Proposal to encode the Khwarezmian script in Unicode

Anshuman Pandey
pandey@umich.edu

March 26, 2018

1 Introduction

This proposal is a substantial revision and expansion of the following:

- L2/17-054R: "Proposal to encode the Khwarezmian script in Unicode"

It incorporates comments provided by the UTC Script Ad Hoc Committee in:

- L2/17-255: Recommendations to UTC #152 July-August 2017 on Script Proposals
- L2/18-039: Recommendations to UTC #154 January 2018 on Script Proposals

Major changes include:

- Definition of the script as a cursive joining *abjad*
- Details on joining properties, shaping behaviors, and usage of ZERO WIDTH JOINER
- Addition of a detached form of *aleph* and a vocalization sign for *waw* and *yodh*
- Encoded representations of examples words and numbers
- Additional specimens showing usage of the script

2 Background

The proposed script was used between the 2nd and 9th centuries CE for writing Khwarezmian (ISO 639-3: xco), a now-extinct Eastern Iranian language that was spoken in the Oxus (Amu Darya) river delta. This region in Central Asia is known in the *Avesta* as ٥ﻝﺍ٤ﻤ *hvâirizem* (*Yašt* 10.5.14). The Achaemenids referred to it as ⟨ᚎ·ᛁᛉ·ᛏ ᛁᛏᛁ ᛁᛁ ᛁᛁ ᚅ *uvârazmiš* (XPh inscription, lines 21–22). It is called خوارزم *xvārazm* in classical Persian, which is transcribed as 'Khwarezm' in English. The Greeks knew it as *Χορασμία*, and this hellenic form entered the English lexicon as 'Chorasmia'. The territory lies across portions of present-day Uzbekistan, Kazakhstan, and Turkmenistan.

Both 'Khwarezm' and 'Chorasmia' are used in modern English, along with the adjectives 'Khwarezmian' and 'Chorasmian', and the variant spellings 'Khwarazmian' and 'Choresmian'. Both adjectival forms are used for the language and script in English scholarship: 'Khwarezmian' by Durkin-Meisterernst (2009),

Federov (2005, 2006), Livshits (1964, 2003), Skjærvø (1996); and 'Chorasmian' by Humbach (1998), MacKenzie (1991), with the variant 'Choresmian' used by Azarpay (1969), Henning (1965). The spellings 'Khwarezm' and 'Khwarezmian' align with 'Хорезм' and 'Хорезмийский', names used by Russian scholars who have conducted extensive archaeological and palaeographical studies of the region, language, and script. Therefore, 'Khwarezmian' has been selected as the identifier for the script in Unicode as it represents nomenclature familiar to scholars around the globe.

The 'Khwarezmian' is a distinctive script derived from Imperial Aramaic. It is the second of three that were used for recording the Khwarezmian language: 1) the Imperial Aramaic script; 2) the indigenous script discussed here; and 3) the Arabic script. The indigenous script is classified by scholars into 'archaic', 'lapidary', and 'cursive' types (Lurje, personal communication, December 2017):

- The 'archaic' occurs, for instance, on silver bowls no. 1 and no. 2 from Isakovka (Исаковка), see fig. 33 here. These inscriptions are dated to the Achaemenid period and appear in a script closely related to Imperial Aramaic. They are the earliest attestations of the Khwarezmian written language (Livshits 2003: 147–148). This type is a non-joining *abjad* (see fig. 35).

- The 'lapidary' is represented, for example, on a flask found in 2005 at Chirik-rabat (Чирик-рабата), described in Ivantchik and Lurje (2013: 286), see fig. 34 here. Similar types occur on ostraca from Koy-krylgan Kala (Кой-крылган-калы). This type is also a non-joining *abjad* (see fig. 35).

The 'cursive' Khwarezmian script differs substantially from the above two types. Graphically, its letters developed features consistent with cursive scripts, such as Arabic and Sogdian, ie. positional forms of letters. Structurally, the cursive connections between letters were governed by rules intended for maintaining distinctions between letters. This type may be considered the 'normative' or 'national' Khwarezmian script. It is attested on at least the following materials:

- Coinage with Khwarezmian legends, which are the earliest attested records in the script, from the 2nd century CE onwards (see Vainberg 1977, Federov 2005). The coins have been classified by Vainberg and are referred to using the Cyrillic and Roman numeral designations БII–ΓVI. Facsimilies of these coins are shown in fig. 2–6 and tracings of inscriptions are shown in fig. 7–13. Some coins are bilingual: class Б coins have inscriptions in Khwarezmian and Greek, and class ΓV have transcriptions of Khwarezmian text in the Sogdian script.

- Inscriptions on wooden items and leather from the palace at Toprak Kala (Топрак-кала), dated to the 3rd century CE.

- Leather inscriptions from a fort at Yakke Parsan (Якке парсан) dated to the 8th century CE (see fig. 14).

- Inscriptions on silver vessels dated between the 6th and 8th centuries CE. Reproductions of nearly all vessels and their inscriptions were published in Smirnov (1909) and republished in Azarpay (1969). Seven bowls and one pitcher, along with their inscriptions, are shown in fig. 16–23.

- Ossuary inscriptions at Tok Kala (Ток-кала), from the 7th and 8th centuries CE. The script of these records represent a development of the style used in the Toprak Kala documents. There are around 100 of these inscriptions, of which nine were initially deciphered by Tolstov and Livshits (1964), shown here in fig. 24–32. Additional inscriptions were deciphered by Lurje (2013).

The Khwarezmian script of these sources is related to other Iranian scripts derived from Imperial Aramaic, such as Inscriptional Parthian; Inscriptional, Psalter, and Book Pahlavi; the Old Sogdian of the 'Ancient Letters' and the later Sogdian 'formal' and 'cursive' scripts (see table 1). However, among these, Khwarezmian was more conservative in its retention of older features of letterforms and it underwent considerably less change than its sister scripts (Tolstov and Livshits 1964: 234).

After continuous usage over 800 years, the Khwarezmian script was replaced by the Arabic script by the turn of the 10th century. The change of orthography for the Khwarezmian language was soon after compounded by a larger linguistic change. By the 14th century, the native language was replaced by Turkic languages. Some insight into the demise of the script, among other aspects of Khwarezmian culture and society, is provided by the medieval scholar named Abū Rayḥān Muḥammad ibn Aḥmad Al-Bīrūnī. Known more commonly as 'Al-Biruni', this native of Khwarezm is considered one of the greatest scholars of the medieval Islamic period. In his *Al-Āthār al-bāqiya ʾan al-qurūn al-ḫāliya* (*The Remaining Signs of Past Centuries*), which was completed in 1000 CE, Al-Biruni wrote:

> When Ḵutaiba ben Muslim had conquered Khwârizm a second time ... [he] had extinguished and ruined in every possible way all those who knew how to write and to read the Khwârizmî writing, who knew the history of the country and who studied their sciences. In consequence these things are involved in so much obscurity, that it is impossible to obtain an accurate knowledge of the history of the country since the time of Islam (not to speak of pre-Muhammadan times).   (Sachau 1879: 41–42)

3   Script Details

Structure   The script is a cursive joining *abjad*. It is written from right to left, with lines that advance from top to bottom. All letters are right-joining, but only some are also left-joining. The joining properties of letters are fairly regular across the sources, with some exceptions (see § 4.1.1).

Repertoire   Of the 22 letters of the Aramaic alphabet, 19 are attested collectively across the relevant sources. Analogues for *teth*, *qoph*, and *sadhe* do not exist. Silver vessel and ossuary inscriptions contain all 19 letters, as well as numerical signs. Coins have a smaller subset of letters, and no numerical signs. The ossuary texts contain additional characters for marking grammatical features. A comparison of the repertoire and letterforms made by Vainberg (1977, plate VIII) is reproduced here in fig. 1. A list of signs used on the silver vessel inscriptions has been produced by Lurje (forthcoming), see fig. 15 here.

Letterforms   A peculiarity of Khwarezmian is that several letters resemble each other in the latest stage of the script. While the nominal form of a letter is distinctive, its contextual form may be similar or identical to the contextual form of another. See § 4.1 for details.

4   Encoding Model

Scope   The encoding is based upon the cursive Khwarezmian script as attested by inscriptions on coinage, silver vessels, and ossuaries. This script differs from the archaic and lapidary type described above, which are non-joining scripts that may be unified with the Imperial Aramaic encoding.

Unification   It is practical to consider the varieties as developmental phases of a distinctive 'Khwarezmian' writing system. For purposes of character encoding the varieties should be unified as a single 'Khwarezmian'

script in Unicode. This approach enables texts to be represented using the same underlying character set, using normative glyphs. The display of the script style of particular records would be managed by custom fonts.

Encoded repertoire   The encoded character repertoire contains 30 characters: 20 letters, 1 vocalization sign, and 7 numbers. The code chart and names list follows p. 8. Naturally, the encoded set may differ from traditional and scholarly inventories of script varieties that occur in written and inscriptional sources. Such differences naturally arise from the requirements for digitally representing a script in plain text and for preserving the semantics of characters.

Representative glyphs   The representative glyphs are based upon the letterforms in the Tok Kala ossuaries. The style used in these sources are the latest development of the script and reflect its distinctiveness.

Character names   Traditional names for Khwarezmian letters are not attested. Therefore, Unicode character names are based upon those of 'Imperial Aramaic' characters. This convention has been followed for Unicode encodings of other Iranian scripts such as 'Inscriptional Parthian', etc. In this document names in italics refer to scholarly names for graphemes while names in small capitals refer to Unicode characters, eg. ⟨ is *aleph* and KHWAREZMIAN LETTER ALEPH. For sake of brevity, the descriptor 'KHWAREZMIAN LETTER' is dropped when refering to Khwarezmian characters, eg. KHWAREZMIAN LETTER ALEPH is referred to as ALEPH. Characters of other scripts are designated by their full Unicode names. Latin transliteration of Khwarezmian follows the current scholarly convention, with Aramaic heterograms given in uppercase letters.

4.1   Letters

The nominal form of each letter is given in the 'X_n' column. The labels 'X_i', 'X_m', 'X_f' refer respectively to the initial, medial, and final forms of letters. The red dash indicates the location on a glyph were connections should occur, while a vertical bar indicates that a connection occurs without an extension.

Character name	X_n	X_f	X_m	X_i	Join	Latin
KHWAREZMIAN LETTER ALEPH	⟨	⟩	(⟨) ⟨	(⟨) ⟨	dual	ʾ
KHWAREZMIAN LETTER DETACHED ALEPH	⟨	—	—	—	non	ʾ
KHWAREZMIAN LETTER BETH	⟩	⟩	⟩	⟩	dual	b
KHWAREZMIAN LETTER GIMEL	⟩	⟩	⟩	⟩	dual	g
KHWAREZMIAN LETTER DALETH	⟩	⟩	⟩	⟩	right	d
KHWAREZMIAN LETTER HE	⟩	⟩	⟩	⟩	right	h
KHWAREZMIAN LETTER WAW	⟩	⟩	⟩	⟩	right	w
KHWAREZMIAN LETTER ZAYIN	⟩	⟩	⟩ , ⟩	⟩ , ⟩	dual*	z
KHWAREZMIAN LETTER HETH	⟩	⟩	⟩	⟩	right	ḥ

KHWAREZMIAN LETTER YODH	ı	ı	ı	ı	right	y
KHWAREZMIAN LETTER KAPH	ㄱ	ァ	ユ	ㄱ	dual	k
KHWAREZMIAN LETTER LAMEDH	⅃	⅃	⅃	⅃	dual	l
KHWAREZMIAN LETTER MEM	ゎ	ゎ	ゎ	ゎ	right	m
KHWAREZMIAN LETTER NUN	⅃	├	⊥	⅃	dual	n
KHWAREZMIAN LETTER SAMEKH	ㅁ	ㅁ	ㅁ	ㅁ	dual	s
KHWAREZMIAN LETTER AYIN	Y	Y	Y	Y	right	ʿ
KHWAREZMIAN LETTER PE	⊃	⅃, ⅃	⅃, ⅃	⅃	dual*	p
KHWAREZMIAN LETTER RESH	Y	Y	Y	Y	right	r
KHWAREZMIAN LETTER SHIN	₥	₥	₥	₥	right	š
KHWAREZMIAN LETTER TAW	ထ	ထ	ထ	ထ	dual	t

4.1.1 Joining behavior

The joining model and shaping requirements for Khwarezmian are similar to that of Arabic. A summary of the joining properties of the letters is given below:

right- & left-joining	*aleph*, *beth*, *gimel*, *zayin*, *kaph*, *lamedh*, *nun*, *samekh*, *pe*, *taw*
right-joining	*daleth*, *he*, *waw*, *heth*, *yodh*, *mem*, *ayin*, *resh*, *shin*
exceptions	*zayin*, *pe*

Throughout this document when a letter is described as joining another letter to the left, it is implied that the joining occurs only if the following letter is right-joining. Equally, a letter described as joining to the right implies that the preceding letter is left-joining. Otherwise, no connections are made between the letters.

The similarities between nominal and non-initial forms of letters led earlier scholars to suggest that joining rules may vary in particular sources. For such cases, it is be useful to consider Henning's advice:

> It becomes then all the more important to observe, in the strictest manner, certain scribal conventions that arise from the material, in particular the rules of linking and separating letters. It seems to me that by refusing any license in such matters we can improve the security of reading [...] Attempts have been made from time to time to arrogate to oneself some license, so as to assert: "in this work *W* has been connected to the left"; in the long run they have invariably been rejected.   (1965: 171)

For instance, the word نسوى in TK no. 69 (fig. 28) was interpreted as *nwšy* by Tolstov and Livshits (1964). This reading of the second letter as *waw* forced an analysis of the right-joining letter as being potentially left-joining in some cases in the Tok Kala texts. A more likely interpretation of نسوى is *NPŠY*. Reading *pe*

instead of *waw* not only provides an accurate reading, but also adheres to the spelling convetions of the script and eliminates the need to complicate the behavior of *waw*. Similarly, the word מחבו was interpreted as *tnbryk* by Tolstov and Livshits. This reading required an analysis of medial *resh* as a potentially left-joining letter, contrary to all other evidence. A reanalysis of the fifth word as medial *kaph* instead of *resh* presented a better option. This approach allowed a more accurate reading of the word as *tpnkwk*. Yet another issue was the joining behaviors of *he* and *kaph*. The word וכ was interpreted as *kw* by Tolstov and Livshits, in which *kaph* was treated as a non-joining letter. This assumption did not align with the features of *kaph* in כאכנ *k ꜣkꜣn-y* (TK 26, fig. 31). Following the behavior of *kaph* in the latter, reading וכ as *hw* provides for more accurate results. Such issues are typical in initial attempts at decipherment. At this point in time, the joining behaviors of Khwarezmian letters have been determined.

There are only two letters whose left-side may be connected in one case and suspended in another are *zayin* and *pe* (see below). The default joining behavior may be modified using ⃞ U+200C ZERO WIDTH NON-JOINER (abbreviated as ZWNJ).

4.1.2   Notes on letters

aleph   The *aleph* is rendered in various ways:

- Initial *aleph* generally connects at the baseline, eg. ꜣ*pbntn*. Examination of its occurrences provides some rules for its positioning. When *aleph* precedes letters that do not join to the left (*waw*, *yodh*, *resh*, etc.), the non-connecting variants of *zayin* and *pe*, the final form of a letter that is not left-joining (*nun*), or a letter that does not connect at the baseline (*gimel*), it connects to these letters at the midpoint at their right edge, eg. ꜣ*ztyk*.

- Medial *aleph* generally connects at the baseline, eg. כאכנ *k ꜣkꜣn-y*, כאעגאמ *pr ꜣny ꜣty*. It connects to following letters according to the rules for the initial form, eg. מחב *tnb ꜣr*, גדדאמ *grdm ꜣn*. Medial *aleph* is rendered as ✕ in some coins and vessels instead of as ⅄. This is a older form and should be treated as a glyphic variant.

- Final *aleph* is written as an elongated stroke ⌐ that does not connect to the preceding letter, eg. ⌐ *YRḤ ꜣ* (TK 25, fig. 25). This is the default behavior for final *aleph* and does not require breaking of the normal cursive behavior. Moreover, final *aleph* triggers rendering of a preceding letter in its final form. For example, the *nun* is written as ן before ⌐ in ⌐ *ZWZN- ꜣ*. In order to produce this behavior, ZWNJ is to be placed after *nun*.

- A detached form ∢ of *aleph* is used for representing possessive forms. It is attested in the ossuary inscriptions. This form occurs in medial and final positions. eg. מחנגמ *tknp ꜣn- ꜣ-k* (TK 52, fig. 24); ∢וכ *hy ꜣn- ꜣ* (TK 26). Usage of ZWNJ to produce this form is not feasible. Therefore, it is encoded as the separate letter ∢ DETACHED ALEPH.

beth, **nun**, **pe**   These letters ⊃ *beth*, ⊐ *nun*, ⊃ *pe* are often written using a shape similar to ⊐ in medial position. Their initial forms are distinguished by the degree of curvature of the primary stroke.

gimel   The *gimel* connects to the left at the top, eg. *gimel* + *waw* as ⊓ in the name of a day, חשמ *gwšt*; *gimel* + *resh* as דר in גדדאמ *grdm ꜣn* 'paradise'. Sequences of *gimel* + *waw* or *yodh* may resemble the letter ⊓ *heth*, but are distinguishable based upon context. It has the archaic form ⊤ in silver vessels.

gimel, **he**, **kaph**   The letters ך *gimel*, ג *he*, כ *kaph* have the same basic structure: a horzontal stroke attached to a descending stroke. The *kaph* is identified by its broad, horizontal top stroke, and the elongated descender of its final form. The *gimel* is written with a 90° angle, while the *he* consistently appears with a descending top stroke and angled bottom stroke. The distinctivenes of *gimel* is evidenced by the archaic form ד used in silver vessels. They are further differentiated by their joining behaviors. The *gimel* and *kaph* are dual joining and *he* is right-joining, as indicated by their interactions with ı *waw* in חשﬨ *gwšt* and רושﬡ *hwnšk* (TK 52, fig. 24), and ﬤﬡﬨ *kwʾn-y* (TK 25, fig. 25).

daleth, **ayin**, **resh**   The letters ד *daleth*, ע *ayin*, ר *resh* have the same basic structure, but are differentiated in terms of their shapes. The *daleth* has a shorter primary stroke than *resh* and a wider top angle than *ayin*. The *resh* and *daleth* are differentiated by the length of the primary stroke, with that of ר *resh* being longer than that of ד *daleth* as shown in רדﬤﬡﬦ *grdm'n* (TK no. 25, fig. 25). Such a distinction appears to be carried over from Imperial Aramaic, where ד *daleth* and ר *resh* differ by the length of the primary stroke. The *ayin* is written with a narrower angle at the left as compared to *daleth* and *resh*, and the left stroke connects at a lower point on the right stroke, eg. ﬠﬨﬦ *'BDt*.

waw, **yodh**   The letters ı *waw* and ı *yodh* are difficult to distinguish in various sources and they have the same joining properties. However, *yodh* has a notched head, while *waw* is typically a vertical stroke, either straight or slightly curved. The two may be written similar to ı *zayin*, but the latter is often distinguished by a curved terminal and its behavior of joining to the left in heterograms in certain sources.

zayin   The joining properties of *zayin* differs across the sources. In some Aramaic heterograms it joins to the left, eg. ﬢﬠ *ZNH*, while in others it does not, ﬧﬦﬠ *ZWZN*. As there is no feasible means for selecting optional connections for a letter, it is necessary to define *zayin* as a dual-joining letter. The ʐwnj may be used to break the connection by placing it after *zayin*. When it connects on the left, the terminal stroke is curved at the baseline to join the following letter.

nun   When *nun* is followed by *aleph* at the end of a word, it does not connect to the latter, and is rendered using its final form because the *aleph* in this position is default non-joining. See details in description of *aleph* above. Word-final ﬢ *nun* is written as ן in the ossuary inscriptions. In coinage and silver vessels, the final form appears as ﬢ. The orientation of the tail may be guided by the vertical constraints of the inscription, ie. being placed on the lip of a vessel or at the edge of a coin. For example, ﬦﬦﬠﬢ *ZWZN-'* occurs on silver vessels, and would be normalized as ﬦﬠﬢ.

ayin   In the available sources this letter occurs in ﬠﬨﬦ *'BDt* 'done, made' (silver bowl #7, fig. 22) and ﬠﬤ *'L*, an Aramaic heterogram for 'to' (various silver bowls and ossuaries). It is only attested in word-initial position. While it appears to join to *lamedh* in ﬠﬤ, the connection is likely a result of letter spacing, not a cursive property of the letter. Given its similar structure to *daleth* and *resh*, it is likely that *ayin* is right-joining. If evidence indicates that it is non-joining, then its property may be modified at that time.

pe   Although it is a default dual-joining letter, in some sources medial כ ᴘᴇ does not connect to the left, compare ﬠﬨﬦﬠשׁ *'pbntn* with ﬡﬠﬨﬦשׁ *'p-bntn*. Such usage is unpredictable and appears to be a scribal convention. As there is no feasible means for selecting optional connections for a letter, it is necessary to define *pe* as a dual-joining letter. The ʐwnj may then be used to break the connection by placing it after *pe*.

shin   The letter ש is represented using the glyphic variant ﬡ on some coins.

taw   In ossuary inscriptions letters that follow ﬨ *taw* join to its left edge without any spacing or extension of the baseline, or are incorporated into the left edge of the glyph, eg. ﬨ is written as ﬨ to accommodate a

following letter. For instance, *taw + yodh* is ‎ﬦ as in ‎ﬧﬨﬦﬡ *ᵓztyk* on TK no. 52 (fig. 24); and *taw + final nun* as ‎ﬦﬢ in ‎ﬣﬤﬥﬦﬧ *ᵓpbntn*. In coins, the point of connection occurs at the head, where the top-stroke of *taw* is extended into that of the following letter, eg. *taw + waw* is written ‎ﬦ *tw*, as in the name ‎ﬦﬨﬦ﬩שׁ *twtwxs* on type BI coins (fig. 8). The *taw* has a glyphic variant form ‎שּׂ that has an open right stroke.

4.2 Vocalization sign

Character name	X_n	X_f	X_m	X_i	Join	Latin
KHWAREZMIAN VOCALIZATION SIGN	٫	—	—	—	non	*-w, -y*

In TK 25 and 52, when *waw* and *yodh* appear in word-final position and indicate a possessive, they are written using the sign ٫, (transliterated as either *-w* or *-y*, depending upon context). This sign is non-joining, therefore, the preceding letter is rendered using its final form. It is described by Henning as a "vocalization mark", which is "a rounded form reminiscent of an Arabic *ḍamm*" and occurs in words, such as ‎ﬡﬢﬣﬤﬥﬦﬧ *wḥwnt'n-w* and ‎ﬨ﬩שׁשׂ *kᵓkᵓn-y*, which are possessives functioning as patronyms (1965: 178). The "*ḍamm*" to which Henning refers is the sign encoded as ٗ U+064F ARABIC DAMMA.

4.3 Numbers

The representative nominal form of each number is given in the 'X_n' column. The red dash shown in the positional forms indicate connection points.

X_n	Character name	Value	X_f	X_m	X_i	Join
۱	KHWAREZMIAN NUMBER ONE	1	۱	۱	۱	non
۱۱	KHWAREZMIAN NUMBER TWO	2	۱۱	۱۱	۱۱	non
۱۱۱	KHWAREZMIAN NUMBER THREE	3	۱۱۱	۱۱۱	۱۱۱	non
۱۱۱۱	KHWAREZMIAN NUMBER FOUR	4	۱۱۱۱	۱۱۱۱	۱۱۱۱	non
ﬤ	KHWAREZMIAN NUMBER TEN	10	ﬤ	ﬤ	ﬤ	right
ﬥ	KHWAREZMIAN NUMBER TWENTY	20	ﬥ	ﬥ	ﬥ	dual
ﬦ	KHWAREZMIAN NUMBER ONE HUNDRED	100	ﬦ	ﬦ	ﬦ	left

Primary units   The primary units are expressed using repetitions of the sign ۱, which is a non-joining character. The numbers 5–9 are written using sequences of ONE arranged in groups containing three or four instances of ۱. See, for example, '۱۱۱ ۱۱۱۱' for 7 and '۱۱۱۱ ۱۱۱۱' for 8 in TK no. 19 (fig. 26), as well as '۱ ۱۱۱ ۱۱۱' for 7 in TK no. 25 (fig. 25). Also, '۱۱ ۱۱۱' for the number 5 in silver bowl #2 (fig. 17). The number 5 in silver bowl #5 appears as '۱۱۱۱۱' without a spaced grouping, but the extended terminal of the third ۱ suggests the

intended grouping '𐾀𐾀 𐾀𐾀𐾀' despite lack of spacing (see fig. 20). Given the grouping behavior of 𐾀, the numbers 𐾀 ONE .. 𐾀𐾀𐾀𐾀 FOUR are encoded atomically. This model for ONE .. FOUR follows the encoding for 'Inscriptional Parthian', eg. 𐡘 U+10B58 INSCRIPTIONAL PARTHIAN NUMBER ONE .. 𐡛 U+10B5B INSCRIPTIONAL PARTHIAN NUMBER FOUR. This model is also used in the encodings for 'Imperial Aramaic', 'Inscriptional Pahlavi', and the forthcoming encoding for 'Old Sogdian'.

Ten   The ᐳ TEN resembles a vertically compressed ⅃ LAMEDH. It is a right-joining character.

Twenty   The sign for ⟨ TWENTY is derived palaeographically from a vertical stack of two instances of ᐳ TEN. The sign is treated as an atomic character. It is a dual joining character.

Hundreds   The number 100 is written using ⟨ ONE HUNDRED. The ONE HUNDRED also functions as a unit mark for the hundreds. Multiples of hundred are indicated using primary numbers placed before ONE HUNDRED. It is a left-joining character.

Higher orders   There are no distinctive signs for orders larger than the hundreds. Such numbers are not attested in the available materials.

4.4   Punctuation

Spaces are commonly used for separating words in the ossuary inscriptions and on some silver bowls. There are no special signs for punctuation.

4.5   Line-breaking

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

4.6   Collation

The sort order of the letters follows the encoded order:

ᴧ ALEPH < • DETACHED ALEPH < ⅃ BETH < ㄱ GIMEL < ㄱ DALETH < ᠈ HE < ᛁ WAW <

ᛁ ZAYIN < ⴖ HETH < ᛁ YODH < ㄱ KAPH < ⅃ LAMEDH < �短 MEM < ⅃ NUN <

ᗡ SAMEKH < ㄚ AYIN < ᠑ PE < ㄚ RESH < ᗐ SHIN < ᗡ TAW < ᠈ VOCALIZATION SIGN

5　Encoded representations

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 Khwarezmian records are given below along with their input strings:

ꞌgꞌdk	ארגריא	\<◢ ALEPH, ר GIMEL, ◣ ALEPH, ר DALETH, ר KAPH\>
ꞌpbntn	מתנשא	\<◢ ALEPH, כ PE, ᴗ BETH, ᒐ NUN, ᴆ TAW, ᒐ NUN\>
ꞌp-bntn	מתנשא	\<◢ ALEPH, כ PE, ⟦ZWNJ⟧ ZWNJ, ᴗ BETH, ᒐ NUN, ᴆ TAW, ᒐ NUN\>
ꞌztyk	אמפיא	\<◢ ALEPH, ᒐ ZAYIN, ᴆ TAW, ᛁ YODH, ר KAPH\>
ꞌrꞌškrk	ארשאכרך	\<◢ ALEPH, ᛉ RESH, ◣ ALEPH, ש SHIN, ר KAPH, ᛉ RESH, ר KAPH\>
ꞌšpynšwk	שנפינשתך	\<◢ ALEPH, ש SHIN, כ PE, ᛁ YODH, ᒐ NUN, ש SHIN, ᴆ TAW, ר KAPH\>
bntk	כתנכ	\<ᴗ BETH, ᒐ NUN, ᴆ TAW, ר KAPH\>
BŠNT	כשנט	\<ᴗ BETH, ש SHIN, ᒐ NUN, ᴆ TAW\>
grdmꞌn	רדדמא	\<ר GIMEL, ᛉ RESH, ר DALETH, ᴝ MEM, ◣ ALEPH, ᒐ NUN\>
gwšt	רושם	\<ר GIMEL, ᛁ WAW, ש SHIN, ᴆ TAW\>
hwnšk	רושך	\<ᛉ HE, ᛁ WAW, ᒐ NUN, ש SHIN, ר KAPH\>
hyꞌn-ꞌ	רומ.	\<ᛉ HE, ᛁ YODH, ◣ ALEPH, ᒐ NUN, ⟦ZWNJ⟧ ZWNJ, ◂ DETACHED ALEPH\>
wḥwntꞌn-w	וחושמא	\<ᛁ WAW, ח HETH, ᛁ WAW, ᒐ NUN, ᴆ TAW, ◣ ALEPH, ᒐ NUN, כ VOCALIZATION SIGN\>
ZNH	נ̤ל	\<ᒐ ZAYIN, ᒐ NUN ᛋ HE\>
ZNH	ננל	\<ᒐ ZAYIN, ⟦ZWNJ⟧ ZWNJ, ᒐ NUN ᛋ HE\>
ZWZN	נונ	\<ᒐ ZAYIN, ⟦ZWNJ⟧ ZWNJ, ᛁ WAW, ᒐ ZAYIN, ⟦ZWNJ⟧ ZWNJ, ᒐ NUN\>
ZWZN-ꞌ	נונ—	\<ᒐ ZAYIN, ⟦ZWNJ⟧ ZWNJ, ᛁ WAW, ᒐ ZAYIN, ⟦ZWNJ⟧ ZWNJ, ᒐ NUN, ⟦ZWNJ⟧ ZWNJ, ◣ ALEPH\>
ḥwsrw	חופצו	\<ח HETH, ᛁ WAW, ᴘ SAMEKH, ᛉ RESH, ᛁ WAW\>
ḥwpsk	חותכ	\<ח HETH, ᛁ WAW, כ PE, ᴘ SAMEKH, ר KAPH\>
YRḤꞌ	ודח—	\<ᛁ YODH, ᛉ RESH, ח HETH, ◣ ALEPH\>
kꞌkꞌn-y	כאכני	\<ר KAPH, ◣ ALEPH, ר KAPH, ᒐ NUN, כ VOCALIZATION SIGN\>
KSP	כסכ	\<ר KAPH, ᴘ SAMEKH, כ PE\>
MLKꞌ	מלכ—	\<ᴝ MEM, ᒐ LAMEDH, ר KAPH, ◣ ALEPH\>

mʾnyʾʾty	𐾀𐾅𐾄𐾄	<𐾀 MEM, 𐾅 ALEPH, 𐾂 NUN, 𐾁 YODH, 𐾅 ALEPH, 𐾅 ALEPH, 𐾂 TAW, 𐾁 YODH>
MRʾY	𐾀𐾄𐾂	<𐾀 MEM, 𐾃 RESH, 𐾅 ALEPH, 𐾁 YODH>
NPŠY	𐾄𐾄𐾄𐾁	<𐾂 NUN, 𐾅 PE, 𐾄 SHIN, 𐾁 YODH>
syʾwršprn	𐾀𐾃𐾄𐾄𐾂𐾄𐾂	<𐾅 SAMEKH, 𐾁 YODH, 𐾅 ALEPH, 𐾁 WAW, 𐾃 RESH, 𐾄 SHIN, 𐾅 PE, 𐾃 RESH, 𐾂 NUN>
ʿBDt	𐾄𐾄𐾂	<𐾃 AYIN, 𐾅 BETH, 𐾃 DALETH, 𐾂 TAW>
ʿL	𐾄𐾂	<𐾃 AYIN, 𐾂 LAMEDH>
prnxwnt	𐾀𐾄𐾂𐾄𐾂	<𐾅 PE, 𐾃 RESH, 𐾅 ALEPH, 𐾂 NUN, 𐾁 WAW, 𐾂 NUN, 𐾂 TAW>
twtwḥs	𐾀𐾄𐾀𐾄	<𐾂 TAW, 𐾁 WAW, 𐾂 TAW, 𐾁 WAW, 𐾃 HETH, 𐾅 SAMEKH>
tnbryk	𐾀𐾄𐾁𐾂	<𐾂 TAW, 𐾂 NUN, 𐾅 BETH, 𐾃 RESH, 𐾁 YODH, 𐾃 KAPH>
tpnkwk	𐾀𐾄𐾂𐾂	<𐾂 TAW, 𐾅 PE, 𐾂 NUN, 𐾃 KAPH, 𐾁 WAW, 𐾃 KAPH>

5.1   Numerical notation

The ordering of numbers follows the right-to-left directionality of the script. The expression of numbers is additive. Compounds of different units are produced by placing larger units first. The exception is the usage of primary units for expressing multiples of hundred, which are placed before the character ONE HUNDRED. Spaces are used for separating groups of primary numbers.

The numbers 5–9 may be represented as shown below. Some numbers have more than one representations, as attested in the available sources:

5	𐾄𐾄 𐾄𐾄𐾄	<𐾄𐾄𐾄 THREE, 𐾄𐾄 TWO>
6	𐾄𐾄𐾄 𐾄𐾄𐾄	<𐾄𐾄𐾄 THREE, 𐾄𐾄𐾄 THREE>
7	𐾄𐾄𐾄 𐾄𐾄𐾄𐾄	<𐾄𐾄𐾄𐾄 FOUR, 𐾄𐾄𐾄 THREE>
	𐾄 𐾄𐾄𐾄 𐾄𐾄𐾄	<𐾄𐾄𐾄 THREE, 𐾄𐾄𐾄 THREE, 𐾄 ONE>
8	𐾄𐾄𐾄𐾄 𐾄𐾄𐾄𐾄	<𐾄𐾄𐾄𐾄 FOUR, 𐾄𐾄𐾄𐾄 FOUR>
9	𐾄𐾄𐾄 𐾄𐾄𐾄 𐾄𐾄𐾄	<𐾄𐾄𐾄 THREE, 𐾄𐾄𐾄 THREE, 𐾄𐾄𐾄 THREE>

Multiples of ten are written using sequences of 𐾄 TEN and 𐾃 TWENTY. Even multiples are expressed with repetitions of TWENTY. Odd multiples are produced by attaching TEN at the end.

10	𐾄	<𐾄 TEN>
20	𐾃	<𐾃 TWENTY>

| 30 | ⹈ | <Ɜ TWENTY, ꓾ TEN> |

| 40 | ⹈ | <Ɜ TWENTY, Ɜ TWENTY> |

| 50 | ⹈⹈ | <Ɜ TWENTY, Ɜ TWENTY, ꓾ TEN> |

| 60 | ⹈⹈⹈ | <Ɜ TWENTY, Ɜ TWENTY, Ɜ TWENTY> |

| 70 | ⹈⹈⹈ | <Ɜ TWENTY, Ɜ TWENTY, Ɜ TWENTY, ꓾ TEN> |

| 80 | ⹈⹈⹈⹈ | <Ɜ TWENTY, Ɜ TWENTY, Ɜ TWENTY, Ɜ TWENTY> |

| 90 | ⹈⹈⹈⹈ | <Ɜ TWENTY, Ɜ TWENTY, Ɜ TWENTY, Ɜ TWENTY, ꓾ TEN> |

Multiples of the hundreds are represented using Ɜ ONE HUNDRED in conjunction with the primary units. The primary units are placed before ONE HUNDRED in the input sequence.

| 100 | Ɜ | <Ɜ ONE HUNDRED> |

| 200 | Ɜ⳽ | <⳽ TWO, Ɜ ONE HUNDRED> |

| 300 | Ɜ⳽⳽ | <⳽⳽ THREE, Ɜ ONE HUNDRED> |

Composite numbers found in the sources are given below along with their encoded representations:

| 570 | ⹈⹈⹈⹈Ɜ ⳽ ⳽⳽ | <⳽⳽ TWO, ⳽⳽ THREE, Ɜ ONE HUNDRED, Ɜ TWENTY, Ɜ TWENTY, Ɜ TWENTY, ꓾ TEN> |

| 678 | ⳽ ⳽⳽ ⳽⳽ ⹈⹈⹈⹈Ɜ ⳽⳽ ⳽⳽ | <⳽⳽ THREE, ⳽⳽ THREE, Ɜ ONE HUNDRED, Ɜ TWENTY, Ɜ TWENTY, Ɜ TWENTY, ꓾ TEN, ⳽ TWO, ⳽⳽ THREE, ⳽⳽ THREE> |

6   Character Properties

6.1   Core data: `UnicodeData.txt`

```
10F00;KHWAREZMIAN LETTER ALEPH;Lo;0;R;;;;;N;;;;;
10F01;KHWAREZMIAN LETTER DETACHED ALEPH;Lo;0;R;;;;;N;;;;;
10F02;KHWAREZMIAN LETTER BETH;Lo;0;R;;;;;N;;;;;
10F03;KHWAREZMIAN LETTER GIMEL;Lo;0;R;;;;;N;;;;;
10F04;KHWAREZMIAN LETTER DALETH;Lo;0;R;;;;;N;;;;;
10F05;KHWAREZMIAN LETTER HE;Lo;0;R;;;;;N;;;;;
10F06;KHWAREZMIAN LETTER WAW;Lo;0;R;;;;;N;;;;;
10F07;KHWAREZMIAN LETTER ZAYIN;Lo;0;R;;;;;N;;;;;
10F08;KHWAREZMIAN LETTER HETH;Lo;0;R;;;;;N;;;;;
10F09;KHWAREZMIAN LETTER YODH;Lo;0;R;;;;;N;;;;;
10F0A;KHWAREZMIAN LETTER KAPH;Lo;0;R;;;;;N;;;;;
```

```
10F0B;KHWAREZMIAN LETTER LAMEDH;Lo;0;R;;;;;N;;;;;
10F0C;KHWAREZMIAN LETTER MEM;Lo;0;R;;;;;N;;;;;
10F0D;KHWAREZMIAN LETTER NUN;Lo;0;R;;;;;N;;;;;
10F0E;KHWAREZMIAN LETTER SAMEKH;Lo;0;R;;;;;N;;;;;
10F0F;KHWAREZMIAN LETTER AYIN;Lo;0;R;;;;;N;;;;;
10F10;KHWAREZMIAN LETTER PE;Lo;0;R;;;;;N;;;;;
10F11;KHWAREZMIAN LETTER RESH;Lo;0;R;;;;;N;;;;;
10F12;KHWAREZMIAN LETTER SHIN;Lo;0;R;;;;;N;;;;;
10F13;KHWAREZMIAN LETTER TAW;Lo;0;R;;;;;N;;;;;
10F14;KHWAREZMIAN VOCALIZATION SIGN;Lo;0;R;;;;;N;;;;;
10F15;KHWAREZMIAN NUMBER ONE;No;0;R;;;;1;N;;;;;
10F16;KHWAREZMIAN NUMBER TWO;No;0;R;;;;2;N;;;;;
10F17;KHWAREZMIAN NUMBER THREE;No;0;R;;;;3;N;;;;;
10F18;KHWAREZMIAN NUMBER FOUR;No;0;R;;;;4;N;;;;;
10F19;KHWAREZMIAN NUMBER TEN;No;0;R;;;;10;N;;;;;
10F1A;KHWAREZMIAN NUMBER TWENTY;No;0;R;;;;20;N;;;;;
10F1B;KHWAREZMIAN NUMBER ONE HUNDRED;No;0;R;;;;100;N;;;;;
```

6.2 Linebreak data: `LineBreak.txt`

```
10F00..10F13;AL   # Lo  [20] KHWAREZMIAN LETTER ALEPH..KHWAREZMIAN LETTER TAW
10F14;            # Lo       KHWAREZMIAN VOCALIZATION SIGN
10F15..10F1B;AL   # No   [7] KHWAREZMIAN NUMBER ONE..KHWAREZMIAN NUMBER ONE HUNDRED
```

6.3 Shaping properties: `ArabicShaping.txt`

```
10F00; KHWAREZMIAN ALEPH; D; No_Joining_Group
10F01; KHWAREZMIAN DETACHED ALEPH; N; No_Joining_Group
10F02; KHWAREZMIAN BETH; D; No_Joining_Group
10F03; KHWAREZMIAN GIMEL; D; No_Joining_Group
10F04; KHWAREZMIAN DALETH; R; No_Joining_Group
10F05; KHWAREZMIAN HE; R; No_Joining_Group
10F06; KHWAREZMIAN WAW; R; No_Joining_Group
10F07; KHWAREZMIAN ZAYIN; R; No_Joining_Group
10F08; KHWAREZMIAN HETH; R; No_Joining_Group
10F09; KHWAREZMIAN YODH; R; No_Joining_Group
10F0A; KHWAREZMIAN KAPH; D; No_Joining_Group
10F0B; KHWAREZMIAN LAMEDH; D; No_Joining_Group
10F0C; KHWAREZMIAN MEM; R; No_Joining_Group
10F0D; KHWAREZMIAN NUN; D; No_Joining_Group
10F0E; KHWAREZMIAN SAMEKH; D; No_Joining_Group
10F0F; KHWAREZMIAN AYIN; R; No_Joining_Group
10F10; KHWAREZMIAN PE; D; No_Joining_Group
10F11; KHWAREZMIAN RESH; R; No_Joining_Group
10F12; KHWAREZMIAN SHIN; R; No_Joining_Group
10F13; KHWAREZMIAN TAW; D; No_Joining_Group
10F14; KHWAREZMIAN VOCALIZATION SIGN; N; No_Joining_Group
10F15; KHWAREZMIAN ONE; N; No_Joining_Group
10F16; KHWAREZMIAN TWO; N; No_Joining_Group
10F17; KHWAREZMIAN THREE; N; No_Joining_Group
10F18; KHWAREZMIAN FOUR; N; No_Joining_Group
10F19; KHWAREZMIAN TEN; R; No_Joining_Group
10F1A; KHWAREZMIAN TWENTY; D; No_Joining_Group
10F1B; KHWAREZMIAN ONE HUNDRED; L; No_Joining_Group
```

7 References

Anderson, Deborah, et. al. 2017 "Recommendations to UTC #152 July-August 2017 on Script Proposals". L2/17-255. https://www.unicode.org/L2/L2017/17255-script-ad-hoc.pdf

———. 2018. "Recommendations to UTC #154 January 2018 on Script Proposals". L2/18-039. https://www.unicode.org/L2/L2018/18039-script-adhoc-rec.pdf

Azarpay, Guitty 1969. "Nine Inscribed Choresmian Bowls". *Artibus Asiae*, vol. 31, no. 2/3, pp. 185–203.

"Coins of Central Asia". http://www.sogdcoins.narod.ru/english/khwarezm/coins.html

Durkin-Meisterernst, Desmond. 2009. "Khwarezmian'. In *The Iranian Languages*, Gernot Windfuhr [ed], pp. 336–376. New York: Routledge.

Federov, Michael 2005. "On Some Articles in the Recent Issues (Nos. V, VI, VII) of the Нумизматика Центральной Азии (Numismatics of Central Asia)". *Central Asiatic Journal*, vol. 49, no. 2 (2005), pp. 175–203.

———. 2006. "First Reported Find of an Early Medieval Khwarezmian Drachm in the Kyrgyz Republic". *American Journal of Numismatics*, vol. 18 (2006), pp. 123–130.

Federov, Michael and Kuznetsov, Andrew. 2008. "A hoard of early mediaeval Khwarezmian drachms from the Kuiuk-kala hill fort". *The Numismatic Chronicle*, vol. 168 (2008), pp. 446–451.

———. 2011. "On some previously unknown Khwarazmian drachms and the names of rulers on them". *Iran*, vol. 49 (2011), pp. 79–88.

———. 2012. "Returning to the Kuiuk-Kala hoard of early Mediaeval Khwarezmian drachms: Khorezm, eastern delta of the Amu Darya, Kuiuk-kala hill fort, 2008". *The Numismatic Chronicle*, vol. 172 (2012), pp. 335–341.

Henning, Walter B. 1965. "The Choresmian Documents". *Asia Major*, vol. 11, no. 2, pp. 166–79.

Humbach, Helmut. 2011 [1998]. "Epigraphy i. Old Persian and Middle Iranian epigraphy". *Encyclopædia Iranica*, vol. VIII, fasc. 5, pp. 478–488. http://www.iranicaonline.org/articles/epigraphy-i

Иванчик, А. И.; Лурье, П. Б. [Ivantchik, A. I.; Lurje, P. B.]. 2013. "Две надписи из Чирик-рабата". *Commentationes Iranicae*. Сборник статей к 90-летию Владимира Ароновича Лившица, Под редакцией С. Р. Тохтасьева и П. Б. Лурье, pp. 286–294. Санкт-Петербург: Нестор-История.

Livshits, Vladimir A. 2003. "Three silver bowls from the Isakovka burial-ground no. 1 with Khwarezmian and Parthian inscriptions". *Ancient Civilizations*, vol. 9, no. 1–2, pp. 147–172.

Лурье, П. Б. [Lurje, Pavel V.] 2013. "Несколько неизданных хорезмийских надписей из Ток-Калы". *Scripta Antiqua*: *Вопросы древней истории, филологии, искусства и материальной культуры* [= *Ancient history, Philology, Arts and Material Culture*], Том третий [= Volume three], К юбилею Эдварда Васильевича Ртвеладзе [= E. v. Rtveladze Felicitation volume], с. 729–740. Москва: Собрание.

———. forthcoming. "Some New Readings of Chorasmian Inscriptions on Silver Vessels And Their Relevance to Chorasmian Era".

MacKenzie, D. N. 2011 [1991]. "Chorasmia iii. The Chorasmian Language." *Encyclopædia Iranica*, vol. V, pp. 517–520. `http://www.iranicaonline.org/articles/chorasmia-iii`

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

Sachau, C. Eduard [tr]. 1897. *The Chronology of Ancient Nations: An English Version of the Arabic Text of the Athâr-ul-Bâkiya of Albîrûnî, or 'Vestiges of the Past', Collected and Reduced to Writing By the Author in A. H. 39–1, A. D. 1000*. London: Oriental Translation Fund of Great Britain & Ireland.

Skjærvø, Prods Oktor. 1996. "Aramaic Scripts for Iranian Languages." *The World's Writing Systems*, edited by Peter T. Daniels and W. Bright, pp. 515–535. New York and Oxford: Oxford University Press.

Смирнов, Я. И. [Smirnov, I. A.] 1909. *Восточное серебро* [= *Vostochnoe Serebro*]. Атлас древней серебряной и золотой посуды восточного происхождения, найденной преимущественно в пределах Российской империи.

Tolstov, S. P. and Livshitz, V. A. 1964. "Decipherment and Interpretation of the Khwarezmian Inscriptions from Tok Kala", *Acta Antiqua Academiae Scientiarum Hungaricae*, vol. 12, pp. 231–251.

Вайнберг, Бэлла Ильинична [Vainberg, B. I.]. 1977. *Монеты Древнего Хорезма* [= *Monety Drevnego Khorezma*]. Москва: Наука, Глав. ред. восточной литераы

8   Acknowledgments

	10F0	10F1
0	ᠠ 10F00	つ 10F10
1	᠘ 10F01	ﻱ 10F11
2	ﻝ 10F02	ﺵ 10F12
3	ヿ 10F03	ㄸ 10F13
4	ﻱ 10F04	ﻭ 10F14
5	ﻝ 10F05	ﺍ 10F15
6	ﻝ 10F06	ﺍﺍ 10F16
7	ﻝ 10F07	ﺍﺍﺍ 10F17
8	ﻥ 10F08	ﺍﺍﺍﺍ 10F18
9	ﻝ 10F09	ﻝ 10F19
A	ﻝ 10F0A	ﻝ 10F1A
B	ﻝ 10F0B	ﻝ 10F1B
C	ﻝ 10F0C	
D	ﻝ 10F0D	
E	ﻝ 10F0E	
F	ﻝ 10F0F	

Also known as 'Chorasmian'.

Letters

10F00	ᠠ	KHWAREZMIAN LETTER ALEPH
10F01	᠘	KHWAREZMIAN LETTER DETACHED ALEPH
10F02	ﻝ	KHWAREZMIAN LETTER BETH
10F03	ヿ	KHWAREZMIAN LETTER GIMEL
10F04	ﻱ	KHWAREZMIAN LETTER DALETH
10F05	ﻝ	KHWAREZMIAN LETTER HE
10F06	ﻝ	KHWAREZMIAN LETTER WAW
10F07	ﻝ	KHWAREZMIAN LETTER ZAYIN
10F08	ﻥ	KHWAREZMIAN LETTER HETH
10F09	ﻝ	KHWAREZMIAN LETTER YODH
10F0A	ﻝ	KHWAREZMIAN LETTER KAPH
10F0B	ﻝ	KHWAREZMIAN LETTER LAMEDH
10F0C	ﻝ	KHWAREZMIAN LETTER MEM
10F0D	ﻝ	KHWAREZMIAN LETTER NUN
10F0E	ﻝ	KHWAREZMIAN LETTER SAMEKH
10F0F	ﻝ	KHWAREZMIAN LETTER AYIN
10F10	つ	KHWAREZMIAN LETTER PE
10F11	ﻱ	KHWAREZMIAN LETTER RESH
10F12	ﺵ	KHWAREZMIAN LETTER SHIN
10F13	ㄸ	KHWAREZMIAN LETTER TAW

Vocalization sign

10F14	ﻭ	KHWAREZMIAN VOCALIZATION SIGN

Numbers

10F15	ﺍ	KHWAREZMIAN NUMBER ONE
10F16	ﺍﺍ	KHWAREZMIAN NUMBER TWO
10F17	ﺍﺍﺍ	KHWAREZMIAN NUMBER THREE
10F18	ﺍﺍﺍﺍ	KHWAREZMIAN NUMBER FOUR
10F19	ﻝ	KHWAREZMIAN NUMBER TEN
10F1A	ﻝ	KHWAREZMIAN NUMBER TWENTY
10F1B	ﻝ	KHWAREZMIAN NUMBER ONE HUNDRED

	Khwarezmian	Old Sogdian	Inscriptional Pahlavi	Inscriptional Parthian	Imperial Aramaic
aleph	⅄	⅄	⅄	⅄	⅄
beth	⅃	⅃	⅃	⅃	⅃
gimel	⅂	⅀	⅃	⅂	⅄
daleth	⅂	(⅄)	ⅈ	⅄	⅄
he	⅄	⅄, ⅃	⅀	⅏	⅄
waw	⅃	⅃	⅂	⅃	⅄
zayin	⅂	⅃	ⅆ	⅄	⅃
heth	⊓	⅄	⅄	⅄	⅄
teth	—	—	⅃	⅄	⅄
yodh	⅃	⅃	⅂	⅃	⅄
kaph	⅃	⅄	ⅈ	⅄	⅄
lamedh	⅃	⅃	⅃	⅃	⅃
mem	⅏	⅄	⅄	⅄	⅄
nun	⅃	⅃	⅃	⅃	⅄
samekh	ⅅ	⅄	⅄	ⅅ	⅄
ayin	⅄	⅄, (⅄)	(⅄)	⅃	ⅴ
pe	⅄	⅄	⅄	⅄	⅄
sadhe	—	⅄	⅄	⅄	⅄
qoph	—	—	(⅄)	⅄	⅄
resh	⅄	⅄	(⅄)	⅄	⅄
shin	ⅳ	⅄	⅄	⅄	ⅴ
taw	⅏	⅄	⅄	⅄	⅃

Table 1: Comparison of Khwarezmian letters with those in Unicode blocks for related Iranian scripts and Aramaic. Parenthesis indicate that a letter has been unified with another in the respective encoding. In Inscriptional Pahlavi, *ayin* and *resh* are unified with *waw*, and *qoph* with *mem*. For Old Sogdian, *daleth* and regular *ayin* are unified with *resh*.

	Khwarezmian	Old Sogdian	Inscriptional Pahlavi	Inscriptional Parthian	Imperial Aramaic
ONE	ɪ	ل	⟩	ل	ɪ
TWO	ɪɪ	لل	⟩⟩	‖	Ⅴ
THREE	ɪɪɪ	للل	⟩⟩⟩	⦀	Ⅶ
FOUR	ɪɪɪɪ	للل	⟩⟩⟩	⦀⦀	—
FIVE	—	للل	—	—	—
TEN	᠌ᠵ	᠌ᠵ	᠃	᠌ᠧ	᠊
TWENTY	�location3	ᢳ	ᢳ	ᠪ	ᠷ
THIRTY	—	ᢶ	—	—	—
ONE HUNDRED	ᢹ	ᢷ	ᢵ	ᢷ	ᠼ
ONE THOUSAND	—	—	ᢿ	ᢶ	ᢏ
TEN THOUSAND	—	—	—	—	ᢅ
ONE HALF	—	ᢠ	—	—	—

Table 2: Comparison of Khwarezmian numerical signs with those in Unicode blocks for related Iranian scripts and Aramaic.

Figure 1: Inventory of characters on Khwarezmian coins (БII–ГVI), Toprak Kala (Топрак-кала),
Yakke Parsan (Якке парсан), Tok Kala (Ток-кала) (from Vainberg 1977: Table 8).
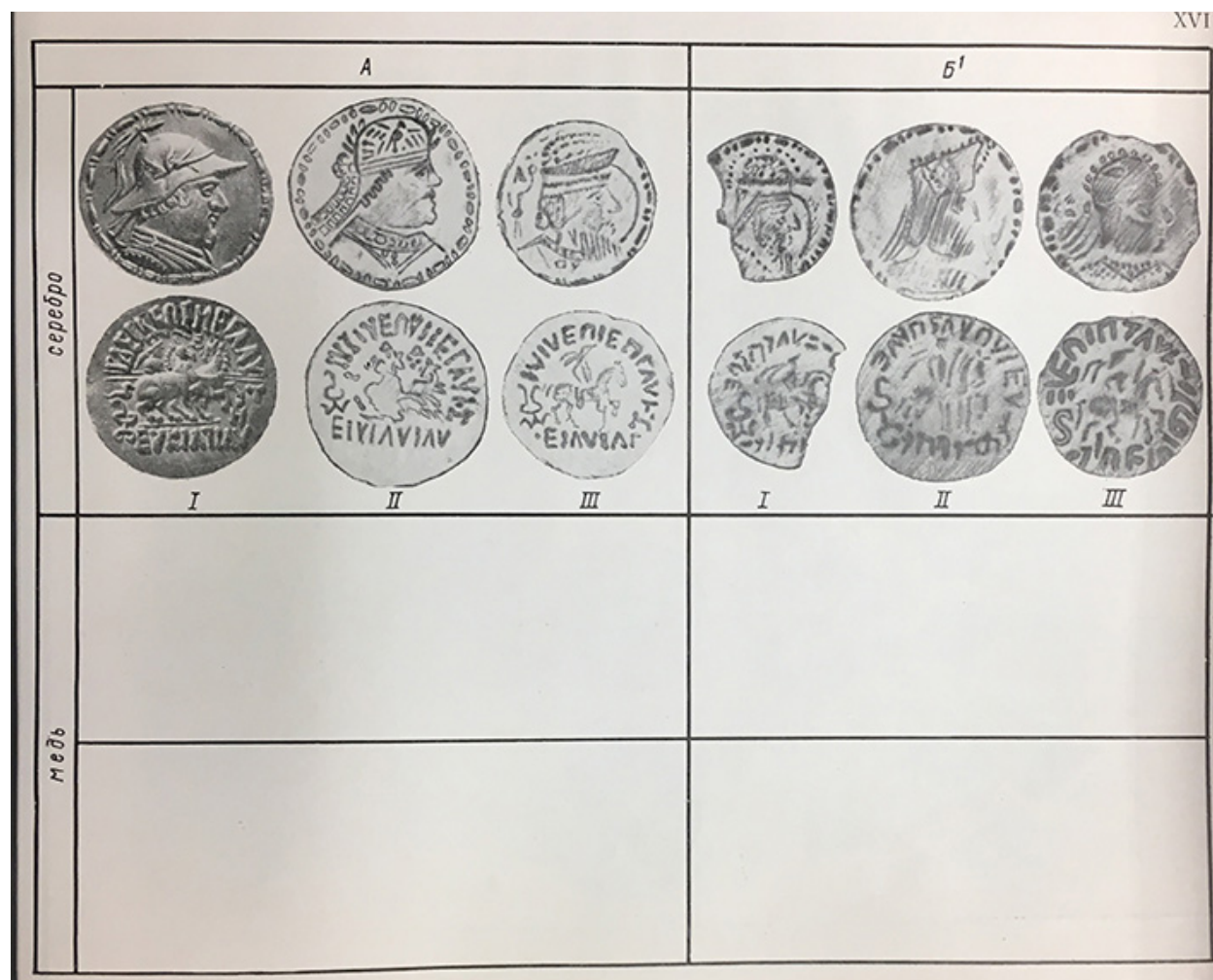
Figure 2: Khwarezmian coins (from Vainberg 1977: Table 16).
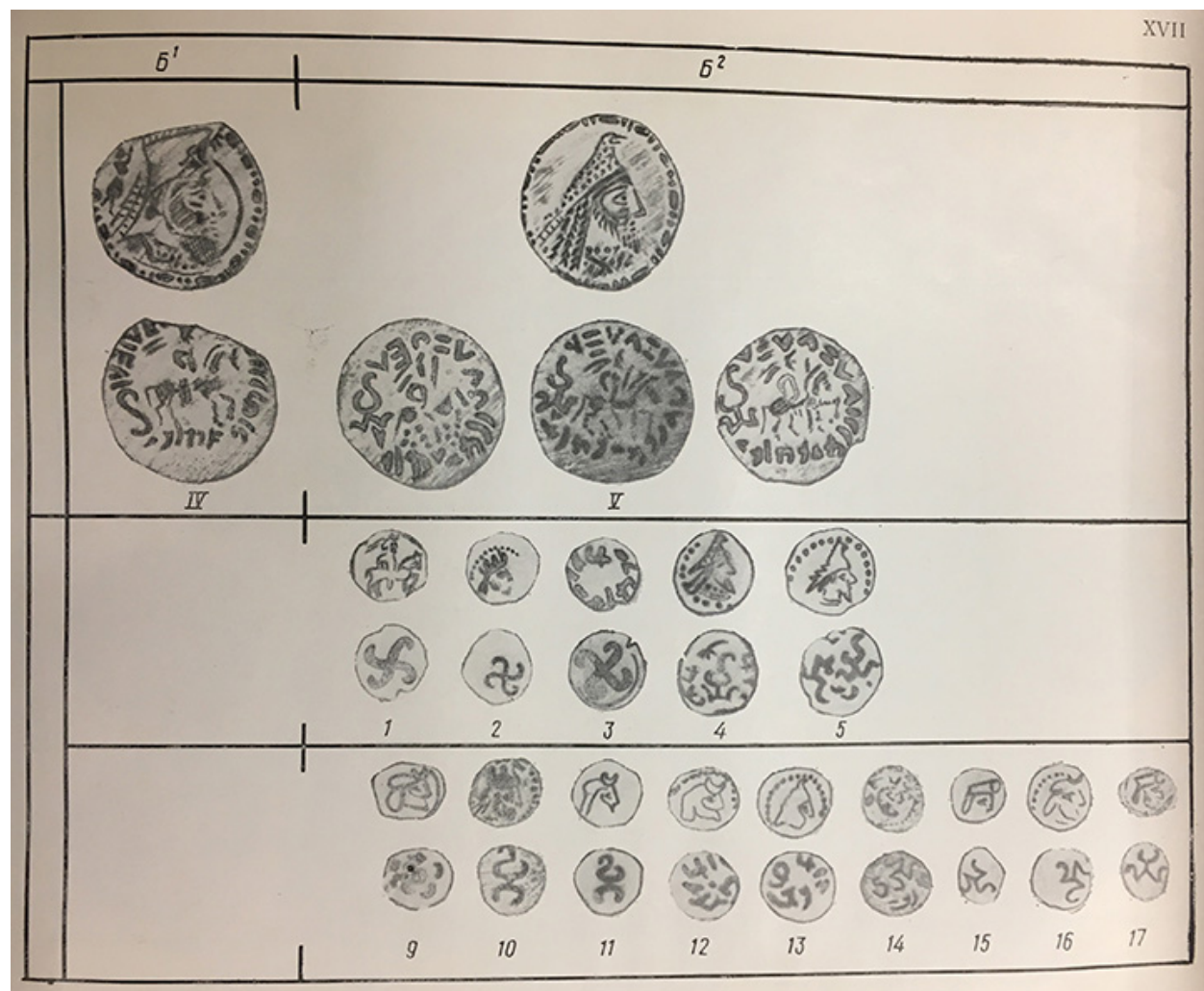
Figure 3: Khwarezmian coins (from Vainberg 1977: Table 17).
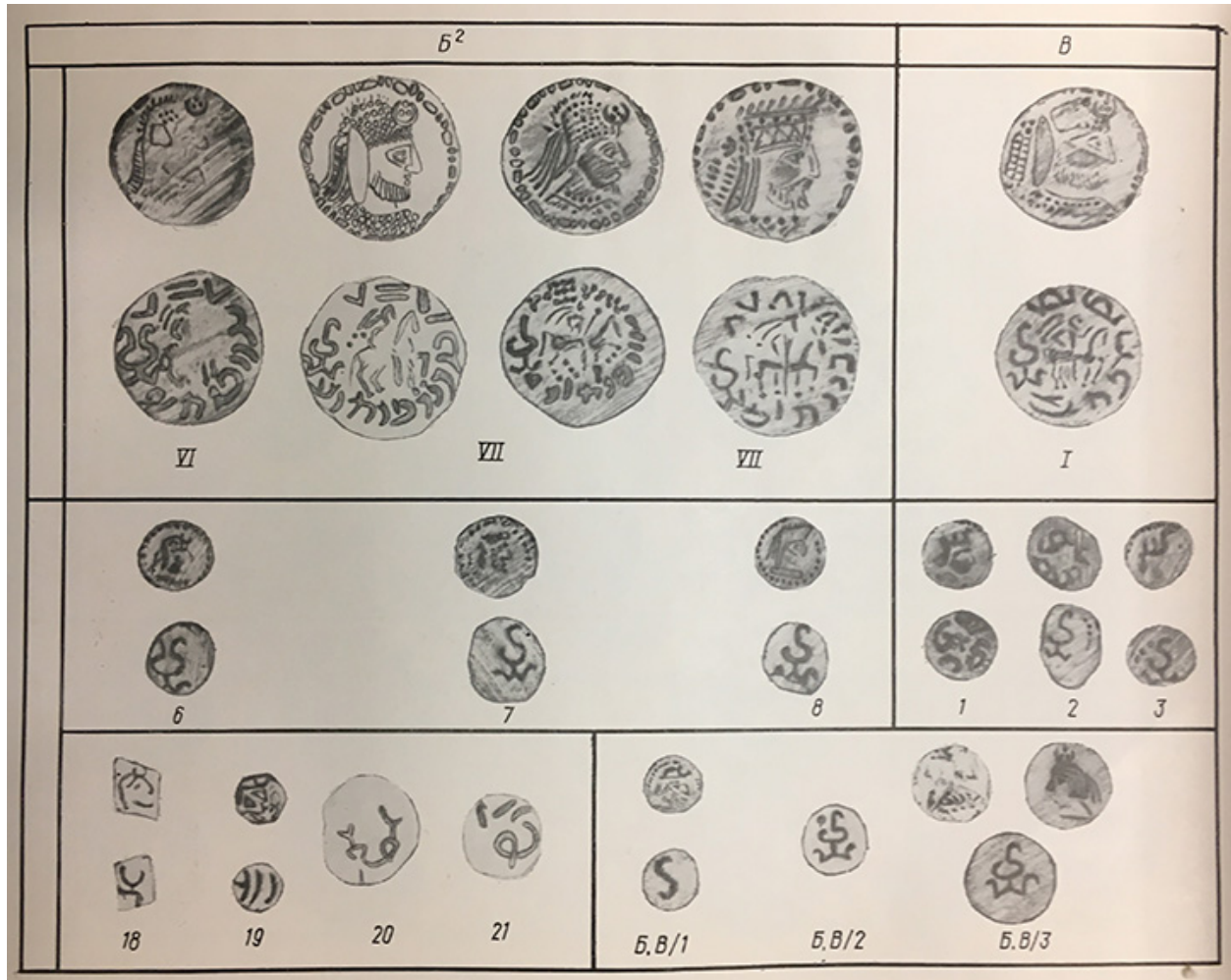
Figure 4: Khwarezmian coins (from Vainberg 1977: Table 18).
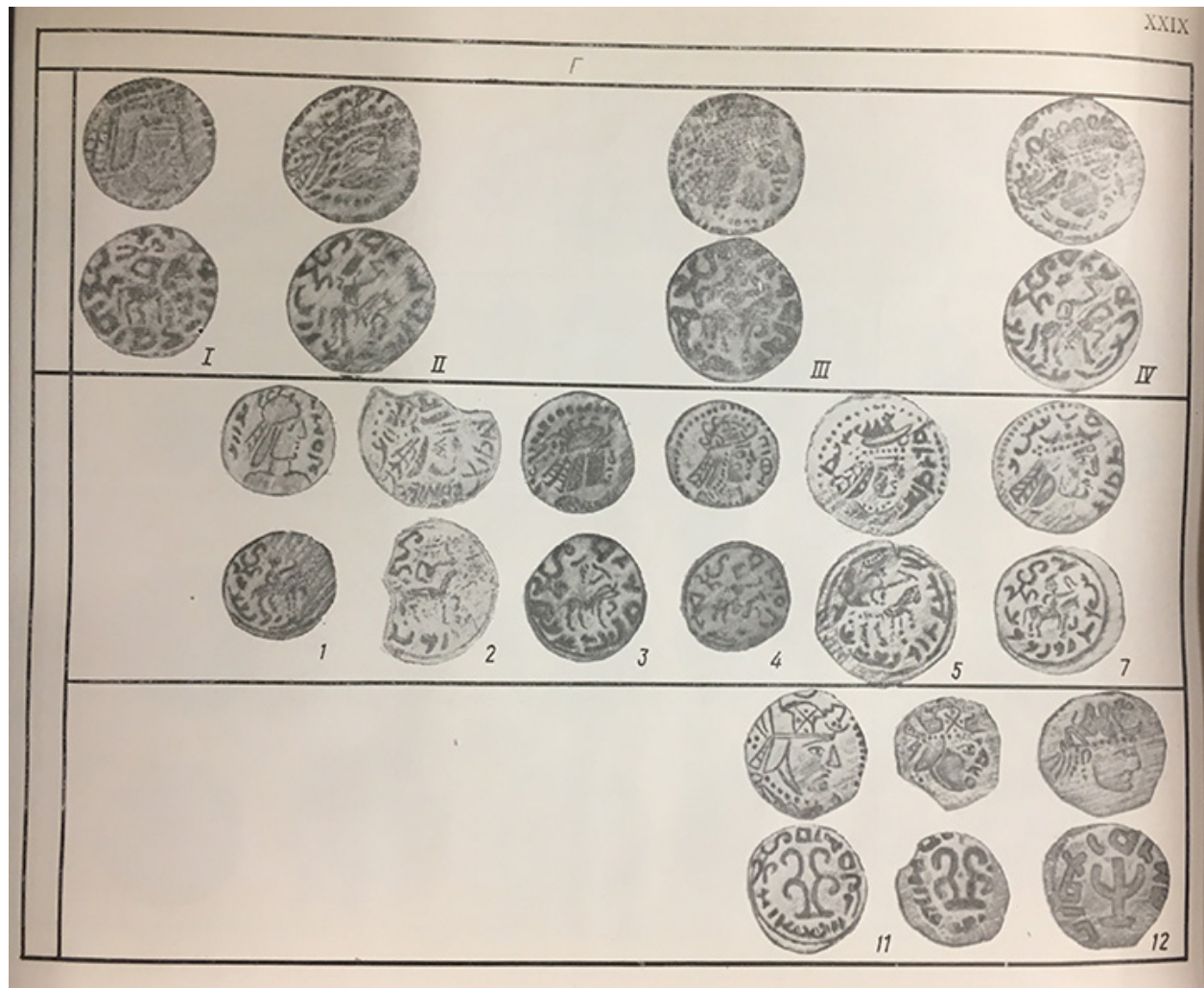
Figure 5: Khwarezmian coins (from Vainberg 1977: Table 19).
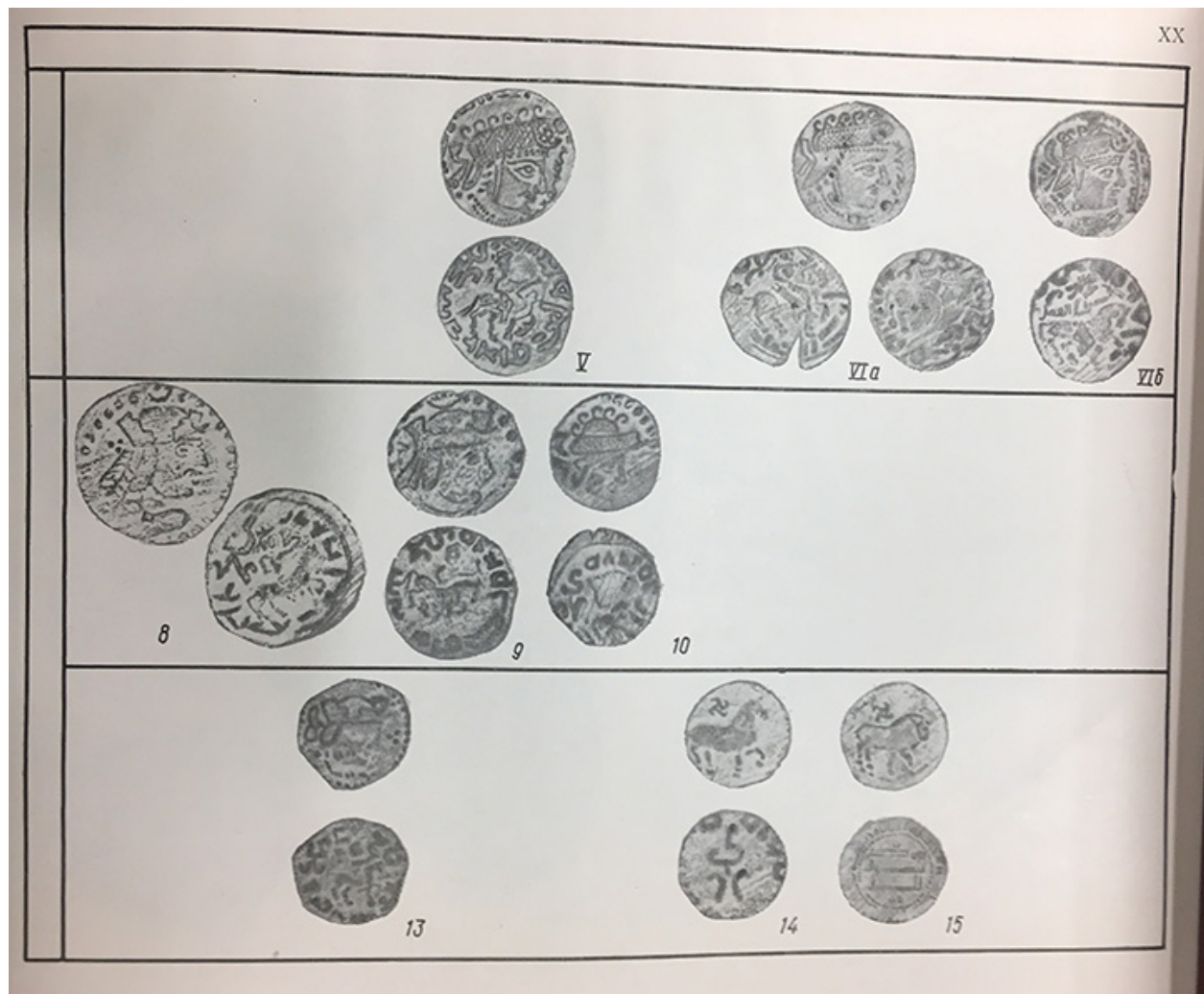
Figure 6: Khwarezmian coins (from Vainberg 1977: Table 20).

Б²Ⅴ/3	№ по каталогу	Б²Ⅴ	№ по каталогу
	130		209
	134	Б²Ⅵ	
	135		194
	137	Б²Ⅶ	
	138		207
	139	Б²Ⅷ	
	143		218
	144	Б²9	
	148		227
	151		228
	154		234
	156		235
	158		236
Б²Ⅴ/4			237
	164		249
	174	Б²12	
Б²Ⅴ			355
	10		356
	11		357
	12		361
			362
			367
			368
			370
			378

Figure 7: Inscriptions on Khwarezmian coins (from Vainberg 1977: Table 1).

Figure 8: Inscriptions on Khwarezmian coins (from Vainberg 1977: Table 2).
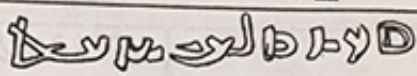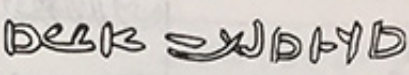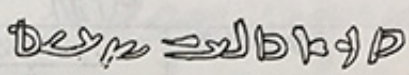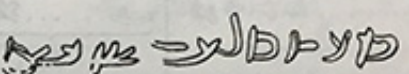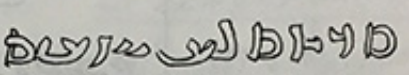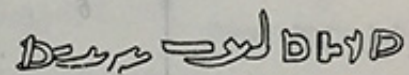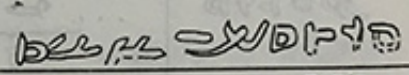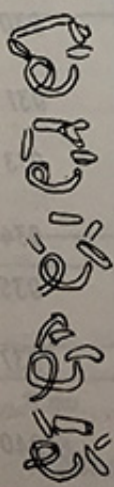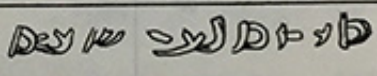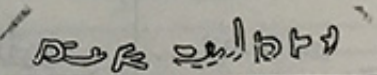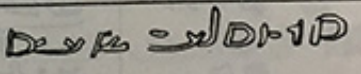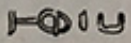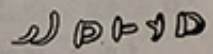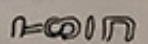
Figure 9: Inscriptions on Khwarezmian coins (from Vainberg 1977: Table 3).

Figure 10: Inscriptions on Khwarezmian coins (from Vainberg 1977: Table 4).

л. ст.		об. ст.	
ГII/2			
ꟾꟾꟾ XbⱵ	ꟾꟾꟾ	.... ꟾⱵ∞ꟾ	959
ГIII/5			
ꟾꟾ ꟾ Ⱶ⅄D	ꟾꟾ	ꟾ ꟾⱵ ∞ꟾ ꟾ	1000
D< ꟾ	D ꟾ	ꟾ ꟾ∞∞ꟾ ꟾꟾ	1001
ꟾ ꟾ	⅄D	ꟾ ꟾ∞∞ꟾ ꟾꟾ	1002
ГIV/6			
ꟾꟾꟾ D ꟾ D	D ꟾ	ꟾ ꟾ∞ꟾ ꟾꟾ ꟾ	1007
ꟾ .... D ꟾ D⅄D	0ꟾ	ꟾ Ⱶ ꟾ.... ꟾ	1008
...... ꟾ D	ꟾ	... ꟾ	1009
ꟾꟾꟾ ꟾ D ꟾꟾ	ꟾ ꟾ	ꟾ ꟾ ꟾꟾ	1010
ꟾꟾꟾ ꟾ D ꟾ D	ꟾꟾ	ꟾ ꟾ ꟾ ꟾ	1011
ꟾꟾꟾ ꟾ D ꟾ ⅄D	ꟾꟾ	ꟾ ꟾ ꟾꟾ Ⱶꟾ	1012
ꟾꟾ⅄ ꟾ ꟾⱵ D Ⱶꟾ	ꟾ⅄	ꟾꟾ ... ꟾ ∞∞∞ ꟾ]	1015
ГIV			
		ꟾ D ꟾ Ⱶ ꟾⱵ ꟾꟾꟾ	1004
		D ꟾ D ꟾⱵ ꟾꟾꟾ	1005
		ꟾ D ꟾ ꟾⱵ ꟾꟾꟾ	1006
ГIV/8			
		⅄ ꟾꟾ Ⱶ ꟾ D Ⱶꟾ	1018
Г13			
ꟾ ∞ꟾ	1365	ꟾ ꟾ∞∞ꟾ ꟾ ꟾ	1364
...ꟾ ꟾꟾ...	1376	ꟾ ꟾ D	1384
... ꟾ ∞	1371	ꟾ ꟾ D Ⱶ ꟾ ꟾ ꟾ	1372
ꟾ ∞ ꟾ	1371	ꟾ ꟾ	1366

Figure 11: Inscriptions on Khwarezmian coins (from Vainberg 1977: Table 5).

Figure 12: Inscriptions on Khwarezmian coins (from Vainberg 1977: Table 6).

Figure 13: Inscriptions on Khwarezmian coins (from Vainberg 1977: Table 7).
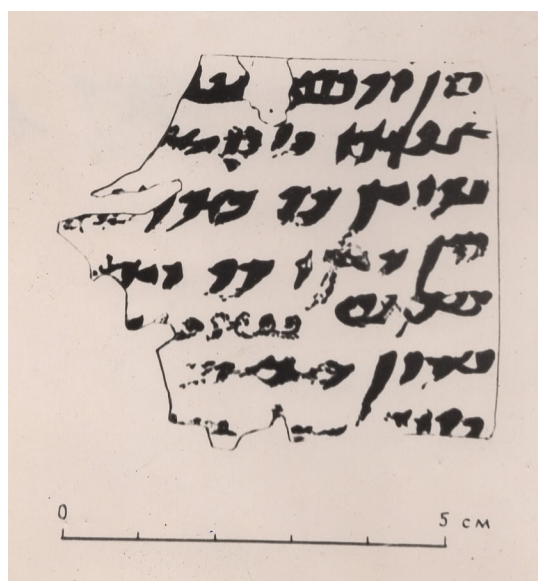
Figure 14: Fragment of a leather document with Khwarezmian inscription from Yakke Parsan. Image courtesy of Lurje.

Letter	initial	internal	final	independent	left join	remarks
'					yes	
b					yes	
g					yes	
d					no	
h					no	
w					no	
z					no	
x					no	
t					yes	joins without horizontal line
y					no	
k					yes | no?	
[l]						ideogram 'L
m					no	
n					yes	
s					yes	
[']					no	ideograms 'L 'BDt
p					yes/no	
r					no	
š					no	
ligatures tn 'n k'n bg by						

Figure 15: Nominal and positional forms of letters in silver vessel inscriptions (from Lurje 2017).
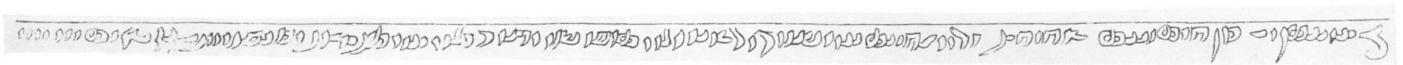
Pl. 1:a   *Choresmian No. 1:* A.D. 658. Inscription from silver phiale in the British Museum, Smirnov, *VS*, pl. XIX:43.

Pl. 1:b   *Choresmian No. 1:* A.D. 658. Silver phiale in the British Museum, see pl. 1:a. Diam. 12.7 cm. Photo courtesy the
Trustees of the British Museum.

Figure 16: Silver vessel #1: 658 CE (from Azarpay 1969: Plate 1:a, b). Silver philae in the British
museum. Original from Smirnov 1909, plate XIX: 43.
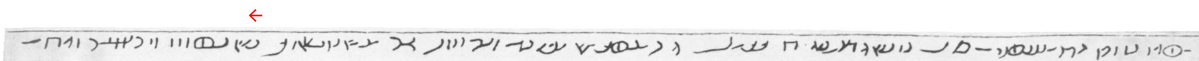
Pl. 3:a *Choresmian No. 2:* A.D. 538 (probably 638). Inscription from silver phiale in the Hermitage Museum, Leningrad, Smirnov, *VS*, pl. XIX: 42.

BŠNT 3 2 100 20 20 20 10 YRX ʾ ʾtry BYWM bgy ʾpbntn yʾ MN bwnšrgk-š x(r)kʾn ʿL ʾtršnky bg ZNH ZWZN 20 10 ʾšpynšwk

Pl. 3:b *Choresmian No. 2:* A.D. 538 (probably 638). Silver phiale in the Hermitage Museum, Leningrad, see pl. 3:a. Diam. 10.6 cm.

Pl. 3:c *Choresmian No. 2:* A.D. 538 (probably 638). Silver phiale in the Hermitage Museum, Leningrad, see pl. 3:b. Smirnov, *VS*, pl. XVIII:42.

Figure 17: Silver bowl #2: either 538 or 638 CE (from Azarpay 1969: Plate 3:a, b, c). Silver philae in the Hermitage Museum (St. Petersburg). Original from Smirnov 1909, plate XIX:42 and XVIII: 42. Transliteration from Lurje (forthcoming).
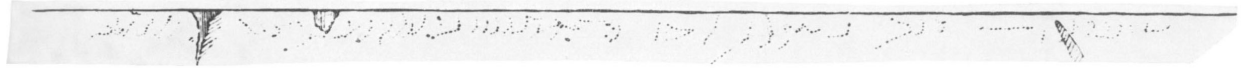
Pl. 5:a   *Choresmian No. 3:* inscription from silver phiale in the Hermitage Museum, Leningrad, Smirnov, *VS*, pl. XIX:44.
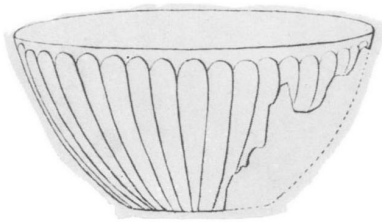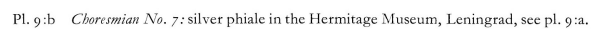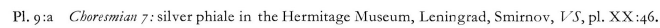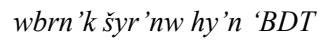

Pl. 5:b   *Choresmian No. 3:* silver phiale in the Hermitage Museum, Leningrad,
see pl. 5:a, diam. 12.5 cm. Smirnov, *VS*, pl. XVIII: 44.
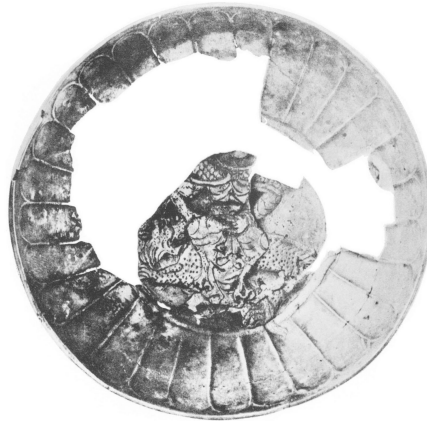
Pl. 5: c   *Choresmian No. 3:* silver phiale in the Hermitage Museum, Leningrad, see pl. 5: b.

Figure 18: Silver vessel #3 (from Azarpay 1969: Plate 5:a, b, c). Silver philae in the Hermitage Museum (St. Petersburg). Original from Smirnov 1909, plate XIX: 44 and XVIII: 44.

ꞇꞇ —ꞇꞇꞇ ꞇꞇ ... ꞇ ꞇ ... ꞇꞇꞇꞇ ꞇ—ꞇ ꞇꞇꞇꞇ

ʾpntn yʾ ... ... zmhy ZWZN-ʾ 20 20

Pl. 8:a
Choresmian No. 4: inscription from silver phiale in the Hermitage
Museum, Leningrad, Smirnov, *VS*, pl. XIX:45.
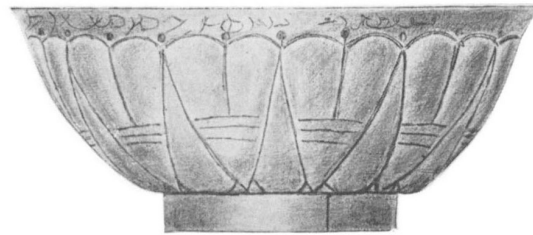
Pl. 8:b   *Choresmian No. 4:* silver phiale in the Hermitage Museum,
Leningrad, see pl. 8:a. Diam. 12.7 cm.

Pl. 8:c   *Choresmian No. 4:* silver phiale in the Hermitage Museum,
Leningrad, see pl. 8:b. Smirnov, *VS*, pl. XVIII:45.

Figure 19: Silver vessel #4 (from Azarpay 1969: Plate 8:a, b, c). Silver philae in the Hermitage
Museum (St. Petersburg). Original from Smirnov 1909, plate XIX: 45 and XVIII: 45.

ᵓpbntn yᵓ MN wrmwzbntk ᵓrᵓškrk ᶜL (g)nyt ᶜL byrty zmhy ZWZN-ᵓ 20 20 3 2

Pl. 11:a, b, c   *Choresmian No. 5:* silver phiale in the Hermitage Museum, Leningrad, Smirnov, *VS*, pl. XIX:47. Diam. 13 cm.

Figure 20: Silver vessel #5 (from Azarpay 1969: Plate 11:a, b, c). Silver philae in the Hermitage Museum (St. Petersburg). Original from Smirnov 1909, plate XIX: 47. Transliteration from Lurje (forthcoming).

רם חוטבֿ נונ— וווו 𝔖𝔷𝔷 ווווווו

gty (xwpsk | xw ksp) ZWZN-' 4 10 20 20 20 7

Figure 21: Silver vessel #6. Original from Smirnov 1909, plate L: 84. Transliteration from Lurje (forthcoming).

וגנׁיﬞﬞﬞﬞ שׁﬞﬞﬞﬞ לﬞﬞﬞ ﬞﬞﬞﬞ

wbrn'k šyr'nw hy'n 'BDT

Pl. 9:a   *Choresmian 7:* silver phiale in the Hermitage Museum, Leningrad, Smirnov, *VS*, pl. XX:46.

Pl. 9:b   *Choresmian No. 7:* silver phiale in the Hermitage Museum, Leningrad, see pl. 9:a.

Figure 22: Silver bowl #7 (from Azarpay 1969: Plate 9:a, b, c). Silver philae in the Hermitage Museum (St. Petersburg). Original from Smirnov 1909, plate XX: 46. Transliteration from Lurje (forthcoming).

ᵓpbntn yᵓ MN ... ... zmhy ZWZN-ᵓ 20 20

Pl. 10:a, b, c   *Choresmian No. 8:* silver phiale in the Hermitage Museum, Leningrad, Smirnov, *VS*, pl. CXIV: 286.

Figure 23: Silver bowl #8 (from Azarpay 1969: Plate 10:a, b, c). Silver philae in the Hermitage Museum (St. Petersburg). Original from Smirnov 1909, plate CXIV: 286.
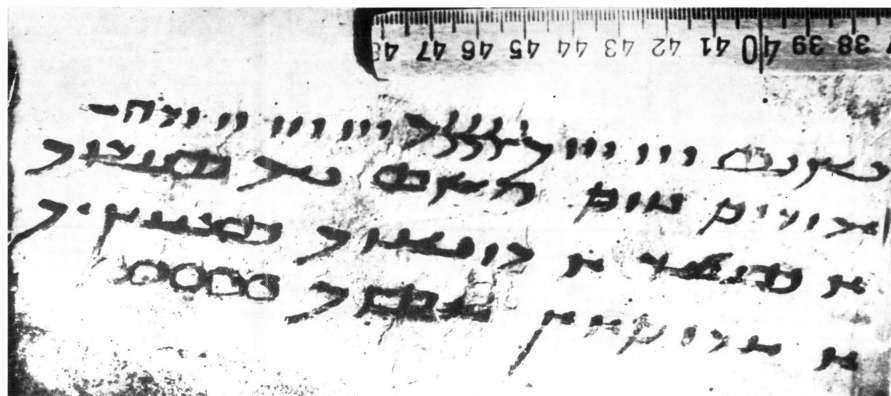
BŠNT III III C XX XX XX X III III II YRḤ ʾ
ʾhwrym BYWM gwšt ZNH tnbryk
ʾy tnbʾr ʾy hwnšk (?) tʾbʾnʾnʾk (?)
ʾy ʾyrw|zmʾwʾn ʾztyk ○○○○

Year 678. Month Ahurem, day Gost. This
ossuary contains the body of hwnsk
tʾbʾnʾnʾk, son of ʾyrw|zmʾwʾn

Figure 24: Tok Kala no. 52, ossuary inscription (from Tolstov and Livshitz 1964: Figure 1).
Transliterations from same; but may be erroneous or outdated.

Tolstov and Livshitz 1964

BŠNT III III I C III III YRḤ'
brwrtn BYWM brwrtn ZN[H] tnbryk
nwšy (?) 'y srywyk tyšy'n'ny 'rw'n
GD kw'n[y] 'y 'rw'n 'L nwš grdm'n
pr'ny'ty

Year 705. Month Rawacina, day Rawacina. This is the ossuary of srywyk [son] of tysy'n, soul [whose] [possesses] kayan farrah. Soul [his] may be sent to the beautiful Paradise.

Henning 1965

BŠNT vii C vi YRḤ'
βrwrtn BYWM βrwrtn ZNH tpnkwk
NPŠY 'y srwywk tyšy'n'n-w 'rw'n
'D hw'n-˙˙ 'y 'rw'n 'L nwš γrδm'n
m'ny'(')ty

In the year 706, on the 19th day of the first month. This chest is the property of the soul of *Sraw-yōk*, the son of *Tiš-yān*. May their souls rest in the eternal Paradise.

Figure 25: Tok Kala no. 25, ossuary inscription (image from Tolstov and Livshitz 1964: Figure 2). Representation in Khwarezmian script based upon Henning 1965.

Fig. 3. Inscription №. 19.

שאוש ווווו וווו ⟨glyph⟩ ווווו וווו חטצויק
ו–– ואיטבדון ניטוי–– דוון ימסא איויק

BŠNT IIII III C XX X IIII IIII tnbryk
y' w'z'sw|ydyn nwšy' ?grn 'rt'w 'rw'n

Year 738. This is the ossuary of w'z'swdyn (?)
[May] in the beautiful Paradise [be sent his] true soul.

Figure 26: Tok Kala no. 19, ossuary inscription (from Tolstov and Livshitz 1964: Figure 3).
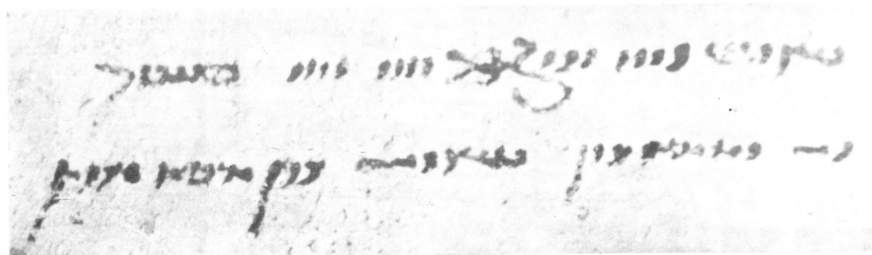Transliterations from same; but may be erroneous or outdated.

Fig. 4. Inscription №. 8.

ZNH tn[b]⌈r⌉yk
'(?)ynšy ⌈šh⌉k '.wn⌈y⌉.
…'y nykšy ? 'YK
MN ty'zhwndy 'L … [

This is the ossuary of woman (? shk, daughter of '.w …
May [soul her be sent] from the [world] of full danger to
(the world of safety?).

Figure 27: Tok Kala no. 8, ossuary inscription (from Tolstov and Livshitz 1964: Figure 4). Transliterations from same; but may be erroneous or outdated.
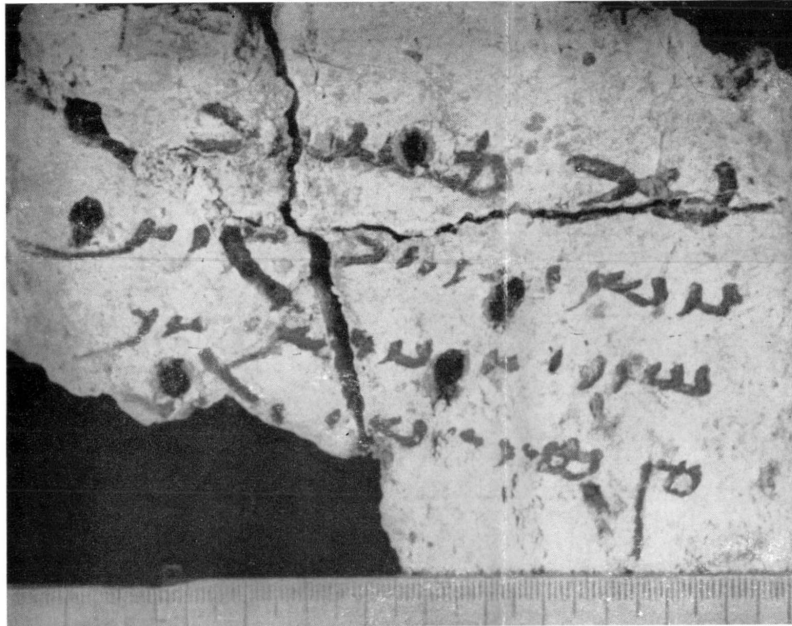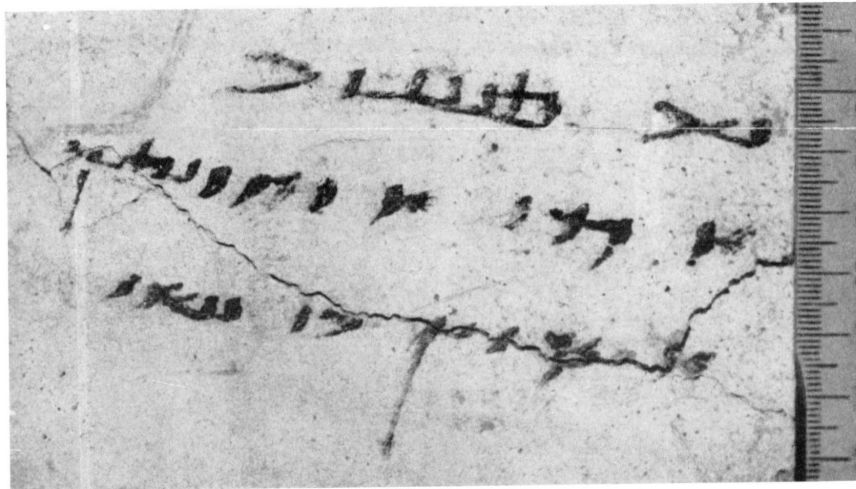
Fig. 5. Inscription N⁰. 69.

Tolstov and Livshitz 1964　　　　　　Reanalysis based upon Henning 1965

ZNH tnbryk　　　　　　　　　　　　*ZNH tpnkwk*
'y gry 'y wḥwntk　　　　　　　　　　　*'y gry 'y wḥwntk*
'y 'rw'n kw nwšy　　　　　　　　　　*'y 'rw'n hw NPŠY*

This ossuary contains the body of wnwnxk
Soul [his may be sent] to the beautiful
[Paradise].

Figure 28: Tok Kala no. 69, ossuary inscription (from Tolstov and Livshitz 1964: Figure 5). Representation in Khwarezmian script based upon reanalysis.

Fig. 6. Inscription No. 39.

ꜯꜯ ...

ᴦBŠNTᴧ III III C XX XX XX X YRḤ'
mᴦtry?ᴧ ᴦBYWᴧM whwmn ZNH
tnᴦbrᴧyk 'y tnb'r 'rw'zd
w … nᴦyᴧ zyt brwrtyk
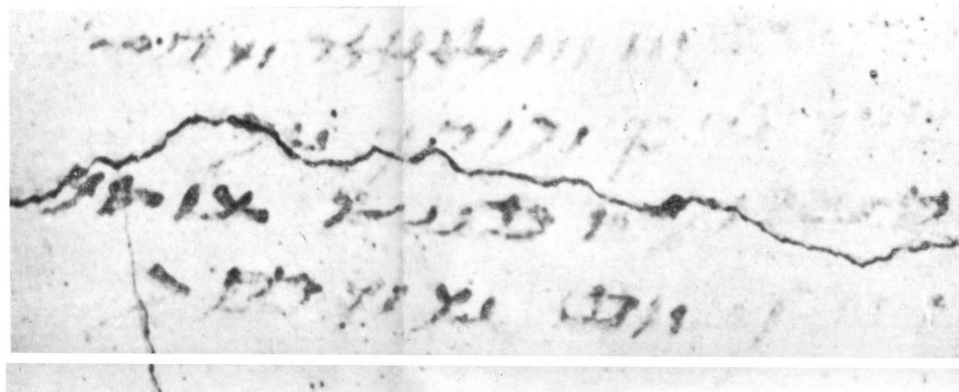
Year 690, month of Miri, day of Ahumen. This ossuary
holds the body of 'rw'zd w…n, son of Hravardik.
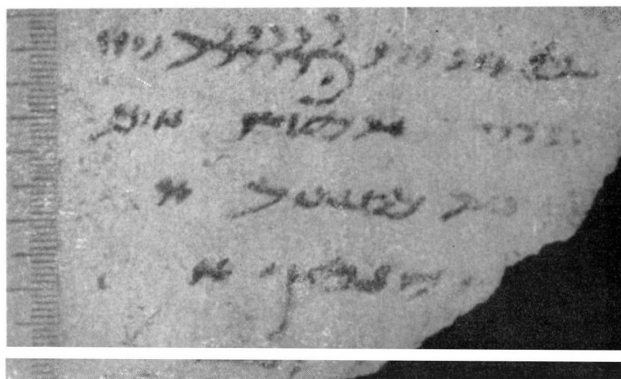
Figure 29: Tok Kala no. 39, ossuary inscription (from Tolstov and Livshitz 1964: Figure 6).
Transliterations from same; but may be erroneous or outdated.

Fig. 7. Inscription № 12

ꗃꘐꚠ ۱۱۱۱ ۱۱۱ ۱۱۱۱ ۞꘠꘠꘠꘠꘠꘡ ۱۱۱۱
ꔅꕷꕞ— ꔆꕸꖢꗈ ꕞꖢꗉ
ꕸꖢ ꕼꕞꕳꕵ ꖴ
ꕼꕷꕧ꘎ ꖴ
[…]

⌐BŠ⌐NT III III C XX XX XX XX X IIII
YRḤ' 'rtwyš BYWM
]. ZNH tnbryk 'y
]. s|ḥnt'ny 'y
]. y'

Year 694, month of Ardwis, day [] This ossuary ...
[of son] of …s|ḥnt…

Figure 30: Tok Kala no. 12, ossuary inscription (from Tolstov and Livshitz 1964: Figure 7). Transliterations from same; but may be erroneous or outdated.
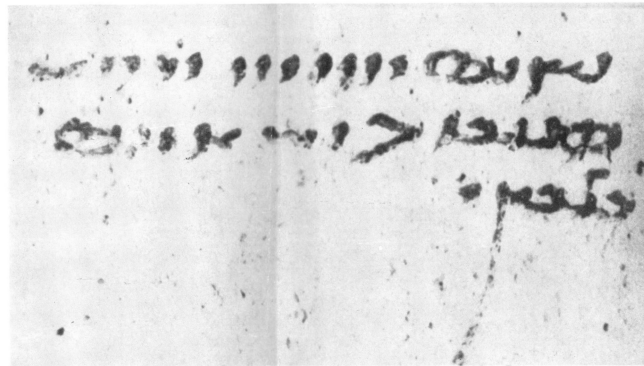
Fig. 8. Inscription No. 26.

כאיⴟ ווווו ווו ווⴟ⌐
ⴟבבול וⴗ ש וⴟⴟ
כⴟⴗⵊ

BŠNT III IIII YRḤ'
tnbryk y' 'y wrt
k'k'ny

Year 7[00]. Month. This is the ossuary of wrt, [of son] of k'k.

Figure 31: Tok Kala no. 26, ossuary inscription (from Tolstov and Livshitz 1964: Figure 8). Transliterations from same; but may be erroneous or outdated.

Fig. 9. Inscription №. 21.

ﬡﬡﬡﬡ
ﬡﬡﬡ
ﬡﬡﬡﬡﬡ
ﬡﬡﬡﬡﬡ
ﬡﬡﬡﬡﬡ

BŠN⌈T⌉ [
YR⌈H'⌉ [
whwmn' [ZHN tnbryk 'y ḥw'r'n [
wḥnwy(?) [

Figure 32: Tok Kala no. 21, ossuary inscription (from Tolstov and Livshitz 1964: Figure 9). Transliterations from same; but may be erroneous or outdated.
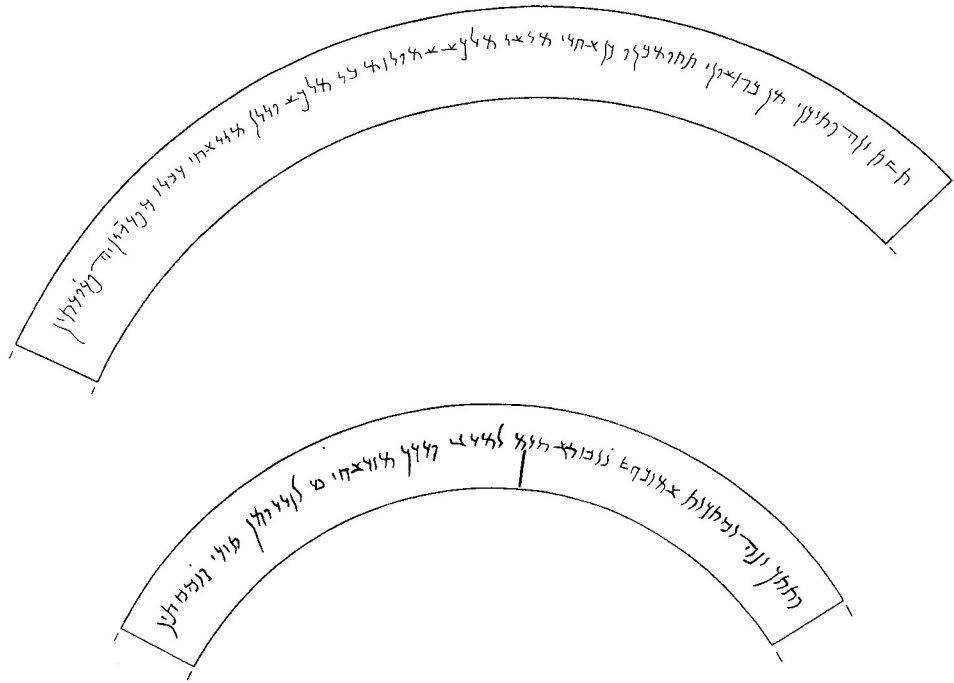
Figure 33: Archaic Khwarezmian inscriptions on silver bowls no. 1 (top) and no. 2 (bottom) from Isakovka from the Achaemenid period (from Livshits 2003: 152, 163). This type of script is out of scope for the present encoding, and may be unified with Imperial Aramaic.
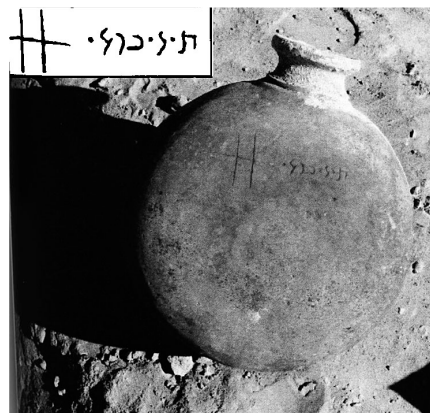

Figure 34: A lapidary Khwarezmian inscription from Chirik-rabat, likely dated between the 2nd and 5th century BCE (from Ivantchik and Lurje 2013: 286). The likely reading is *tyrybwdy*. This type of script is out of scope for the present encoding, and may be unified with Imperial Aramaic.

	Имперский арамейский		Парфянский		Бактр. (?)	Согдийский				Хорезмийский						
	Бактрия 1)	Арахосия2)	Ниса 3)	Авроман 4)	Ай-Ханум 5)	Афрасиаб 6)	Куль-тобе 7)	Ст. письма 8)	Шатиал 9)	Айбуйир 10)	Исаковка 11)	Бурлы-кала 12)	Калалы-гыр 13)	Кой-Крылган 14)	Чирик-рабат	
ʾ																ʾ
b																b
g																g
d																d
h																h
w																w
z																z
ḥ																ḥ
t																t
y																y
k																k
l																l
m																m
n																n
s																s
ʿ																ʿ
p																p
ṣ																ṣ
q																q
r																r
š																š
t																t

1 — *Shaked, Naveh* 2012, Doc. A1;
2–5 В.А. Лившиц по *Расторгуева, Молчанова* 1981;
6 — *Grenet* 2006;
7 — *Sims-Williams, Grenet*, 2006;
8, 9 — *Исхаков* 2008. Табл. XI, XV;

10 — В.А. Лившиц по *Мамбетуллаев* 1979;
11 — *Лившиц* 2002;
12 — *Лившиц, Мамбетуллаев* 1985;
13 — *Лившиц* 2004;
14 — *Толстов, Вайнберг* 1967. С. 220.

Илл. 2. Знаки чирик-рабатской надписи в сравнении с другими письменностями древней Средней Азии

Figure 35: Comparison of early Iranian lapidary script types derived from Imperial Aramaic (from Ivantchik and Lurje 2013: 290).