The Unicode Standard
Version 6.0 – Core Specification

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

The Unicode Standard / the Unicode Consortium ; edited by Julie D. Allen ... [et al.]. — Version 6.0.
Includes bibliographical references and index.
ISBN 978-1-936213-01-6 (http://www.unicode.org/versions/Unicode6.0.0/)
1. Unicode (Computer character set) I. Allen, Julie D. II. Unicode Consortium.
QA268.U545 2011

ISBN 978-1-936213-01-6
Published in Mountain View, CA
February 2011

Chapter 14

Ancient and Historic Scripts

Unicode encodes a number of ancient scripts, which have not been in normal use for a millennium or more, as well as historic scripts, whose usage ended in recent centuries. Although they are no longer used to write living languages, documents and inscriptions using these scripts exist, both for extinct languages and for precursors of modern languages. The primary user communities for these scripts are scholars interested in studying the scripts and the languages written in them. A few, such as Coptic, also have contemporary use for liturgical or other special purposes. Some of the historic scripts are related to each other as well as to modern alphabets.

The following ancient and historic scripts are encoded in this version of the Unicode Standard and described in this chapter:

Ogham	*Ancient Anatolian Alphabets*	*Inscriptional Pahlavi*
Old Italic	*Old South Arabian*	*Avestan*
Runic	*Phoenician*	*Ugaritic*
Gothic	*Imperial Aramaic*	*Old Persian*
Old Turkic	*Mandaic*	*Sumero-Akkadian*
Linear B	*Inscriptional Parthian*	*Egyptian Hieroglyphs*
Cypriot Syllabary		

The following ancient and historic scripts are also encoded in this version of the Unicode Standard, but are described in other chapters for consistency with earlier versions of the Unicode Standard, and due to their close relationship with other scripts described in those chapters:

Coptic *Glagolitic* *Phags-pa* *Kaithi* *Kharoshthi* *Brahmi*

The Ogham script is indigenous to Ireland. While its originators may have been aware of the Latin or Greek scripts, it seems clear that the sound values of Ogham letters were suited to the phonology of a form of Primitive Irish.

Old Italic was derived from Greek and was used to write Etruscan and other languages in Italy. It was borrowed by the Romans and is the immediate ancestor of the Latin script now used worldwide. Old Italic had other descendants, too: The Alpine alphabets seem to have been influential in devising the Runic script, which has a distinct angular appearance owing to its use in carving inscriptions in stone and wood. Gothic, like Cyrillic, was developed on the basis of Greek at a much later date than Old Italic.

The two historic scripts of northwestern Europe, Runic and Ogham, have a distinct appearance owing to their primary use in carving inscriptions in stone and wood. They are conventionally rendered from left to right in scholarly literature, but on the original stone carvings often proceeded in an arch tracing the outline of the stone.

The Old Turkic script is known from eighth-century Siberian stone inscriptions, and is the oldest known form of writing for a Turkic language. Also referred to as Turkic Runes due to

its superficial resemblance to Germanic Runes, it appears to have evolved from the Sogdian script, which is in turn derived from Aramaic.

Both Linear B and Cypriot are syllabaries that were used to write Greek. Linear B is the older of the two scripts, and there are some similarities between a few of the characters that may not be accidental. Cypriot may descend from Cypro-Minoan, which in turn may descend from Linear B.

The ancient Anatolian alphabets Lycian, Carian, and Lydian all date from the first millennium BCE, and were used to write various ancient Indo-European languages of western and southwestern Anatolia. All are closely related to the Greek script.

The elegant 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. Used over a wide region beginning in the eighth century BCE as Aramaic became the principal administrative language of the Assyrian empire and then the official language of the Achaemenid Persian empire, it was the source of many other scripts, such as the square Hebrew script and the Arabic script. The Mandaic script was probably derived from a cursive form of Aramaic, and was used in southern Mesopotamia for liturgical texts by adherents of the Mandaean gnostic religion. Inscriptional Parthian, Inscriptional Pahlavi, and Avestan are also derived from Imperial Aramaic, and were used to write various Middle Persian languages.

Three ancient cuneiform scripts are described in this chapter: Ugaritic, Old Persian, and Sumero-Akkadian. The largest and oldest of these is Sumero-Akkadian. The other two scripts are not derived directly from the Sumero-Akkadian tradition but had common writing technology, consisting of wedges indented into clay tablets with reed styluses. Ugaritic texts are about as old as the earliest extant Biblical texts. Old Persian texts are newer, dating from the fifth century BCE.

Egyptian Hieroglyphs were used for more than 3,000 years from the end of the fourth millennium BCE.

14.1 Ogham

Ogham: U+1680–U+169F

Ogham is an alphabetic script devised to write a very early form of Irish. Monumental Ogham inscriptions are found in Ireland, Wales, Scotland, England, and on the Isle of Man. Many of the Scottish inscriptions are undeciphered and may be in Pictish. It is probable that Ogham (Old Irish "Ogam") was widely written in wood in early times. The main flowering of "classical" Ogham, rendered in monumental stone, was in the fifth and sixth centuries CE. Such inscriptions were mainly employed as territorial markers and memorials; the more ancient examples are standing stones.

The script was originally written along the edges of stone where two faces meet; when written on paper, the central "stemlines" of the script can be said to represent the edge of the stone. Inscriptions written on stemlines cut into the face of the stone, instead of along its edge, are known as "scholastic" and are of a later date (post-seventh century). Notes were also commonly written in Ogham in manuscripts as recently as the sixteenth century.

Structure. The Ogham alphabet consists of 26 distinct characters (*feda*), the first 20 of which are considered to be primary and the last 6 (*forfeda*) supplementary. The four primary series are called *aicmí* (plural of *aicme*, meaning "family"). Each *aicme* was named after its first character, (*Aicme Beithe*, *Aicme Uatha*, meaning "the B Family," "the H Family," and so forth). The character names used in this standard reflect the spelling of the names in modern Irish Gaelic, except that the acute accent is stripped from *Úr*, *Éabhadh*, *Ór*, and *Ifín*, and the mutation of *nGéadal* is not reflected.

Rendering. Ogham text is read beginning from the bottom left side of a stone, continuing upward, across the top, and down the right side (in the case of long inscriptions). Monumental Ogham was incised chiefly in a bottom-to-top direction, though there are examples of left-to-right bilingual inscriptions in Irish and Latin. Manuscript Ogham accommodated the horizontal left-to-right direction of the Latin script, and the vowels were written as vertical strokes as opposed to the incised notches of the inscriptions. Ogham should therefore be rendered on computers from left to right or from bottom to top (never starting from top to bottom).

Forfeda (Supplementary Characters). In printed and in manuscript Ogham, the fonts are conventionally designed with a central stemline, but this convention is not necessary. In implementations without the stemline, the character U+1680 OGHAM SPACE MARK should be given its conventional width and simply left blank like U+0020 SPACE. U+169B OGHAM FEATHER MARK and U+169C OGHAM REVERSED FEATHER MARK are used at the beginning and the end of Ogham text, particularly in manuscript Ogham. In some cases, only the *Ogham feather mark* is used, which can indicate the direction of the text.

The word *latheirt* shows the use of the feather marks. This word was written in the margin of a ninth-century Latin grammar and means "massive hangover," which may be the scribe's apology for any errors in his text.

14.2 Old Italic

Old Italic: U+10300–U+1032F

The Old Italic script unifies a number of related historical alphabets located on the Italian peninsula. Some of these were used for non-Indo-European languages (Etruscan and probably North Picene), and some for various Indo-European languages belonging to the Italic branch (Faliscan and members of the Sabellian group, including Oscan, Umbrian, and South Picene). The ultimate source for the alphabets in ancient Italy is Euboean Greek used at Ischia and Cumae in the bay of Naples in the eighth century BCE. Unfortunately, no Greek abecedaries from southern Italy have survived. Faliscan, Oscan, Umbrian, North Picene, and South Picene all derive from an Etruscan form of the alphabet.

There are some 10,000 inscriptions in Etruscan. By the time of the earliest Etruscan inscriptions, circa 700 BCE, local distinctions are already found in the use of the alphabet. Three major stylistic divisions are identified: the Northern, Southern, and Caere/Veii. Use of Etruscan can be divided into two stages, owing largely to the phonological changes that occurred: the "archaic Etruscan alphabet," used from the seventh to the fifth centuries BCE, and the "neo-Etruscan alphabet," used from the fourth to the first centuries BCE. Glyphs for eight of the letters differ between the two periods; additionally, neo-Etruscan abandoned the letters KA, KU, and EKS.

The unification of these alphabets into a single Old Italic script requires language-specific fonts because the glyphs most commonly used may differ somewhat depending on the language being represented.

Most of the languages have added characters to the common repertoire: Etruscan and Faliscan add LETTER EF; Oscan adds LETTER EF, LETTER II, and LETTER UU; Umbrian adds LETTER EF, LETTER ERS, and LETTER CHE; North Picene adds LETTER UU; and Adriatic adds LETTER II and LETTER UU.

The Latin script itself derives from a south Etruscan model, probably from Caere or Veii, around the mid-seventh century BCE or a bit earlier. However, because there are significant differences between Latin and Faliscan of the seventh and sixth centuries BCE in terms of formal differences (glyph shapes, directionality) and differences in the repertoire of letters used, this warrants a distinctive character block. Fonts for early Latin should use the *uppercase* code positions U+0041..U+005A. The unified Alpine script, which includes the Venetic, Rhaetic, Lepontic, and Gallic alphabets, has not yet been proposed for addition to the Unicode Standard but is considered to differ enough from both Old Italic and Latin to warrant independent encoding. The Alpine script is thought to be the source for Runic, which is encoded at U+16A0..U+16FF. (See *Section 14.3, Runic*.)

Character names assigned to the Old Italic block are unattested but have been reconstructed according to the analysis made by Sampson (1985). While the Greek character names (ALPHA, BETA, GAMMA, and so on) were borrowed directly from the Phoenician names (modified to Greek phonology), the Etruscans are thought to have abandoned the Greek names in favor of a phonetically based nomenclature, where stops were pronounced with a following -e sound, and liquids and sibilants (which can be pronounced more or less on their own) were pronounced with a leading e- sound (so [k], [d] became [ke:], [de:] became [l:], [m:] became [el], [em]). It is these names, according to Sampson, which were borrowed by the Romans when they took their script from the Etruscans.

Directionality. Most early Etruscan texts have right-to-left directionality. From the third century BCE, left-to-right texts appear, showing the influence of Latin. Oscan, Umbrian, and Faliscan also generally have right-to-left directionality. Boustrophedon appears rarely, and not especially early (for instance, the Forum inscription dates to 550–500 BCE). Despite this, for reasons of implementation simplicity, many scholars prefer left-to-right presentation of texts, as this is also their practice when transcribing the texts into Latin script. Accordingly, the Old Italic script has a default directionality of strong left-to-right in this standard. If the default directionality of the script is overridden to produce a right-to-left presentation, the glyphs in Old Italic fonts should also be mirrored from the representative glyphs shown in the code charts. This kind of behavior is not uncommon in archaic scripts; for example, archaic Greek letters may be mirrored when written from right to left in boustrophedon.

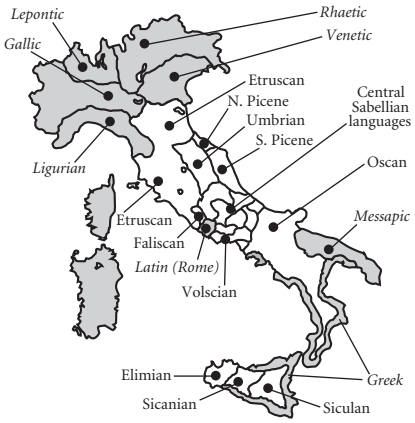
Punctuation. The earliest inscriptions are written with no space between words in what is called *scriptio continua*. There are numerous Etruscan inscriptions with dots separating word forms, attested as early as the second quarter of the seventh century BCE. This punctuation is sometimes, but only rarely, used to separate syllables rather than words. From the sixth century BCE, words were often separated by one, two, or three dots spaced vertically above each other.

Numerals. Etruscan numerals are not well attested in the available materials, but are employed in the same fashion as Roman numerals. Several additional numerals are attested, but as their use is at present uncertain, they are not yet encoded in the Unicode Standard.

Glyphs. The default glyphs in the code charts are based on the most common shapes found for each letter. Most of these are similar to the Marsiliana abecedary (mid-seventh century BCE). Note that the phonetic values for U+10317 OLD ITALIC LETTER EKS [ks] and U+10319 OLD ITALIC LETTER KHE [kh] show the influence of western, Euboean Greek; eastern Greek has U+03A7 GREEK CAPITAL LETTER CHI [x] and U+03A8 GREEK CAPITAL LETTER PSI [ps] instead.

The geographic distribution of the Old Italic script is shown in *Figure 14-1*. In the figure, the approximate distribution of the ancient languages that used Old Italic alphabets is shown in white. Areas for the ancient languages that used other scripts are shown in gray, and the labels for those languages are shown in oblique type. In particular, note that the ancient Greek colonies of the southern Italian and Sicilian coasts used the Greek script proper. Also, languages such as Ligurian, Venetic, and so on, of the far north of Italy made use of alphabets of the Alpine script. Rome, of course, is shown in gray, because Latin was written with the Latin alphabet, now encoded in the Latin script.

Figure 14-1. Distribution of Old Italic



14.3 Runic

Runic: U+16A0–U+16F0

The Runic script was historically used to write the languages of the early and medieval societies in the German, Scandinavian, and Anglo-Saxon areas. Use of the Runic script in various forms covers a period from the first century to the nineteenth century. Some 6,000 Runic inscriptions are known. They form an indispensable source of information about the development of the Germanic languages.

Historical Script. The Runic script is an historical script, whose most important use today is in scholarly and popular works about the old Runic inscriptions and their interpretation. The Runic script illustrates many technical problems that are typical for this kind of script. Unlike many other scripts in the Unicode Standard, which predominantly serve the needs of the modern user community—with occasional extensions for historic forms—the encoding of the Runic script attempts to suit the needs of texts from different periods of time and from distinct societies that had little contact with one another.

Direction. Like other early writing systems, runes could be written either from left to right or from right to left, or moving first in one direction and then the other (*boustrophedon*), or following the outlines of the inscribed object. At times, characters appear in mirror image, or upside down, or both. In modern scholarly literature, Runic is written from left to right. Therefore, the letters of the Runic script have a default directionality of strong left-to-right in this standard.

The Runic Alphabet. Present-day knowledge about runes is incomplete. The set of graphemically distinct units shows greater variation in its graphical shapes than most modern scripts. The Runic alphabet changed several times during its history, both in the number

and the shapes of the letters contained in it. The shapes of most runes can be related to some Latin capital letter, but not necessarily to a letter representing the same sound. The most conspicuous difference between the Latin and the Runic alphabets is the order of the letters.

The Runic alphabet is known as the *futhark* from the name of its first six letters. The original *old futhark* contained 24 runes:

ᚠ ᚢ ᚦ ᚨ ᚱ ᚲ ᚷ ᚹ ᚺ ᚾ ᛁ ᛃ ᛇ ᛈ ᛉ ᛊ ᛏ ᛒ ᛖ ᛗ ᛚ ᛜ ᛞ ᛟ

They are usually transliterated in this way:

f u þ a r k g w h n i j ï p z s t b e m l ŋ d o

In England and Friesland, seven more runes were added from the fifth to the ninth century.

In the Scandinavian countries, the *futhark* changed in a different way; in the eighth century, the simplified younger *futhark* appeared. It consists of only 16 runes, some of which are used in two different forms. The long-branch form is shown here:

ᚠ ᚢ ᚦ ᚬ ᚱ ᚴ ᚼ ᚾ ᛁ ᛅ ᛋ ᛏ ᛒ ᛘ ᛚ ᛦ

f u þ o r k h n i a s t b m l ʀ

The use of runes continued in Scandinavia during the Middle Ages. During that time, the *futhark* was influenced by the Latin alphabet and new runes were invented so that there was full correspondence with the Latin letters.

Representative Glyphs. The known inscriptions can include considerable variations of shape for a given rune, sometimes to the point where the nonspecialist will mistake the shape for a different rune. There is no dominant main form for some runes, particularly for many runes added in the Anglo-Friesian and medieval Nordic systems. When transcribing a Runic inscription into its Unicode-encoded form, one cannot rely on the idealized *representative glyph* shape in the character charts alone. One must take into account to which of the four Runic systems an inscription belongs and be knowledgeable about the permitted form variations within each system. The representative glyphs were chosen to provide an image that distinguishes each rune visually from all other runes in the same system. For actual use, it might be advisable to use a separate font for each Runic system. Of particular note is the fact that the glyph for U+16C4 ᛄ RUNIC LETTER GER is actually a rare form, as the more common form is already used for U+16E1 ᛇ RUNIC LETTER IOR.

Unifications. When a rune in an earlier writing system evolved into several different runes in a later system, the unification of the earlier rune with one of the later runes was based on similarity in graphic form rather than similarity in sound value. In cases where a substantial change in the typical graphical form has occurred, though the historical continuity is undisputed, unification has not been attempted. When runes from different writing systems have the same graphic form but different origins and denote different sounds, they have been coded as separate characters.

Long-Branch and Short-Twig. Two sharply different graphic forms, the *long-branch* and the *short-twig* form, were used for 9 of the 16 Viking Age Nordic runes. Although only one form is used in a given inscription, there are runologically important exceptions. In some cases, the two forms were used to convey different meanings in later use in the medieval system. Therefore the two forms have been separated in the Unicode Standard.

Staveless Runes. Staveless runes are a third form of the Viking Age Nordic runes, a kind of Runic shorthand. The number of known inscriptions is small and the graphic forms of many of the runes show great variability between inscriptions. For this reason, staveless runes have been unified with the corresponding Viking Age Nordic runes. The correspond-

ing Viking Age Nordic runes must be used to encode these characters—specifically the short-twig characters, where both short-twig and long-branch characters exist.

Punctuation Marks. The wide variety of Runic punctuation marks has been reduced to three distinct characters based on simple aspects of their graphical form, as very little is known about any difference in intended meaning between marks that look different. Any other punctuation marks have been unified with shared punctuation marks elsewhere in the Unicode Standard.

Golden Numbers. Runes were used as symbols for Sunday letters and golden numbers on calendar staves used in Scandinavia during the Middle Ages. To complete the number series 1–19, three more calendar runes were added. They are included after the punctuation marks.

Encoding. A total of 81 characters of the Runic script are included in the Unicode Standard. Of these, 75 are Runic letters, 3 are punctuation marks, and 3 are Runic symbols. The order of the Runic characters follows the traditional *futhark* order, with variants and derived runes being inserted directly after the corresponding ancestor.

Runic character names are based as much as possible on the sometimes several traditional names for each rune, often with the Latin transliteration at the end of the name.

14.4 Gothic

Gothic: U+10330–U+1034F

The Gothic script was devised in the fourth century by the Gothic bishop, Wulfila (311–383 CE), to provide his people with a written language and a means of reading his translation of the Bible. Written Gothic materials are largely restricted to fragments of Wulfila's translation of the Bible; these fragments are of considerable importance in New Testament textual studies. The chief manuscript, kept at Uppsala, is the Codex Argenteus or "the Silver Book," which is partly written in gold on purple parchment. Gothic is an East Germanic language; this branch of Germanic has died out and thus the Gothic texts are of great importance in historical and comparative linguistics. Wulfila appears to have used the Greek script as a source for the Gothic, as can be seen from the basic alphabetical order. Some of the character shapes suggest Runic or Latin influence, but this is apparently coincidental.

Diacritics. The tenth letter U+10339 GOTHIC LETTER EIS is used with U+0308 COMBINING DIAERESIS when word-initial, when syllable-initial after a vowel, and in compounds with a verb as second member as shown below:

swe gameliþ ïst ïn esaïïn praufetau

"as is written in Isaiah the prophet"

To indicate contractions or omitted letters, U+0305 COMBINING OVERLINE is used.

Numerals. Gothic letters, like those of other early Western alphabets, can be used as numbers; two of the characters have only a numeric value and are not used alphabetically. To indicate numeric use of a letter, it is either flanked on one side by U+00B7 MIDDLE DOT or followed by both U+0304 COMBINING MACRON and U+0331 COMBINING MACRON BELOW, as shown in the following example:

•𐌵• or 𐌵̲ means "5"

Punctuation. Gothic manuscripts are written with no space between words in what is called *scriptio continua*. Sentences and major phrases are often separated by U+0020 SPACE, U+00B7 MIDDLE DOT, or U+003A COLON.

14.5 Old Turkic

Old Turkic: U+10C00–U+10C4F

The origins of the Old Turkic script are unclear, but it seems to have evolved from a non-cursive form of the Sogdian script, one of the Aramaic-derived scripts used to write Iranian languages, in order to write the Old Turkish language. Old Turkic is attested in stone inscriptions from the early eighth century CE found around the Orkhon River in Mongolia, and in a slightly different version in stone inscriptions of the later eighth century found in Siberia near the Yenisei River and elsewhere. These inscriptions are the earliest written examples of a Turkic language. By the ninth century the Old Turkic script had been supplanted by the Uighur script.

Because Old Turkic characters superficially resemble Germanic runes, the script is also known as Turkic Runes and Turkic Runiform, in addition to the names Orkhon script, Yenisei script, and Siberian script.

Where the Orkhon and Yenisei versions of a given Old Turkic letter differ significantly, each is separately encoded.

Structure. Old Turkish vowels can be classified into two groups based on their front or back articulation. A given word uses vowels from only one of these groups; the group is indicated by the form of the consonants in the word, because most consonants have separate forms to match the two vowel types. Other phonetic rules permit prediction of rounded and unrounded vowels, and high, medium or low vowels within a word. Some consonants also indicate that the preceding vowel is a high vowel. Thus, most initial and medial vowels are not explicitly written; only vowels that end a word are always written, and there is sometimes ambiguity about whether a vowel precedes a given consonant.

Directionality. For horizontal writing, the Old Turkic script is written from right to left within a row, with rows running from bottom to top. Conformant implementations of Old Turkic script must use the Unicode Bidirectional Algorithm (see Unicode Standard Annex #9, "Unicode Bidirectional Algorithm").

In some cases, under Chinese influence, the layout was rotated 90° counterclockwise to produce vertical columns of text in which the characters are read top to bottom within a column, and the columns are read right to left.

Punctuation. Word division and some other punctuation functions are usually indicated by a two-dot mark similar to a colon; U+205A TWO DOT PUNCTUATION may be used to represent this punctuation mark. In some cases a mark such as U+2E30 RING POINT is used instead.

14.6 Linear B

Linear B Syllabary: U+10000–U+1007F

The Linear B script is a syllabic writing system that was used on the island of Crete and parts of the nearby mainland to write the oldest recorded variety of the Greek language. Linear B clay tablets predate Homeric Greek by some 700 years; the latest tablets date from

the mid- to late thirteenth century BCE. Major archaeological sites include Knossos, first uncovered about 1900 by Sir Arthur Evans, and a major site near Pylos. The majority of currently known inscriptions are inventories of commodities and accounting records.

Early attempts to decipher the script failed until Michael Ventris, an architect and amateur decipherer, came to the realization that the language might be Greek and not, as previously thought, a completely unknown language. Ventris worked together with John Chadwick, and decipherment proceeded quickly. The two published a joint paper in 1953.

Linear B was written from left to right with no nonspacing marks. The script mainly consists of phonetic signs representing the combination of a consonant and a vowel. There are about 60 known phonetic signs, in addition to a few signs that seem to be mainly free variants (also known as Chadwick's optional signs), a few unidentified signs, numerals, and a number of ideographic signs, which were used mainly as counters for commodities. Some ligatures formed from combinations of syllables were apparently used as well. Chadwick gives several examples of these ligatures, the most common of which are included in the Unicode Standard. Other ligatures are the responsibility of the rendering system.

Standards. The catalog numbers used in the Unicode character names for Linear B syllables are based on the Wingspread Convention, as documented in Bennett (1964). The letter "B" is prepended arbitrarily, so that name parts will not start with a digit, thus conforming to ISO/IEC 10646 naming rules. The same naming conventions, using catalog numbers based on the Wingspread Convention, are used for Linear B ideograms.

Linear B Ideograms: U+10080–U+100FF

The Linear B Ideograms block contains the list of Linear B signs known to constitute ideograms (logographs), rather than syllables. When generally agreed upon, the names include the meaning associated with them—for example, U+10080 𐂀 LINEAR B IDEOGRAM B100 MAN. In other instances, the names of the ideograms simply carry their catalog number.

Aegean Numbers: U+10100–U+1013F

The signs used to denote Aegean whole numbers (U+10107..U+10133) derive from the non-Greek Linear A script. The signs are used in Linear B. The Cypriot syllabary appears to use the same system, as evidenced by the fact that the lower digits appear in extant texts. For measurements of agricultural and industrial products, Linear B uses three series of signs: liquid measures, dry measures, and weights. No set of signs for linear measurement has been found yet. Liquid and dry measures share the same symbols for the two smaller subunits; the system of weights retains its own unique subunits. Though several of the signs originate in Linear A, the measuring system of Linear B differs from that of Linear A. Linear B relies on units and subunits, much like the imperial "quart," "pint," and "cup," whereas Linear A uses whole numbers and fractions. The absolute values of the measurements have not yet been completely agreed upon.

14.7 Cypriot Syllabary

Cypriot Syllabary: U+10800–U+1083F

The Cypriot syllabary was used to write the Cypriot dialect of Greek from about 800 to 200 BCE. It is related to both Linear B and Cypro-Minoan, a script used for a language that has not yet been identified. Interpretation has been aided by the fact that, as use of the Cypriot syllabary died out, inscriptions were carved using both the Greek alphabet and the Cypriot syllabary. Unlike Linear B and Cypro-Minoan, the Cypriot syllabary was usually written

from right to left, and accordingly the characters in this script have strong right-to-left directionality.

Word breaks can be indicated by spaces or by separating punctuation, although separating punctuation is also used between larger word groups.

Although both Linear B and the Cypriot syllabary were used to write Greek dialects, Linear B has a more highly abbreviated spelling. Structurally, the Cypriot syllabary consists of combinations of up to 12 initial consonants and 5 different vowels. Long and short vowels are not distinguished. The Cypriot syllabary distinguishes among a different set of initial consonants than Linear B; for example, unlike Linear B, Cypriot maintained a distinction between [l] and [r], though not between [d] and [t], as shown in *Table 14-1*. Not all of the 60 possible consonant-vowel combinations are represented. As is the case for Linear B, the Cypriot syllabary is well understood and documented.

Table 14-1. Similar Characters in Linear B and Cypriot

Linear B		Cypriot	
da	𐀁	*ta*	𐠁
na	𐀍	*na*	𐠍
pa	𐀖	*pa*	𐠖
ro	𐀬	*lo*	𐠬
se	𐀮	*se*	𐠮
ti	𐀯	*ti*	𐠯
to	𐀰	*to*	𐠰

For Aegean numbers, see the subsection "Aegean Numbers: U+10100–U+1013F" in *Section 14.6, Linear B*.

14.8 Ancient Anatolian Alphabets

Lycian: U+10280–U+1029F
Carian: U+102A0–U+102DF
Lydian: U+10920–U+1093F

The Anatolian scripts described in this section all date from the first millennium BCE, and were used to write various ancient Indo-European languages of western and southwestern Anatolia (now Turkey). All are closely related to the Greek script and are probably adaptations of it. Additional letters for some sounds not found in Greek were probably either invented or drawn from other sources. However, development parallel to, but independent of, the Greek script cannot be ruled out, particularly in the case of Carian.

Lycian. Lycian was used from around 500 BCE to about 200 BCE. The term "Lycian" is now used in place of "Lycian A" (a dialect of Lycian, attested in two texts in Anatolia, is called "Lycian B", or "Milyan", and dates to the first millennium BCE). The Lycian script appears on some 150 stone inscriptions, more than 200 coins, and a few other objects.

Lycian is a simple alphabetic script of 29 letters, written left-to-right, with frequent use of word dividers. The recommended word divider is U+205A TWO DOT PUNCTUATION. *Scrip-*

tio continua (a writing style without spaces or punctuation) also occurs. In modern editions U+0020 SPACE is sometimes used to separate words.

Carian. The Carian script is used to write the Carian language, and dates from the first millennium BCE. While a few texts have been found in Caria, most of the written evidence comes from Carian communities in Egypt, where they served as mercenaries. The repertoire of the Carian texts is well established. Unlike Lycian and Lydian, Carian does not use a single standardized script, but rather shows regional variation in the repertoire of signs used and their form. Although some of the values of the Carian letters remain unknown or in dispute, their distinction from other letters is not. The Unicode encoding is based on the standard "Masson set" catalog of 45 characters, plus 4 recently-identified additions. Some of the characters are considered to be variants of others—and this is reflected in their names—but are separately encoded for scholarly use in discussions of decipherment.

The primary direction of writing is left-to-right in texts from Caria, but right-to-left in Egyptian Carian texts. However, both directions occur in the latter, and left-to-right is favored for modern scholarly usage. Carian is encoded in Unicode with left-to-right directionality. Word dividers are not regularly employed; *scriptio continua* is common. Word dividers which are attested are U+00B7 MIDDLE DOT (or U+2E31 WORD SEPARATOR MIDDLE DOT), U+205A TWO DOT PUNCTUATION, and U+205D TRICOLON. In modern editions U+0020 SPACE may be found.

Lydian. While Lydian is attested from inscriptions and coins dating from the end of the eighth century (or beginning of the seventh) until the third century BCE, the longer well-preserved inscriptions date to the fifth and fourth centuries BCE.

Lydian is a simple alphabetic script of 26 letters. The vast majority of Lydian texts have right-to-left directionality (the default direction); a very few texts are left-to-right and one is boustrophedon. Most Lydian texts use U+0020 SPACE as a word divider. Rare examples have been found which use *scriptio continua* or which use dots to separate the words. In the latter case, U+003A COLON and U+00B7 MIDDLE DOT (or U+2E31 WORD SEPARATOR MIDDLE DOT) can be used to represent the dots. U+1093F LYDIAN TRIANGULAR MARK is thought to indicate quotations, and is mirrored according to text directionality.

14.9 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 13.1, 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 diacritic 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 14-2*—four of these are letters that double as numeric values, and two are characters not used as letters.

Table 14-2. 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 14-3*. When 10, 50, or 100 occur preceding 1000 they serve to indicate multiples of 1000. The example numbers shown in *Table 14-3* are rendered in a right-to-left direction in the last column.

Table 14-3. 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

Table 14-3. Number Formation in Old South Arabian (Continued)

Value	Schematic	Character Sequence	Display
16	10 + 5 + 1	10A72 10A6D 10A7D	𐩲𐩬𐩺
1000	1000	10A71	𐩱
3000	1000 + 1000 + 1000	10A71 10A71 10A71	𐩱𐩱𐩱
10000	10 × 1000	10A72 10A71	𐩲𐩱
11000	10 × 1000 + 1000	10A72 10A71 10A71	𐩲𐩱𐩱
30000	(10 + 10 + 10) × 1000	10A72 10A72 10A72 10A71	𐩲𐩲𐩲𐩱
30001	(10 + 10 + 10) × 1000 + 1	10A72 10A72 10A72 10A71 10A7D	𐩲𐩲𐩲𐩱𐩺

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

14.10 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 Palaeo-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 Palaeo-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 Palaeo-Hebrew forms, eventually developing them into the distinct Samaritan script. (See *Section 8.4, Samaritan*.) The Jews in exile gave up the Palaeo-Hebrew alphabet and instead adopted Imperial Aramaic writing, which was a descendant of the Early Aramaic form of the Phoenician script. (See *Section 14.11, 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 8.1, Hebrew*.)

Some scholars conceive of the language written in the Palaeo-Hebrew form of the Phoenician script as being quintessentially Hebrew and consistently transliterate it into Square Hebrew. In such contexts, Palaeo-Hebrew texts are often considered to simply *be* Hebrew, and because the relationship between the Palaeo-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: 𐤙𐤚𐤚𐤛 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.

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

14.11 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 14-4*. The example numbers shown in the last column are rendered in a right-to-left direction.

Table 14-4. 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
500	(3 + 2) × 100	1085A 10859 1085D	𐡝𐡙𐡚
3000	3 × 1000	1085A 1085E	𐡞𐡚
30000	3 × 10000	1085A 1085F	𐡟𐡚

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 and Inscriptional Pahlavi scripts use a similar system for forming numeric values.

14.12 Mandaic

Mandaic: U+0840—U+085F

The origins of the Mandaic script are unclear, but it is thought to have evolved between the 2nd and 7th century CE from a cursivized form of the Aramaic script (as did the Syriac script) or from the Parthian chancery script. It was developed by adherents of the Mandaean gnostic religion of southern Mesopotamia to write the dialect of Eastern Aramaic they used for liturgical purposes, which is referred to as Classical Mandaic.

The religion has survived into modern times, with more than 50,000 Mandaeans in several communities worldwide (most having left what is now Iraq). In addition to the Classical Mandaic still used within some of these communities, a variety known as Neo-Mandaic or Modern Mandaic has developed and is spoken by a small number of people. Mandaeans consider their script sacred, with each letter having specific mystic properties, and the script has changed very little over time.

Structure. Mandaic is unusual among Semitic scripts in being a true alphabet; the letters *halqa*, *ushenna*, *aksa*, and *in* are used to write both long and short forms of vowels, instead of functioning as consonants also used to write long vowels (*matres lectionis*), in the manner characteristic of other Semitic scripts. This is possible because some consonant sounds represented by the corresponding letters in other Semitic scripts are not used in the Mandaic language.

Two letters have morphemic function. U+0847 MANDAIC LETTER IT is used only for the third person singular suffix. U+0856 MANDAIC LETTER DUSHENNA, also called *adu*, is used to write the relative pronoun and the genitive exponent *di*, and is a digraph derived from an old ligature for *ad* + *aksa*. It is thus an addition to the usual Semitic set of 22 characters.

The Mandaic alphabet is traditionally represented as the 23 letters *halqa* through *dushenna*, with *halqa* appended again at the end to form a symbolically-important cycle of 24 letters. Two additional Mandaic characters are encoded in the Unicode Standard: U+0857 MANDAIC LETTER KAD is derived from an old ligature of *ak* + *dushenna*; it is a digraph used to write the word *kd*, which means "when, as, like". The second additional character, U+0858 MANDAIC LETTER AIN, is a borrowing from U+0639 ARABIC LETTER AIN.

Three diacritical marks are used in teaching materials to differentiate vowel quality; they may be omitted from ordinary text. U+0859 MANDAIC AFFRICATION MARK is used to extend the character set for foreign sounds (whether affrication, lenition, or another sound). U+085A MANDAIC VOCALIZATION MARK is used to distinguish vowel quality of *halqa*, *ushenna*, and *aksa*. U+085B MANDAIC GEMINATION MARK is used to indicate what native writers call a "hard" pronunciation.

Punctuation. Sentence punctuation is used sparsely. A single script-specific punctuation mark is encoded: U+085E MANDAIC PUNCTUATION. It is used to start and end text sections, and is also used in colophons—the historical lay text added to the religious text—where it is typically displayed in a smaller size.

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

Shaping and Layout Behavior. Mandaic has fully-developed joining behavior, with forms as shown in *Table 14-5* and *Table 14-6*. In these tables, X_n, X_r, X_m, and X_l designate the

Table 14-5. Dual-Joining Mandaic Characters

Character	X_n	X_r	X_m	X_l
AB	ﬕ	ﬕ	ﬕ	ﬕ
AG	ﬗ	ﬗ	ﬗ	ﬗ
AD	ﬓ	ﬓ	ﬓ	ﬓ
AH	ﬧ	ﬧ	ﬧ	ﬧ
USHENNA	ﬔ	ﬔ	ﬔ	ﬔ
IT	ﬖ	ﬖ	ﬖ	ﬖ
ATT	﬘	﬘	﬘	﬘
AK	﬙	﬙	﬚	﬚
AL	﬛	﬛	﬛	﬛
AM	﬜	﬜	﬜	﬜
AN	יִ	יִ	ﬞ	ﬞ

Table 14-5. Dual-Joining Mandaic Characters (Continued)

Character	X_n	X_r	X_m	X_l
AS	ـصـ	ـصـ	ـصـ	ـصـ
AP	ـפ	ـפ	ـפ	ـפ
ASZ	ـצ	ـצ	ـצ	ـצ
AQ	ـק	ـק	ـק	ـק
AR	ـר	ـר	ـר	ـר
AT	ـת	ـת	ـת	ـת

nominal, right-joining, dual-joining (medial), and left-joining forms respectively, just as in *Table 8-7*, *Table 8-8*, and *Table 8-9*.

Table 14-6. Right-Joining Mandaic Characters

Character	X_n	X_r
HALQA	ـه	ـه
AZ	ـا	ـا
AKSA	ـح	ـح
IN	ـע	ـע
ASH	ـפ	ـפ

Linebreaking. Spaces provide the primary line break opportunity. When text is fully justified, words may be stretched as in Arabic. U+0640 ARABIC TATWEEL may be inserted for this purpose.

14.13 Inscriptional Parthian and Inscriptional Pahlavi

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

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 Sassanid Persian 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. Other varieties of Pahlavi script include Psalter Pahlavi and the later Book Pahlavi.

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 Bidirec-

tional 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 14-7* shows these behaviors.

Table 14-7. Inscriptional Parthian Shaping Behavior

Character Sequence	Glyph Sequence	Resulting Display	Transcription
gimel-waw	𐭂 𐭅	𐭂	gw
heth-waw	𐭇 𐭅	𐭇	xw
yodh-waw	𐭉 𐭅	𐭉	yw
nun-waw	𐭍 𐭅	𐭍	nw
ayin-lamedh	𐭏 𐭋	𐭏	ʻl
resh-waw	𐭓 𐭅	𐭓	rw
taw-waw	𐭕 𐭅	𐭕	tw
nun-nun	𐭍 𐭍	𐭍	nn
nun-daleth	𐭍 𐭃	𐭍	nd

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 they way numbers are built up for Imperial Aramaic; see *Table 14-4*. 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.

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".

14.14 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 14-8*. U+200C ZERO WIDTH NON-JOINER can be used to prevent ligature formation.

Table 14-8. Avestan Shaping Behavior

Character Sequence	Display	Transcription
<10B31 𐬱 she, 10B00 𐬀 a>	𐬱𐬀	ša
<10B31 𐬱 she, 10B17 𐬗 ce>	𐬱𐬗	šc
<10B31 𐬱 she, 10B19 𐬙 te>	𐬱𐬙	št
<10B00 𐬀 a, 10B35 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.

14.15 Ugaritic

Ugaritic: U+10380–U+1039F

The city state of Ugarit was an important seaport on the Phoenician coast (directly east of Cyprus, north of the modern town of Minet el-Beida) from about 1400 BCE until it was completely destroyed in the twelfth century BCE. The site of Ugarit, now called Ras Shamra (south of Latakia on the Syrian coast), was apparently continuously occupied from Neolithic times (circa 5000 BCE). It was first uncovered by a local inhabitant while plowing a field in 1928 and subsequently excavated by Claude Schaeffer and Georges Chenet beginning in 1929, in which year the first of many tablets written in the Ugaritic script were discovered. They later proved to contain extensive portions of an important Canaanite mythological and religious literature that had long been sought and that revolutionized

Biblical studies. The script was first deciphered in a remarkably short time jointly by Hans Bauer, Edouard Dhorme, and Charles Virolleaud.

The Ugaritic language is Semitic, variously regarded by scholars as being a distinct language related to Akkadian and Canaanite, or a Canaanite dialect. Ugaritic is generally written from left to right horizontally, sometimes using U+1039F 𐎟 UGARITIC WORD DIVIDER. In the city of Ugarit, this script was also used to write the Hurrian language. The letters U+1039B 𐎛 UGARITIC LETTER I, U+1039C 𐎜 UGARITIC LETTER U, and U+1039D 𐎝 UGARITIC LETTER SSU are used for Hurrian.

Variant Glyphs. There is substantial variation in glyph representation for Ugaritic. Glyphs for U+10398 𐎘 UGARITIC LETTER THANNA, U+10399 𐎙 UGARITIC LETTER GHAIN, and U+1038F 𐎏 UGARITIC LETTER DHAL differ somewhat between modern reference sources, as do some transliterations. U+10398 𐎘 UGARITIC LETTER THANNA is most often displayed with a glyph that looks like an occurrence of U+10393 𐎓 UGARITIC LETTER AIN overlaid with U+10382 𐎂 UGARITIC LETTER GAMLA.

Ordering. The ancient Ugaritic alphabetical order, which differs somewhat from the modern Hebrew order for similar characters, has been used to encode Ugaritic in the Unicode Standard.

Character Names. Some of the Ugaritic character names have been reconstructed; others appear in an early fragmentary document.

14.16 Old Persian

Old Persian: U+103A0–U+103DF

The Old Persian script is found in a number of inscriptions in the Old Persian language dating from the Achaemenid Empire. Scholars today agree that the character inventory of Old Persian was invented for use in monumental inscriptions of the Achaemenid king, Darius I, by about 525 BCE. Old Persian is an alphabetic writing system with some syllabic aspects. While the shapes of some Old Persian letters look similar to signs in Sumero-Akkadian Cuneiform, it is clear that only one of them, U+103BE 𐎾 OLD PERSIAN SIGN LA, was actually borrowed. It was derived from the New Assyrian historic variant 𒆷 of Sumero-Akkadian U+121B7 𒉺 CUNEIFORM SIGN LA, because *la* is a foreign sound not used in the Old Persian language.

Directionality. Old Persian is written from left to right.

Repertoire. The repertoire contains 36 signs. These represent consonants, vowels, or consonant plus vowel syllables. There are also five numbers, one word divider, and eight ideograms. It is considered unlikely that any additional characters will be discovered.

Numerals. The attested numbers are built up by stringing the base numbers (1, 2, 10, 20, and 100) in sequences.

Variants. The signs U+103C8 OLD PERSIAN SIGN AURAMAZDAA and U+103C9 OLD PERSIAN SIGN AURAMAZDAA-2, and the signs U+103CC OLD PERSIAN SIGN DAHYAAUSH and U+103CD OLD PERSIAN SIGN DAHYAAUSH-2, have been encoded separately because their conventional attestation in the corpus of Old Persian texts is quite limited and scholars consider it advantageous to distinguish the forms in plain text representation.

14.17 Sumero-Akkadian

Cuneiform: U+12000–U+123FF

Sumero-Akkadian Cuneiform is a logographic writing system with a strong syllabic component. It was written from left to right on clay tablets.

Early History of Cuneiform. The earliest stage of Mesopotamian Cuneiform as a complete system of writing is first attested in Uruk during the so-called Uruk IV period (circa 3500–3200 BCE) with an initial repertoire of about 700 characters or "signs" as Cuneiform scholars customarily call them.

Late fourth millennium ideographic tablets were also found at Susa and several other sites in western Iran, in Assyria at Nineveh (northern Iraq), at Tell Brak (northwestern Syria), and at Habuba Kabira in Syria. The writing system developed in Sumer (southeastern Iraq) was repeatedly exported to peripheral regions in the third, second, and first millennia BCE. Local variations in usage are attested, but the core of the system is the Sumero-Akkadian writing system.

Writing emerged in Sumer simultaneously with a sudden growth in urbanization and an attendant increase in the scope and scale of administrative needs. A large proportion of the elements of the early writing system repertoire was devised to represent quantities and commodities for bureaucratic purposes.

At this earliest stage, signs were mainly pictographic, in that a relatively faithful facsimile of the thing signified was traced, although some items were strictly ideographic and represented by completely arbitrary abstractions, such as the symbol for sheep ⊕. Some scholars believe that the abstract symbols were derived from an earlier "token" system of accounting, but there is no general agreement on this point. Where the pictographs are concerned, interpretation was relatively straightforward. The head of a bull was used to denote "cattle"; an ear of barley was used to denote "barley." In some cases, pictographs were also interpreted logographically, so that meaning was derived from the symbol by close conceptual association. For example, the representation of a bowl might mean "bowl," but it could indicate concepts associated with bowls, such as "food." Renditions of a leg might variously suggest "leg," "stand," or "walk."

By the next chronological period of south Mesopotamian history (the Uruk III period, 3200–2900 BCE), logographic usage seems to have become much more widespread. In addition, individual signs were combined into more complex designs to express other concepts. For example, a head with a bowl next to it was used to denote "eat" or "drink." This is the point during script development at which one can truly speak of the first Sumerian texts. In due course, the early graphs underwent change, conditioned by factors such as the most widely available writing medium and writing tools, and the need to record information more quickly and efficiently from the standpoint of the bureaucracy that spawned the system.

Clay was the obvious writing medium in Sumer because it was widely available and easily molded into cushion- or pillow-shaped tablets. Writing utensils were easily made for it by sharpening pieces of reed. Because it was awkward and slow to inscribe curvilinear lines in a piece of clay with a sharpened reed (called a *stylus*), scribes tended to approximate the pictographs by means of short, wedge-shaped impressions made with the edge of the stylus. These short, mainly straight shapes gave rise to the modern word "cuneiform" from the Latin *cuneus*, meaning "wedge." Cuneiform proper was common from about 2700 BCE, although experts use the term "cuneiform" to include the earlier forms as well.

Geographic Range. The Sumerians did not live in complete isolation, and there is very early evidence of another significant linguistic group in the area immediately north of Sumer known as Agade or Akkad. Those peoples spoke a Semitic language whose dialects are subsumed by scholars under the heading "Akkadian." In the long run, the Akkadian speakers became the primary users and promulgators of Cuneiform script. Because of their trade involvement with their neighbors, Cuneiform spread through Babylonia (the umbrella term for Sumer and Akkad) to Elam, Assyria, eastern Syria, southern Anatolia, and even Egypt. Ultimately, many languages came to be written in Cuneiform script, the most notable being Sumerian, Akkadian (including Babylonian, Assyrian, Eblaite), Elamite, Hittite, and Hurrian.

Periods of script usage are defined according to geography and primary linguistic representation, as shown in *Table 14-9*.

Table 14-9. Cuneiform Script Usage

<table>
<tr><td colspan="2">Archaic Period
(to 2901 BCE)</td></tr>
<tr><td colspan="2">Early Dynastic
(2900–2335 BCE)</td></tr>
<tr><td colspan="2">Old Akkadian
(2334–2154 BCE)</td></tr>
<tr><td colspan="2">Ur III (NeoSumerian)
(2112–2095 BCE)</td><td rowspan="6">Elamite
(2100–360 BCE)</td></tr>
<tr><td>Old Assyrian
(1900–1750 BCE)</td><td>Old Babylonian
(2004–1595 BCE)</td></tr>
<tr><td>Middle Assyrian
(1500-1000 BCE)</td><td rowspan="2">Middle Babylonian
(1595–627 BCE)</td></tr>
<tr><td>Neo-Assyrian
(1000-609 BCE)</td></tr>
<tr><td></td><td>Neo-Babylonian
(626–539 BCE)</td></tr>
<tr><td colspan="2">Hittite
(1570-1220 BCE)</td></tr>
</table>

Sources and Coverage. The base character repertoire for the Cuneiform block was distilled from the list of Ur III signs compiled by the Cuneiform Digital Library Initiative (UCLA) in union with the list constructed independently by Miguel Civil. This repertoire is comprehensive from the Ur III period onward. Old Akkadian, Early Dynastic, and Archiac Cuneiform are not covered by this repertoire.

Simple Signs. Most Cuneiform signs are simple units; each sign of this type is represented by a single character in the standard.

Complex and Compound Signs. Some Cuneiform signs are categorized as either complex or compound signs. Complex signs are made up of a primary sign with one of more secondary signs written within it or conjoined to it, such that the whole is generally treated by scholars as a unit; this includes linear sequences of two or more signs or wedge-clusters where one or more of those clusters have not been clearly identified as characters in their own right. Complex signs, which present a relative visual unity, are assigned single individual code points irrespective of their components.

Compound signs are linear sequences of two or more signs or wedge-clusters generally treated by scholars as a single unit, when each and every such wedge-cluster exists as a clearly identified character in its own right. Compound signs are encoded as sequences of their component characters. Signs that shift from compound to complex, or vice versa, generally have been treated according to their Ur III manifestation.

Mergers and Splits. Over the long history of Cuneiform, an number of signs have simplified and merged; in other cases, a single sign has diverged and developed into more than one distinct sign. The choice of signs for encoding as characters was made at the point of maximum differentiation in the case of either mergers or splits to enable the most comprehensive set for the representation of text in any period.

Fonts for the representation of Cuneiform text may need to be designed distinctly for optimal use for different historic periods. Fonts for some periods will contain duplicate glyphs depending on the status of merged or split signs at that point of the development of the writing system.

Glyph Variants Acquiring Independent Semantic Status. Glyph variants such as U+122EC CUNEIFORM SIGN TA ASTERISK, a Middle Assyrian form of the sign U+122EB CUNEIFORM SIGN TA, which in Neo-Assyrian usage has its own logographic interpretation, have been assigned separate code positions. They are to be used only when the new interpretation applies.

Formatting. Cuneiform was often written between incised lines or in blocks surrounded by drawn boxes known as *case rules*. These boxes and lines are considered formatting and are not part of the script. Case ruling and the like are not to be treated as punctuation.

Ordering. The characters are encoded in the Unicode Standard in Latin alphabetical order by primary sign name. Complex signs based on the primary sign are organized according to graphic principles; in some cases, these correspond to the native analyses.

Other Standards. There is no standard legacy encoding of Cuneiform primarily because it was not possible to encode the huge number of characters in the pre-Unicode world of 8-bit fonts.

Cuneiform Numbers and Punctuation: U+12400–U+1247F

Cuneiform Punctuation. A small number of signs are occasionally used in Cuneiform to indicate word division, repetition, or phrase separation.

Cuneiform Numerals. In general, numerals have been encoded separately from signs that are visually identical but semantically different (for example, U+1244F CUNEIFORM NUMERIC SIGN ONE BAN2, U+12450 CUNEIFORM NUMERIC SIGN TWO BAN2, and so on, versus U+12226 CUNEIFORM SIGN MASH, U+1227A CUNEIFORM SIGN PA, and so on).

14.18 Egyptian Hieroglyphs

Egyptian Hieroglyphs: U+13000–U+1342F

Hieroglyphic writing appeared in Egypt at the end of the fourth millennium BCE. The writing system is pictographic: the glyphs represent tangible objects, most of which modern scholars have been able to identify. A great many of the pictographs are easily recognizable even by nonspecialists. Egyptian hieroglyphs represent people and animals, parts of the bodies of people and animals, clothing, tools, vessels, and so on.

Hieroglyphs were used to write Egyptian for more than 3,000 years, retaining characteristic features such as use of color and detail in the more elaborated expositions. Throughout the Old Kingdom, the Middle Kingdom, and the New Kingdom, between 700 and 1,000 hieroglyphs were in regular use. During the Greco-Roman period, the number of variants, as distinguished by some modern scholars, grew to somewhere between 6,000 and 8,000.

Hieroglyphs were carved in stone, painted on frescos, and could also be written with a reed stylus, though this cursive writing eventually became standardized in what is called hieratic writing. The hieratic forms are not separately encoded; they are simply considered cursive forms of the hieroglyphs encoded in this block.

The Demotic script and then later the Coptic script replaced the earlier hieroglyphic and hieratic forms for much practical writing of Egyptian, but hieroglyphs and hieratic continued in use until the fourth century CE. An inscription dated August 24, 394 CE has been found on the Gateway of Hadrian in the temple complex at Philae; this is thought to be among the latest examples of Ancient Egyptian writing in hieroglyphs.

Structure. Egyptian hieroglyphs made use of 24 letters comprising a true alphabet. In addition to these phonetic characters, Egyptian hieroglyphs made use of a very large number of logographic characters (called "logograms" or "ideograms" by Egyptologists), some of which could be read as a word, and some of which had only a semantic determinative function, to enable the reader to distinguish between words which were otherwise written the same. Within a word, characters were arranged together to form an aesthetically-pleasing arrangement within a notional square.

Directionality. Characters may be written left-to-right or right-to-left, generally in horizontal lines, but often—especially in monumental texts—in vertical columns. Directionality of a text is usually easy to determine because one reads a line facing into the glyphs depicting the faces of people or animals.

Egyptian hieroglyphs are given strong left-to-right directionality in the Unicode Standard, because most Egyptian editions are published in English, French, or German, and left-to-right directionality is the conventional presentation mode. When left-to-right directionality is overridden to display Egyptian hieroglyphic text right to left, the glyphs should be mirrored from those shown in the code charts.

Rendering. The encoded characters for Egyptian hieroglyphs in the Unicode Standard simply represent basic text elements, or *signs*, of the writing system. A higher-level protocol is required to represent the arrangement of signs into notional squares and for effects involving mirroring or rotation of signs within text. This approach to encoding the hieroglyphs works well in the context of pre-existing conventions for the representation of Egyptian text, which use simple markup schemes to indicate such formatting.

The most prominent example of such conventions in use since computers were introduced into Egyptology in the late 1970s and the early 1980s is called the *Manuel de Codage* (MdC), published in 1988. The MdC conventions make use of ASCII characters to separate hieroglyphic signs and to indicate the organization of the elements in space—that is, the position of each sign, as arranged in a block. For example, the hyphen-minus "-" is used to separate adjacent hieroglyphic blocks. The colon ":" indicates the superposition of one hieroglyphic sign over another. The asterisk "*" indicates the left-right juxtaposition of two hieroglyphic signs within a visual block.

For example, using the MdC conventions, the hieroglyphic representation of the name Amenhotep would be transliterated as <i-mn:n-R4:t*p>. The lowercase letters represent transliterations of alphabetic or other phonetic signs, whereas "R4" is the catalog label for one of the logograms in the standard Gardiner list. The "-", ":", and "*" characters provide the markup showing how the individual signs are visually arranged. The "<" and ">" bracket characters indicate a cartouche, often used for a king's name. The Unicode representation of the same hieroglyphic name, using MdC conventions, but substituting Unicode characters for the transliterations and catalog numbers is shown in *Table 14-10*.

The interpretation of these MdC markup conventions in text is not part of plain text. Ordinary word processors and plain text display would not be expected to be able to interpret those conventions to render sequences of Egyptian hieroglyphic signs stacked correctly into

Table 14-10. Hieroglyphic Character Sequence

U+003C	LESS-THAN SIGN
U+131CB	EGYPTIAN HIEROGLYPH M017 (= y, i)
U+002D	HYPHEN-MINUS
U+133E0	EGYPTIAN HIEROGLYPH Y005 (= mn)
U+003A	COLON
U+13216	EGYPTIAN HIEROGLYPH N035 (= n)
U+002D	HYPHEN-MINUS
U+132B5	EGYPTIAN HIEROGLYPH R004 (= R4)
U+003A	COLON
U+133CF	EGYPTIAN HIEROGLYPH X001 (= t)
U+002A	ASTERISK
U+132AA	EGYPTIAN HIEROGLYPH Q003 (= p)
U+003E	GREATER-THAN SIGN

blocks. Instead, such display would require a specialized rendering process familiar with the layout of Egyptian hieroglyphs. This distinction is illustrated in *Figure 14-2*. The first line shows the marked-up MdC sequence for the name of the king, Amenhotep. The second line shows the Unicode hieroglyphic version of that sequence, as interpreted by an ordinary Unicode plain text rendering process. The third line shows a rendering by a specialized hieroglyphic rendering process, which can interpret the markup and render a cartouche.

Figure 14-2. Interpretion of Hieroglyphic Markup

Manuel de Codage: <i-mn:n-R4:t*p>

Unicode Plain Text:

Interpreted Markup:

Other markup schemes have been proposed, which attempt to provide greater flexibility than MdC by use of more elaborate encodings. XML has also been used to represent Egyptian texts. Such representations also require specialized rendering systems to lay out hieroglyphic text.

Hieratic Fonts. In the years since Champollion published his decipherment of Egyptian in 1824, Egyptologists have shown little interest in typesetting hieratic text. Consequently, there is no tradition of hieratic fonts in either lead or digital formats. Because hieratic is a cursive form of the underlying hieroglyphic characters, hieratic text is normally rendered using the more easily legible hieroglyphs. In principle a hieratic font could be devised for specialist applications, but as for fonts for other cursive writing systems, it would require very large ligature tables—even larger than usual, because of the great many hieroglyphic signs involved.

Repertoire. The set of hieroglyphic characters encoded in this block is loosely referred to as "the Gardiner set." However, the Gardiner set was not actually exhaustively described and enumerated by Gardiner, himself. The chief source of the repertoire is Gardiner's Middle Egyptian sign list as given in his *Egyptian Grammar* (Gardiner 1957). That list is supplemented by additional characters found in his font catalogues (Gardiner 1928, Gardiner 1929, Gardiner 1931, and Gardiner 1953), and by a collection of signs found in the Griffith Institute's *Topographical Bibliography*, which also used the Gardiner fonts.

A few other characters have been added to this set, such as entities to which Gardiner gave specific catalog numbers. They are retained in the encoding for completeness in representation of Gardiner's own materials. A number of positional variants without catalog numbers were listed in Gardiner 1957 and Gardiner 1928.

Character Names. Egyptian hieroglyphic characters have traditionally been designated in several ways:

- By complex description of the pictographs: GOD WITH HEAD OF IBIS, and so forth.
- By standardized sign number: C3, E34, G16, G17, G24.
- For a minority of characters, by transliterated sound.

The characters in the Unicode Standard make use of the standard Egyptological catalog numbers for the signs. Thus, the name for U+13049 EGYPTIAN HIEROGLYPH E034 refers uniquely and unambiguously to the Gardiner list sign E34, described as a "DESERT HARE" and used for the sound "wn". The catalog values are padded to three places with zeros.

Names for hieroglyphic characters identified explicitly in Gardiner 1953 or other sources as variants for other hieroglyphic characters are given names by appending "A", "B", ... to the sign number. In the sources these are often identified using asterisks. Thus Gardiner's G7, G7*, and G7** correspond to U+13146 EGYPTIAN SIGN G007, U+13147 EGYPTIAN SIGN G007A, and U+13148 EGYPTIAN SIGN G007B, respectively.

Sign Classification. In Gardiner's identification scheme, Egyptian hieroglyphs are classified according to letters of the alphabet, so A000 refers to "Man and his occupations," B000 to "Woman and her occupations," C000 to "Anthropomorphic deities," and so forth. The order of signs in the code charts reflects this classification. The Gardiner categories are shown in headers in the names list accompanying the code charts.

Some individual characters may have been identified as belonging to other classes since their original category was assigned, but the ordering in the Unicode Standard simply follows the original category and catalog values.

Enclosures. The two principal names of the king, the nomen and prenomen, were normally written inside a cartouche: a pictographic representation of a coil of rope, as shown in *Figure 14-2*.

In the Unicode representation of hieroglyphic text, the beginning and end of the cartouche are represented by separate paired characters, somewhat like parentheses. Rendering of a full cartouche surrounding a name requires specialized layout software.

There are a several characters for these start and end cartouche characters, reflecting various styles for the enclosures.

Numerals. Egyptian numbers are encoded following the same principles used for the encoding of Aegean and Cuneiform numbers. Gardiner does not supply a full set of numerals with catalog numbers in his *Egyptian Grammar*, but does describe the system of numerals in detail, so that it is possible to deduce the required set of numeric characters.

Two conventions of representing Egyptian numerals are supported in the Unicode Standard. The first relates to the way in which hieratic numerals are represented. Individual signs for each of the 1s, the 10s, the 100s, the 1000s, and the 10,000s are encoded, because in hieratic these are written as units, often quite distinct from the hieroglyphic shapes into which they are transliterated. The other convention is based on the practice of the *Manual de Codage*, and is comprised of five basic text elements used to build up Egyptian numerals. There is some overlap between these two systems.

