The Unicode® Standard
Version 14.0 – Core Specification

To learn about the latest version of the Unicode Standard, see https://www.unicode.org/versions/latest/.

The Unicode Standard / the Unicode Consortium; edited by the Unicode Consortium. — Version 14.0.

Includes index.

ISBN 978-1-936213-29-0 (https://www.unicode.org/versions/Unicode14.0.0/)

1. Unicode (Computer character set) I. Unicode Consortium.

QA268.U545 2021

ISBN 978-1-936213-29-0

Published in Mountain View, CA

September 2021

Chapter 10

Middle East-II

Ancient Scripts

This chapter covers a number of ancient scripts of the Middle East. All of these scripts were written right to left.

Old North Arabian	***Manichaean***	***Elymaic***
Old South Arabian	***Parthian and Pahlavi***	***Nabataean***
Phoenician	***Avestan***	***Palmyrene***
Imperial Aramaic	***Chorasmian***	***Hatran***

Old North Arabian and Old South Arabian are two branches of the South Semitic script family used in and around Arabia from about the tenth century BCE to the sixth century CE. The Old South Arabian script was used around the southwestern part of the Arabian peninsula for 1,200 years beginning around the 8th century BCE. Carried westward, it was adapted for writing the Ge'ez language, and evolved into the root of the modern Ethiopic script.

The Phoenician alphabet was used in various forms around the Mediterranean. It is ancestral to Latin, Greek, Hebrew, and many other scripts—both modern and historical.

The Imperial Aramaic script evolved from Phoenician and was the source of many other scripts, such as the square Hebrew and the Arabic script. Imperial Aramaic was used to write the Aramaic language beginning in the eighth century BCE, and was the principal administrative language of the Assyrian empire and then the official language of the Achaemenid Persian empire. Inscriptional Parthian, Inscriptional Pahlavi, and Avestan are also derived from Imperial Aramaic, and were used to write various Middle Persian languages.

Psalter Pahlavi is a cursive alphabetic script used to write the Middle Persian language during the 6th or 7th century CE. It is a historically conservative variety of Pahlavi used by Christians in the Neo-Persian empire.

The Chorasmian script was used between the 2nd century and 8th to 9th centuries CE primarily to write the Chorasmian language, an Eastern Iranian language. The script was derived from Imperial Aramaic and is related to Parthian, Inscriptional Pahlavi, Psalter Pahlavi, Book Pahlavi, and Old Sogdian.

The Manichaean script is a cursive alphabetic script related to Syriac, as well as Palmyrene Aramaic. The script was used by those practicing the Manichaean religion, which was founded during the third century CE in Babylonia, and spread widely over the next four centuries before later vanishing.

The Elymaic script was used to write Achaemenid Aramaic in the state of Elymais, which flourished from the second century BCE to the early third century CE and was located in the southwestern portion of modern-day Iran. Elymaic derives from the Aramaic script and is closely related to Parthian and Mandaic.

The Nabataean script developed from the Aramaic script and was used to write the language of the Nabataean kingdom. The script was in wide use from the second century BCE to the fourth century CE. It is generally considered the precursor of the Arabic script.

The Palmyrene script was derived from the customary forms of Aramaic developed during the Achaemenid empire. The script was used for writing the Palmyrene dialect of West Aramaic, and is known from inscriptions and documents found mainly in the city of Palmyra and other cities in the region of Syria, dating from 44 BCE to about 280 CE.

The Hatran script belongs to the North Mesopotamian branch of the Aramaic scripts, and was used for writing a dialect of the Aramaic language. The script is known from inscriptions discovered in the ancient city of Hatra, in present-day Iraq, dating from 98–97 BCE until circa 241 CE.

10.1 Old North Arabian

Old North Arabian: U+10A80–U+10A9F

Old North Arabian, or Ancient North Arabian, refers to a group of scripts used in the western two-thirds of Arabia and the Levant, from Syria to the borders of Yemen. Old North Arabian is a member of the South Semitic script family, which was used exclusively in Arabia and environs, and is a relative of the Old South Arabian script. The earliest datable Old North Arabian texts are from the mid-sixth century BCE. The script is thought to have fallen out of use after the fourth century CE. The encoding of Old North Arabian is based on the Dadanitic form, which is attested in many formal inscriptions on stelae and rock-faces, and hundreds of graffiti used in the oasis of Dadan (Dedān, modern al-ʻUlā) in northwest Saudi Arabia.

Other forms of the Old North Arabian script, such as Minaic, Safaitic, Hismaic, Taymanitic and Thamudic B, have many variant forms of the letters. Dialect-specific fonts can be used to render these variant forms.

Structure. Old North Arabian is an alphabetic script consisting only of consonants; vowels are not indicated in the script, though some Dadanitic texts do make limited use of consonant letters to write long vowels (*matres lectionis*). The script has been encoded with right-to-left directionality, which is typical for Dadanitic. Glyphs may be mirrored in lines when they have left-to-right directionality.

Ordering. Traditional sorting orders are poorly attested. Modern scholars specializing in Old North Arabian prefer the South Semitic alphabetical order shown in the code charts.

Numbers. Three numbers are attested in Old North Arabian: one, ten, and twenty. The numbers have right-to-left directionality.

Punctuation. A vertical word separator is usually used between words in Dadanitic, but this is not widely used in the other Old North Arabian alphabets. U+10A9D OLD NORTH ARABIAN NUMBER ONE is used to represent both this punctuation and the digit one.

10.2 Old South Arabian

Old South Arabian: U+10A60–U+10A7F

The Old South Arabian script was used on the Arabian peninsula (especially in what is now Yemen) from the 8th century BCE to the 6th century CE, after which it was supplanted by the Arabic script. It is a consonant-only script of 29 letters, and was used to write the southwest Semitic languages of various cultures: Minean, Sabaean, Qatabanian, Hadramite, and Himyaritic. Old South Arabian is thus known by several other names including Mino-Sabaean, Sabaean and Sabaic. It is attested primarily in an angular form ("Musnad") in monumental inscriptions on stone, ceramic material, and metallic surfaces; however, since the mid 1970s examples of a more cursive form ("Zabur") have been found on softer materials, such as wood and leather.

Around the end of the first millennium BCE, the westward migration of the Sabaean people into the Horn of Africa introduced the South Arabic script into the region, where it was adapted for writing the Ge'ez language. By the 4th century CE the script for Ge'ez had begun to change, and eventually evolved into a left-to-right syllabary with full vowel representation, the root of the modern Ethiopic script (see *Section 19.1, Ethiopic*).

Directionality. The Old South Arabian script is typically written from right to left. Conformant implementations of Old South Arabian script must use the Unicode Bidirectional Algorithm (see Unicode Standard Annex #9, "Unicode Bidirectional Algorithm"). However, some older examples of the script are written in boustrophedon style, with glyphs mirrored in lines with left-to-right directionality.

Structure. The character repertoire of Old South Arabian corresponds to the repertoire of Classical Arabic, plus an additional letter presumed analogous to the letter *samekh* in West Semitic alphabets. This results in four letters for different kinds of "s'"sounds. While there is no general system for representing vowels, the letters U+10A65 OLD SOUTH ARABIAN LETTER WAW and U+10A7A OLD SOUTH ARABIAN LETTER YODH can also be used to represent the long vowels *u* and *i*. There is no evidence of any kind of diacritical marks; geminate consonants are indicated simply by writing the corresponding letter twice, for example.

Segmentation. Letters are written separately, there are no connected forms. Words are not separated with space; word boundaries are instead marked with a vertical bar. The vertical bar is indistinguishable from U+10A7D "𐩽" OLD SOUTH ARABIAN NUMBER ONE—only one character is encoded to serve both functions. Words are broken arbitrarily at line boundaries in attested materials.

Monograms. Several letters are sometimes combined into a single group, in which the glyphs for the constituent characters are overlaid and sometimes rotated to create what appears to be a single unit. These combined units are traditionally called *monograms* by scholars of this script.

Numbers. Numeric quantities are differentiated from surrounding text by writing U+10A7F 𐩿 OLD SOUTH ARABIAN NUMERIC INDICATOR before and after the number. Six characters have numeric values as shown in *Table 10-1*—four of these are letters that double as numeric values, and two are characters not used as letters.

Table 10-1. Old South Arabian Numeric Characters

Code Point	Glyph	Numeric function	Other function
10A7F	𐩿	numeric separator	
10A7D	𐩽	1	word separator
10A6D	𐩭	5	kheth
10A72	𐩲	10	ayn
10A7E	𐩾	50	
10A63	𐩣	100	mem
10A71	𐩱	1000	alef

Numbers are built up through juxtaposition of these characters in a manner similar to that of Roman numerals, as shown in *Table 10-2*. When 10, 50, or 100 occur preceding 1000 they serve to indicate multiples of 1000. The example numbers shown in *Table 10-2* are rendered in a right-to-left direction in the last column.

Table 10-2. Number Formation in Old South Arabian

Value	Schematic	Character Sequence	Display
1	1	10A7D	𐩽
2	1 + 1	10A7D 10A7D	𐩽𐩽
3	1 + 1 + 1	10A7D 10A7D 10A7D	𐩽𐩽𐩽
5	5	10A6D	𐩭
7	5 + 1 + 1	10A6D 10A7D 10A7D	𐩽𐩽𐩭
16	10 + 5 + 1	10A72 10A6D 10A7D	𐩽𐩭𐩲
1000	1000	10A71	𐩱
3000	1000 + 1000 + 1000	10A71 10A71 10A71	𐩱𐩱𐩱
10000	10 × 1000	10A72 10A71	𐩱𐩲
11000	10 × 1000 + 1000	10A72 10A71 10A71	𐩱𐩱𐩲
30000	(10 + 10 + 10) × 1000	10A72 10A72 10A72 10A71	𐩱𐩲𐩲𐩲
30001	(10 + 10 + 10) × 1000 + 1	10A72 10A72 10A72 10A71 10A7D	𐩽𐩱𐩲𐩲𐩲

Character Names. Character names are based on those of corresponding letters in northwest Semitic.

10.3 Phoenician

Phoenician: U+10900–U+1091F

The Phoenician alphabet and its successors were widely used over a broad area surrounding the Mediterranean Sea. Phoenician evolved over the period from about the twelfth century BCE until the second century BCE, with the last neo-Punic inscriptions dating from about the third century CE. Phoenician came into its own from the ninth century BCE. An older form of the Phoenician alphabet is a forerunner of the Greek, Old Italic (Etruscan), Latin, Hebrew, Arabic, and Syriac scripts among others, many of which are still in modern use. It has also been suggested that Phoenician is the ultimate source of Kharoshthi and of the Indic scripts descending from Brahmi.

Phoenician is an historic script, and as for many other historic scripts, which often saw continuous change in use over periods of hundreds or thousands of years, its delineation as a script is somewhat problematic. This issue is particularly acute for historic Semitic scripts, which share basically identical repertoires of letters, which are historically related to each other, and which were used to write closely related Semitic languages.

In the Unicode Standard, the Phoenician script is intended for the representation of text in Paleo-Hebrew, Archaic Phoenician, Phoenician, Early Aramaic, Late Phoenician cursive, Phoenician papyri, Siloam Hebrew, Hebrew seals, Ammonite, Moabite, and Punic. The line from Phoenician to Punic is taken to constitute a single continuous branch of script evolution, distinct from that of other related but separately encoded Semitic scripts.

The earliest Hebrew language texts were written in the Paleo-Hebrew alphabet, one of the forms of writing considered to be encompassed within the Phoenician script as encoded in the Unicode Standard. The Samaritans who did not go into exile continued to use Paleo-Hebrew forms, eventually developing them into the distinct Samaritan script. (See *Section 9.4, Samaritan*.) The Jews in exile gave up the Paleo-Hebrew alphabet and instead adopted Imperial Aramaic writing, which was a descendant of the Early Aramaic form of the Phoenician script. (See *Section 10.4, Imperial Aramaic*.) Later, they transformed Imperial Aramaic into the "Jewish Aramaic" script now called (Square) Hebrew, separately encoded in the Hebrew block in the Unicode Standard. (See *Section 9.1, Hebrew*.)

Some scholars conceive of the language written in the Paleo-Hebrew form of the Phoenician script as being quintessentially Hebrew and consistently transliterate it into Square Hebrew. In such contexts, Paleo-Hebrew texts are often considered to simply *be* Hebrew, and because the relationship between the Paleo-Hebrew letters and Square Hebrew letters is one-to-one and quite regular, the transliteration is conceived of as simply a font change. Other scholars of Phoenician transliterate texts into Latin. The encoding of the Phoenician script in the Unicode Standard does not invalidate such scholarly practice; it is simply intended to make it possible to represent Phoenician, Punic, and similar textual materials directly in the historic script, rather than as specialized font displays of transliterations in modern Square Hebrew.

Directionality. Phoenician is written horizontally from right to left. The characters of the Phoenician script are all given strong right-to-left directionality.

Punctuation. Inscriptions and other texts in the various forms of the Phoenician script generally have no space between words. Dots are sometimes found between words in later exemplars—for example, in Moabite inscriptions—and U+1091F PHOENICIAN WORD SEPARATOR should be used to represent this punctuation. The appearance for this word separator is somewhat variable; in some instances it may appear as a short vertical bar, instead of a rounded dot.

Stylistic Variation. The letters for Phoenician proper and especially for Punic have very exaggerated descenders. These descenders help distinguish the main line of Phoenician script evolution toward Punic, as contrasted with the Hebrew forms, where the descenders instead grew shorter over time.

Numerals. Phoenician numerals are built up from six elements used in combination. These include elements for one, two, and three, and then separate elements for ten, twenty, and one hundred. Numerals are constructed essentially as tallies, by repetition of the various elements. The numbers for two and three are graphically composed of multiples of the tally mark for one, but because in practice the values for two or three are clumped together in display as entities separate from one another they are encoded as individual characters. This same structure for numerals can be seen in some other historic scripts ultimately descendant from Phoenician, such as Imperial Aramaic and Inscriptional Parthian.

Like the letters, Phoenician numbers are written from right to left: 𐤛𐤚𐤚𐤙 means 143 (100 + 20 + 20 + 3). This practice differs from modern Semitic scripts like Hebrew and Arabic, which use decimal numbers written from left to right.

Character Names. The names used for the characters here are those reconstructed by Theodor Nöldeke in 1904, as given in Powell (1996).

10.4 Imperial Aramaic

Imperial Aramaic: U+10840–U+1085F

The Aramaic language and script are descended from the Phoenician language and script. Aramaic developed as a distinct script by the middle of the eighth century BCE and soon became politically important, because Aramaic became first the principal administrative language of the Assyrian empire, and then the official language of the Achaemenid Persian empire beginning in 549 BCE. The Imperial Aramaic script was the source of many other scripts, including the square Hebrew script, the Arabic script, and scripts used for Middle Persian languages, including Inscriptional Parthian, Inscriptional Pahlavi, and Avestan.

Imperial Aramaic is an alphabetic script of 22 consonant letters but no vowel marks. It is written either in *scriptio continua* or with spaces between words.

Directionality. The Imperial Aramaic script is written from right to left. Conformant implementations of the script must use the Unicode Bidirectional Algorithm. For more information, see Unicode Standard Annex #9, "Unicode Bidirectional Algorithm".

Punctuation. U+10857 IMPERIAL ARAMAIC SECTION SIGN is thought to be used to mark topic divisions in text.

Numbers. Imperial Aramaic has its own script-specific numeric characters with right-to-left directionality. Numbers are built up using sequences of characters for 1, 2, 3, 10, 20, 100, 1000, and 10000 as shown in *Table 10-3*. The example numbers shown in the last column are rendered in a right-to-left direction.

Table 10-3. Number Formation in Aramaic

Value	Schematic	Character Sequence	Display
1	1	10858	𐡘
2	2	10859	𐡙
3	3	1085A	𐡚
4	3 + 1	1085A 10858	𐡘𐡚
5	3 + 2	1085A 10859	𐡙𐡚
9	3 + 3 + 3	1085A 1085A 1085A	𐡚𐡚𐡚
10	10	1085B	𐡛
11	10 + 1	1085B 10858	𐡘𐡛
12	10 + 2	1085B 10859	𐡙𐡛
20	20	1085C	𐡜
30	20 + 10	1085C 1085B	𐡛𐡜
55	20 + 20 + 10 + 3 + 2	1085C 1085C 1085B 1085A 10859	𐡙𐡚𐡛𐡜𐡜
70	20 + 20 + 20 + 10	1085C 1085C 1085C 1085B	𐡛𐡜𐡜𐡜
100	1 × 100	10858 1085D	𐡝𐡘
200	2 × 100	10859 1085D	𐡝𐡙

Table 10-3. Number Formation in Aramaic (Continued)

Value	Schematic	Character Sequence	Display
500	(3 + 2) × 100	1085A 10859 1085D	𐡚𐡹𐡹
3000	3 × 1000	1085A 1085E	𐡚𐡹
30000	3 × 10000	1085A 1085F	𐡚𐡹

Values in the range 1-99 are represented by a string of characters whose values are in the range 1-20; the numeric value of the string is the sum of the numeric values of the characters. The string is written using the minimum number of characters, with the most significant values first. For example, 55 is represented as 20 + 20 + 10 + 3 + 2. Characters for 100, 1000, and 10000 are prefixed with a multiplier represented by a string whose value is in the range 1-9. The Inscriptional Parthian, Inscriptional Pahlavi, Nabataean, Palmyrene, and Hatran scripts use a similar system for forming numeric values.

10.5 Manichaean

Manichaean U+10AC0–U+10AFF

The Manichaean religion was founded during the third century CE in Babylonia, then part of the Sassanid Persian empire. It spread widely over the next four centuries, as far west as north Africa and as far east as China, but had mostly vanished by the fourteenth century. From 762 until around 1000 it was a state religion in the Uyghur kingdom.

The Manichaean script was used by adherents of Manichaeism, and was based on or influenced by the Estrangela form of Syriac, as well as Palmyrene Aramaic. It is said to have been invented by Mani, but may be older. Because of the wide spread of Manichaeism and Mani's decision to spread his teachings in any language available, the Manichaean script was used to write a variety of languages with some variation in character repertoire: the Iranian languages Middle and Early Modern Persian, Parthian, Sogdian, and Bactrian, as well as the Turkic language Uyghur and, to a lesser extent, the Indo-European language Tocharian.

Directionality. The Manichaean script is written from right to left. Conformant implementations of Manichaean script must use the Unicode Bidirectional Algorithm (see Unicode Standard Annex #9, "Unicode Bidirectional Algorithm").

Structure. Manichaean is alphabetic, written with spaces between words. The alphabet includes 24 base letters, two more than Aramaic. There are a total of 36 letters. Ten of these are formed by adding one or two dots above the base letter to represent a spirant or other modified sound. There is also a sign representing the conjunction *ud*.

In addition, two diacritical marks are used to indicate abbreviations, elisions, or plural forms. Manichaean text paid careful attention to the layout of characters, often stretching or shrinking letters, using abbreviations, or eliminating vowels (indicated with elision dots) to achieve desired line widths and to avoid breaking words across lines. Sogdian written in Manichaean script also sometimes shows the use of doubled vowels to fill out a line.

To graphically extend a word, U+0640 ARABIC TATWEEL may be used.

Shaping. Manichaean has shaping rules and rendering requirements that are similar to those for Syriac and Arabic, with joining forms as shown in *Table 10-4*, *Table 10-5*, *Table 10-6* and *Table 10-7*. In these tables, X_n, X_r, X_m, and X_l designate the isolated, final, medial, and initial forms respectively. The dotted letters are not shown separately, because their joining behavior is the same as the corresponding un-dotted letter. Note that Manichaean has two letters with the rare Joining_Type of Left_Joining.

Five Manichaean letters—*daleth*, *he*, *mem*, *nun*, *resh*—have alternate forms whose occurrence cannot be predicted from context, although the alternate forms tend to occur most often at the end of lines. These forms are represented using standardized variation sequences and are shown in the tables that follow.

Table 10-4 lists the dual-joining letters Manichaean. In this and the following tables, the standardized variation sequences are indicated in the joining group column in separate rows showing the relevant joining group plus the variation selector.

Table 10-4. Dual-Joining Manichaean Letters

Joining Group	X_n	X_r	X_m	X_l
ALEPH	𐫀	𐫀	𐫀	𐫀
BETH	𐫁	𐫁	𐫁	𐫁
GIMEL	𐫂	𐫂	𐫂	𐫂
GHIMEL	𐫃	𐫃	𐫃	𐫃
LAMEDH	𐫄	𐫄	𐫄	𐫄
DHAMEDH	𐫅	𐫅	𐫅	𐫅
THAMEDH	𐫆	𐫆	𐫆	𐫆
MEM	𐫇	𐫇	𐫇	𐫇
MEM + VS-1	𐫇	𐫇	𐫇	𐫇
SAMEKH	𐫈	𐫈	𐫈	𐫈
AYIN	𐫉	𐫉	𐫉	𐫉
PE	𐫊	𐫊	𐫊	𐫊
QOPH	𐫋	𐫋	𐫋	𐫋

Table 10-5 lists the right-joining letters for Manichaean.

Table 10-5. Right-Joining Manichaean Letters

Joining Group	X_n	X_r
DALETH	𐫌	𐫌
DALETH + VS-1	𐫌	𐫌
WAW	𐫍	𐫍
ZAYIN	𐫎	𐫎
TETH	𐫏	𐫏
YODH	𐫐	𐫐
KAPH	𐫑	𐫑
SADHE	𐫒	𐫒
RESH	𐫓	𐫓
RESH + VS-1	𐫓	𐫓
TAW	𐫔	𐫔

Table 10-6 lists the left-joining letters for Manichaean.

Table 10-6. Left-Joining Manichaean Letters

Joining Group	X_n	X_l
HETH	𐫚	𐫚
NUN	𐫛	𐫛
NUN + VS-1	𐫛	𐫛

Table 10-7 lists the non-joining letters for Manichaean

Table 10-7. Non-Joining Manichaean Letters

Joining Group	X_n
HE	𐫜
HE + VS-1	𐫜
JAYIN	𐫝
SHIN	𐫞

Manichaean has two obligatory ligatures for *sadhe* followed by *yodh* or *nun*. These are shown in *Table 10-8*.

Table 10-8. Manichaean Ligatures

Character Sequence	X_n	X_r
SADHE + YODH	𐫟	𐫟
SADHE + NUN	𐫟	𐫟

Numbers. Manichaean has script-specific numeric characters with right-to-left directionality. Numbers are built up using sequences of characters for 1, 5, 10, 20, and 100 in a manner which appears similar to Imperial Aramaic number formation (see *Table 10-3*); however, very few numeric values are attested in Manichaean sources. Manichaean numeric characters exhibit contextual joining behavior, as with letters, but the existing sources do not demonstrate all of the forms.

Punctuation. Manichaean consistently uses a number of script-specific punctuation marks. U+10AF0 MANICHAEAN PUNCTUATION STAR is used to mark the beginning and end of headlines; U+10AF1 MANICHAEAN PUNCTUATION FLEURON and U+10AF5 MANICHAEAN PUNCTUATION TWO DOTS are used to mark the beginning and end of headlines and captions. U+10AF6 MANICHAEAN PUNCTUATION LINE FILLER is used as a sort of ellipsis to fill out a line.

U+10AF2 MANICHAEAN PUNCTUATION DOUBLE DOT WITHIN DOT is used to indicate larger units of text in a prose text or the end of a strophe in a verse text. U+10AF3 MANICHAEAN

PUNCTUATION DOT WITHIN DOT is used to indicate smaller units of text in a prose text or the end of a half-verse in a verse text. U+10AF4 MANICHAEAN PUNCTUATION DOT is used to indicate sub-units of text, logical parts of a sentence or units in a list.

10.6 Pahlavi and Parthian

The Inscriptional Parthian script was used to write Parthian and other languages. It had evolved from the Imperial Aramaic script by the second century CE, and was used as an official script during the first part of the Neo-Persian (Sasanian) empire. It is attested primarily in surviving inscriptions, the last of which dates from 292 CE. Inscriptional Pahlavi also evolved from the Aramaic script during the second century CE during the late period of the Parthian Persian empire in what is now southern Iran. It was used as a monumental script to write Middle Persian until the fifth century CE.

Psalter Pahlavi is a cursive alphabetic script that was used to write the Middle Persian language during the 6th or 7th century CE. It is a historically conservative variety of Pahlavi used by Christians in the Neo-Persian empire. The name of the script is based on its main attestation in a fragmentary manuscript of the Psalms of David, known as the Pahlavi Psalter. The later Book Pahlavi is another variety of the script.

Inscriptional Parthian: U+10B40–U+10B5F
Inscriptional Pahlavi: U+10B60–U+10B7F

Inscriptional Parthian and Inscriptional Pahlavi are both alphabetic scripts and are usually written with spaces between words. Inscriptional Parthian has 22 consonant letters but no vowel marks, while Inscriptional Pahlavi consists of 19 consonant letters; two of which are used for writing multiple consonants, so that it can be used for writing the usual Phoenician-derived 22 consonants.

Directionality. Both the Inscriptional Parthian script and the Inscriptional Pahlavi script are written from right to left. Conformant implementations must use the Unicode Bidirectional Algorithm. For more information, see Unicode Standard Annex #9, "Unicode Bidirectional Algorithm."

Shaping and Layout Behavior. Inscriptional Parthian makes use of seven standard ligatures. Ligation is common, but not obligatory; U+200C ZERO WIDTH NON-JOINER can be used to prevent ligature formation. The same glyph is used for both the *yodh-waw* and *nun-waw* ligatures. The letters *sadhe* and *nun* have swash tails which typically trail under the following letter; thus two *nun*s will nest, and the tail of a *nun* that precedes a *daleth* may be displayed between the two parts of the *daleth* glyph. *Table 10-9* shows these behaviors.

In Inscriptional Pahlavi, U+10B61 INSCRIPTIONAL PAHLAVI LETTER BETH has a swash tail which typically trails under the following letter, similar to the behavior of U+10B4D INSCRIPTIONAL PARTHIAN LETTER NUN.

Numbers. Inscriptional Parthian and Inscriptional Pahlavi each have script-specific numeric characters with right-to-left directionality. Numbers in both are built up using sequences of characters for 1, 2, 3, 4, 10, 20, 100, and 1000 in a manner similar to the way numbers are built up for Imperial Aramaic; see *Table 10-3*. In Inscriptional Parthian the units are sometimes written with strokes of the same height, or sometimes written with a longer ascending or descending final stroke to show the end of the number.

Table 10-9. Inscriptional Parthian Shaping Behavior

𐭣	*(gimel)*	+	𐭥	*(waw)*	→	𐭧	*(gw)*
𐭧	*(heth)*	+	𐭥	*(waw)*	→	𐭨	*(xw)*
𐭩	*(yodh)*	+	𐭥	*(waw)*	→	𐭪	*(yw)*
𐭭	*(nun)*	+	𐭥	*(waw)*	→	𐭮	*(nw)*
𐭯	*(ayin)*	+	𐭶	*(lamedh)*	→	𐭰	*(ʿl)*
𐭱	*(resh)*	+	𐭥	*(waw)*	→	𐭲	*(rw)*
𐭵	*(taw)*	+	𐭥	*(waw)*	→	𐭶	*(tw)*
𐭭	*(nun)*	+	𐭭	*(nun)*	→	𐭻	*(nn)*
𐭭	*(nun)*	+	𐭣	*(daleth)*	→	𐭼	*(nd)*

Heterograms. As scripts derived from Aramaic (such as Inscriptional Parthian and Pahlavi) were adapted for writing Iranian languages, certain words continued to be written in the Aramaic language but read using the corresponding Iranian-language word. These are known as heterograms or xenograms, and were formerly called "ideograms".

Psalter Pahlavi: U+10B80–U+10BAF

Structure. Psalter Pahlavi is an alphabetic script written right-to-left. It uses spaces between words. The script has fully-developed cursive joining behavior. To graphically extend a word, U+0640 ARABIC TATWEEL may be used.

Numbers. Psalter Pahlavi has its own numbers, which also have right-to-left directionality. Numbers are built up out of 1, 2, 3, 4, 10, 20, and 100. Some Psalter Pahlavi numbers have joining behavior, and can join with letters as well as numbers.

Punctuation. There are four types of large section-ending punctuation. The most common is U+10B99 PSALTER PAHLAVI SECTION MARK, which is written with red dots in the vertical position and black dots in the horizontal position; the red dots are often written as rings. Less common but found together with this is U+10B9A PSALTER PAHLAVI TURNED SECTION MARK, which is written with black dots in the vertical position and red dots in the horizontal position. More rare are U+10B9B PSALTER PAHLAVI FOUR DOTS WITH CROSS (sometimes found immediately following the *section mark*), and U+10B9C PSALTER PAHLAVI FOUR DOTS WITH DOT.

10.7 Avestan

Avestan: U+10B00–U+10B3F

The Avestan script was created around the fifth century CE to record the canon of the Avesta, the principal collection of Zoroastrian religious texts. The Avesta had been transmitted orally in the Avestan language, which was by then extinct except for liturgical purposes. The Avestan script was also used to write the Middle Persian language, which is called Pazand when written in Avestan script. The Avestan script was derived from Book Pahlavi, but provided improved phonetic representation by adding consonants and a complete set of vowels—the latter probably due to the influence of the Greek script. It is an alphabetic script of 54 letters, including one that is used only for Pazand.

Directionality. The Avestan script is written from right to left. Conformant implementations of Avestan script must use the Unicode Bidirectional Algorithm. For more information, see Unicode Standard Annex #9, "Unicode Bidirectional Algorithm".

Shaping Behavior. Four ligatures are commonly used in manuscripts of the Avesta, as shown in *Table 10-10*. U+200C ZERO WIDTH NON-JOINER can be used to prevent ligature formation.

Table 10-10. Avestan Shaping Behavior

𐬱	(š)	+	𐬀	(a)	→	𐬱𐬀	(ša)
𐬱	(š)	+	𐬌	(ce)	→	𐬱𐬌	(šc)
𐬱	(š)	+	𐬍	(te)	→	𐬱𐬍	(št)
𐬀	(a)	+	𐬵	(he)	→	𐬀𐬵	(ah)

Punctuation. Archaic Avestan texts use a dot to separate words. The texts generally use a more complex grouping of dots or other marks to indicate boundaries between larger units such as clauses and sentences, but this is not systematic. In contemporary critical editions of Avestan texts, some scholars have systematized and differentiated the usage of various Avestan punctuation marks. The most notable example is Karl F. Geldner's 1880 edition of the Avesta.

The Unicode Standard encodes a set of Avestan punctuation marks based on the system established by Geldner. U+10B3A TINY TWO DOTS OVER ONE DOT PUNCTUATION functions as an Avestan colon, U+10B3B SMALL TWO DOTS OVER ONE DOT PUNCTUATION as an Avestan semicolon, and U+10B3C LARGE TWO DOTS OVER ONE DOT PUNCTUATION as an Avestan end of sentence mark; these indicate breaks of increasing finality. U+10B3E LARGE TWO RINGS OVER ONE RING PUNCTUATION functions as an Avestan end of section, and may be doubled (sometimes with a space between) for extra finality. U+10B39 AVESTAN ABBREVIATION MARK is used to mark abbreviation and repetition. U+10B3D LARGE ONE DOT OVER

TWO DOTS PUNCTUATION and U+10B3F LARGE ONE RING OVER TWO RINGS PUNCTUATION are found in Avestan texts, but are not used by Geldner.

Minimal representation of Avestan requires two separators: one to separate words and a second mark used to delimit larger units, such as clauses or sentences. Contemporary editions of Avestan texts show the word separator dot in a variety of vertical positions: it may appear in a midline position or on the baseline. Dots such as U+2E31 WORD SEPARATOR MIDDLE DOT, U+00B7 MIDDLE DOT, or U+002E FULL STOP can be used to represent this.

10.8 Chorasmian

Chorasmian: U+10FB0–U+10FDF

The Chorasmian script was derived from Imperial Aramaic and is related to Parthian, Inscriptional Pahlavi, Psalter Pahlavi, Book Pahlavi, and Old Sogdian. It was used between the 2nd century and the 8th to 9th centuries CE primarily to write the Chorasmian language, a now-extinct Eastern Iranian language. The script and language were used in a region in Central Asia situated at the delta of the Amu Darya river, classically known as the Oxus, which today is spread across Uzbekistan, Kazakhstan, and Turkmenistan. The name of the territory was first mentioned in the Avesta; it is found inscribed at Persepolis and referenced in classical Persian. The name was once transcribed in English as *Khwarezm*, however, the Greek form entered the English lexicon as *Chorasmian*, and this name is used here.

The Chorasmian script is classified into lapidary and cursive forms. The lapidary form is non-joining and occurs on certain specific items, such as a few silver bowls and a flask found in 2005. The cursive Chorasmian form is derived from the lapidary form, and is found on coinage, wooden items, leather, other silver vessels, and ossuaries, and is the form encoded in the Unicode Standard.

Chorasmian contains 21 letters and 7 numbers. The Unicode character names are based on those of Imperial Aramaic characters.

Directionality. The Chorasmian script is a cursively joining *abjad*, most commonly written from right to left, with lines that advance from top to bottom. Some inscriptions are written vertically and read top to bottom with lines that advance from left to right.

Joining Behavior. Letters are classified as dual-joining, right-joining, and non-joining. Dual-joining and right-joining letters have contextual shapes that are determined by adjacent letters. In some cases, a ZWNJ is used to prevent the left-side connection of a dual-joining letter from joining.

Punctuation and Line Breaking. Spaces are used to separate words. There are no special punctuation marks. There are no formal rules to break words at the end of line.

Numbers. The primary numbers one to four are encoded atomically. The numbers five to nine are expressed using combinations of one to four. This model aligns with Imperial Aramaic and related scripts.

10.9 Elymaic

Elymaic: U+10FE0–U+10FFF

The Elymaic script, also called "Elymaean," was used to write Achaemenid Aramaic in the ancient state of Elymais, which flourished from the second century BCE to the early third century CE and was located in the southwestern portion of modern-day Iran. Elymaic derives from the Aramaic script and is closely related to Parthian and Mandaic. The script is found on inscriptions and coins.

Directionality. The Elymaic script is written from right to left. Conformant implementations of the Elymaic script must use the Unicode Bidirectional Algorithm. For more information, see Unicode Standard Annex #9, "Unicode Bidirectional Algorithm."

Structure. Elymaic is encoded as a non-joining *abjad*. Although some sources show adjacent letters connecting or overlapping, the overall script does not contain intrinsic cursive behavior. However, Elymaic includes one ligature: U+10FF6 ELYMAIC LIGATURE ZAYIN-YODH.

Character Names and Glyphs. The Elymaic character names are based on those for Imperial Aramaic because the native names for the characters are unknown. The representative glyphs in the code charts are based on the stone inscriptions at Tang-e Sarvak in southwest Iran.

Punctuation. There is no script-specific punctuation for Elymaic. Although word boundaries are not generally indicated, some inscriptions have spaces between words. Modern editors tend to use U+0020 SPACE for word separation.

Numerals. There are no known script-specific numerals.

10.10 Nabataean

Nabataean U+10880–U+108AF

The Nabataean script developed from the Aramaic script and was used to write the language of the Nabataean kingdom. The script was in wide use from the second century BCE to the fourth century CE, well after the Roman province of Arabia Petraea was formed.

Nabataean is generally considered to be the precursor of the Arabic script. The Namara inscription, dating from the fourth century CE and believed to be one of the oldest Arabic texts, was written in the Nabataean script.

The glyphs of the Nabataean script are more ornate than those of other scripts derived from Aramaic, and flourishes can be found in some inscriptions. As the script evolved, a range of ligatures was introduced. Because their usage is irregular, no joining behavior is specified for Nabataean.

Structure. The Nabataean script consists of 22 consonants. Nine consonants have final forms and are treated similarly to the final letters of the Hebrew script. The final forms are encoded separately because their occurrence in text is not predictable. For more information about the use of distinctly encoded final consonants in Semitic scripts, see *Section 9.1, Hebrew*.

Directionality. Both words and numbers in the Nabataean script are written from right to left in horizontal lines. Conformant implementations of the script must use the Unicode Bidirectional Algorithm. For more information on bidirectional layout, see Unicode Standard Annex #9, "Unicode Bidirectional Algorithm."

Numerals. Nabataean has script-specific numeral characters, with strong right-to-left directionality. Nabataean numbers are built up using sequences of characters for 1, 2, 3, 4, 5, 10, 20, and 100 in a manner similar to the way numbers are built up for Imperial Aramaic, which is shown in *Table 10-3*. A cruciform variant of the numeral 4 is encoded separately at U+108AB.

Punctuation. There is no script-specific punctuation in Nabataean. The inscriptions usually have no space between words, but modern editors tend to use U+0020 SPACE for word separation.

10.11 Palmyrene

Palmyrene U+10860–U+1087F

The Palmyrene script was derived by modification of the customary forms of Aramaic developed during the Achaemenid empire. The script was used for writing the Palmyrene dialect of West Aramaic, and is known from inscriptions and documents found mainly in the city of Palmyra and other cities in the region of Syria, dating from 44 BCE to about 280 CE.

Palmyrene has both a monumental and a cursive form. Earlier inscriptions show more rounded forms, while later inscriptions tend to regularize the letterforms. Most pre-Unicode fonts for Palmyrene have followed the monumental style. Ligatures exist in both forms of the script, but are not used consistently.

At a certain point, some Palmyrene letterforms became confused and a distinguishing diacritical dot was introduced, although not regularly or systematically, as seen in the glyphic variation of consonants *daleth* and *resh* across the various styles of the script. Sometimes the two glyphs appear with different skeletons, which is sufficient to distinguish them; sometimes they have the same skeleton and are differentiated by a dot; and sometimes they appear with the same skeleton and no dot, in which case they are indistinguishable. In the Unicode code charts, a dot distinguishes the *daleth* and *resh* glyphs.

Structure. The Palmyrene script consists of 22 consonants. The consonant *nun* has a final form variant, encoded as a separate character, U+1086D PALMYRENE LETTER FINAL NUN, and used similarly to the counterpart Hebrew consonant. For information about the use of distinctly encoded final consonants in Semitic scripts, see *Section 9.1, Hebrew*.

Directionality. Both words and numbers in the Palmyrene script are written from right to left in horizontal lines. Conformant implementations of the script must use the Unicode Bidirectional Algorithm. For more information on bidirectional layout, see Unicode Standard Annex #9, "Unicode Bidirectional Algorithm."

Numerals. Palmyrene has script-specific numeral characters, with strong right-to-left directionality. Palmyrene numbers are built up using sequences of characters for 1, 2, 3, 4, 5, 10, 20, and 100 in a manner similar to the way numbers are built up for Imperial Aramaic, which is shown in *Table 10-3*. The glyphs for the numerals 10 and 100, which had been distinct in Aramaic, coalesced into the same glyph in Palmyrene. The two numerals are generally distinguished by their position in sequences representing numbers rather than their shape. A single character is encoded at U+1087E PALMYRENE NUMBER TEN and should be used for both numerals.

Symbols. Two symbols are encoded at U+10877 PALMYRENE LEFT-POINTING FLEURON and U+10878 PALMYRENE RIGHT-POINTING FLEURON. They usually appear next to numbers.

Punctuation. There is no script-specific punctuation in Palmyrene. The inscriptions usually have no space between words, but modern editors tend to use U+0020 SPACE for word separation.

10.12 Hatran

Hatran: U+108E0–U+108FF

The Hatran *abjad* belongs to the North Mesopotamian branch of the Aramaic scripts, and was used for writing a dialect of the Aramaic language. Hatran writing was discovered in the ancient city of Hatra in present-day Iraq. The inscriptions found there date from 98–97 BCE until circa 241 CE, when the city of Hatra was destroyed. Many of the known texts in Hatran are graffiti, but there are some longer texts.

Structure. The Hatran script consists of 22 consonants, encoded as 21 characters. The consonants *daleth* and *resh* are indistinguishable by shape and are encoded as a single character, U+108E3 HATRAN LETTER DALETH-RESH. Ligatures can occur—for example, the letter *beth* often joins or touches the letter following it—but are not used consistently.

Directionality. Both words and numbers in the Hatran script are written from right to left in horizontal lines. Conformant implementations of the script must use the Unicode Bidirectional Algorithm. For more information on bidirectional layout, see Unicode Standard Annex #9, "Unicode Bidirectional Algorithm."

Numerals. Hatran has script-specific characters for numerals, with strong right-to-left directionality. Hatran numbers are built up using sequences of characters for 1, 5, 10, 20, and 100 in a manner similar to the way numbers are built up for Imperial Aramaic, which is shown in *Table 10-3*. The numbers 2, 3, and 4 are formed from sequences of repeated characters for the numeral 1, and are not separately encoded.

Punctuation. There is no script-specific punctuation encoded for Hatran. The inscriptions sometimes have spaces between words; modern editors tend to insert U+0020 SPACE for word separation even if there were no spaces in the original text.