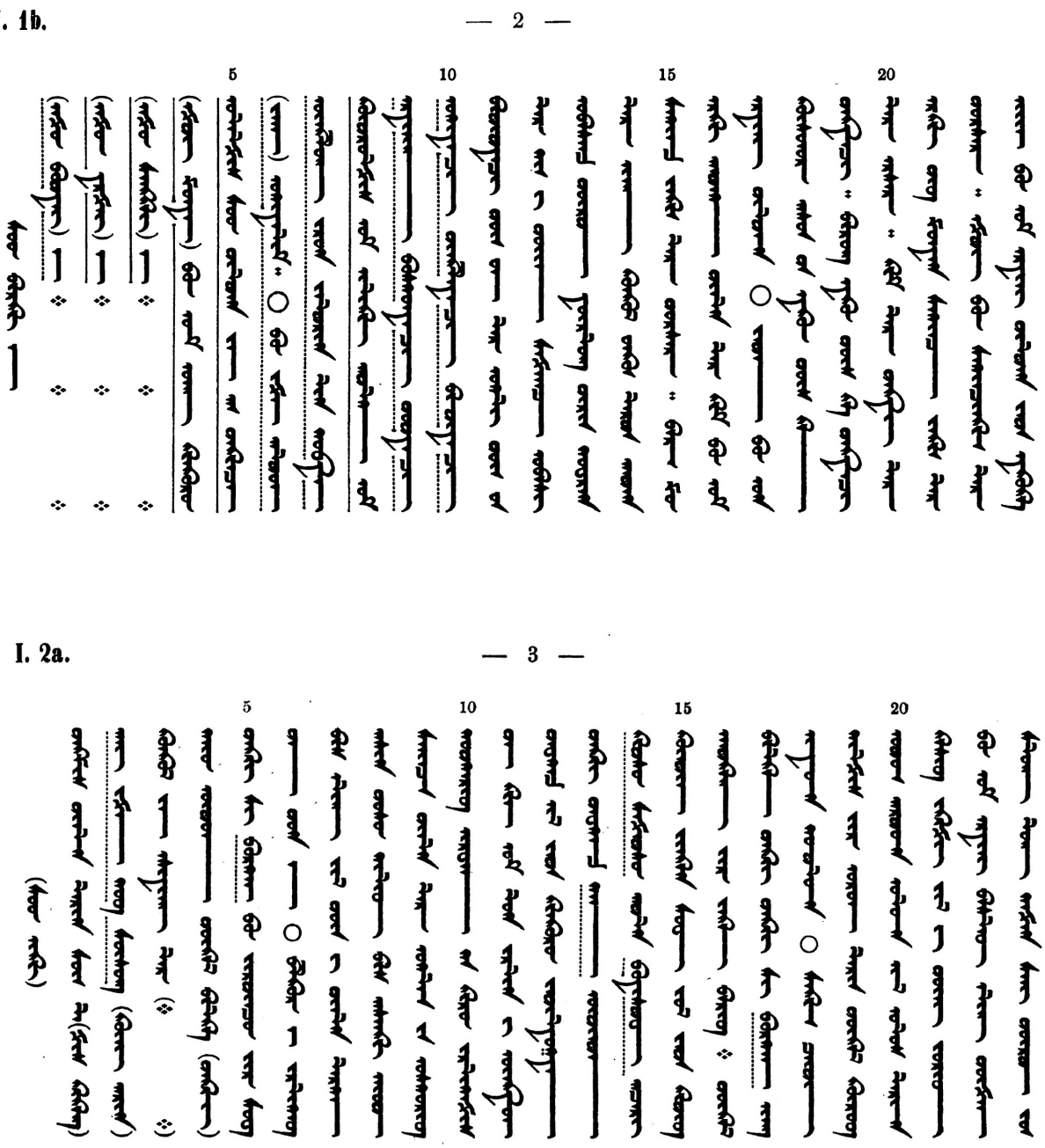
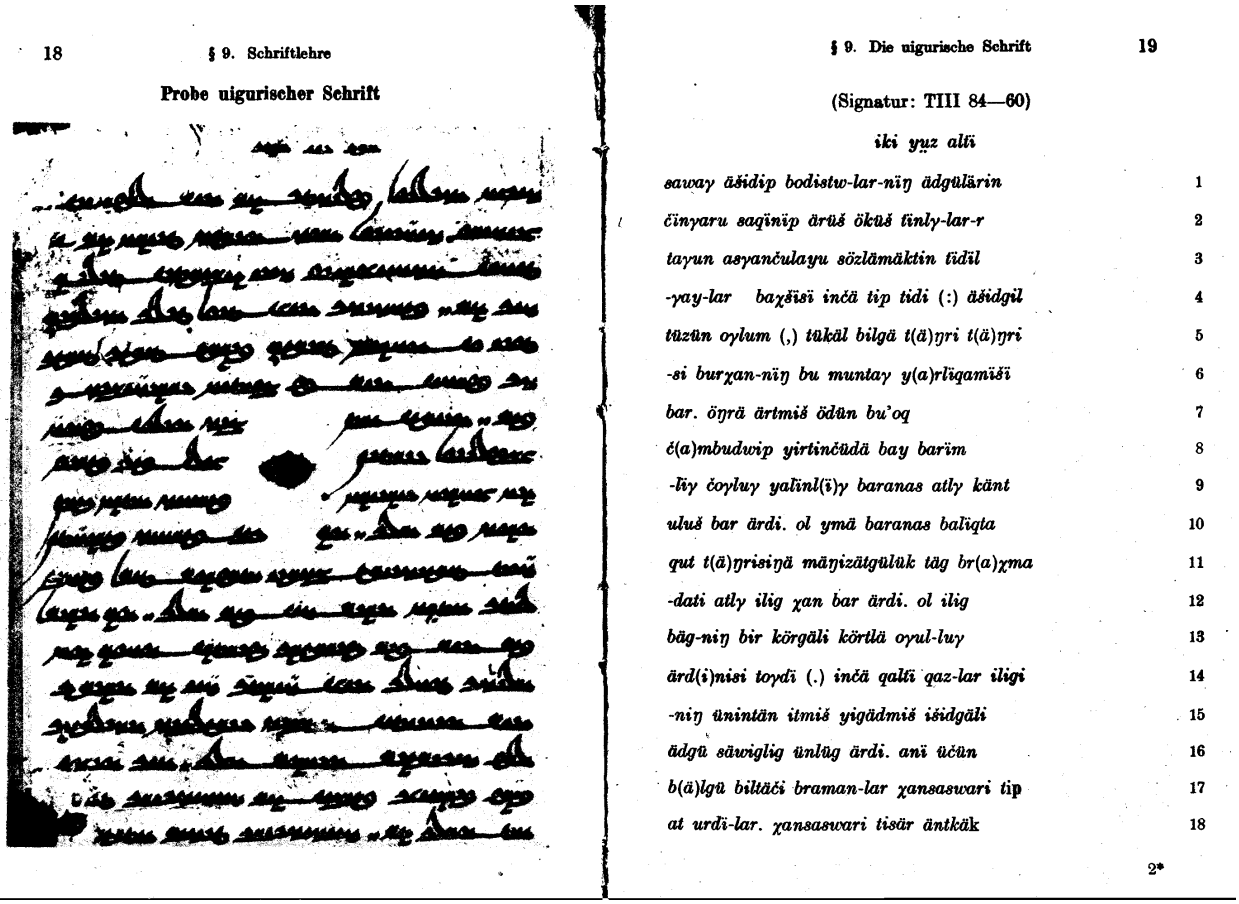


Figure 44: Printed edition of *Suvarṇaprabhāsa*, a Mahayana Buddhist text, in the Old Uyghur script (from Radlov and Malov 1913: 2–3).



18 § 9. Schriftlehre

Probe uigurischer Schrift

§ 9. Die uigurische Schrift 19

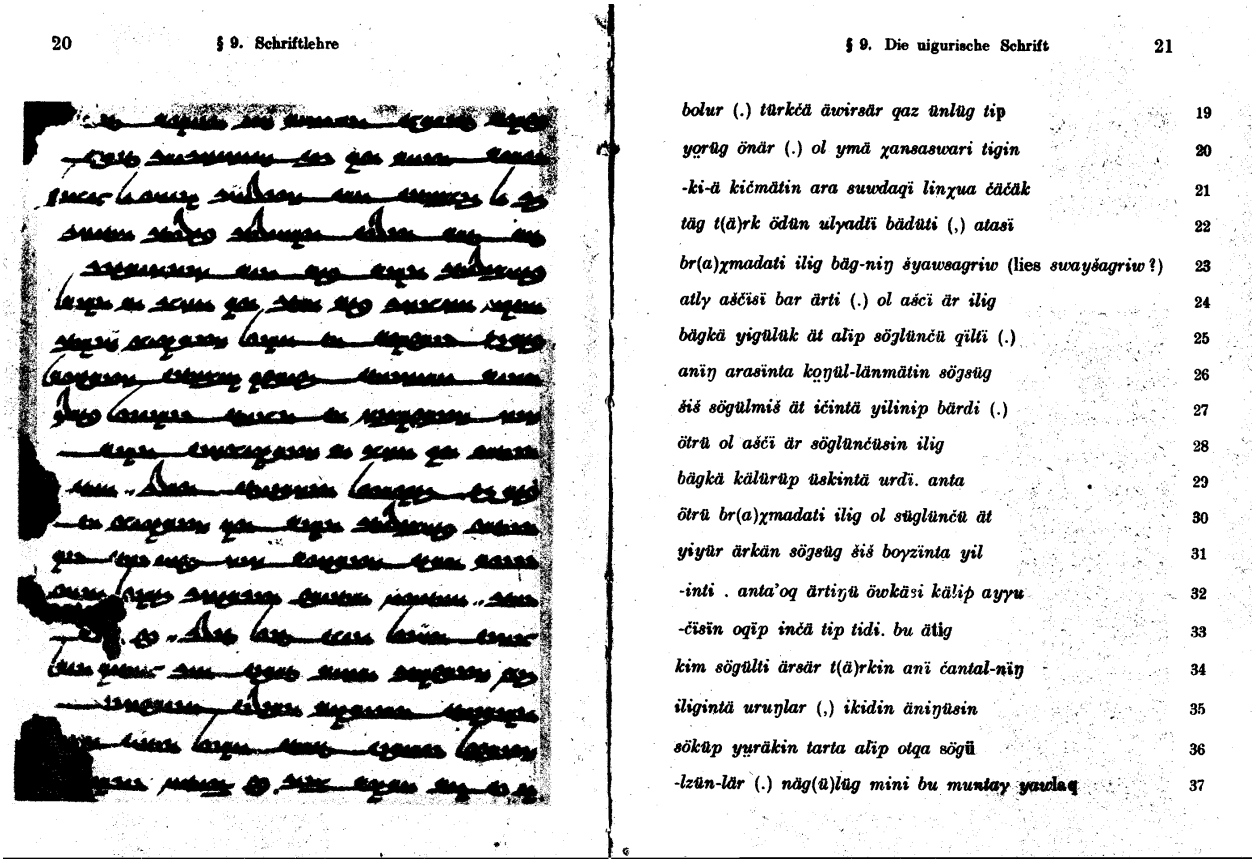
(Signatur: TIII 84—60)

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 45: Transcription of an Old Uyghur manuscript (from von Gabain 1950: 18–19). Continued in fig. 46.

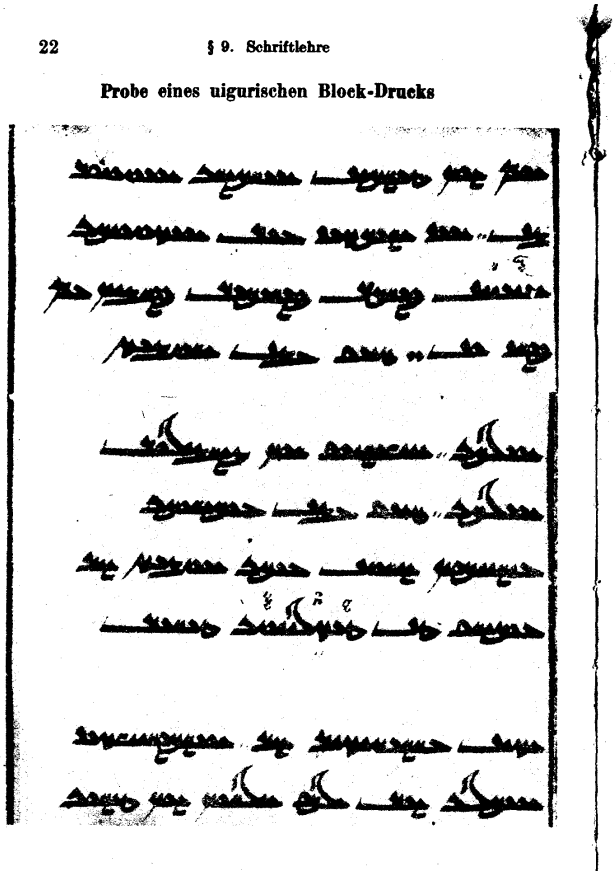


20 § 9. Schriftlehre

§ 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 46: Transcription of an Old Uyghur manuscript in a grammar of Old Turkic (from von Gabain 1950: 20–21). Continued from fig. 45.



22 § 9. Schriftlehre

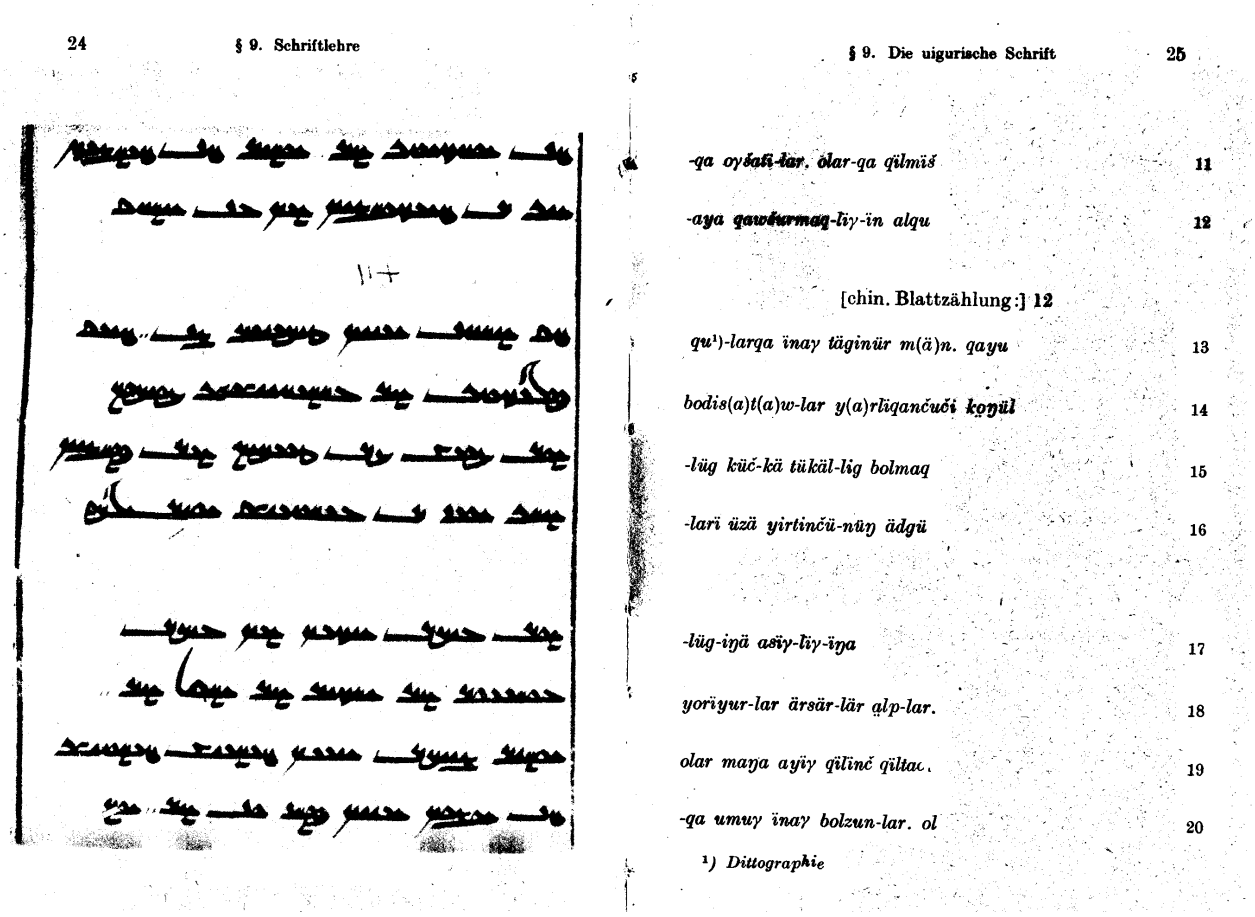
Probe eines uigurischen Block-Drucks

§ 9. Die uigurische Schrift 23

(Signatur: TIIIM 195)

nom-luγ tilgän äwirgäli ötünür	1
m(ä)n. üzäliksiz yig üstünki	2
inyana [darüber in *brāhmī*: *jñāna*] *bilgä biligig bulmaqïm*	3
bolzun. qayu ymä ärtmiš	4
ödki. ančulayu oq kälmädük	5
ödki. qayu ymä yuγärüki	6
yalaŋuq-larta yigi utmiš-lar	7
yuγärü-tä tiṯdanti [darüber in *brāhmī*: *dhiṣṭhanti*] *tirig*	8
äsän y(a)rlïqasar-lar. ülgüläńčsiz	9
ögdi-lig ädgü adruq-luγ taluγ	10

Figure 47: Transcription of an Old Uyghur manuscript in a grammar of Old Turkic (from von Gabain 1950: 22–23). Continued in fig. 48.



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

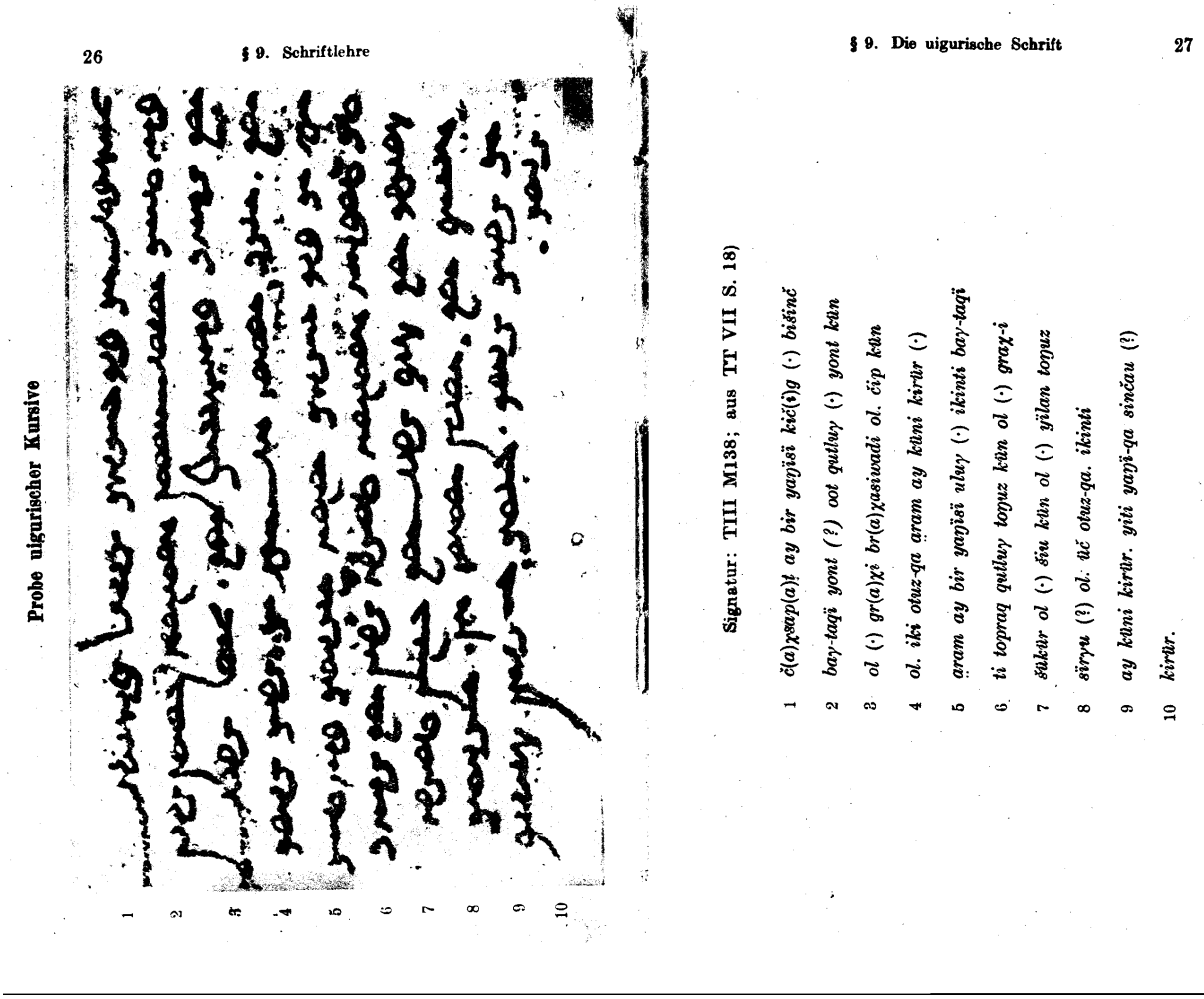
yoriyur-lar ärsär-lär alp-lar. 18

olar maŋa ayïγ qïlïnč qïltac. 19

-qa umuγ ïnaγ bolzun-lar. ol 20

[1]) *Dittographie*

Figure 48: Transcription of an Old Uyghur manuscript in a grammar of Old Turkic (from von Gabain 1950: 24–25). Continued from fig. 47.



26 § 9. Schriftlehre

Probe uigurischer Kursive

§ 9. Die uigurische Schrift 27

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

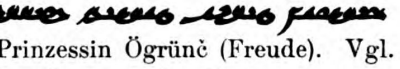
1 *č(a)ɣap(a)ɣ ay bir yapïsï kič(i)g (·) bišinč*
2 *bay-taqï yont (?) oot qulwɣ (·) yont kün*
3 *ol (·) gr(a)χï br(a)χasiwadi ol. čïp kün*
4 *ol. iki otuz-qa aram ay künü kïrtïr (·)*
5 *aram ay bir yapïsï ulwɣ (·) ikinti bay-taqï*
6 *ti topraq qulwɣ toɣuz kün ol (·) graχ-i*
7 *šükür ol (·) šu kün ol (·) ülen toɣuz*
8 *šïryu (?) ol. üč otuz-qa. ikinti*
9 *ay künü kïrtïr. yiti yapï-qa inčau (?)*
10 *kïrtïr.*

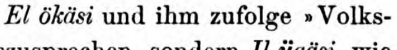
Figure 49: Transcription of an Old Uyghur manuscript in a grammar of Old Turkic (from von Gabain 1950: 26–27).

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 50: 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.

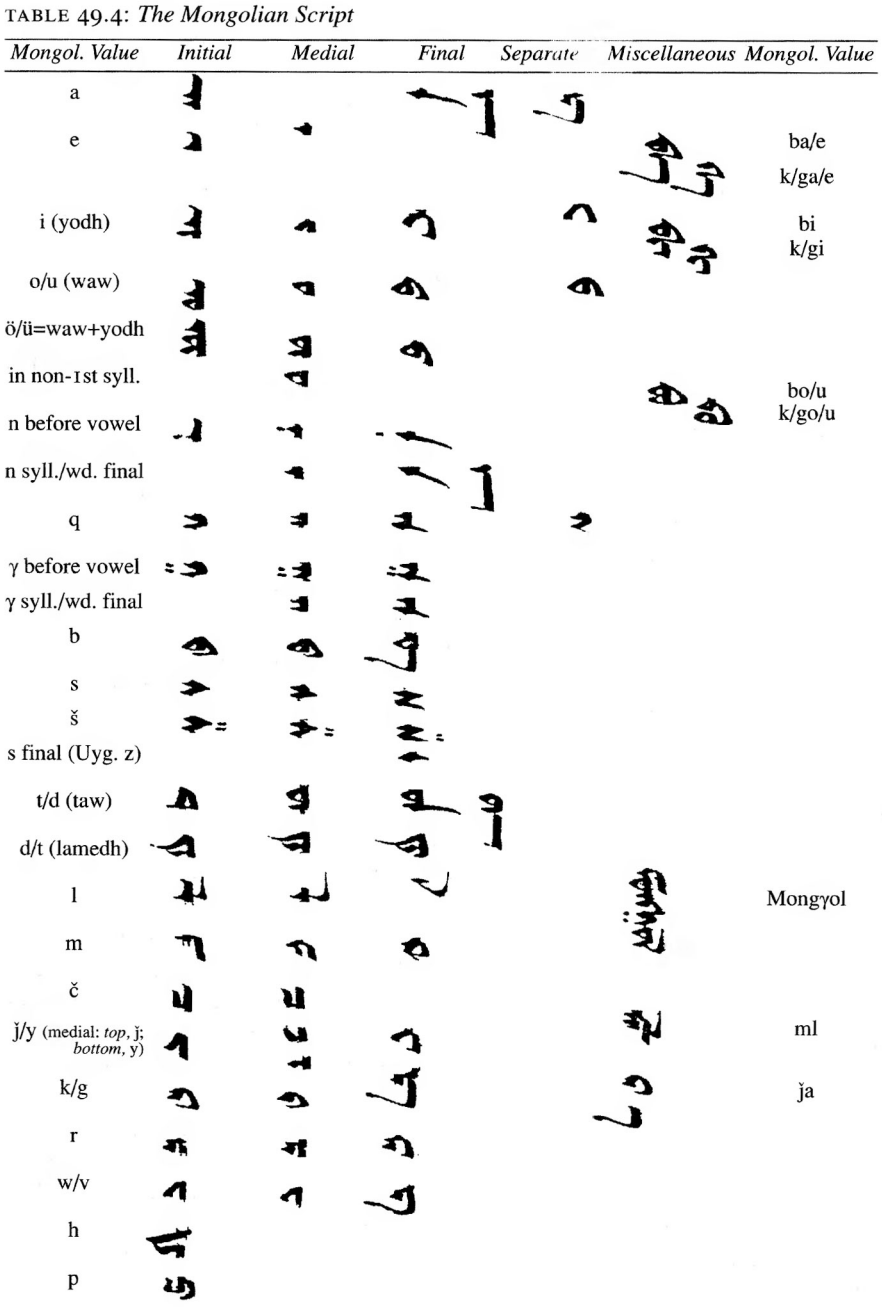
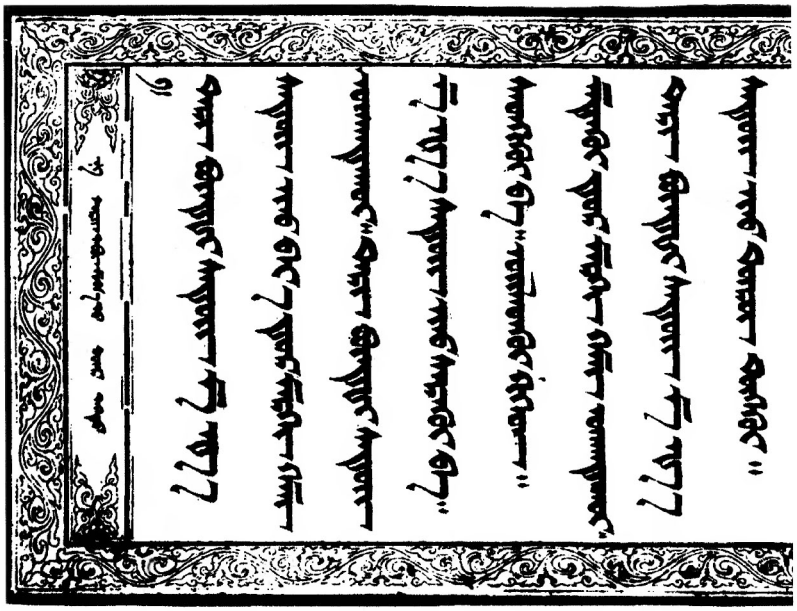


Figure 51: Table showing letters of the Mongolian script (from Kara 1996: 545). See table of Old Uyghur letters from the same source in fig. 6.

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 52: Sample Mongolian text (from Kara 1996: 546). Compare the Mongolian block print with the Old Uyghur block print in fig. 36.

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: *2020-01-15*
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: *37*
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? *Dr. Dai Matsui <dmatsui@let.osaka-u.ac.jp>*
   Dr. Mehmet Ölmez <olmez.mehmet@gmail.com>
   Dr. Yukiyo Kasai <yukiyo.kasai@ruhr-uni-bochum.de>
   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: