L2/20-003
2020-01-15

Revised proposal to encode Old Uyghur in Unicode

Anshuman Pandey

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

January 15, 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”

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

- Revision of the proposed repertoire on the basis of printed specimens (§ 5)
- Additional details of the script repertoire and its evolution (§ 4.3)
- Revision of the encoding model for *aleph* and *nun* (§ 6.1.1)
- Redefinition of *zayin* from a dual-joining to right-joining letter (§ 6.1.4)
- Notes on representing ambiguous readings of pairs of letters (§ 6.1.12)
- Addition of a character used for producing an ornamental terminal (§ 6.5)
- Tables showing comparisons of letterforms from various sources (tables 1–3)

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 '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, Sanskrit, Sogdian, and Tibetan. It was the basis of vibrant scribal and block-print cultures across Central Asia. Four main styles of the script are observed in the attested records: square, semi-square, semi-cursive, and cursive. The 'square' style is the formal style used for religious and literary manuscripts. The 'cursive' style occurs in numerous civil and administrative documents from the 12th through 15th centuries. The script was developed further through the usage of block printing. This advancement established a style that may be considered a 'print standard', although one that reflects the phase of the script used in the 14th century. 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. 47).

The writing system is situated in the middle of a script continuum that originates from the Sogdian script of the 'Ancient Letters' and terminates at modern Mongolian. The Uyghur script developed from the 'cursive' style of the Sogdian script during the 8th–9th century (Kara 1996: 539). Just as speakers of Turkic languages adopted the Sogdian script, speakers of other languages in Central Asia turned to the Uyghur script to develop new orthographies. A popular narrative states that in the 13th century, the scholar and former chancellor of the Naiman Khanate known as Tata Tonga 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 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.

At the beginning of the 20th century, 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. 48–50). 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.

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

2 Script identifier

The name 'Uyghur' (/ʊjɣʊr/) has variant transliterations, transcriptions, and spellings in English, such as 'Uighur', 'Uigur', 'Uygur', 'Uyğur', as well as 'Ouïgour' in French, 'Uigurisch' in German, etc. The term is used for referring to the script discussed here. However, as the modern Uyghur language is written using an orthography based upon the Arabic script, the phrase 'Uyghur script' may refer to both the historical Sogdian-based script and the Arabic-based script. In order to differentiate between the two, the descriptor 'Old Uyghur' is often used as a matter of convenience 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. For purposes of identifying the script in Unicode, the adjective 'Old' is appended to 'Uyghur' in order to distinguish the script from the later Arabic orthography used for writing the modern Uyghur language, which is not directly related to the Uyghur language of the 8th century. Given the polysemia of 'Uyghur', the term 'Old Uyghur' has become common for referring to the script, even if it is imprecise.

The proposed Unicode identifier for the script is 'Old Uyghur', which is a scholarly designation. The name applies specifically to the script within the context of Unicode, and it does not apply to any language, culture, or community. The script is also known generically as 'Uyghur', without the descriptor 'Old'.

3 Encoding history

3.1 Justification for encoding

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

- The repertoire, order, and names of Uyghur letters is based upon that of Sogdian. The proposed encoding for Old Uyghur retains these attributes. The Mongolian encoding uses different names and ordering for letters, which reflect Mongolian preferences and pronunciations. Mongolian letter names do not correspond to Uyghur values.

- Following from the above, a separate encoding preserves the glyphic distinctions of formal Uyghur in multilingual contexts that include Sogdian and Mongolian text. In particular, Mongolian glyphs do not adequately transmit the aesthetic and orthographic features of Uyghur letters.

- The proposed encoding for Uyghur is based upon a palaeographic and graphetic model. The Unicode encoding for Mongolian is based upon a phonetic model, which presents several issues and is unsuitable for the Uyghur script. The proposed model for Uyghur offers a practical implementation for a vertical script that avoids the complications of the Mongolian model.

3.2 Previously proposed encodings for Unicode

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, with features of an *abjad*, and is characterized by its normatively vertical orientation. Its historical repertoire consists of 18 letters, which are derived from Sogdian, and ultimately from Imperial Aramaic. The letters represent consonantal sounds, while three are used for expressing vowels, following the Semitic convention inherited from Sogdian: *aleph*, *waw*, *yodh*. The rich vocalic repertoire of Turkic languages is represented using combinations of these letters in diagraphs and trigraphs (see § 8.1).

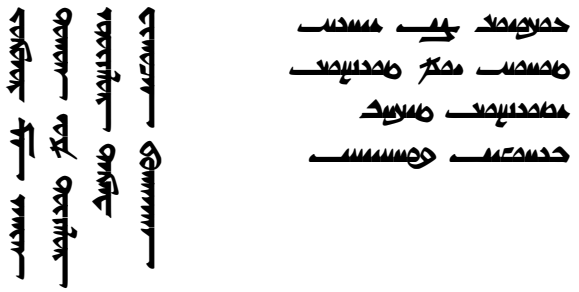
Diacritics are used for diambiguating letters with similar appearances and for representing sounds for which distinctive letters do not exist (see § 8.2).

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

4.2 Directionality

The conventional direction of writing for Old Uyghur is vertical, from top to bottom in columns that run from left to right. The vertical orientation is confirmed by biscriptal documents containing Han characters and Central Asian Brahmi. In some Iranian documents from the 14th century, the script is written horizontally. This may be influenced the Arabic script. When Old Uyghur texts were begun to be printed in the 20th century, publishers maintained fidelity to the standard vertical orientation (see fig. 49, 50). 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. 56). 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.

4.3 Repertoire

The traditional Old Uyghur alphabet consists of 18 letters. The repertoire appears in the margin of the manuscript U 40 (see fig. 1), which contains a Manichaean text and is dated to the 9th century:

The inventory contains 21 characters, to be read from left to right. The first 17 are the basic letters of the script: *aleph*, *beth*, *gimel*, *waw*, *zayin*, *heth*, *yodh*, *kaph*, *lamedh*, *mem*, *nun*, *samekh*, *pe*, *sadhe*, *resh*, *shin*, *taw*. The names and order follow the scholarly convention based upon Aramaic names; however in this inventory, the glyphs for *samekh* and *shin* are swapped. 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*; however, he does not offer an explanation for the presence of #18 and #19 in this list.[1]

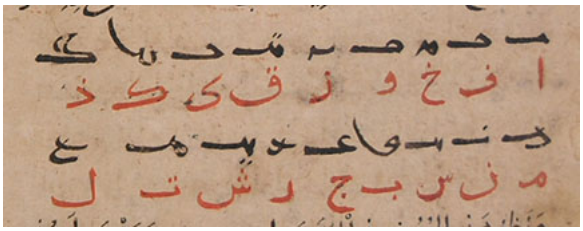
The inventory is important in that it provides:

- attestation for the full repertoire and order of the alphabet.
- evidence for the independent forms of letters, as attested by the inclusion of final *mem* (#19) as well as its independent form.
- distinctive shapes for *aleph* (#1) and *nun* (#11), in which the former is written with an initial horizontal stroke before the triangle, while the latter begins with a rounded stroke. These letters are also distinguished from *zayin* (#5) on account of their terminals.
- distinctive shapes for *beth* (#2) and *yodh* (#7), which are distinguished both by the curvature of their bodies and the length of their terminals.
- distinctive shapes for *gimel* (#3) and *heth* (#6), which are identified by their terminals.
- distinctive shapes for *samekh* (#12) and *shin* (#16), that being the presence of the elongated down- and rightward initial stroke in the latter, from which the second stroke merges at the midpoint.
- evidence for the usage of diacritics to expand the alphabet. The two-dotted *heth* is a common character used for representing /x/ or /q/.

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

The attestation of the complete repertoire in U 40 is also significant for palaeographical reasons. After the 9th century, writing practices led to the merger of some letters, resulting in an abridgement of the script by the 14th century. Therefore, U 40 provides insight into the original shapes and arrangement of the letters.

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*, *heth*, *yodh*, *kaph*, *lamedh*, *mem*, *nun*, *samekh*, *pe*, *sadhe*, *resh*, *shin*, *taw*, 'hooked r'. The inventory 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 with Arabic text during this period.

The following description of the changes to the Old Uyghur repertoire is based upon Clauson (1969: 109–110) and details provided by Dai Matsui (personal communication, August 2018–January 2019):

9th century

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

11th century

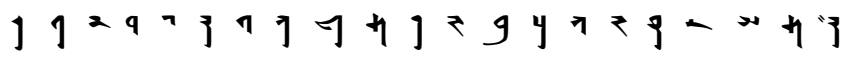
- *samekh* and *shin* become indistinguishable and represented using *shin*
- when necessary, two dots beneath *samekh* for representing *shin*
- *aleph* and *nun* become difficult to distinguish
- final *nun* indistinguishable from *zayin* without a dot over the former

14th century

- only *kaph*, *lamedh*, *mem*, *pe*, 'hooked' *resh* remain distinctive
- *beth* and *yodh* begin to merge and become indistinguishable
- in some instances *sadhe* is indistinguishable from *beth* / *yodh*
- *gimel*/*heth* is indistinct from consecutive *aleph*/*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*
- *resh* begins to become indistinguishable from consecutive *aleph* and/or *nun*;

Identifiying a complete repertoire of 'the' Old Uyghur script in order to develop a Unicode encoding that may be used for representing all attested texts requires an understanding of the periodization of the script and its development. Given the ambiguities of letterforms in the script by the 14th century, the repertoire and letterforms for the Unicode repertoire for Old Uyghur should be based upon the earliest sources in order to enable the complete representation of texts in the script.

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

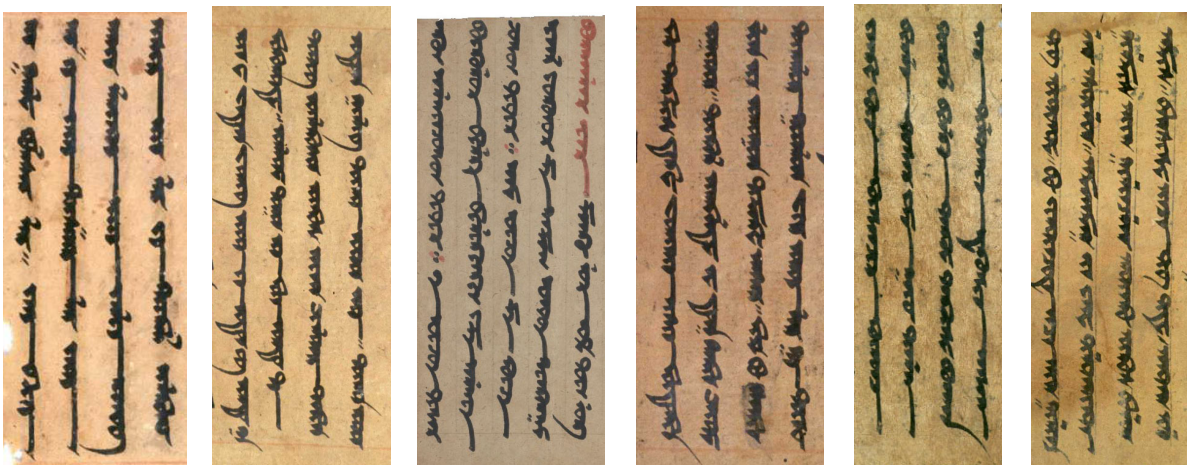


4.4 Styles of the script

The styles of the Uyghur script are classified into three broad categories: square, semi-square, semi-cursive, and cursive. The square styles were the basis for the block-print styles.

Uyghur 'square' style

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



Uyghur ‘cursive’ style

U 499 U 560 U 456 U 558

Uyghur block-print styles

U 387 U 7008 Mainz 801 U 343 U 496 PEALD 6r

5 Encoding model

5.1 Repertoire

The proposed encoding for Old Uyghur provides a unified block for encoding typical texts in the script. The encoded repertoire is based upon the full attested alphabet in manuscripts such as U 40 and Kashgari, and also upon the repertoire used in modern printed facsimilies of Uyghur manuscripts, such as Müller 1908 (see table 3). Although the letters used in block prints may be considered a 'standard' following the implications of 'printing', such documents were produced after the 14th century. Thus, these documents represent the later Uyghur script and do not maintain the palaeographic distinctions between all letters. Therefore, it is not practical to rely upon block prints repertoires as the basis for a complete encoding for Old Uyghur.

The proposed repertoire contains 37 characters: 21 letters, 7 combining signs, 6 punctuation signs, 1 stem-extending sign, 1 ornamental-terminal sign, and 1 editorial sign. The code chart and names list follows p. 11.

It provides a means for handling some of the complexities introduced by the ambiguous usage of a single glyph for multiple letters in block prints and cursive texts. For example, it includes distinctive characters for *aleph* and *nun*, as well as a merged form that may be used when a single letter is used for both. That said, the encoding does not aim to reconcile all ambiguities that result from cursive writing, but rather is guided by the attestations of carefully-written texts, as well as modern printed facsimiles.

Alternate forms of some letters have been encoded as atomic characters on the basis of semantic distinctions. This approach eliminates the need for using variation selectors or font changes for such contrastive usage. Diacritics are encoded as combining signs to be used with base letters.

5.2 Representative glyphs

In Unicode, the representative glyphs for most cursive joining scripts are based upon the final form of the letters. However, the Uyghur repertoires in U 40 and by Kashgari show the independent forms of the letters, which suggests a tradition of representing the script in this fashion. Therefore, the independent forms of the proposed letters are selected as the representative glyphs. The printed facsimilies in Muller (1908) contains nearly all contextual forms of the letters (see table 3) The representative glyphs have been designed to reflect the general aesthetics of the block-print style.

5.3 Joining model

The model is based upon that used for cursive joining scripts in Unicode. Each 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.

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.

5.4 Character names

The names of Old Uyghur letters are based upon 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 𐽁 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 'below-base' sign.

6 Proposed repertoire

6.1 Letters

Character name	Glyph	Joining	Latin
OLD UYGHUR LETTER ALEPH	𐽀	dual	ʾ
OLD UYGHUR LETTER ALEPH WITH LEFT TAIL	𐽁	right	-ʾ
OLD UYGHUR LETTER ALEPH WITH RIGHT TAIL	𐽂	right	-ʾ
OLD UYGHUR LETTER ALEPH-NUN	𐽃	dual	-ʾ, -n
OLD UYGHUR LETTER BETH	𐽄	dual	β
OLD UYGHUR LETTER GIMEL-HETH	𐽅	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, š
OLD UYGHUR LETTER PE	𐽌	dual	p
OLD UYGHUR LETTER SADHE	𐽍	dual	c
OLD UYGHUR LETTER RESH	𐽎	dual	r
OLD UYGHUR LETTER SHIN	𐽏	dual	š
OLD UYGHUR LETTER TAW	𐽐	dual	t
OLD UYGHUR LETTER LESH	𐽑	dual	l

6.1.1 *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. 48). 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. 14). It is written using the regular independent form 𐽀 or the alternate 𐽁 with a vertical terminal. In some cases, the two are used concurrently

for distinguishing between final *a* (𐽰) and *ä* or *e* (𐽱), see fig. 4; also see forms used for *-a* in fig. 7. The 𐽱 is not used for *nun*.

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

- *Initial* The initial 𐽀 *aleph* and 𐽀 *nun* are preserved in carefully written texts, such as Mainz 126 and Pelliot Ouïgour 13 (see fig. 15), 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 𐽀.

- *Medial* In Müller (1908), there is a clear distinction between the medial 𐽁 *aleph* and medial 𐽁 *nun*, where the former is more hooked and shorter than the latter (see fig. 12). It appears that the letters are contrasted (see fig. 16) in some documents. However, the medial forms are not contrasted consistently. Some ambiguity may be ascribed to the thick strokes that are characteristic of some scribal practices, or to the cursive aspects of such practices where there is less consideration for producing letters carefully. 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 𐽁.

- *Final* In Müller (1908), there is a clear distinction between the final 𐽂 *aleph* and final 𐽂 *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 the forms 𐽂 or 𐽂 or a swash variant in which the body of the letter forms a curved stroke with the terminal. However, in some block prints it appears that the finals are differentiated: the body of final 𐽲 *aleph* is triangular with points at east and south (from a vertical perspective), while the final 𐽲 *nun* is slanted eastward (see fig. 17). 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 𐽲 is in fact distinct from 𐽴, the highly similar structures of these glyphs lend themselves to being interpreted as the same letter, and in several documents, that is the case.

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

  – Following *kaph* or *pe* In these contexts *aleph* is written using the independent form 𐽰, eg. 𐽸𐽰 *k ›*, as is the convention for penultimate *kaph* (see § 7.1.1). This final form occurs concurrently with the regular final *aleph*, and is attested in manuscripts and block prints (see fig. 18). In documents where the 𐽲 / 𐽳 'toothed' form of independent *aleph*, described above, is used instead of the regular independent form 𐽰, it is used with *kaph* and *pe* as well: 𐽸𐽲 / 𐽸𐽳 *k ›*, 𐽹𐽲 / 𐽹𐽳 *p ›* (see fig. 18), compare to 𐽸𐽰 *k ›*, 𐽹𐽰 *p ›*. Such contextual glyph variation should be considered conventional behavior.

  – The *aleph* is also written as 𐽉 when final, and occurs concurrently with the regular final *aleph* (𐽰, 𐽲) 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. 14). 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 𐽰 *aleph*.

- *Disambiguation* Due to the ambiguity of these two letters in some documents, the ◌́ is written above *nun* in order to distinguish it from *aleph*, compare 𐽺́ vs 𐽰 for /n/ and /a/, respectively (see § 8.2).

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

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

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 𐽰 *aleph* and 𐽺 *nun*
- the distinctive initial forms 𐽰 *aleph* and 𐽺 *nun*
- the distinctive medial forms 𐽰 *aleph* and 𐽺 *nun*
- the shared initial form 𐽰 for *aleph* and *nun*
- the shared medial form 𐽰 for *aleph* and *nun*
- the shared final form 𐽰 for *aleph* and *nun*
- the alternate independent form 𐽰 of *aleph*
- the contextual substitution for 𐽰 *aleph* following penultimate *kaph* and *pe*
- the alternate final form 𐽰 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	𐽀	𐽀	𐽁	𐽁
ALEPH WITH LEFT TAIL	right	𐽂	𐽂	—	—
ALEPH WITH RIGHT TAIL	right	𐽃	𐽃	—	—
NUN	dual	𐽄	𐽄	𐽅	𐽅
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 'toothed' forms 𐽈 / 𐽉 of *aleph* are to be treated as stylistic variants of 𐽁 / 𐽅.
- The final forms 𐽊 *aleph* and 𐽋 *nun* used in block prints are to be treated as stylistic variants.

6.1.2 *beth*

The regular final form of *beth* is 𐽌, however, the final is also written as 𐽍 (see fig. 19). The left-ward orientation of the tail is used likely for distinguishing 𐽌 *beth* from 𐽎 *yodh* when there is a limitation of space for extending the final stroke of the former. Such distinctions are necessary in block-print styles, where non-final forms of *beth* and *yodh* are highly identical. However, the 𐽍 is a glyphic variant and is not proposed for separate encoding. In plain-text representations, usage of 𐽌 for final *beth* is sufficient.

Although, *beth* and *yodh* have distinctive shapes (see fig. 20), in some documents they may be written with a similar glyph in medial position. In such cases, the value of the ambiguous sign may be determined by morphological contexts. But a note regardingg the handling of such ambiguity has been given in § 6.1.12.

6.1.3 *gimel* and *heth*

As evidenced by the inventory in U 40, these two letters are distinguished in independent and final positions using the glyphs 𐽏 and 𐽐, respectively, but they have the same 𐽑 initial and 𐽒 medial forms (also see

fig. 21). For this reason, the following model is proposed for representing these letters:

- The letters *gimel* and *heth* are unified as the dual-joining letter 𐽼 GIMEL-HETH.
- To represent final *heth* when it is distinguished from *gimel*, the right-joining letter 𐽽 FINAL HETH is proposed for encoding.
- The diacritics ◌́ and ◌̋ may be placed above 𐽼 and 𐽽 for representing the sounds /q/ and /x/, eg. 𐽼́, 𐽼̋, 𐽽́, 𐽽̋ (see § 8.2).

6.1.4 *zayin*

The representative form 𐽻 of *zayin* is based upon the shape used in formal and block-print styles (see fig. 22). The glyphic variant 𐽻 'sawtooth' form occurs in some documents (see fig. 11).

In some sources *zayin* is distinguished using the diacritics ◌́ and ◌̋, eg. 𐽻́ and 𐽻̋, in order to indicate /ž/ (see § 8.2).

6.1.5 *kaph*

The regular final form of *kaph* is 𐽾, however, the final is also written as 𐽿 (see fig. 23). Both forms may occur concurrently within a document, but as there is no semantic difference between 𐽾 and 𐽿, the latter should be considered a stylistic variant. There is no need to encode it as a separate character. In digital plain-text representations, usage of the regular 𐽾 in place of 𐽿 is acceptable.

6.1.6 *mem*

As attested in the inventory in U 40, the *mem* has two distinctive graphemes: 𐾀 and 𐾁. 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.

6.1.7 *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) (see fig. 24). An example of distinctive forms of the letters in final position is shown in fig. 25. 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 § 8.2).

6.1.8 *pe*

In various manuscripts and block prints, final 𐾅 *pe* is rendered as the ornamental form 𐾆 (see fig. 26). 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 𐾆, which would mean that 𐾅𐾆 is actually

a sequence of non-final *pe* and the space-filling sign 𐽙. This sign has been encoded as a combining sign (see § 6.5).

6.1.9 *sadhe*

The regular final form of *sadhe* is 𐽆, however, the final is also written as 𐽅 (see fig. 27). Both forms may occur concurrently within a document, but, as there is no semantic difference between 𐽆 and 𐽅, the latter should be considered a stylistic variant. There is no need to encode it as a separate character. In digital plain-text representations, usage of 𐽆 for final *sadhe* is sufficient.

6.1.10 *taw*

The body of the initial form 𐽇 of *taw* sits below the baseline, as compared to its medial 𐽇 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. 28). Both forms may occur concurrently within a document, but as As there is no known semantic difference between the forms, the variant 𐽉 should be considered a stylistic variant. There is no need to encode it separately

6.1.11 *lesh*

The letter 𐽊 represents the sound /l/. It is derived from 𐼴 U+10F44 SOGDIAN LETTER LESH, which is known as 'hooked *r*' (see Pandey 2016b for details). The Uyghur 𐽊 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.

6.1.12 Note on handling ambiguity for pairs of letters

The following pairs of letters have similar forms in some texts. They should be represnted in encoded text as follows:

- *aleph* and *nun* Use the regular letters ALEPH and NUN when letterforms can be distinguished, and the unified character ALEPH-NUN when the letterform is ambiguous, eg. in medial or final position.

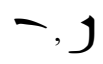
- *beth* and *yodh* Use the regular letters BETH and YODH when letterforms can be distinguished; use the character that best resembles the glyph used for both letters in the given source.

- *gimel* and *heth* Use the unified letter GIMEL-HETH, but use FINAL HETH for representing the alternate form as necessary.

- *samekh* and *shin* Use the regular letters SAMEKH and SHIN when letterforms can be distinguished; use the character that best resembles the glyph used for both letters in the given source.

6.1.13 Note on variation in terminal orientation

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

	regular	alternate
aleph		
beth		
kaph		
pe		
sadhe		
taw		

There are various possible explanations for such variation:

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

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

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

At present only the 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.

6.2 Combining signs

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

Character name	Glyph
OLD UYGHUR COMBINING DOT RIGHT	
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	

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. 29). 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 ◌∘ ('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.

6.3 Punctuation signs

The following signs are used for punctuation (see fig. 31 for examples):

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	𐽞

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. 36, 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. 37. 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.

6.4 Stem extender

The following character is used for extending the baseline (see § 8.3 for details). It is used as a typographic filler and also for indicating a suffix that is separated from the stem. The stem-extending sign is defined as a left-joining character.

Character name	Glyph
OLD UYGHUR STEM EXTENDER	𐽙

6.5 Ornamental terminal

The following character is used for representing an ornamental terminal, eg. the final 𐽂 *pe* written at the end of a line (see fig. 26). In the available materials, this terminal occurs only with *pe*. 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.

Character name	Glyph
OLD UYGHUR ORNAMENTAL TERMINAL	𐽆

6.6 Editoral sign

The following editorial sign is used in manuscripts:

Character name	Glyph
OLD UYGHUR DELETION MARK	◌𐽐

When written beneath a word or letter, this sign indicates that the respective text is an error and is to be omitted. The correct word is generally written after the mispelled word (see fig. 30).

6.7 Word boundaries

There is clear demarcation of word boundaries using spaces in block prints and manuscripts. In

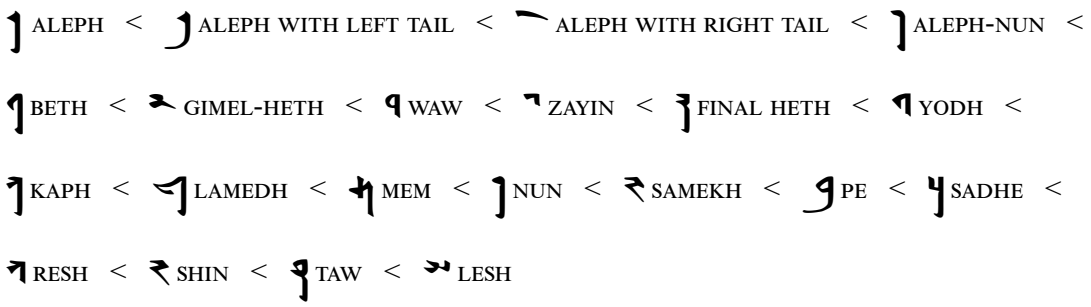
This generally applies to manuscripts, as well. However, in numerous manuscripts the terminal of a final letter may connect with the initial letter of the following word. In such cases, the word boundary is identifiable by the elongation of the final stroke. Such stroke elongation may be a space-filling calligraphic technique; there is no joining behavior between such a final letter and the following initial letter. In plain encoded text, a space is expected after the final letter in such cases.

6.8 Line-breaking

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

6.9 Collation

The sort order for Old Uyghur follows the encoded order:



7 Joining behavior

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

Dual-joining letters

	independent	final	medial	inital
ALEPH	𐽀	𐽀	𐽀	𐽀
ALEPH-NUN	𐽀	𐽀	𐽀	𐽀
BETH	𐽁	𐽁	𐽁	𐽁
GIMEL-HETH	𐽂	𐽂	𐽂	𐽂
WAW	𐽃	𐽃	𐽃	𐽃
YODH	𐽄	𐽄	𐽄	𐽄
KAPH	𐽅	𐽅	𐽅	𐽅
LAMEDH	𐽆	𐽆	𐽆	𐽆
MEM	𐽇	𐽇	𐽇	𐽇
NUN	𐽈	𐽈	𐽈	𐽈
SAMEKH	𐽉	𐽉	𐽉	𐽉
PE	𐽊	𐽊	𐽊	𐽊
SADHE	𐽋	𐽋	𐽋	𐽋
RESH	𐽌	𐽌	𐽌	𐽌
SHIN	𐽍	𐽍	𐽍	𐽍
TAW	𐽎	𐽎	𐽎	𐽎
LESH	𐽏	𐽏	𐽏	𐽏

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

Right-joining letters

	independent	final
ALEPH WITH LEFT TAIL	𐽀	𐽀
ALEPH WITH RIGHT TAIL	𐽁	𐽁
FINAL HETH	𐽂	𐽂
ZAYIN	𐽃	𐽃

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

ʾʾltwn
‘altun’

𐽀 ALEPH, 𐽀 ALEPH, 𐽐 LESH, 𐽆 TAW, 𐽅 WAW, 𐽉 NUN

bʾmsʾn
‘vamsan’

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

wyγwr
‘üigür’

𐽅 WAW, 𐽅 YODH, 𐽇 GIMEL-HETH, 𐽅 WAW, 𐽄 RESH

qwtlwγ
‘qutlug’

𐽇 GIMEL-HETH, ◌𐽑 COMBINING TWO DOTS LEFT, 𐽆 TAW, 𐽐 LESH, 𐽅 WAW, 𐽇 GIMEL-HETH

mncwšry
‘mancusari’

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

swδwr
‘sutur’

𐽍 SAMEKH, 𐽅 WAW, 𐽊 LAMEDH, 𐽅 WAW, 𐽄 RESH

pwδystb
‘bodisatav’

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

pwrγʾn
‘burxan’

PE, WAW, RESH, GIMEL-HETH, ALEPH, NUN

pwlz-wn
‘bolz-un’

PE, WAW, LESH, ZAYIN, WAW, NUN

pylyk
‘bilig’

PE, YODH, LESH, YODH, KAPH

twykʾl
‘tükäl’

TAW, WAW, YODH, KAPH, ALEPH, LESH

tnkry
‘tängri’

TAW, NUN, KAPH, RESH, YODH

7.1 Glyph interactions

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

7.1.1 *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>		

7.1.2 *waw*

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

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

7.1.3 *mem*

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

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

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

7.1.4 *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’		
kˀlmˀdwk ‘kälmädük’		

8 Encoded representations

8.1 Vowels

In general, all vowels are indicated in Old Uyghur. The representation of vowels follows the basic '*matres lectionis*' pattern for Semitic scripts, in which 𐽰 *aleph*, 𐽳 *waw*, and 𐽶 *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. 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 𐾀𐽺𐽲𐽾𐽶, without an explicit *aleph* for /ä/.

	Initial		Medial	
ä	𐽰	𐽰 ALEPH	𐽰	𐽰 ALEPH
a, *e*	𐽰𐽰	𐽰 ALEPH, 𐽰 ALEPH	𐽰	𐽰 ALEPH
i, *ï*	𐽰𐽶	𐽰 ALEPH, 𐽶 YODH	𐽶	𐽶 YODH
ī, *ï̄*	𐽰𐽶𐽶	𐽰 ALEPH, 𐽶 YODH, 𐽶 YODH	𐽶𐽶	𐽶 YODH, 𐽶 YODH
o, *u*	𐽰𐽳	𐽰 ALEPH, 𐽳 WAW	𐽳	𐽳 WAW
ö, *ü*	𐽰𐽳𐽶	𐽰 ALEPH, 𐽳 WAW, 𐽶 YODH	𐽳	𐽳 WAW
ö, *ü*	𐽳𐽶	𐽳 WAW, 𐽶 YODH	𐽳𐽶	𐽳 WAW, 𐽶 YODH
ō, *ȫ*, *ū*, *ǖ*	𐽰𐽳𐽳	𐽰 ALEPH, 𐽳 WAW, 𐽳 WAW	𐽳𐽳	𐽳 WAW, 𐽳 WAW

The final forms of all vowels are represented using the regular final form of *aleph*, *waw*, *yodh*, respectively.

8.2 Disambiguation and extension of letters

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

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

8.3 Stem extension

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

tynly lr-r
‘tinlag-lar-r’

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

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

9 Character Properties

9.1 Core data: `UnicodeData.txt`

```
10F70;OLD UYGHUR LETTER ALEPH;Lo;0;AL;;;;;N;;;;;
10F71;OLD UYGHUR LETTER ALEPH WITH LEFT TAIL;Lo;0;AL;;;;;N;;;;;
10F72;OLD UYGHUR LETTER ALEPH WITH RIGHT 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;;;;;
```

9.2 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
```

9.3 Property list: `PropList.txt`

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

9.4 Shaping properties: `ArabicShaping.txt`

```
10F70; OLD UYGHUR ALEPH; D; No_Joining_Group
10F71; OLD UYGHUR ALEPH WITH LEFT TAIL; R; No_Joining_Group
10F72; OLD UYGHUR ALEPH WITH RIGHT 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
```

10 References

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

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

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.

11 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.

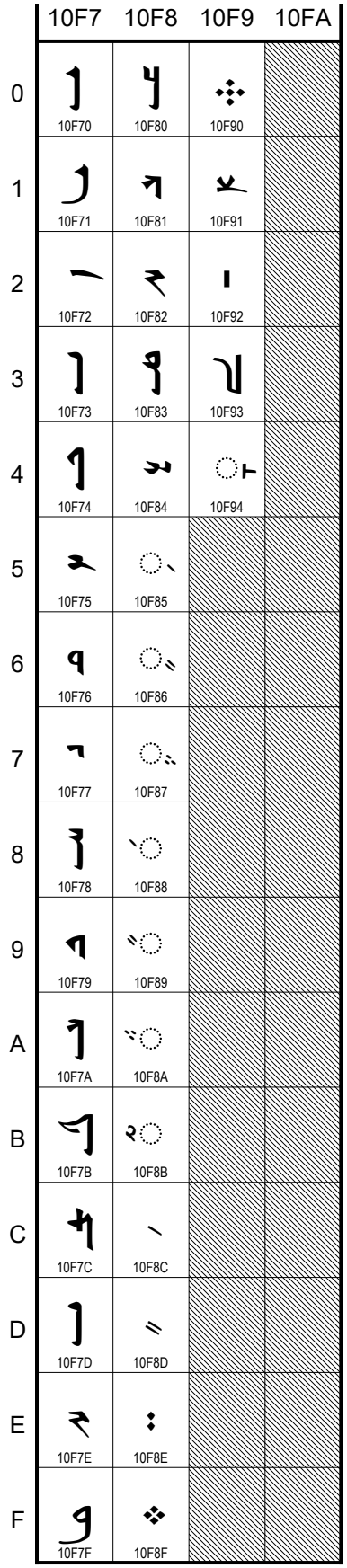
I also would like to thank Andrew West for providing reproductions and transcriptions of the Old Uyghur inscriptions on the walls of the Cloud Platform at Juyong Guan.

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 ALEPH WITH RIGHT TAIL
10F73 OLD UYGHUR LETTER ALEPH-NUN
10F74 OLD UYGHUR LETTER BETH
10F75 OLD UYGHUR LETTER GIMEL-HETH
10F76 OLD UYGHUR LETTER WAW
10F77 OLD UYGHUR LETTER ZAYIN
10F78 OLD UYGHUR LETTER FINAL HETH
10F79 OLD UYGHUR LETTER YODH
10F7A OLD UYGHUR LETTER KAPH
10F7B OLD UYGHUR LETTER LAMEDH
10F7C OLD UYGHUR LETTER MEM
10F7D OLD UYGHUR LETTER NUN
10F7E OLD UYGHUR LETTER SAMEKH
10F7F OLD UYGHUR LETTER PE
10F80 OLD UYGHUR LETTER SADHE
10F81 OLD UYGHUR LETTER RESH
10F82 OLD UYGHUR LETTER SHIN
10F83 OLD UYGHUR LETTER TAW
10F84 OLD UYGHUR LETTER LESH
• hooked r

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

Stem extender

10F92 OLD UYGHUR STEM EXTENDER

Ornamental terminal

10F93 OLD UYGHUR ORNAMENTAL TERMINAL

Editorial mark

10F94 OLD UYGHUR DELETION MARK

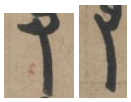
	PO 13 (fig. 32)			Mainz 126 (fig. 33)		
aleph						
beth	—	—		—	—	—
gimel / heth						
waw			—			
zayin		—	—		—	—
yodh						
kaph						
lamedh	—		—			—
mem						
nun						
samekh	—	—	—	—	—	—
pe	—	—			—	
sadhe						
resh			—			—
shin	—					
taw	—					
lesh						

Table 1: Specimens of letters from manuscripts

	U 387 – U 390 (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 block prints

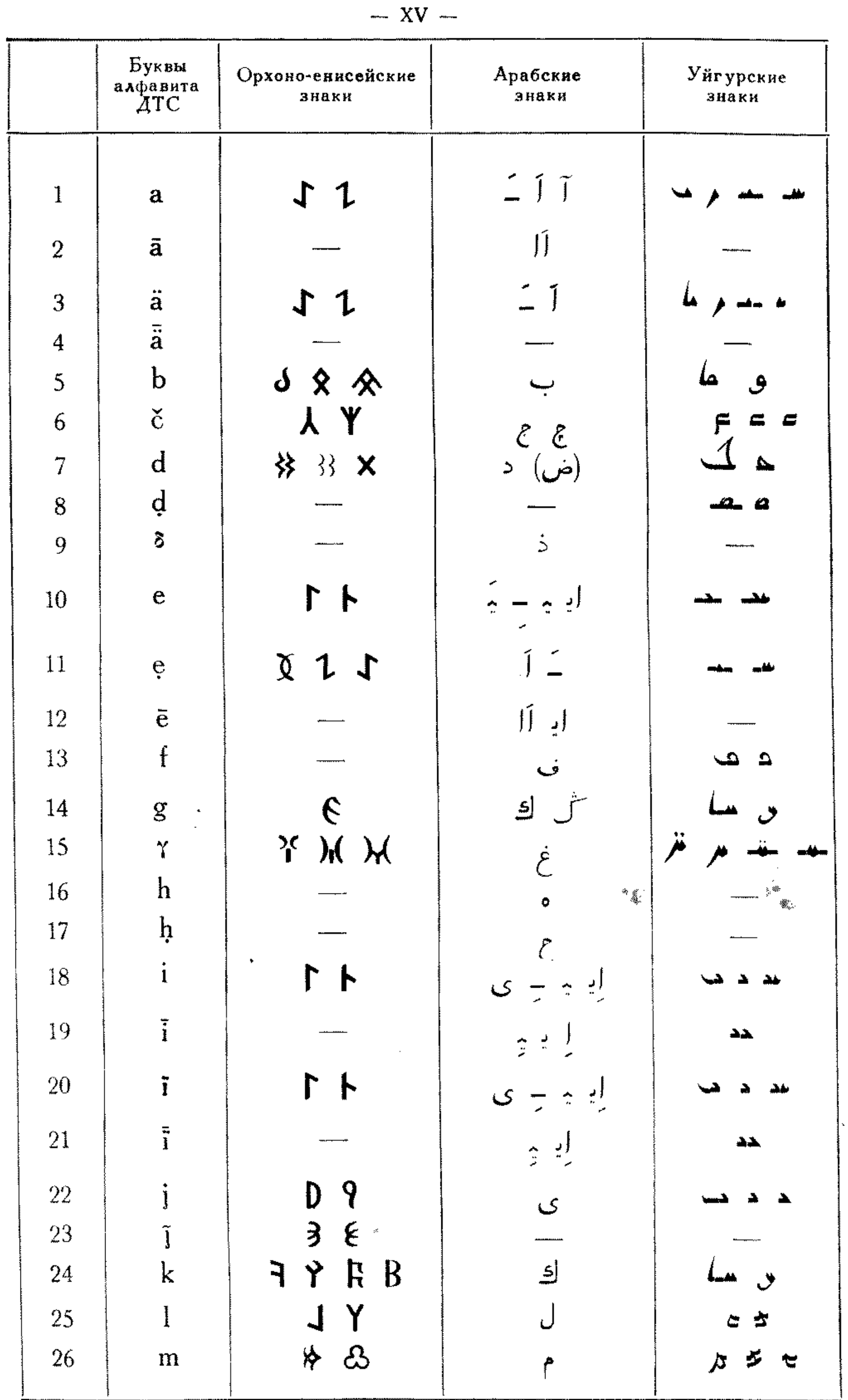
	Muller (1908: 36–44)			
aleph				
beth	—			
gimel / heth				
waw				
zayin		—	—	
yodh				
kaph				
lamedh				
mem				
nun				
samekh				—
pe				
sadhe				
resh				
shin				
taw				
lesh				

Table 3: 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.

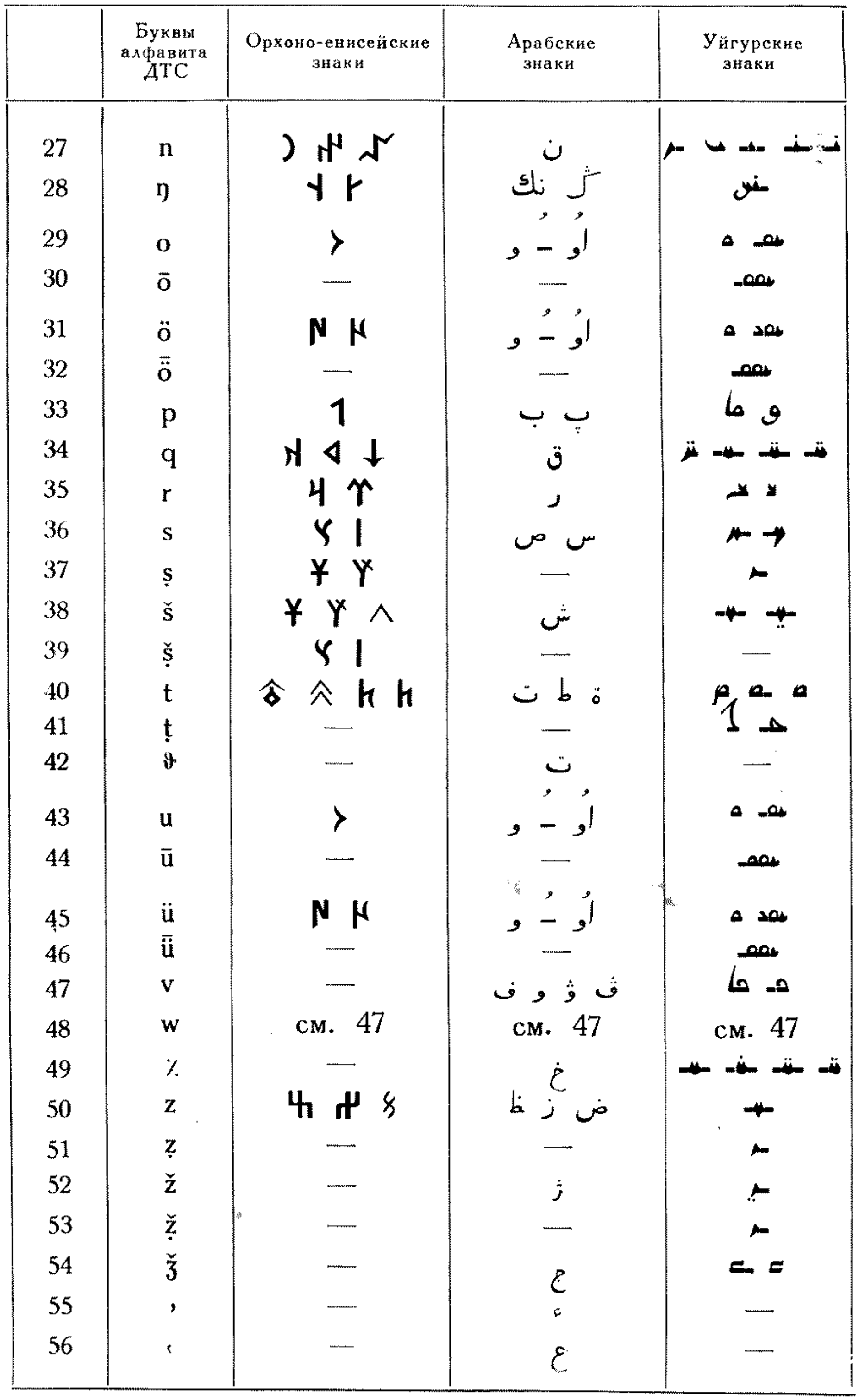


— XV —

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

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	[illegible]	ن	[illegible]
28	ŋ	[illegible]	ڭ نك	[illegible]
29	o	[illegible]	اُو ـُـ و	[illegible]
30	ō	—	—	[illegible]
31	ö	[illegible]	اُو ـُـ و	[illegible]
32	ȫ	—	—	[illegible]
33	p	[illegible]	پ ب	[illegible]
34	q	[illegible]	ق	[illegible]
35	r	[illegible]	ر	[illegible]
36	s	[illegible]	س ص	[illegible]
37	ṣ	[illegible]	—	[illegible]
38	š	[illegible]	ش	[illegible]
39	ṣ̌	[illegible]	—	—
40	t	[illegible]	ت ط ة	[illegible]
41	ṭ	—	—	[illegible]
42	ϑ	—	ث	—
43	u	[illegible]	اُو ـُـ و	[illegible]
44	ū	—	—	[illegible]
45	ü	[illegible]	اُو ـُـ و	[illegible]
46	ǖ	—	—	[illegible]
47	v	—	ڤ ۋ و ف	[illegible]
48	w	см. 47	см. 47	см. 47
49	γ	—	غ	[illegible]
50	z	[illegible]	ض ز ظ	[illegible]
51	ẓ	—	—	[illegible]
52	ž	—	ژ	[illegible]
53	ẓ̌	—	—	[illegible]
54	ǯ	—	ج	[illegible]
55	ʼ	—	ء	—
56	ʻ	—	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. 57.

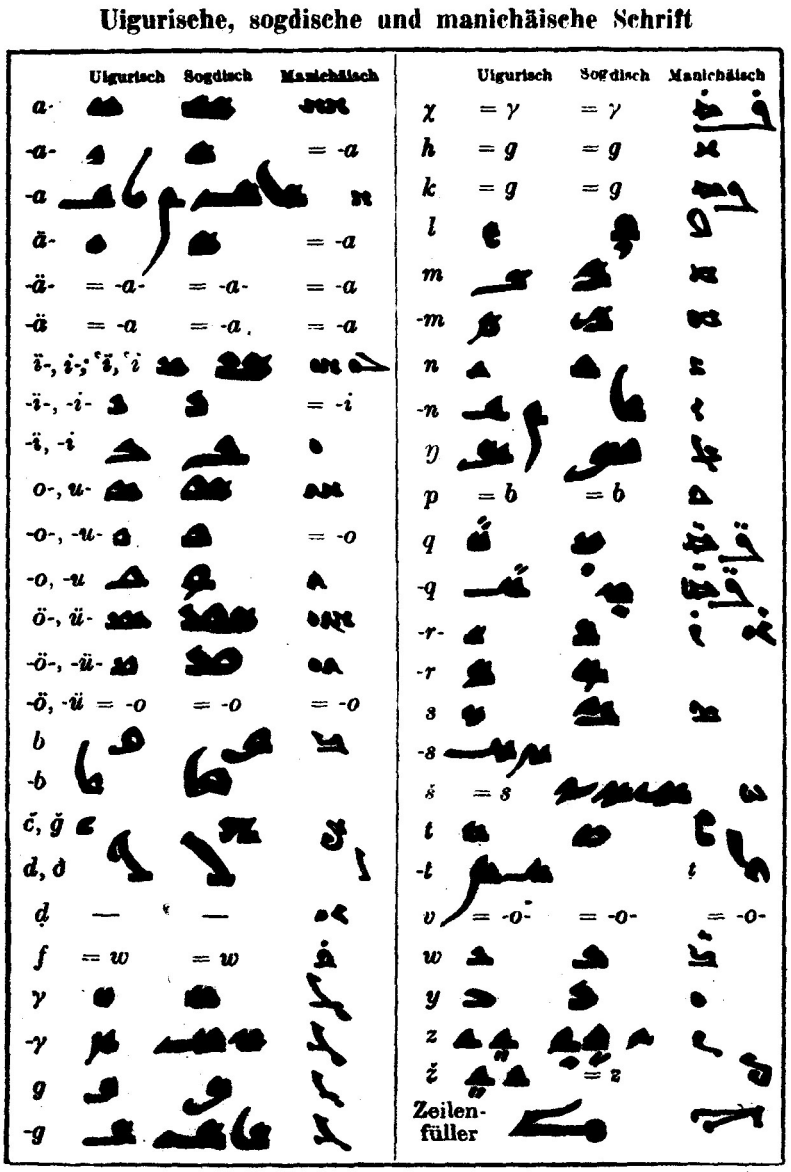


Figure 7: Comparison of Old Uyghur, Sogdian, and Manichaean letters (from von Gabain 1950: 17). For clearer examples of Old Uyghur letterforms referenced by von Gabain see the three Old Uyghur manuscripts, two in the formal script and the third in the cursive script, illustrated and transcribed in her work, reproduced here in fig. 51–55.

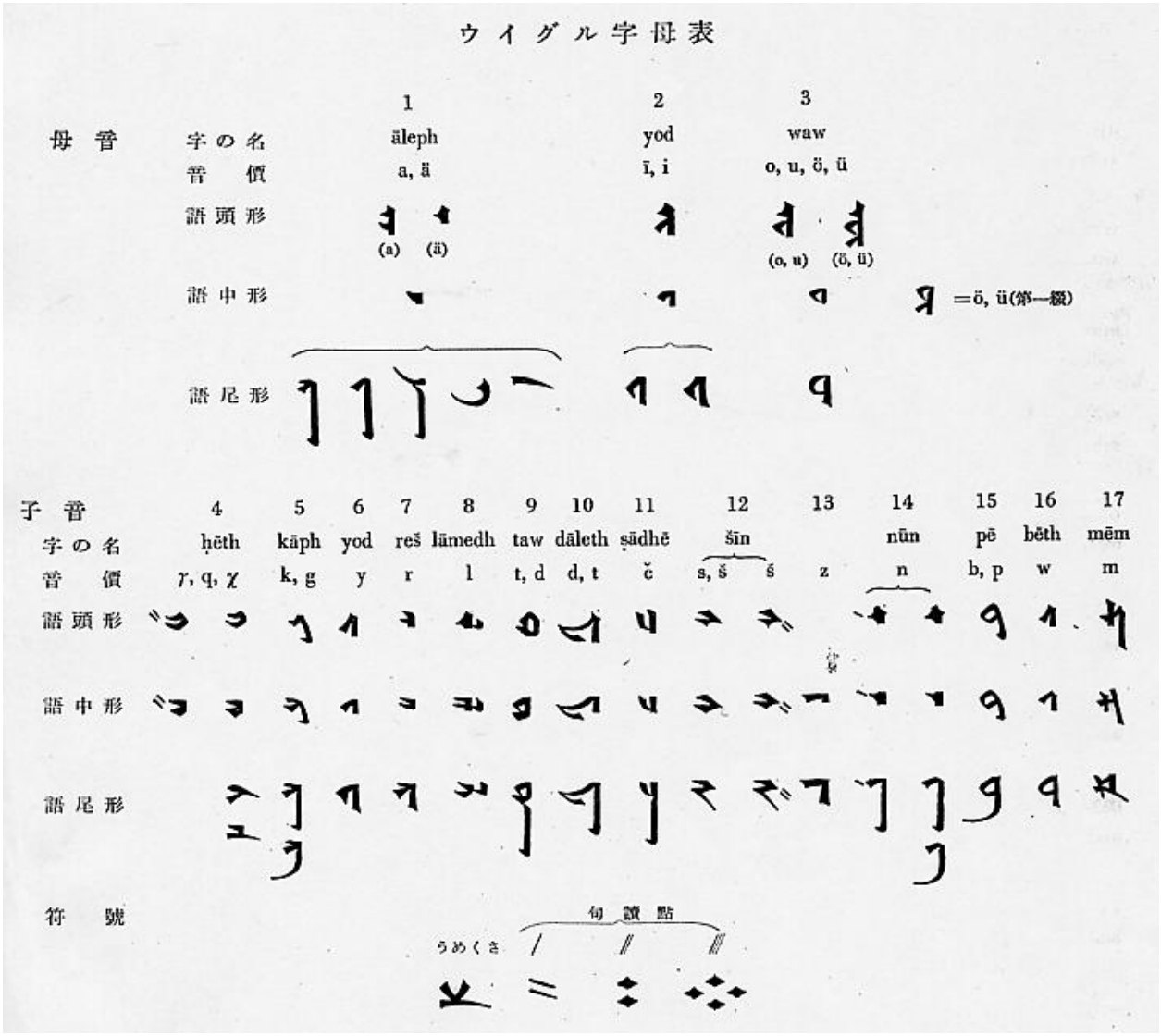


Figure 8: Table of Old Uyghur characters used in the Uyghur inscription in the multi-script Yuan dynasty inscriptions at Juyong Guan 居庸關 pass at the Great Wall northwest of Beijing (from Chü-Yung-Kuan 居庸關, "The Buddhist Arch of the Fourteenth Century A.D. at the Pass of the Great Wall Northwest of Peking", vol. 1, p. 165; reproduced from West 2006). See photograph containing an excerpt of the inscription in fig. 47.

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				

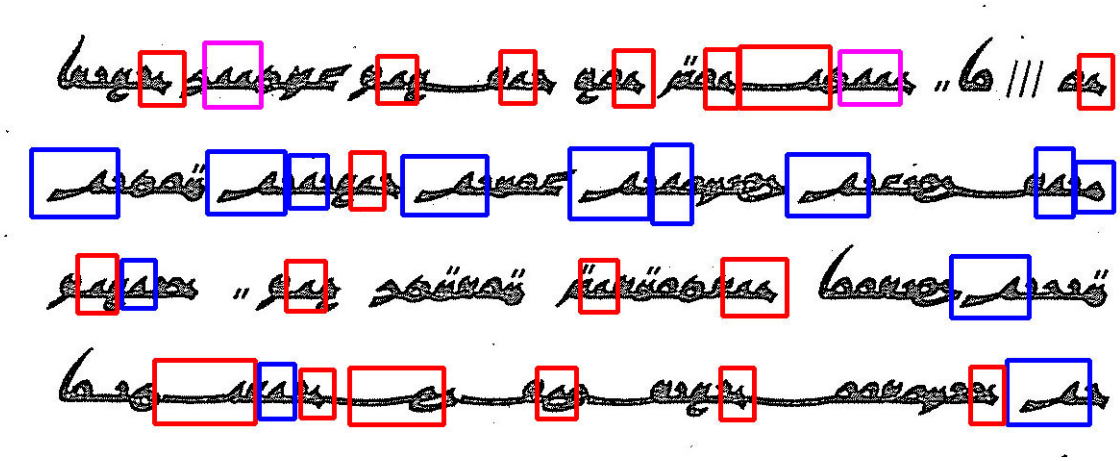
Table 2 Various forms of the Old Uighur alphabet from texts dating between the fourteenth and the tenth centuries BCE
Source: *adapted from Zieme 1991*

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

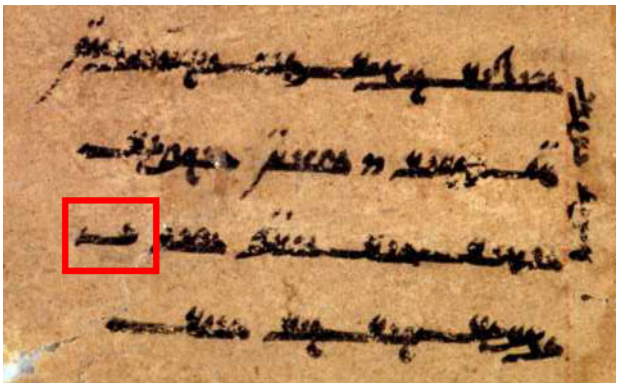
	Berliner Transkription system	Turkey	transcription at *Uigurisches Wörterbuch*	transliteration at *Uigurisches Wörterbuch*
𐽀	a, ạ	a, ạ	a	’’ / ’
𐽁	b	b	b	P
𐽂	č	ç	č	Č
𐽃	d, ṭ	d, ṭ	d, ḍ	D, T
𐽄	ä, ’ä	e, ’e	ä	’
𐽅	[e] i	ė / i	e	Y / ’Y
𐽆	g	g	g	K
𐽇	γ / γ́	g / ġ	g	Q, Q̈, Q̇
𐽈	h / χ, x, ẍ	h / ẖ, ḥ	h	H / X
𐽉	ï	ı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

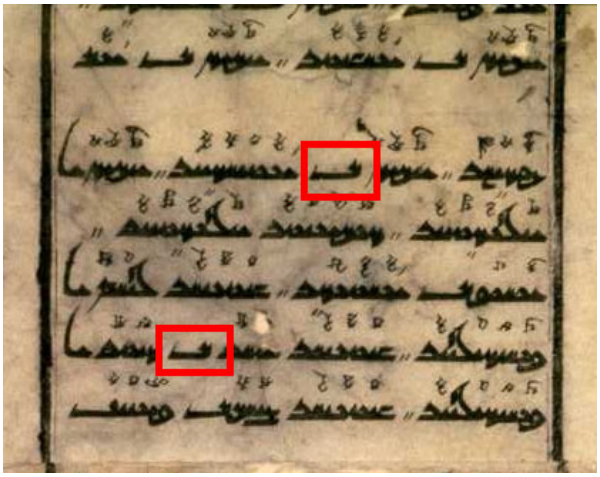


u . . . p ○○ anta oq̈ ol yäk-lär čštani ilig
? als nun jene Dämonen des Königs Tschastani
ning küčin küsünin čoγïn yalï[nïn] . . q̈utïn
Kraft, Macht und Majestät
q̈ïvïn körüp artuq̈raq̈ q̈orq̈tï-lar ○○ ünlär-
sahen, noch mehr gerieten sie in Furcht. Ihre Stimme
in ösürüp ilig bäg-kä inčä tip
anschwellen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

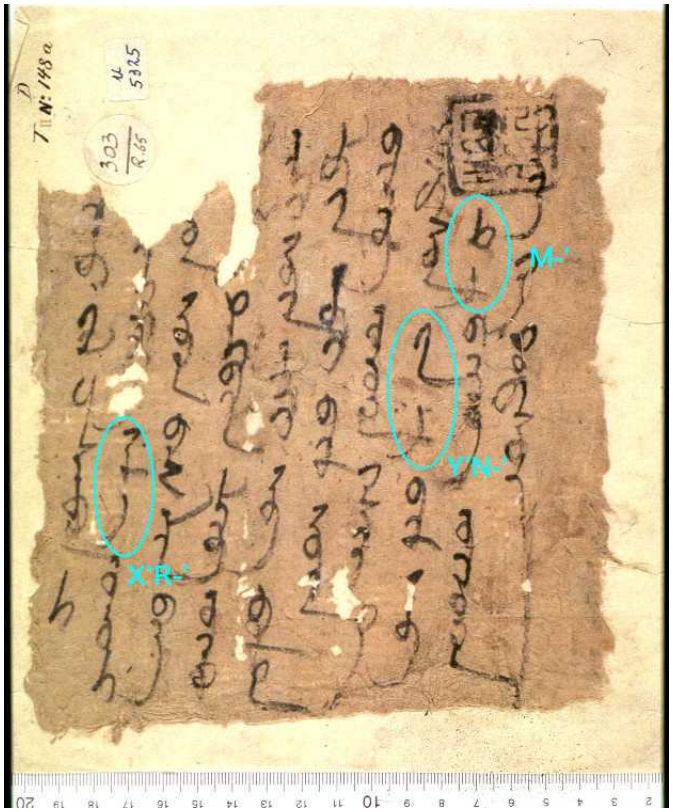


Example of the regular 𐽀 independent *aleph* (excerpt from U 2215).



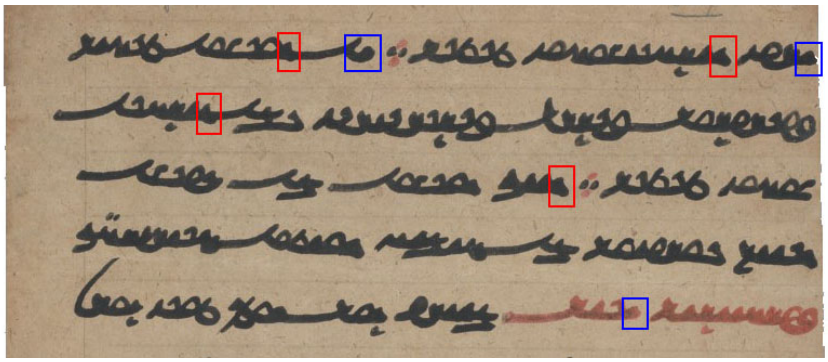
Examples of the alternate form 𐽁 of independent *aleph* (excerpt from Mainz 801).

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

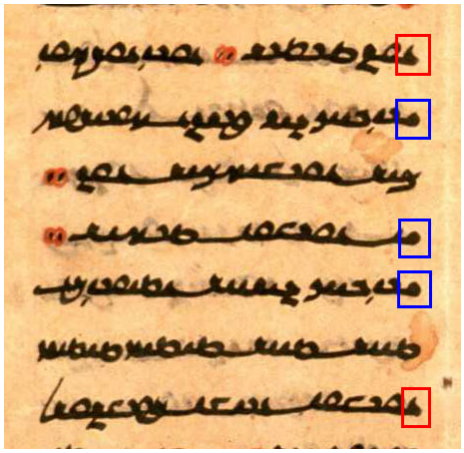


Usage of ﹷ ALEPH WITH RIGHT TAIL and ﹹ ALEPH WITH LEFT TAIL in U 5325. Annotations produced by Dai Matsui, November 2018.

Figure 14: Alternate forms of final *aleph*.

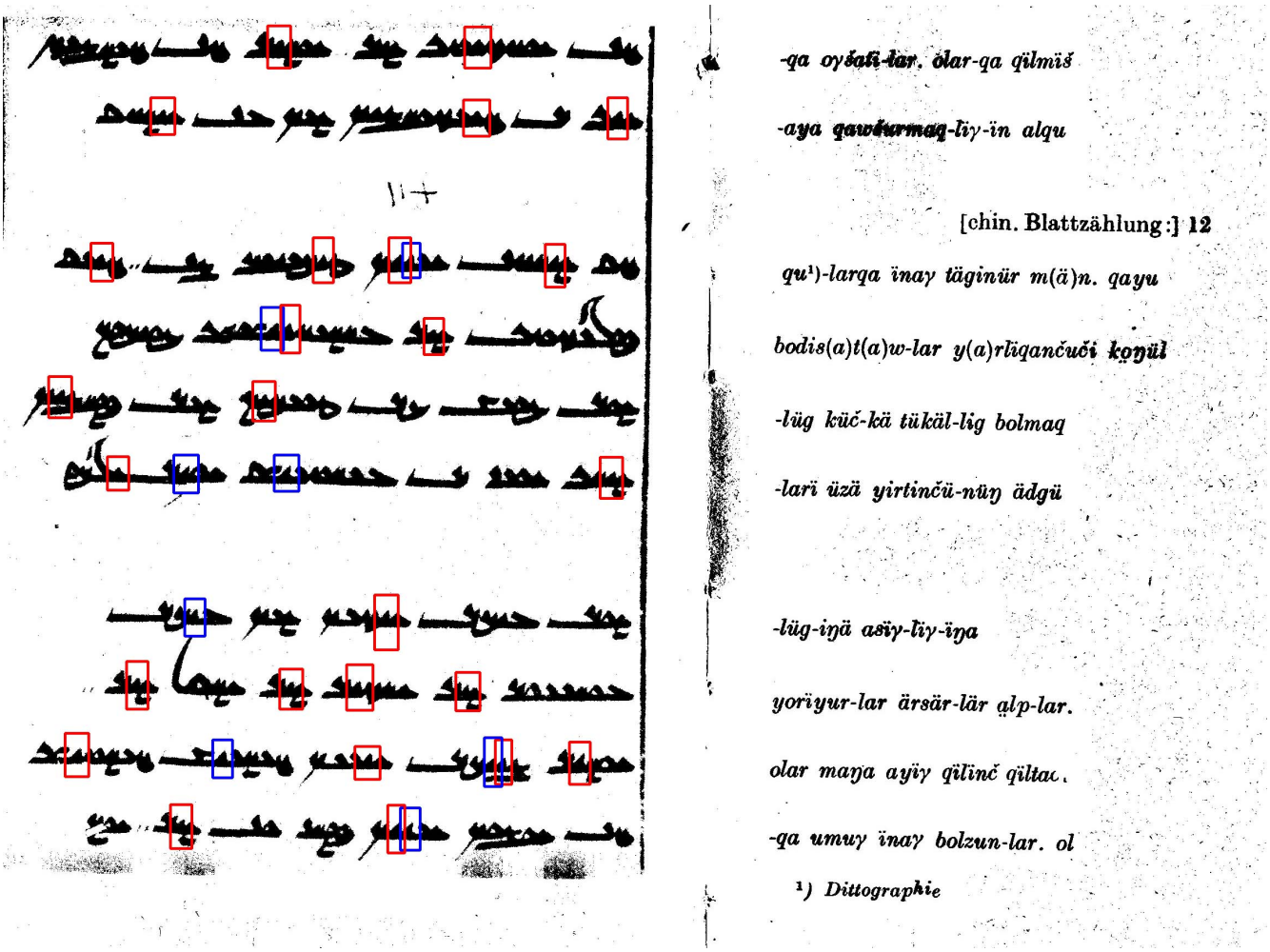


Excerpt from Pelliot Ouïgour 13 showing initial forms of 𐽀 *aleph* (red) and 𐽂 *nun* (blue).



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

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



-qa oγšatï-lar. ölar-qa qïlmïš

-aya qawšurmaq-lïγ-ïn alqu

[chin. Blattzählung:] 12

qu[1])-larqa ïnaγ täginür m(ä)n. qayu

bodis(a)t(a)w-lar y(a)rlïqančučï kǫŋül

-lüg küč-kä tükäl-lig bolmaq

-larï üzä yirtinčü-nüŋ ädgü

-lüg-iŋä asïγ-lïγ-ïŋa

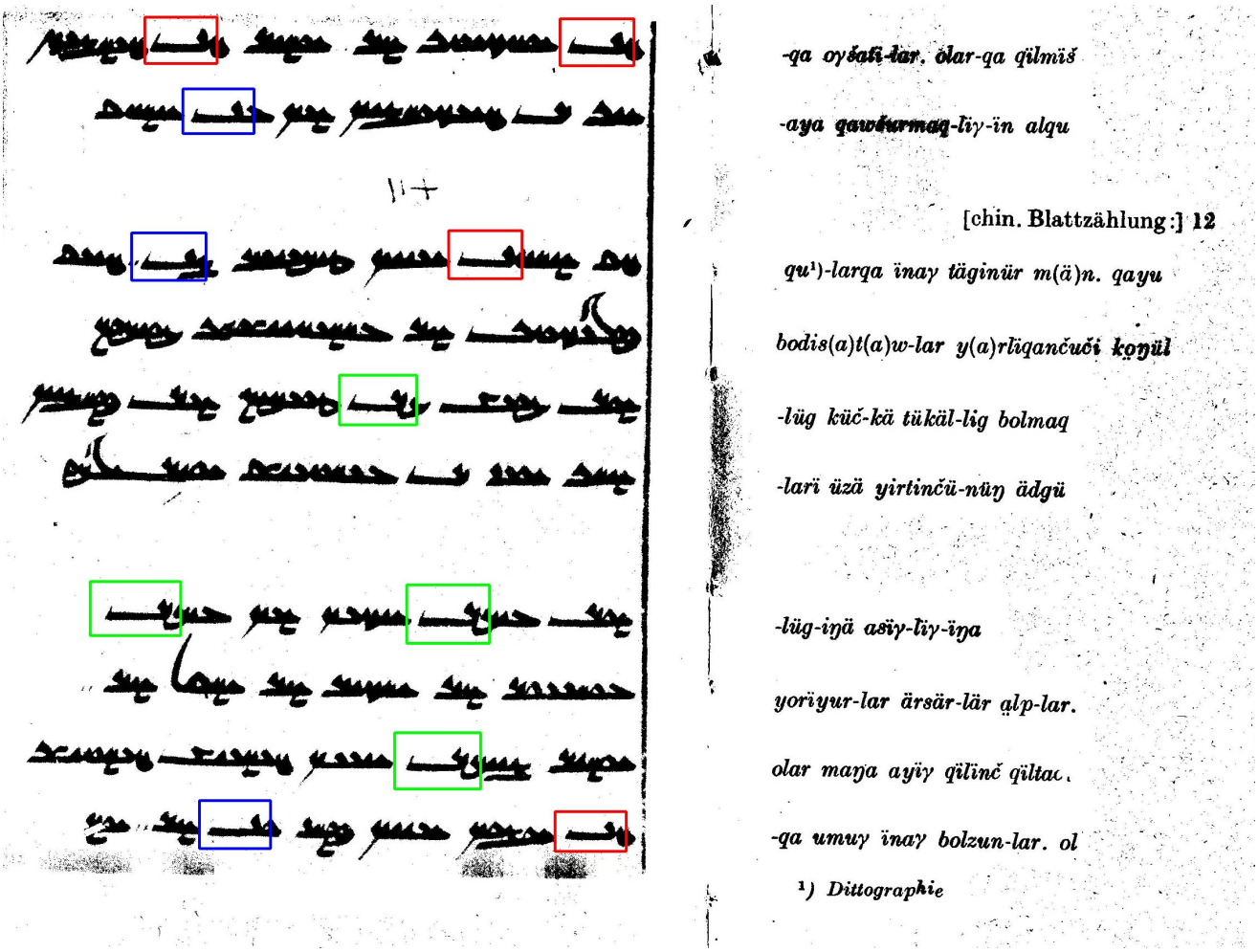
yorïyur-lar ärsär-lär alp-lar.

olar maŋa ayïγ qïlïnč qïltac.

-qa umuγ ïnaγ bolzun-lar. ol

[1]) Dittographie

Figure 16: Examples showing forms of ◂ medial *aleph* (red) and ◂ medium *nun* (nun), in which the letters are contrasted to some extent (from von Gabain 1950: 24–25).



-qa oγšati-lar. ōlar-qa qïlmïš

-aya qawšurmaq-liγ-in alqu

[chin. Blattzählung:] 12

qu[1])-larqa ïnaγ täginür m(ä)n. qayu

bodis(a)t(a)w-lar y(a)rlïqančuči kǫnül

-lüg küč-kä tükäl-lig bolmaq

-lari üzä yirtinčü-nüŋ ädgü

-lüg-iŋä asïγ-liγ-ïŋa

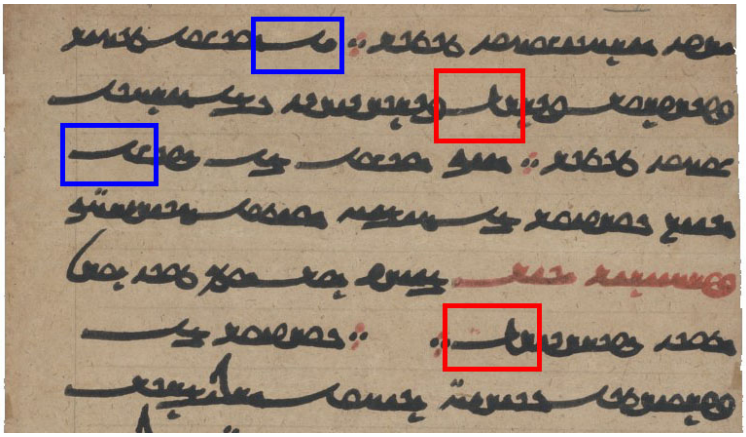
yorïyur-lar ärsär-lär ạlp-lar.

olar maŋa ayïγ qïlınč qïltac.

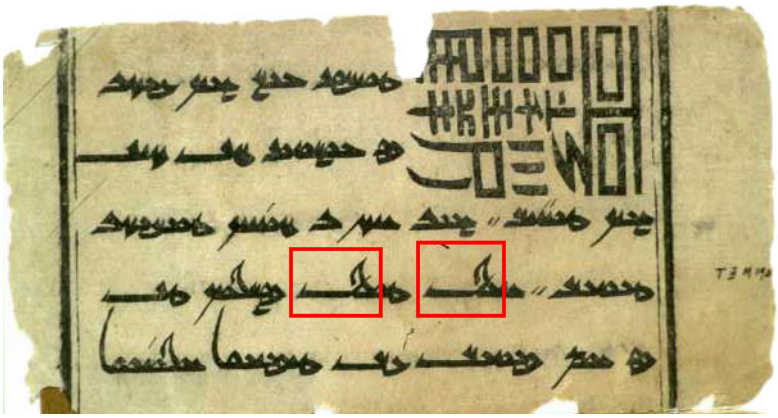
-qa umuγ ïnaγ bolzun-lar. ol

[1]) Dittographie

Figure 17: Examples showing contrastive forms of ـــك final *aleph* (red) and ـــك final *nun* (nun); with contextual variants of final *aleph* after *kaph* written as ـــك (green) (from von Gabain 1950: 24–25).

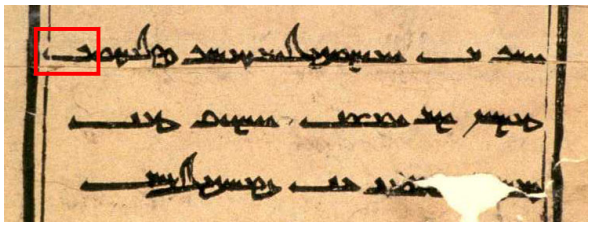


Excerpt from Pelliot Ouïgour 13 showing the final form of *aleph* used with *kaph* (red), compared with the regular form (blue).

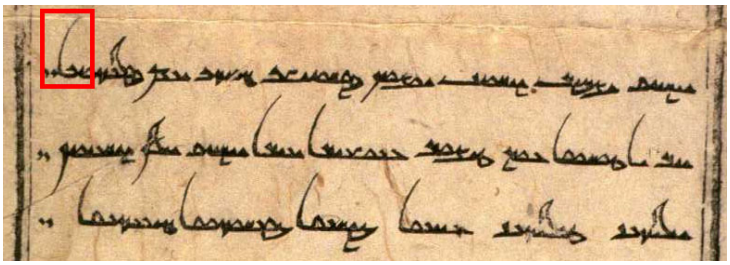


Folio from U 4960 showing the contextual form of *aleph* used with *lamedh*.

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



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



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

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

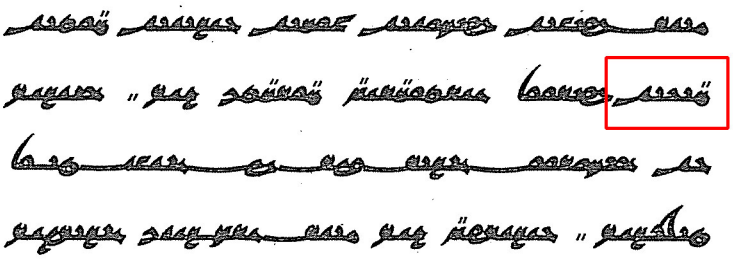
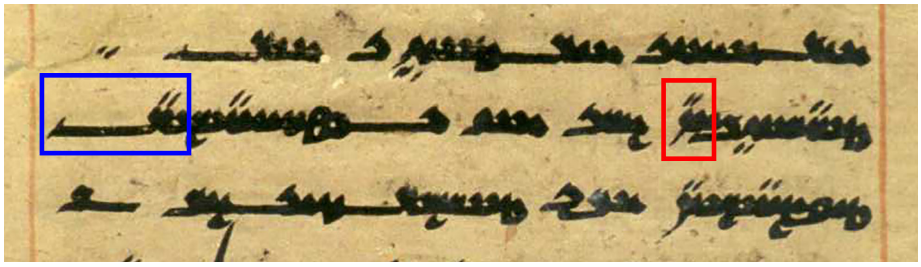
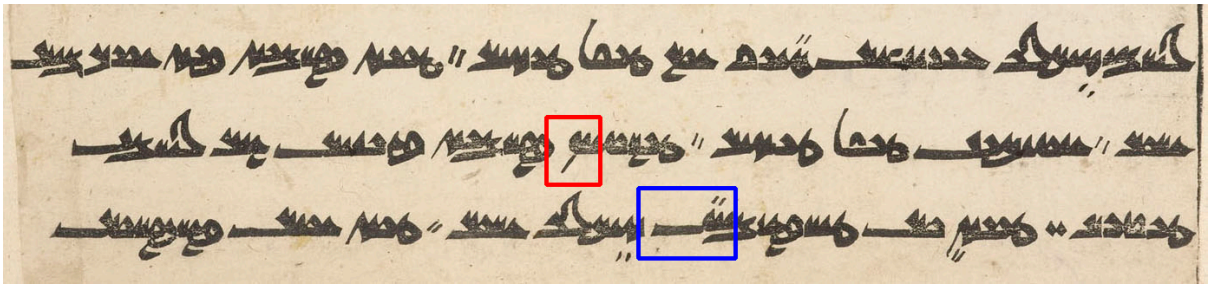


Figure 20: Contrastive representation of *beth* and *yodh* (from Müller 1908: 42, 43). The highlighted word shows the a sequence of medial *yodh*, *beth*, *yodh*.

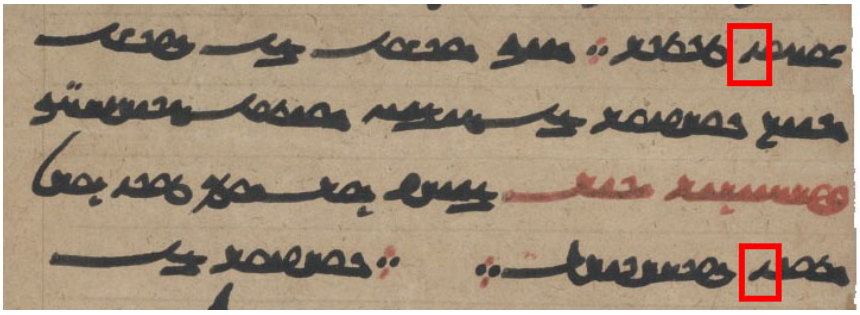


Excerpt from U 924 showing distinctive usage final forms of 𐽂 *gimel* (red) and 𐽆 *heth* (blue).

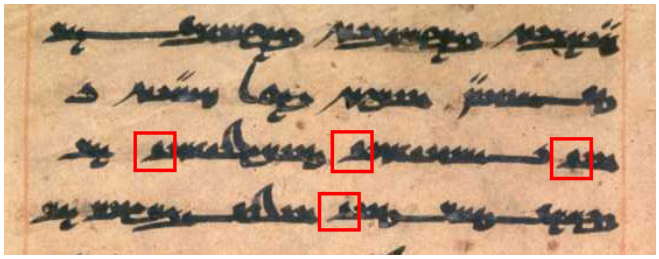


Excerpt from PEALD 6a showing final forms of 𐽂 *gimel* (red) and 𐽆 *heth* (blue).

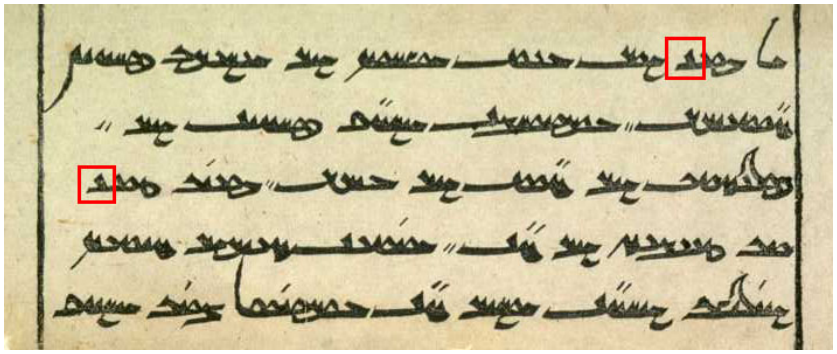
Figure 21: Examples of *gimel* and *heth*. Images have been rotated 90 degrees counter-clockwise for layout purposes.



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

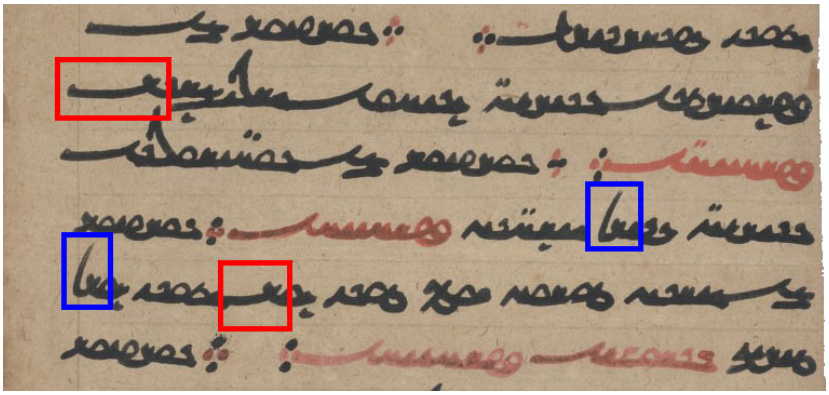


Excerpt from Mainz 119 showing a hand-written form of ـز *zayin*.

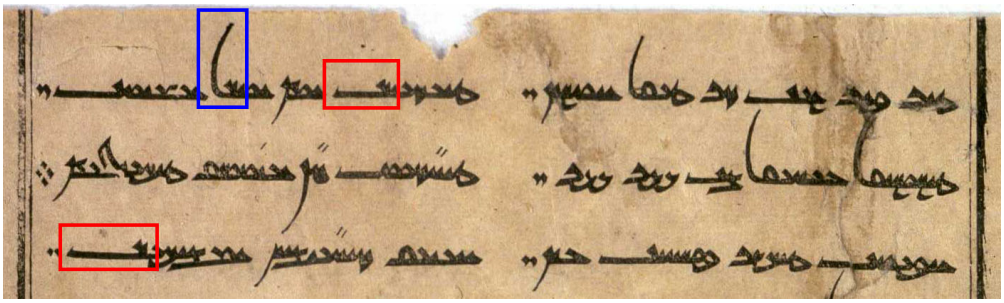


Excerpt from U 387 showing a block-print form of ـز *zayin*.

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



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



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

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

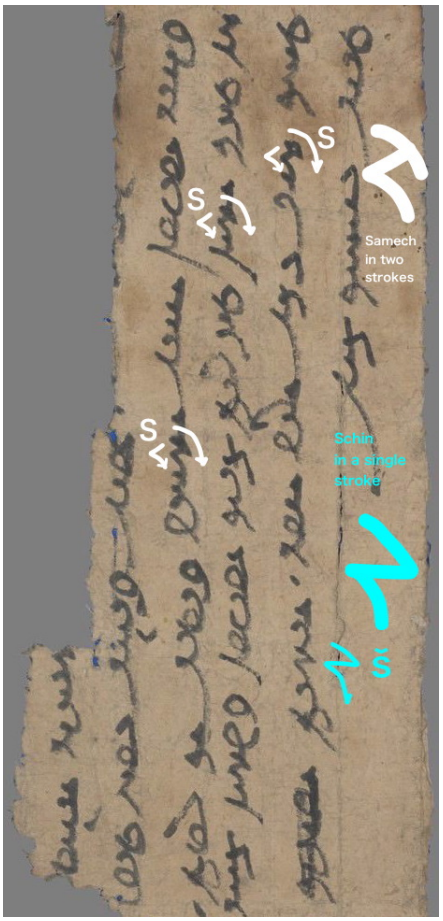


Figure 24: Excerpt from Pelliot Ouïgour 5 (9th–10th c.) showing the distinction between 𐽎 SAMEKH and 𐽏 SHIN. Annotations produced by Dai Matsui, August 2018.

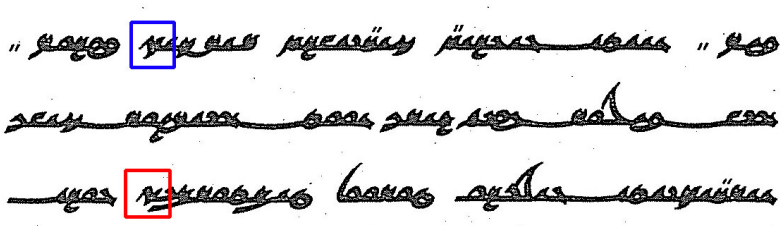
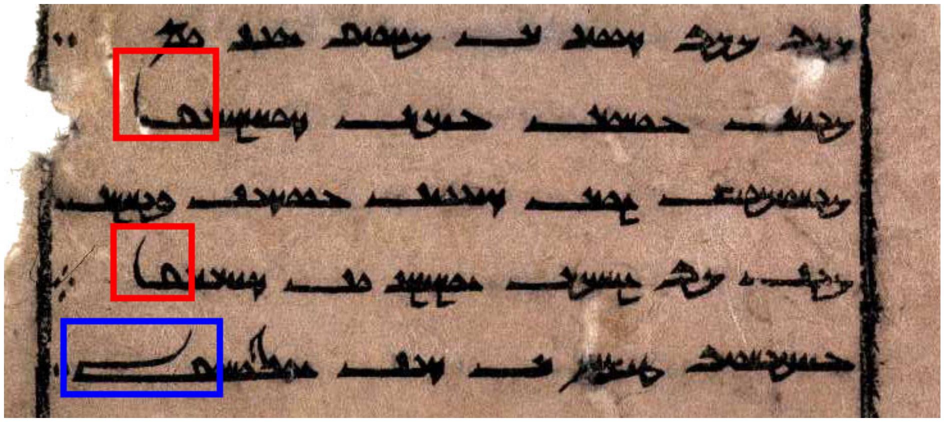
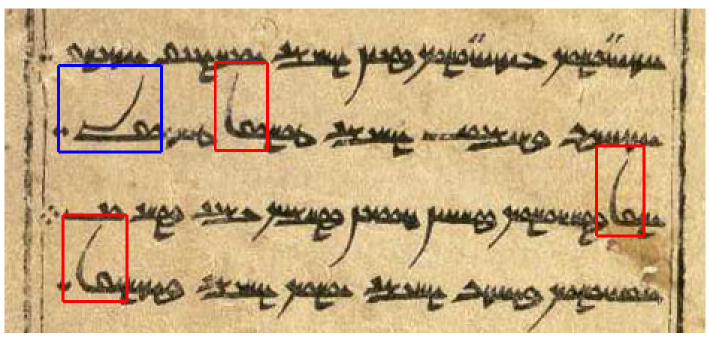


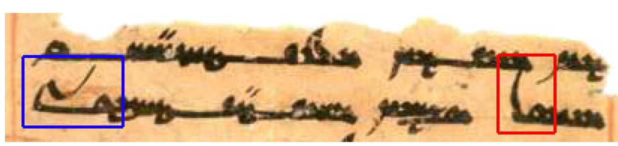
Figure 25: Distinctive final forms of 𐽆 SAMEKH and 𐽈 SHIN (from Müller 1908: 44



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

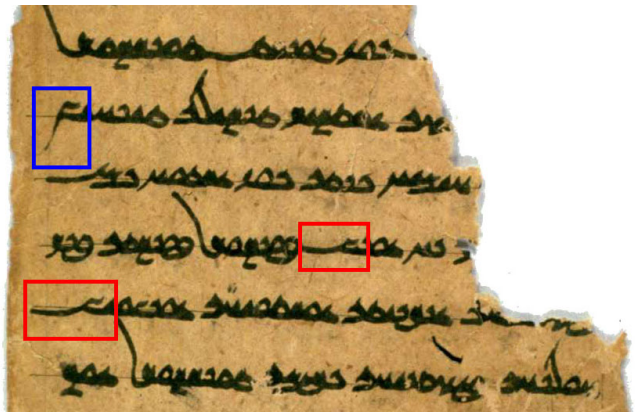


Regular (red) and ornamental (blue) forms of final *pe* (excerpt from U 4162)

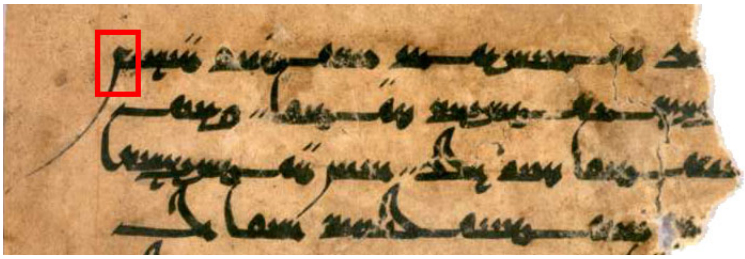


Regular (red) and ornamental (blue) forms of final *pe* (excerpt from Mainz 34)

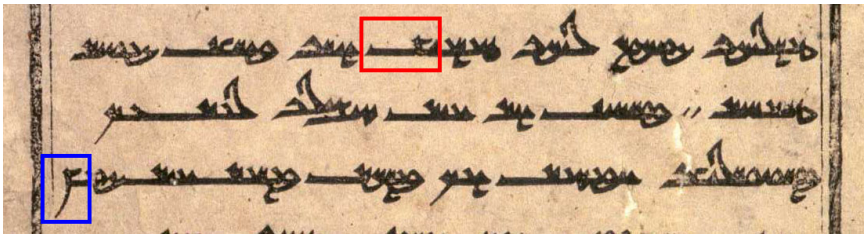
Figure 26: Examples of *pe*. Images have been rotated 90 degrees counter-clockwise for layout purposes.



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

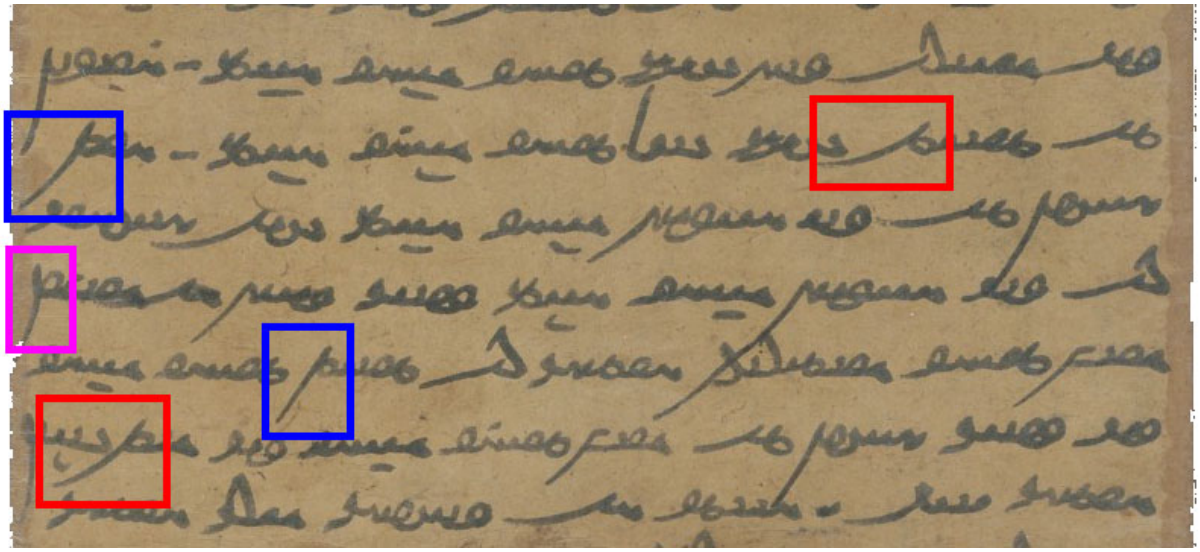


Usage of the alternate 𐽆 *sadhe* in a manuscript (excerpt from Mainz 393)



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

Figure 27: Examples of final *sadhe*. Images have been rotated 90 degrees counter-clockwise for layout purposes.



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

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

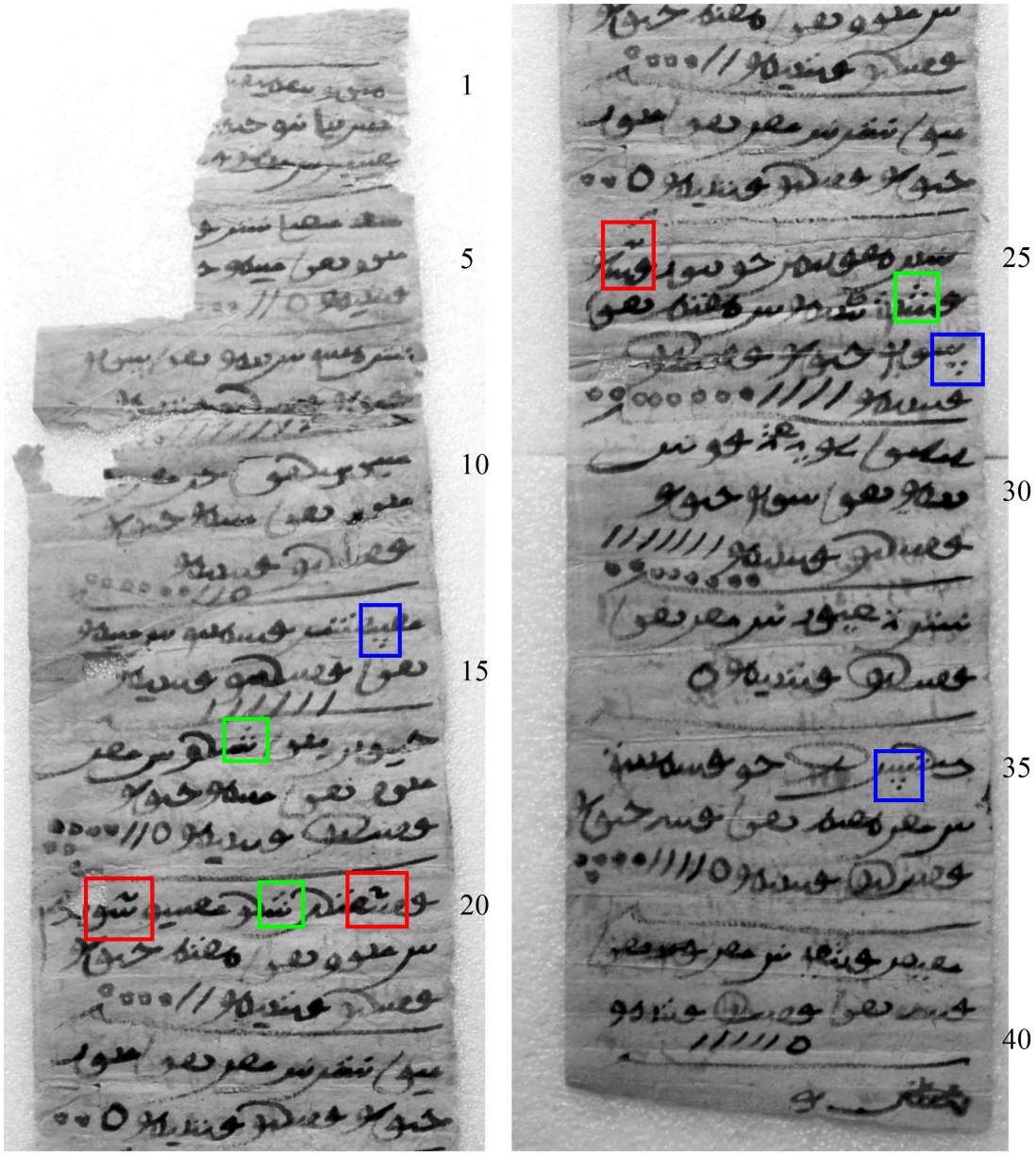
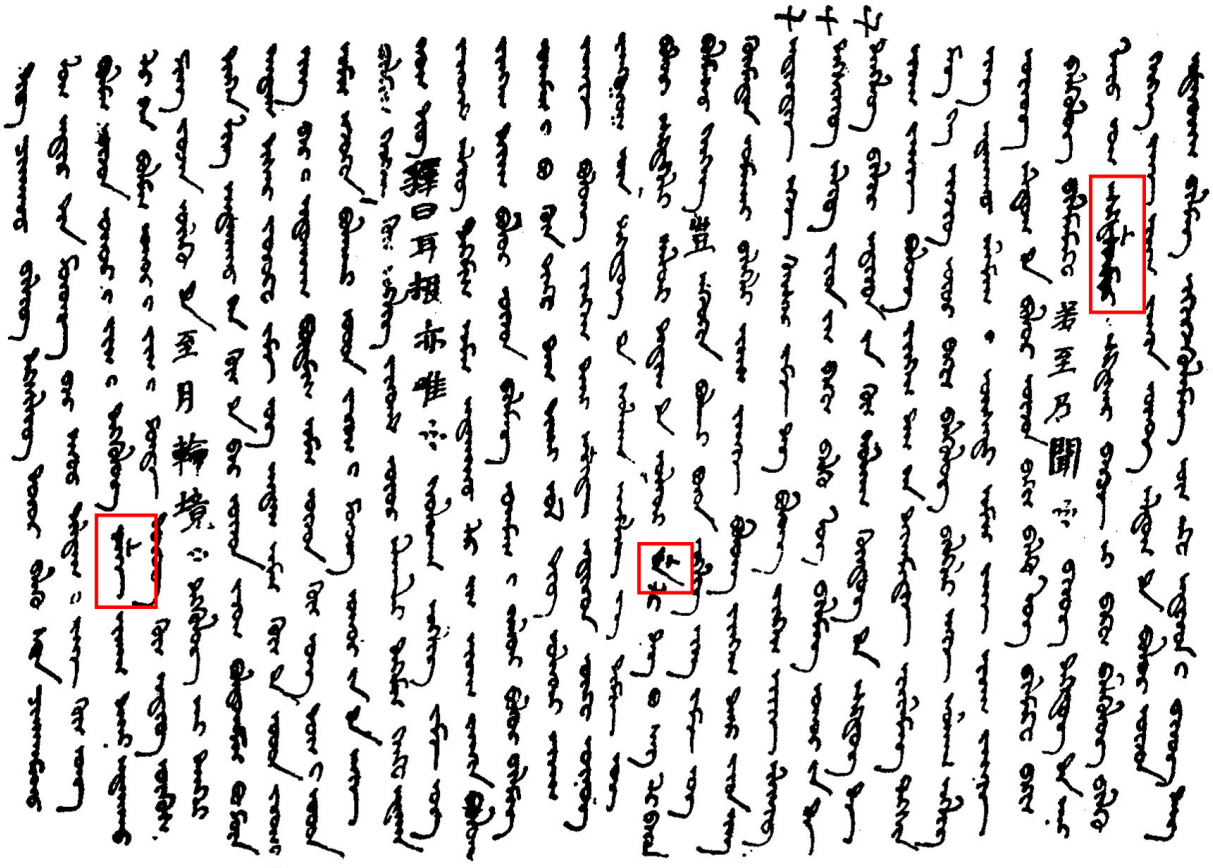
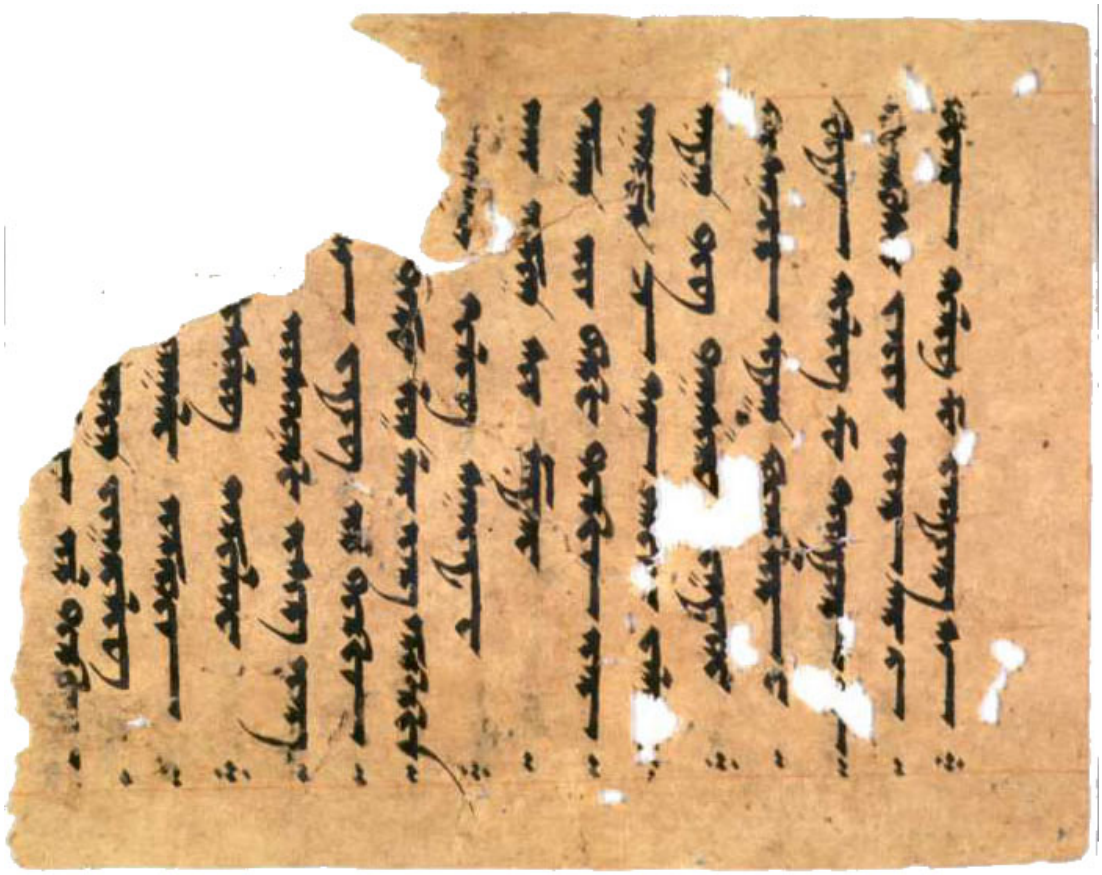


Figure 29: Usage of ◌ (blue), ◌ٔ (green), and ◌ٕ (red) for transcribing Arabic in a Old Uyghur administrative document (from Israpil 2014: plate I).



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

Figure 30: Usage of the ◌꡷ deletion mark for indicating error correction in Or. 8212/75, an Old Uyghur manuscript containing passages of the of the Buddhist text *Abhidharma-nyāyānusāra-śāstra* (from Shōgaito 1988: 207). Note the intralinear text in Han characters.



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

Figure 31: Examples of punctuation signs

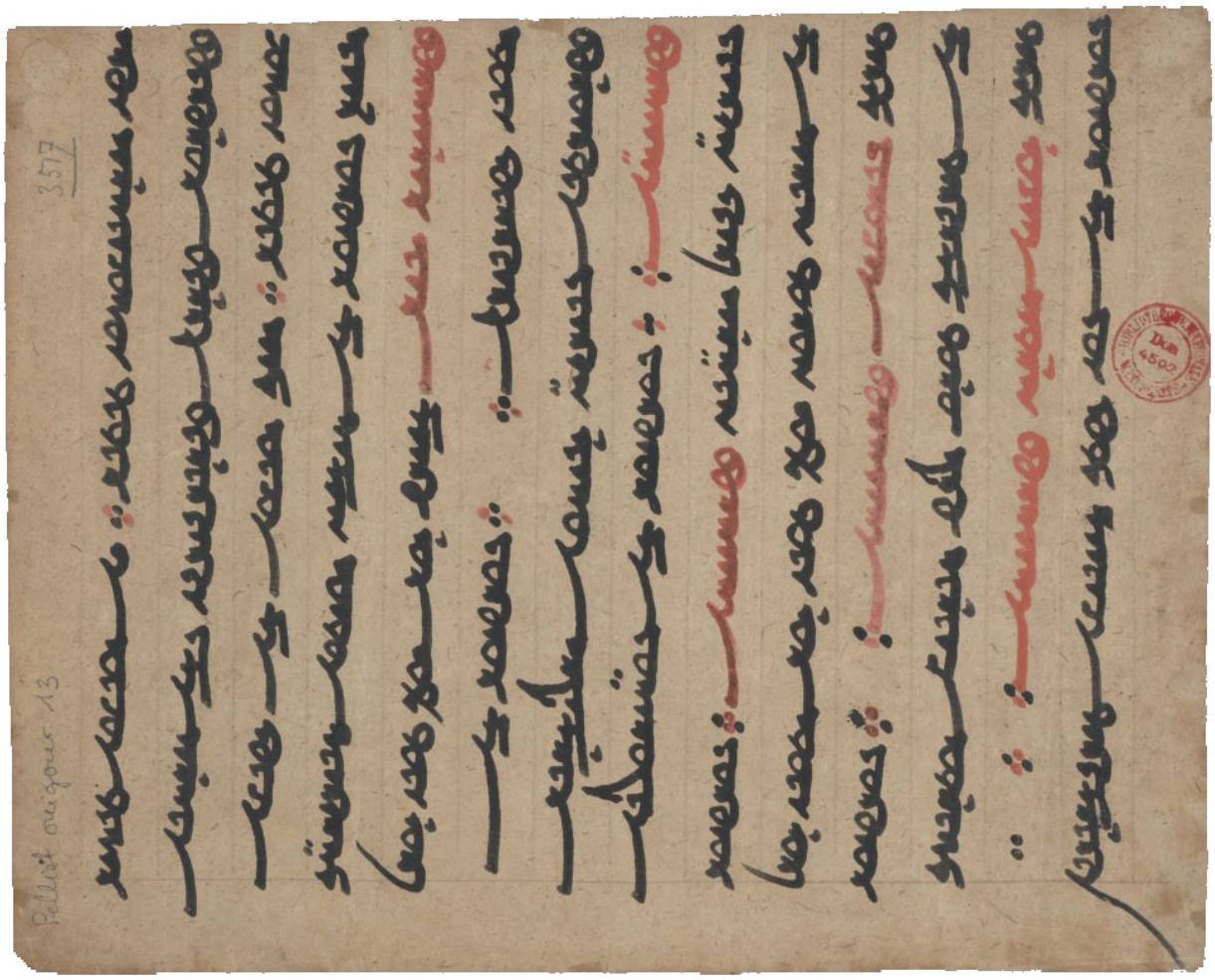


Figure 32: Pelliot Ouïgour 13.

Figure 33: Mainz 126.

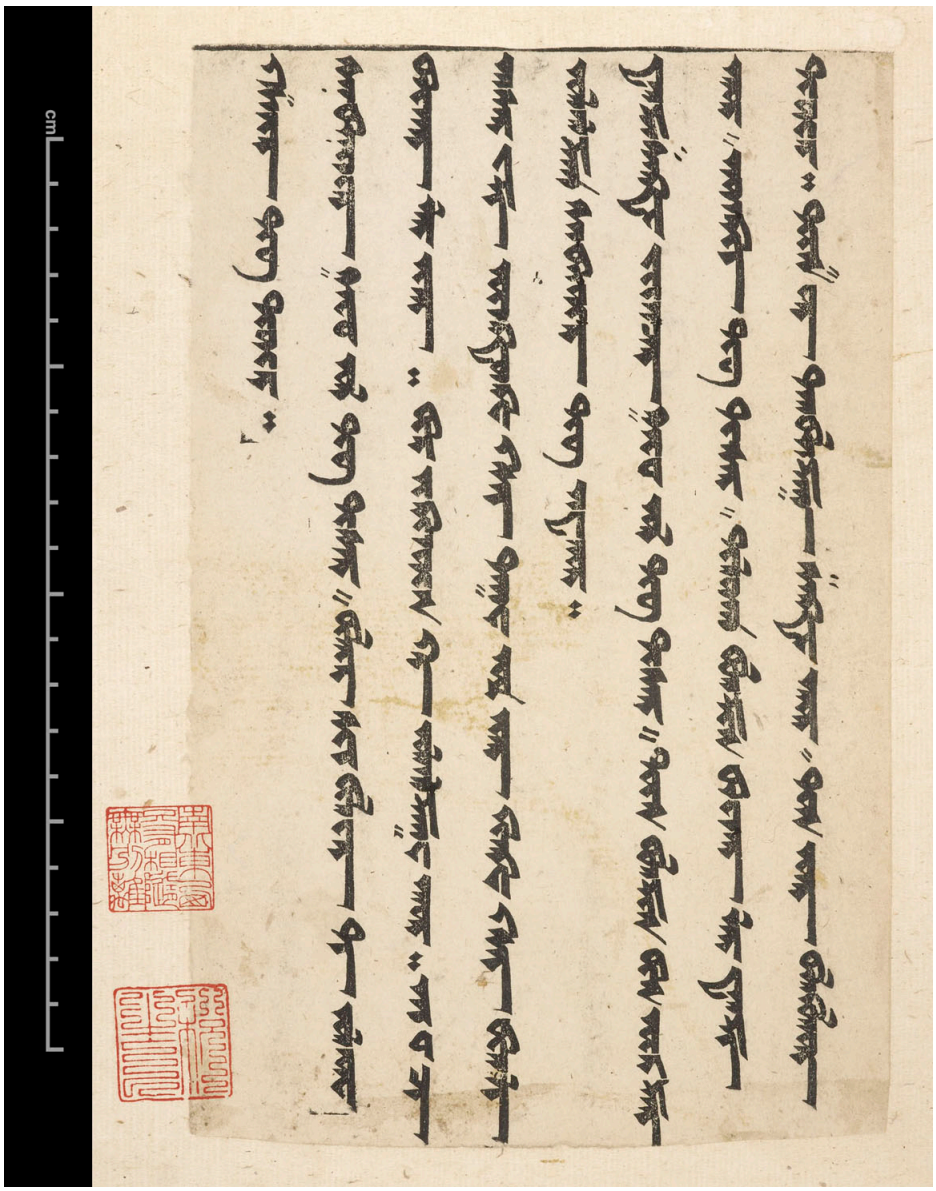


Figure 34: Princeton East Asian Library, PEALD 6a, recto. Block print.

Figure 35: BBAW, U 387 & U 388, recto. Block print.

Figure 36: BBAW, U 4960, folio 1, recto. Block print. Seal in Han characters.

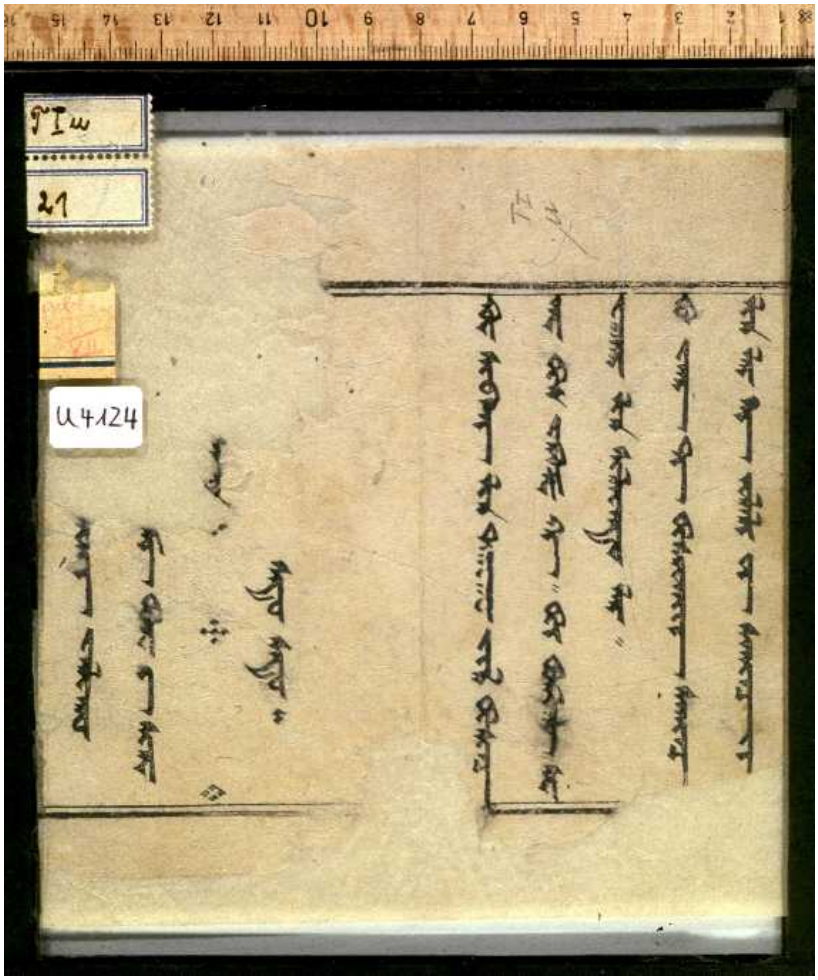


Figure 37: BBAW, U 4124. Block print.

Figure 38: BBAW, U 343, folio 1, recto. Block print.

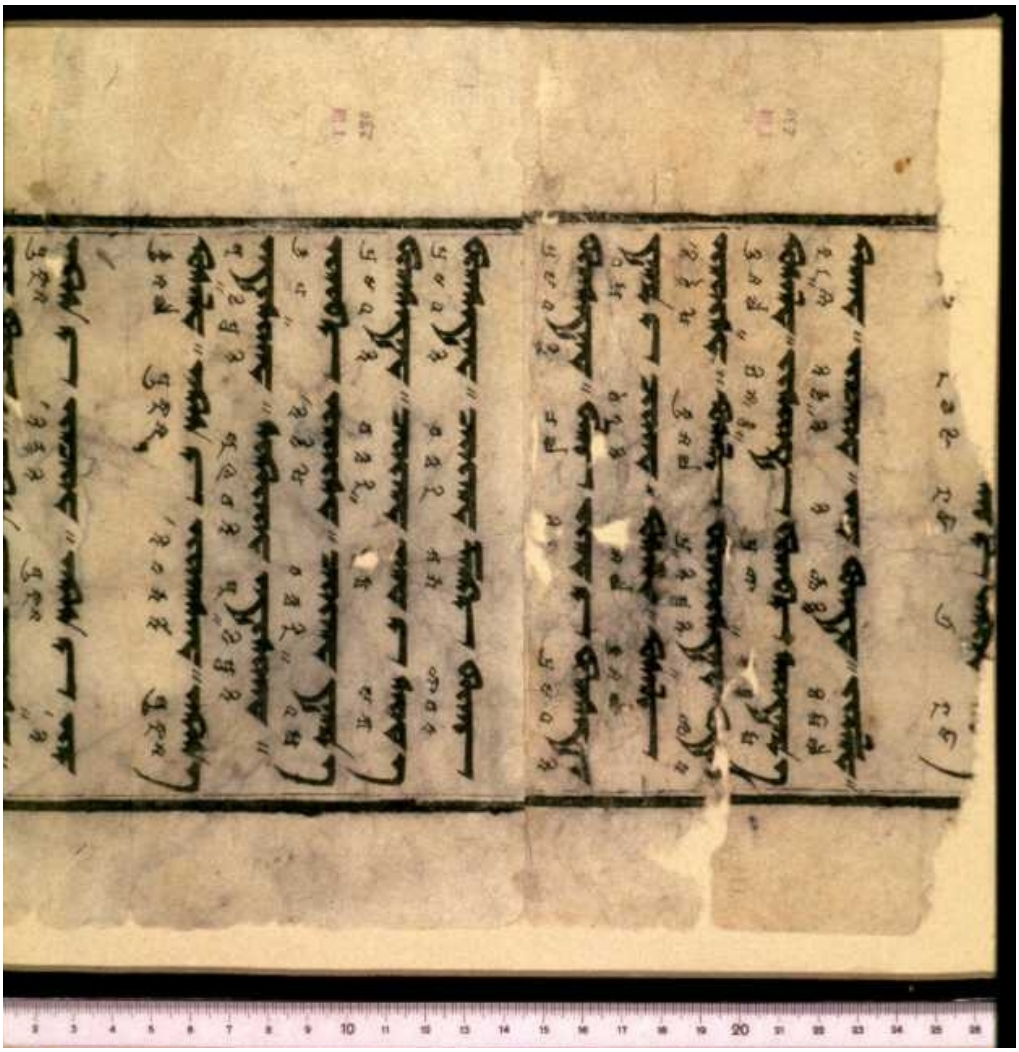


Figure 39: BBAW, Mainz 801, middle portion. Block print. Annotations in Central Asian Brahmi.

Figure 40: BBAW, U 7008. Block print.

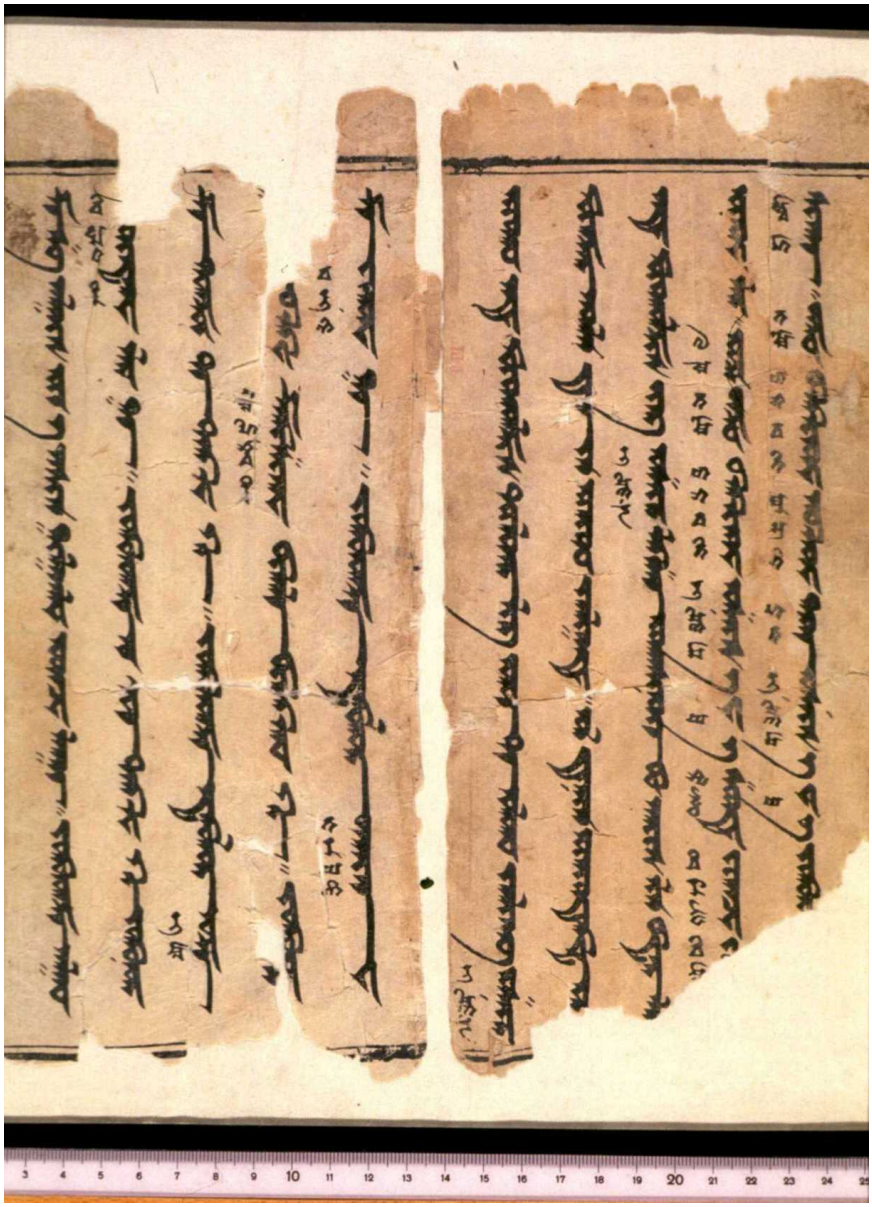


Figure 41: BBAW, Mainz 764, middle. Formal script. Annotations in Central Asian Brahmi.

Figure 42: BBAW, U 3832, folio 1. Formal script.

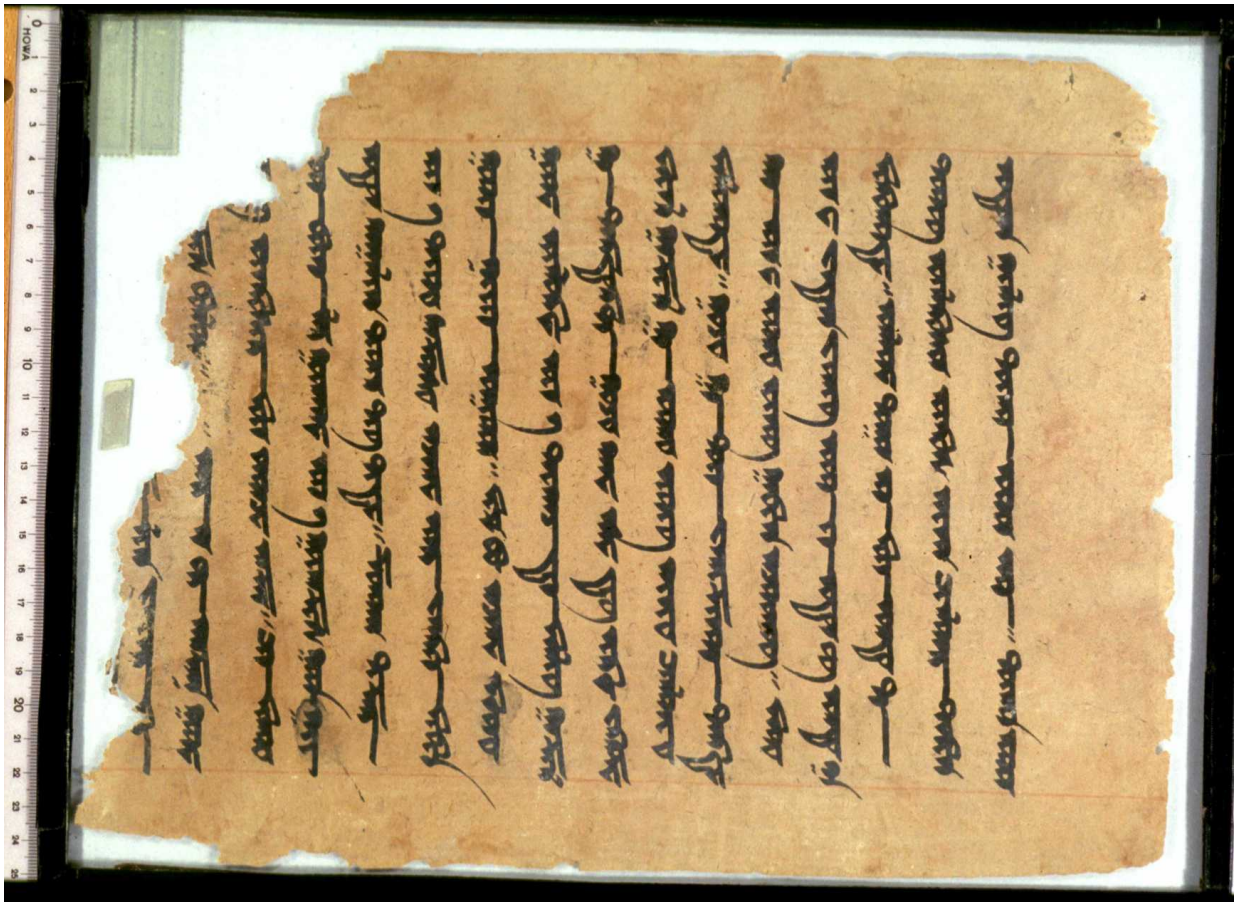


Figure 43: BBAW, Mainz 841, folio 2.

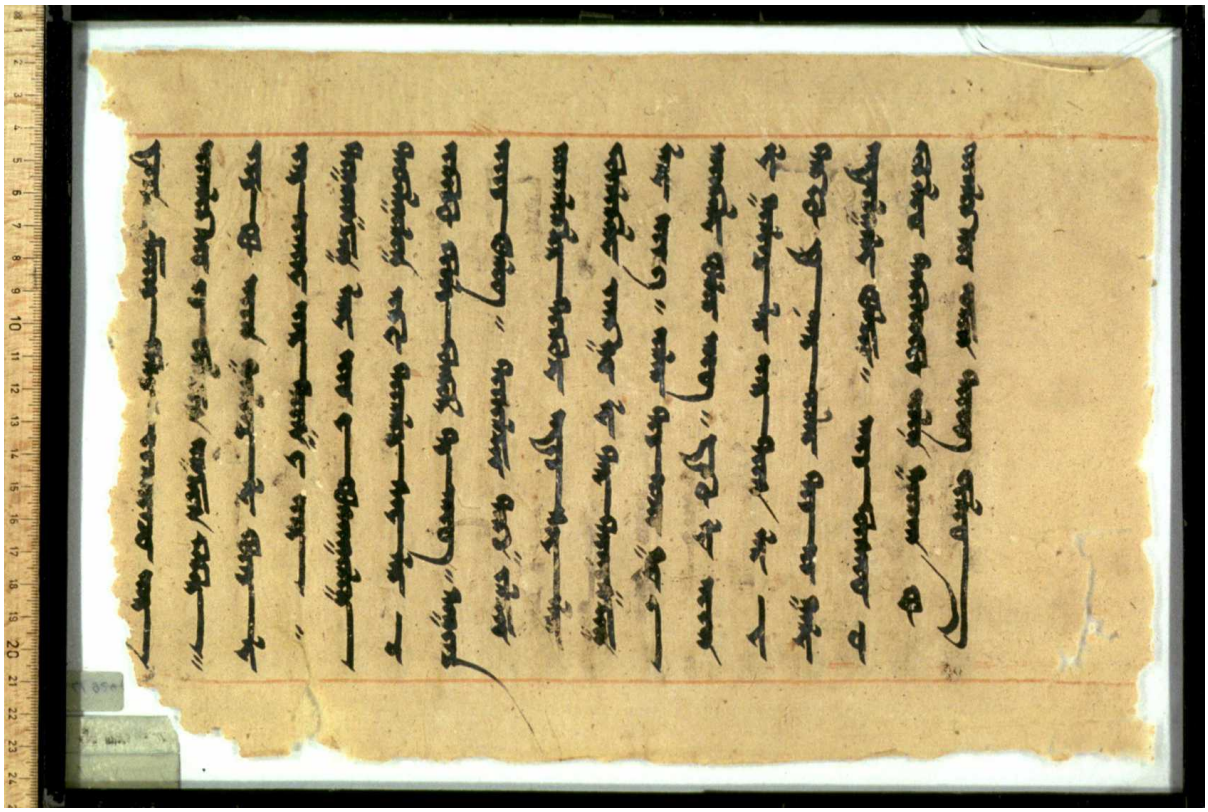


Figure 44: BBAW, U 924, folio 2.

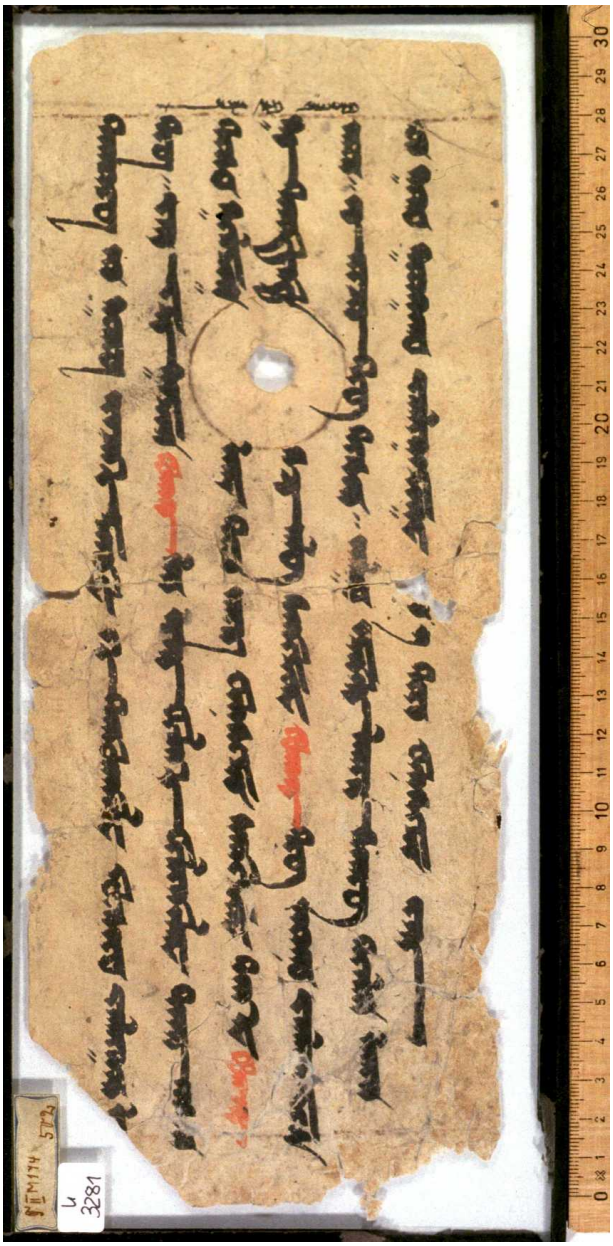


Figure 45: BBAW, U 3281, folio 1.

Figure 46: BBAW, U 456, folio 1.

Figure 47: 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.

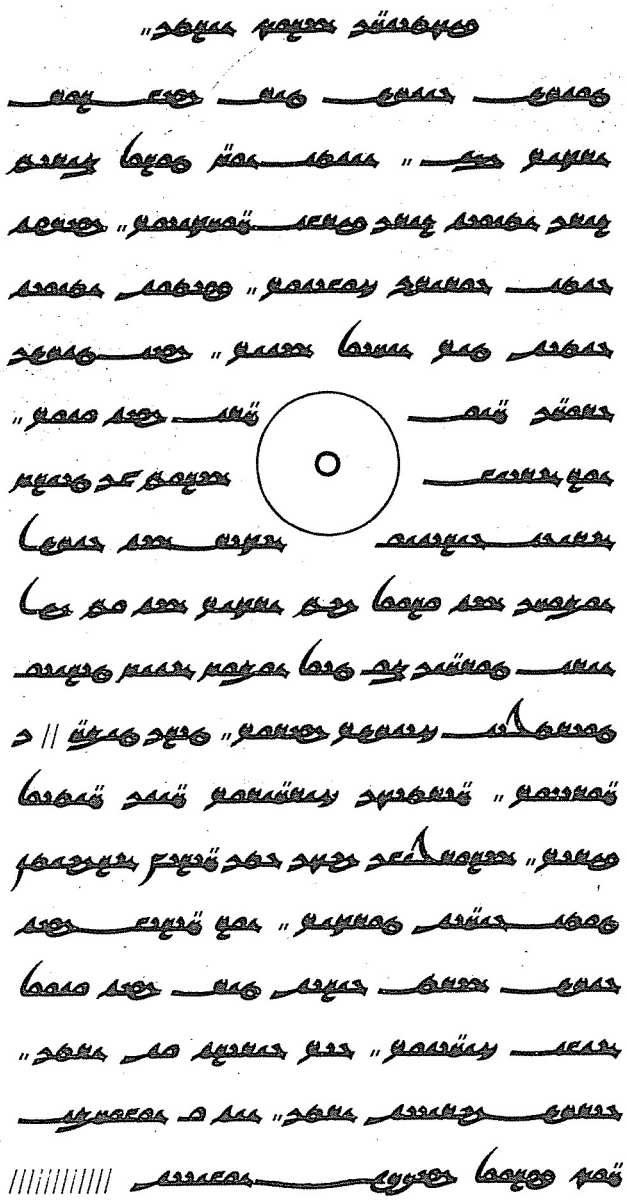


Figure 48: Excerpt from a printed edition of *Altun Yaruq* in the Old Uyghur script (from Müller 1908: 36).

Figure 49: Excerpt from a printed edition of *Ṭišastvustik* in the Old Uyghur script (from Radloff 1910: 3).

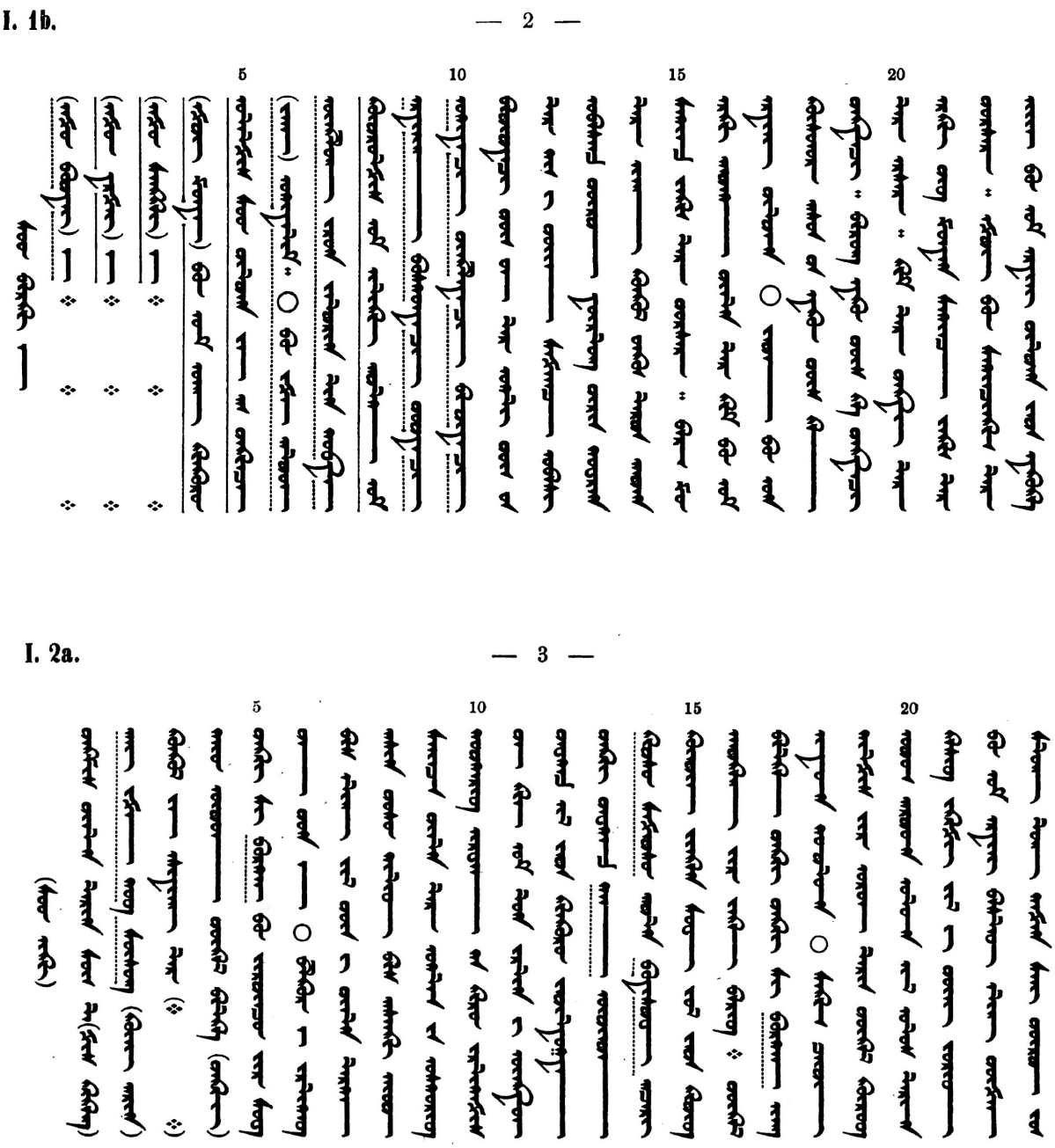
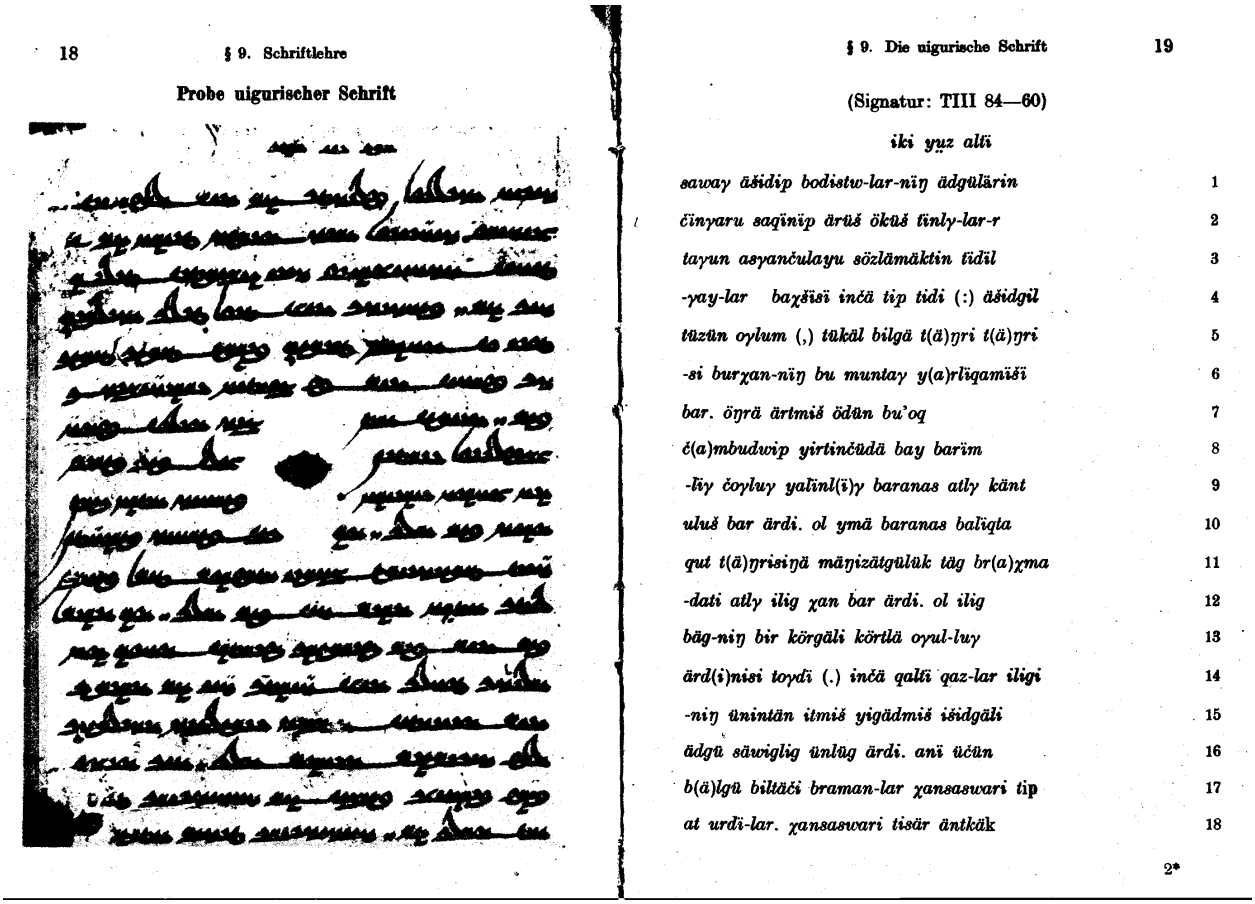


Figure 50: Printed edition of *Suvarṇaprabhāsa*, a Mahayana Buddhist text, in the Old Uyghur script (from Radlov and Malov 1913: 2–3).



18 § 9. Schriftlehre

Probe uigurischer Schrift

§ 9. Die uigurische Schrift 19

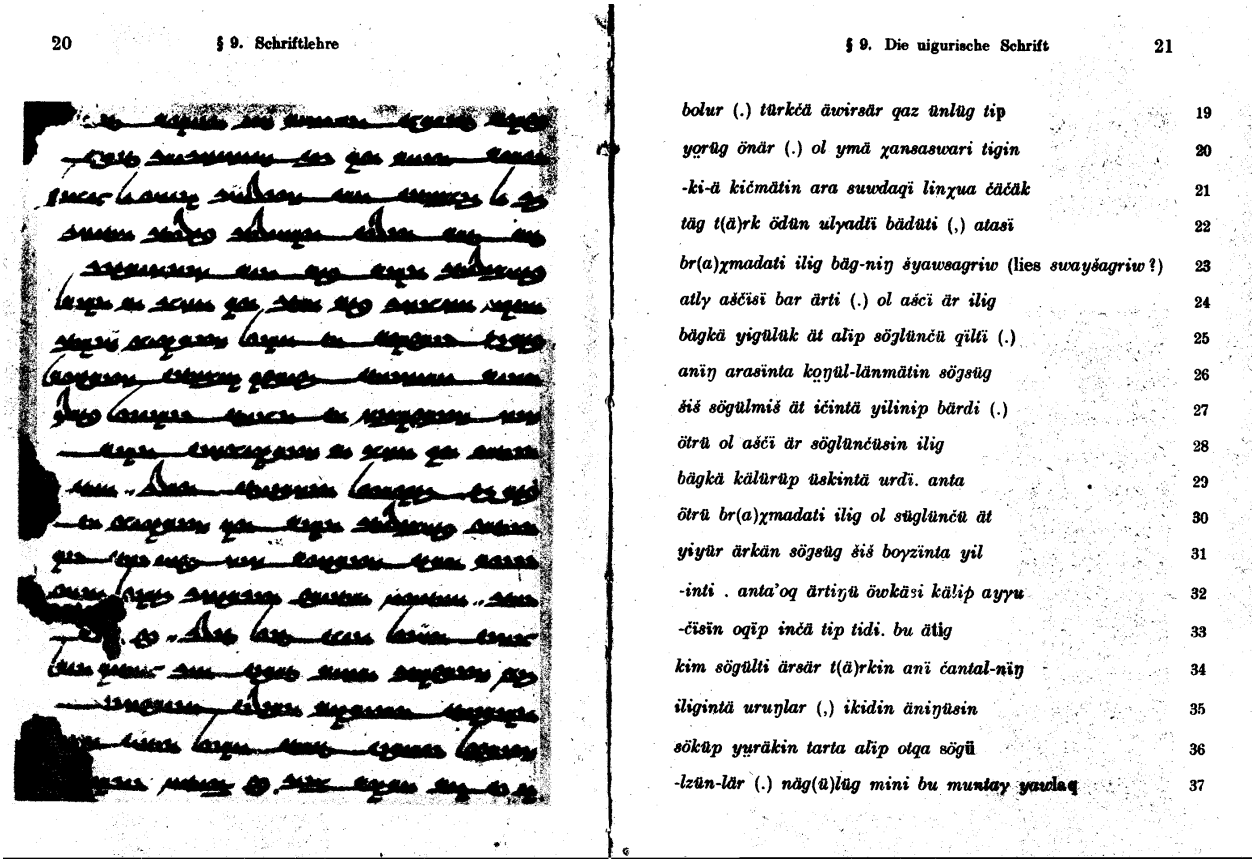
(Signatur: TIII 84—60)

iki yuz altï

sawaγ ädšidip bodistw-lar-nïŋ ädgülärin	1
čïnγaru saqïnïp ärüš öküš tïnlγ-lar-r	2
tayun asγančulayu sözlämäktin tïdïl	3
-γay-lar baχšïsï inčä tip tidi (:) ädšidgil	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 51: Transcription of an Old Uyghur manuscript (from von Gabain 1950: 18–19). Continued in fig. 52.

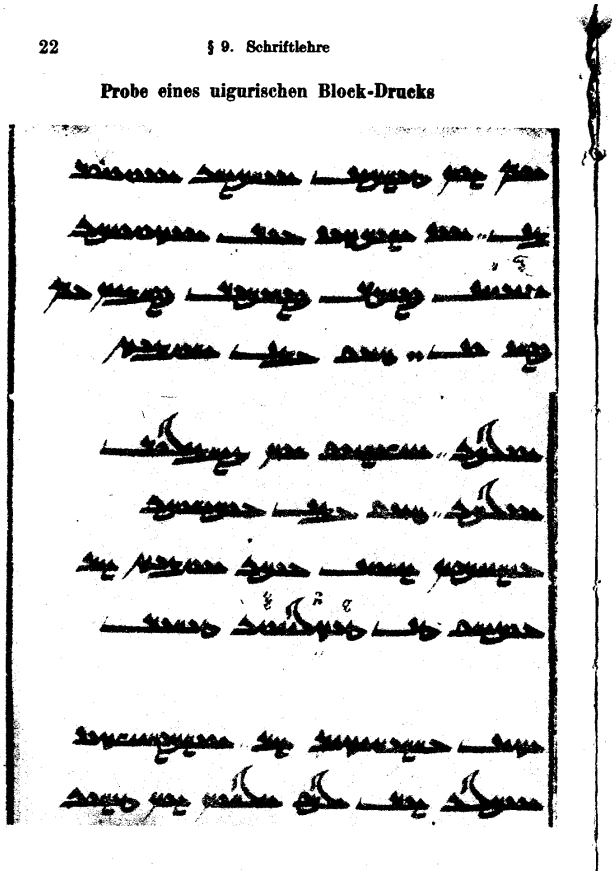


20 § 9. Schriftlehre

§ 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 52: Transcription of an Old Uyghur manuscript in a grammar of Old Turkic (from von Gabain 1950: 20–21). Continued from fig. 51.



22 § 9. Schriftlehre

Probe eines uigurischen Block-Drucks

§ 9. Die uigurische Schrift 23

(Signatur: TIIIM 195)

nom-luγ tilgän äwirgäli ötünür 1

m(ä)n. üzäliksiz yig üstünki 2

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

bolzun. qayu ymä ärtmiš 4

ödki. ančulayu oq kälmädük 5

ödki. qayu ymä yuγärüki 6

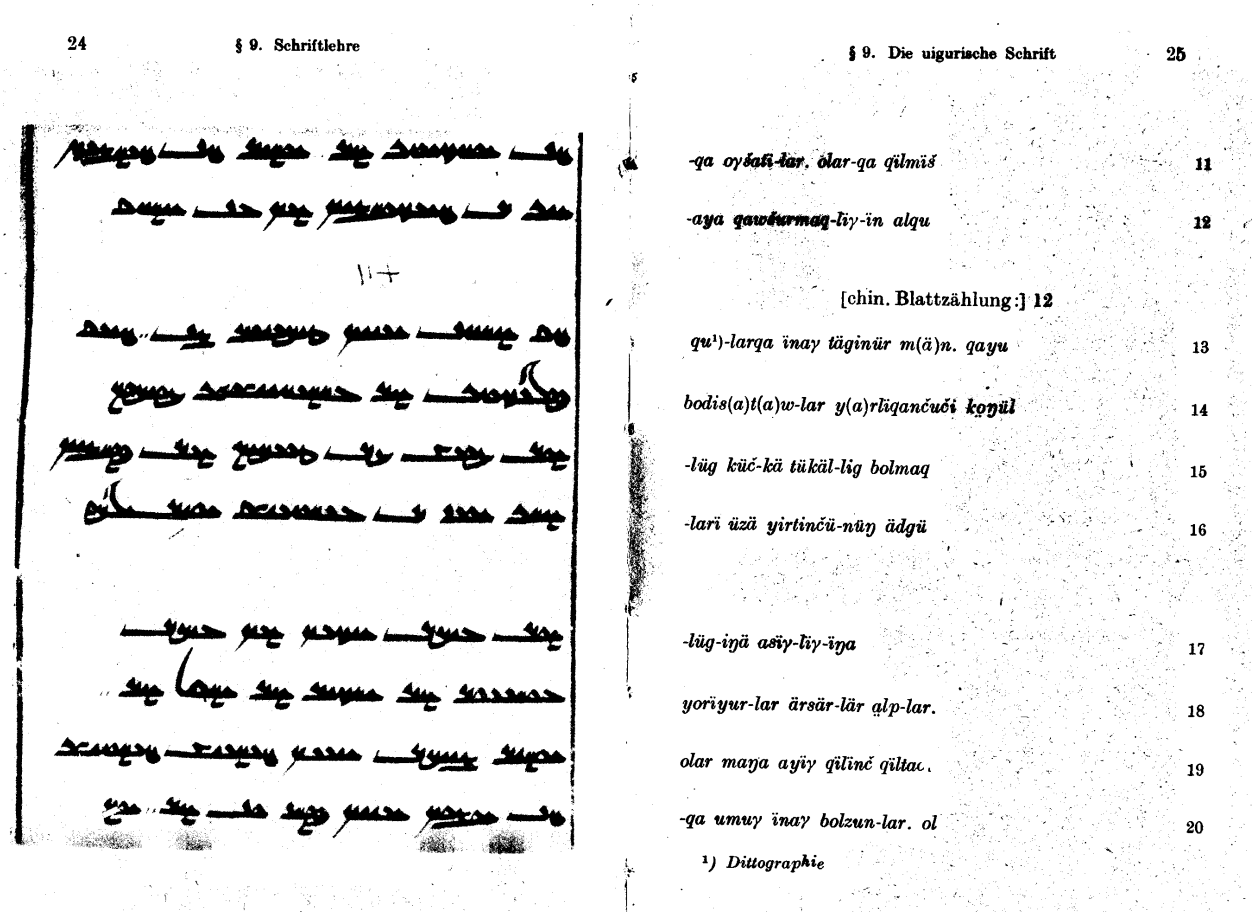
yalaγuq-larta yigi utmiš-lar 7

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

äsän y(a)rlïqasar-lar. ülgüläńčsiz 9

ögdi-lig ädgü adruq-luγ taluγ 10

Figure 53: Transcription of an Old Uyghur manuscript in a grammar of Old Turkic (from von Gabain 1950: 22–23). Continued in fig. 54.



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

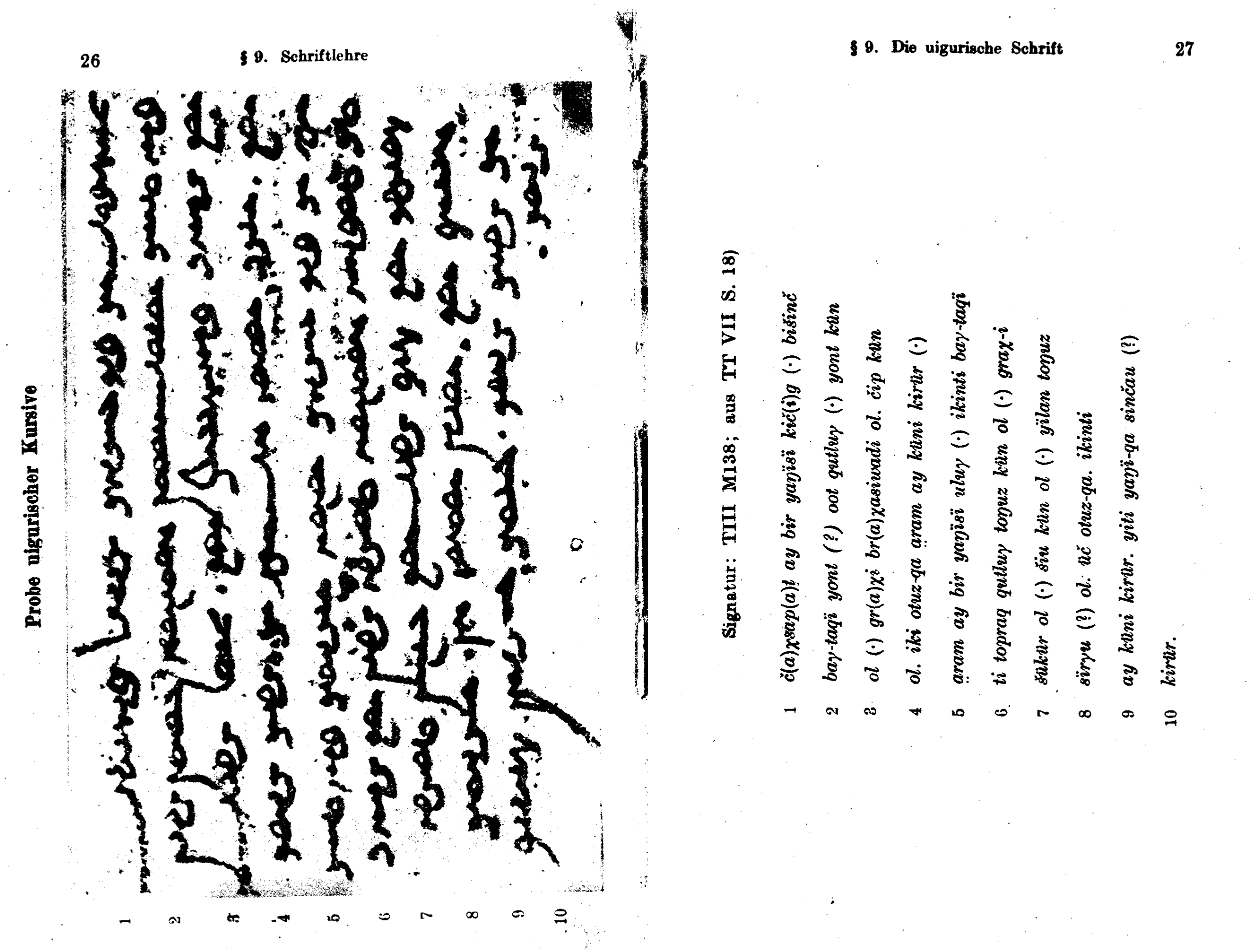
yoriyur-lar ärsär-lär alp-lar. 18

olar maŋa ayïγ qïlïnč qïltac. 19

-qa umuγ ïnaγ bolzun-lar. ol 20

[1]) *Dittographie*

Figure 54: Transcription of an Old Uyghur manuscript in a grammar of Old Turkic (from von Gabain 1950: 24–25). Continued from fig. 53.



26 § 9. Schriftlehre

Probe uigurischer Kursive

§ 9. Die uigurische Schrift 27

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

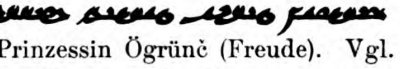
1 *č(a)χsap(a)t ay bir yaŋïsï kič(i)g (·) bišinč*
2 *bay-taqï yont (?) oot qutluγ (·) yont kün*
3 *ol (·) gr(a)χi br(a)χasiwadi ol. čip kün*
4 *ol. iki otuz-qa aram ay küni kirür (·)*
5 *aram ay bir yaŋïsï uluγ (·) ikinti bay-taqï*
6 *ti topraq qutluγ toŋuz kün ol (·) graχ-i*
7 *šükür ol (·) šiu kün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

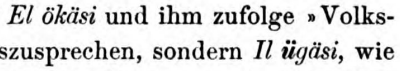
Figure 55: Transcription of an Old Uyghur manuscript in a grammar of Old Turkic (from von Gabain 1950: 26–27).

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 56: 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.

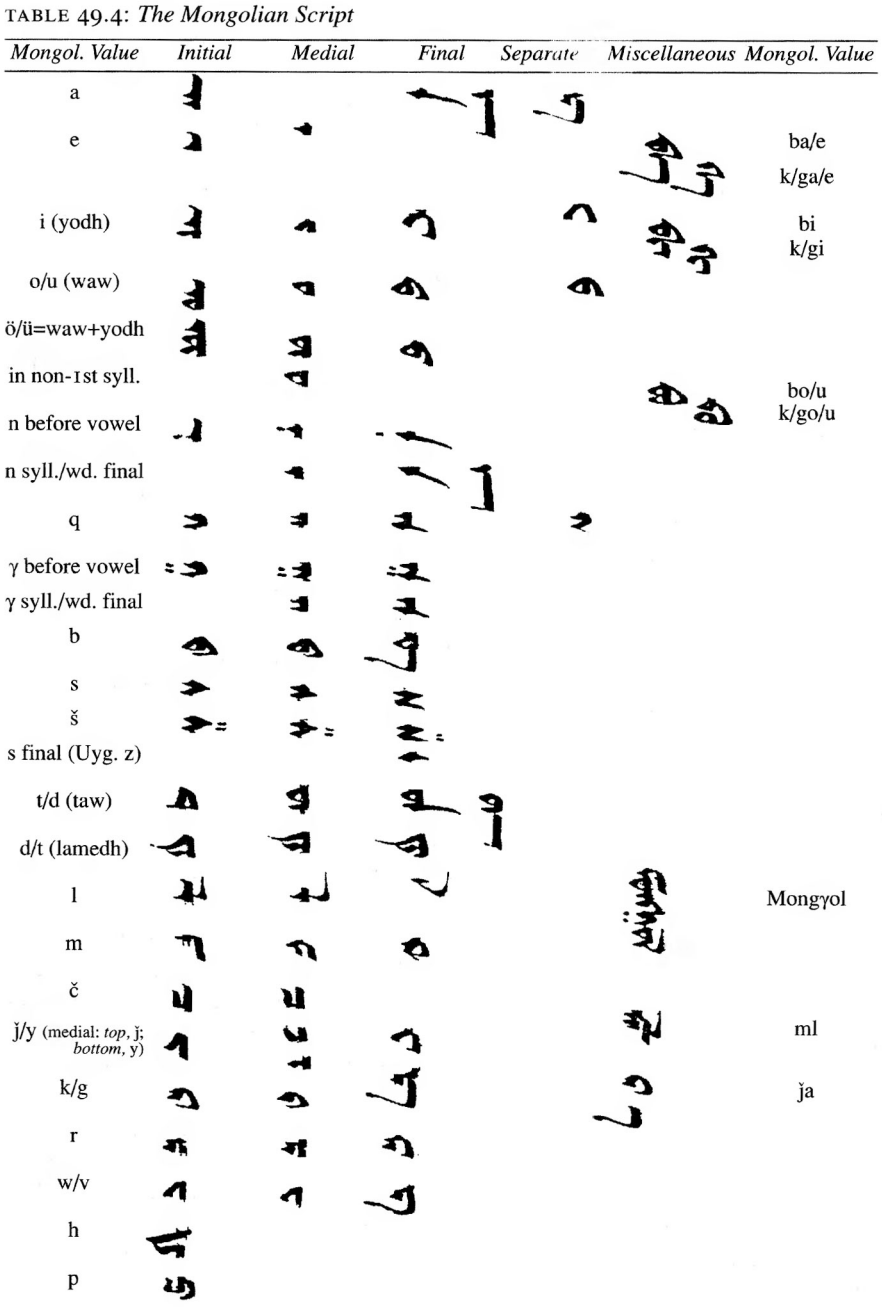
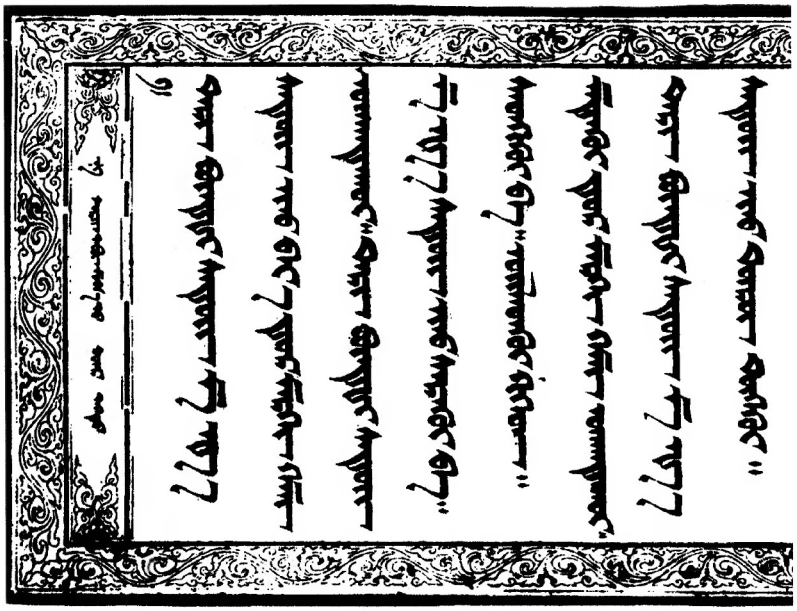


Figure 57: Table showing letters of the Mongolian script (from Kara 1996: 545). See table of Old Uyghur letters from the same source in fig. 6.

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 58: Sample Mongolian text (from Kara 1996: 546). Compare the Mongolian block print with the Old Uyghur block print in fig. 34.

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: