The Unicode® Standard
Version 13.0 – 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 the Unicode Consortium. — Version 13.0.

Includes index.

ISBN 978-1-936213-26-9 (http://www.unicode.org/versions/Unicode13.0.0/)

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

QA268.U545 2020

ISBN 978-1-936213-26-9
Published in Mountain View, CA
March 2020

Chapter 11

Cuneiform and Hieroglyphs

The following scripts are described in this chapter:

Sumero-Akkadian Cuneiform	***Old Persian***	***Meroitic***
Ugaritic	***Egyptian Hieroglyphs***	***Anatolian Hieroglyphs***

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.

Meroitic hieroglyphs and Meroitic cursive were used from around the second century BCE to the fourth century CE to write the Meroitic language of the Nile valley kingdom known as Kush or Meroë. Meroitic cursive was for general use, and its appearance was based on Egyptian demotic. Meroitic hieroglyphs were used for inscriptions, and their appearance was based on Egyptian hieroglyphs.

Anatolian Hieroglyphs date to the second and first millennia BCE, and were used to write the Luwian language, an Indo-European language, in the area of present-day Turkey and environs.

11.1 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 11-1*.

Table 11-1. Cuneiform Script Usage

<table>
<tr><td colspan="3">Archaic Period
(to 2901 BCE)</td></tr>
<tr><td colspan="3">Early Dynastic
(2900–2335 BCE)</td></tr>
<tr><td colspan="3">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 and Archaic Cuneiform are not covered by the repertoire in this block. Signs specific to the Early Dynastic period are encoded separately in the Early Dynastic Cuneiform block.

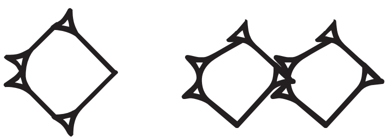
Simple Signs. Most Cuneiform signs are simple units; each sign of this type is represented by a single character in the standard.

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, a 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. Fonts for the representation of Cuneiform text may need to be designed distinctly for optimal use for different historic periods. For example, in the late third millennium BCE, the head of the glyph of the lower right-hand stroke in a ring of four strokes changed its orientation. In earlier times it sloped down to the left, as shown in the glyph for U+1212D, but was later replaced by a stroke in which the head sloped up to the right, as shown in the glyph for U+12423. The glyphs in the code charts do not use a consistent style for these kinds of historic features.



U+1212D U+12423

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 𒋬 CUNEIFORM SIGN TA ASTERISK, a Middle Assyrian form of the sign U+122EB 𒋫 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 𒑏 CUNEIFORM NUMERIC SIGN ONE BAN2, U+12450 𒑐 CUNEIFORM NUMERIC SIGN TWO BAN2, and so on, versus U+12226 𒈦 CUNEIFORM SIGN MASH, U+1227A 𒉺 CUNEIFORM SIGN PA, and so on).

Early Dynastic Cuneiform: U+12480–U+1254F

This block contains characters covering extensions for Cuneiform for the Early Dynastic period, 2900-2335 BCE. The writing of this period is attested primarily from two sites, Fāra and Tell Abū-Ṣalābīkh, both located in the southern part of Iraq. The attestations include administrative, legal, lexical, and literary texts.

The repertoire in this block is compiled primarily from the modern Assyriological sign list of the Early Dynastic period, *Liste der archaischen Keilschriftzeichen aus Fara* (abbreviated LAK), with a few additions derived from other sources. Only Early Dynastic signs not already included in the main Cuneiform block have been added here.

11.2 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.

11.3 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 of Sumero-Akkadian U+121B7 𒆷 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.

11.4 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. To represent the arrangement of signs in notional squares, or quadrats, a set of format controls should be employed (see Egyptian Hieroglyph Format Controls).

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 use the standard Egyptological catalog numbers for the signs. Thus, the name for U+130F9 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.

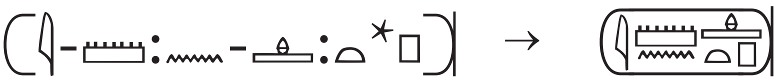
Names for hieroglyphic characters identified explicitly in Gardiner 1953 or other sources as variants for other hieroglyphic characters are given names by appending “A”, “B”, ... to the sign number. In the sources these are often identified using asterisks. Thus Gardiner’s G7, G7*, and G7** correspond to U+13146 EGYPTIAN HIEROGLYPH G007, U+13147 EGYPTIAN HIEROGLYPH G007A, and U+13148 EGYPTIAN HIEROGLYPH 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 11-1*.

Figure 11-1. Representation of Egyptian Hieroglyph Cartouches



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 is handled by the font.

There are a several pairs of characters for the different types of enclosures used in Egyptian Hieroglyphic texts.

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.

Egyptian Hieroglyph Format Controls: U+13430–U+1343F

The structural arrangement of Egyptian Hieroglyphs in notional squares or quadrats is handled by format control characters in the range U+13430..U+13438. Seven of the format characters control the basic placement of hieroglyphs in quadrats. They are used to join hieroglyphs vertically, horizontally, as an overlay, or to insert signs into one of the four corners of a quadrat. Two format controls are used for grouping signs in complex combinations.

Prior to Version 12.0 of Unicode, many Egyptologists used simple markup conventions to indicate formatting, notably the scheme published in the Manuel de Codage (MdC). MdC used ASCII characters to indicate the spatial organization of hieroglyphs. Four of the Egyptian Hieroglyph format controls derive from MdC usage:

- U+13430 EGYPTIAN HIEROGLYPH VERTICAL JOINER indicates a vertical join, and corresponds to MdC use of a colon.
- U+13431 EGYPTIAN HIEROGLYPH HORIZONTAL JOINER indicates a horizontal join, and corresponds to MdC use of an asterisk.
- U+13437 EGYPTIAN HIEROGLYPH BEGIN SEGMENT and U+13438 EGYPTIAN HIEROGLYPH END SEGMENT indicate grouping, and correspond to MdC use of opening and closing parentheses, respectively.

A quadrat layout of one hieroglyph above another is represented by inserting U+13430 EGYPTIAN HIEROGLYPH VERTICAL JOINER between two hieroglyphs, where the first logical glyph in the sequence is the upper of the two hieroglyphs as shown in the first example of *Figure 11-2*. Similarly, U+13431 EGYPTIAN HIEROGLYPH HORIZONTAL JOINER joins two adjacent hieroglyphs horizontally. The horizontal ordering of the joined glyphs matches the logical ordering of the two hieroglyphs, as shown in the second example in *Figure 11-2*.

Figure 11-2. Vertical and Horizontal Formatting of Hieroglyphs

Image	Symbolic	Character Sequence
𓀀𓐰𓉐	A1 $\boxed{:}$ O1	<13000, 13430, 13250>
𓏌𓐱𓏲	W24 $\boxed{*}$ Z7	<133CC, 13431, 133F2>

The column labeled "Symbolic" in *Figure 11-2* (and subsequent figures) emulates the way such quadrats are represented using the MdC conventions. Thus "A1" is the symbolic abbreviation used in MdC for U+13000 EGYPTIAN HIEROGLYPH A001 (a seated man). MdC simply uses a few ASCII characters (":", "*", "+") for the operators that combine signs into sequences expressing the full quadrats. So, the MdC representation of the first example in *Figure 11-2* would be "A1:O1". The symbolic representation in *Figure 11-2* instead uses the dotted box glyph convention to represent the actual Unicode Egyptian Hieroglyph format controls, as for example, U+13430 $\boxed{:}$ EGYPTIAN HIEROGLYPH VERTICAL JOINER.

Four control characters are used in similar fashion to insert a following hieroglyph into the corner of a preceding hieroglyph. The control U+13432 EGYPTIAN HIEROGLYPH INSERT AT TOP START places a following hieroglyph within the frame of the preceding hieroglyph in the corner at the top edge and starting side, as shown in the first example of *Figure 11-3*. Similarly, U+13433 EGYPTIAN HIEROGLYPH INSERT AT BOTTOM START causes a following hieroglyph to display in the bottom-starting corner within the frame of the preceding hieroglyph.

U+13434 EGYPTIAN HIEROGLYPH INSERT AT TOP END causes a following hieroglyph to display in the top-ending corner within the frame of the preceding hieroglyph. U+13435 EGYPTIAN HIEROGLYPH INSERT AT BOTTOM END causes a following hieroglyph to display in the bottom-ending corner within the frame of the preceding hieroglyph. *Figure 11-3* shows examples of this use.

Hieroglyphs may also overlay other hieroglyphs. This arrangement is controlled by U+13436 EGYPTIAN HIEROGLYPH OVERLAY MIDDLE. This control character causes a following hieroglyph to overlay on top of a preceding hieroglyph, as shown in the last example in *Figure 11-3*.

Figure 11-3. Insertion and Overlay Formatting of Hieroglyphs

Image	Symbolic	Character Sequence
	F4 [⌟] X1	<13102, 13432, 133CF>
	I10 [⌝] A1	<13193, 13433, 13000>
	D17 [⌞] X1	<13087, 13434, 133CF>
	G25 [⌜] X1	<1315C, 13435, 133CF>
	D36 [+] V28	<1309D, 13436, 1339B>

Complex Clusters. The basic joining controls may be used in conjunction with one another to render more complex clusters, as shown in the first example in *Figure 11-4*.

The two characters, U+13437 EGYPTIAN HIEROGLYPH BEGIN SEGMENT and U+13438 EGYPTIAN HIEROGLYPH END SEGMENT, are used to group signs in complex clusters comprising different levels of joining controls, as shown in the second example in *Figure 11-4*.

Some rendering systems may support multiple levels of the segment controls for use in the most complex hieroglyphic sign arrangements, as shown in the third example in *Figure 11-4*.

Figure 11-4. Complex Cluster Formatting of Hieroglyphs

Image	Symbolic	Character Sequence
	I9 [:] N35 [:] F20 [⌝] A1	<13191, 13430, 13216, 13430, 13113, 13433, 13000>
	G9 [⌞] [(] N27 [:] N27 [)]	<1314A, 13433, 13437, 1320C, 13430, 1320C, 13438>
	J15 [:] Z11 [*] [(] D2 [*] [(] D21 [:] X1 [)] [:] N25 [)]	<1341D, 13430, 133F6, 13431, 13437, 13077, 13431, 13437, 1308B, 13430, 133CF, 13438, 13430, 13209, 13438>

Some Egyptian hieroglyphs with complex structures have previously been encoded as single, atomically encoded characters. These atomically encoded characters should be used instead of sequences of hieroglyphs with appropriate joining controls, as demonstrated by the examples in *Table 11-2*. Sequence checking is comparable to spell-checking in alphabetic languages and is the responsibility of a higher-level protocol.

Table 11-2. Complex Hieroglyphs and Nonequivalent Sequences

For	Use	Do Not Use
	13217	<13216, 13430, 13216, 13430, 13216>
	130C1	<130C0, 13436, 1309D>
	13196	<13193, 13433, 13437, 133CF, 13430, 131FF, 13438>

Sometimes a portion of a graphically complex quadrat could be identified as an atomically encoded character. However, in cases where the use of that atomically encoded character as a component of a quadrat sequence would cause ambiguities or uneven distribution in the structure, then a sequence of simpler hieroglyphs should be used instead, with the appropriate joining controls. This principle is exemplified in the first example in *Table 11-3*.

Table 11-3. Complex Quadrat Representation

Display		Character Sequence
	✓	<13216, 13430, 13216, 13430, 13216, 13430, 13216>
		<13216, 13430, 13217>
		<13217, 13430, 13216>
	✓	<13283, 13430, 130C0, 13431, 13437, 13216, 13430, 13216, 13430, 13216, 13438>
		<13283, 13430, 130C0, 13431, 13217>

In the first example shown in *Table 11-3*, a nonce quadrat consisting of four stacked water signs would be rendered with the signs evenly spaced when represented by a sequence of four single water signs, U+13216 EGYPTIAN HIEROGLYPH N035, joined by U+13430 EGYPTIAN HIEROGLYPH VERTICAL JOINER inserted between each sign. This would be the preferred representation, indicated with a check mark in the first line of this example. Alternative representations, combining the triple water sign U+13217 EGYPTIAN HIEROGLYPH N035A and a single water sign, U+13216 EGYPTIAN HIEROGLYPH N035, in either order, would result in uneven display, as shown in the second and third lines of the example.

In complex quadrat structures, such as the nonce quadrat in the second example in *Table 11-3*, the longer sequence composed of simpler signs and format controls should also be employed, as indicated with a check mark, instead of using a graphically complex sign as a component of the larger quadrat.

11.5 Meroitic

Meroitic Hieroglyphs: U+10980–U+1099F
Meroitic Cursive: U+109A0–U+109FF

Meroitic hieroglyphs and Meroitic cursive were used from around the second century BCE to the fourth century CE to write the Meroitic language of the Nile valley kingdom known as Kush or Meroë. The kingdom originated south of Egypt around 850 BCE, with its capital at Napata, located in modern-day northern Sudan. At that time official inscriptions used the Egyptian language and script. Around 560 BCE the capital was relocated to Meroë, about 600 kilometers upriver. As the use of Egyptian language and script declined with the greater distance from Egypt, two native scripts developed for writing Meroitic:

- Meroitic cursive was for general use, and its appearance was based on Egyptian demotic.
- Meroitic hieroglyphs were used for inscriptions on royal monuments and temples, and their appearance was based on Egyptian hieroglyphs. (See *Section 11.4, Egyptian Hieroglyphs* for more information.)

After the fourth century CE, the Meroitic language was gradually replaced by Nubian, and by the sixth century the Meroitic scripts had been superseded by the Coptic script, which picked up three additional symbols from Meroitic cursive to represent Nubian.

Although the values of the script characters were deciphered around 1911 by the English Egyptologist F. L. Griffith, the Meroitic language is still not understood except for names and a few other words. It is not known to be related to any other language. It may be related to Nubian.

Structure. Unlike the Egyptian scripts, the Meroitic scripts are almost purely alphabetic. There are 15 basic consonants; if not followed by an explicit vowel letter, they are read with an inherent *a*. There are four vowels: *e*, *i*, *o*, and *a*. The *a* vowel is only used for initial *a*. In addition, for unknown reasons, there are explicit letters for the syllables *ne*, *te*, *se*, and *to*. This may have been due to dialect differences, or to the possible use of *n*, *t*, and *s* as final consonants in some cases.

Meroitic cursive also uses two logograms for *rmt* and *imn*, derived from Egyptian demotic.

Directionality. Horizontal writing is almost exclusively right-to-left, matching the direction in which the hieroglyphs depicting people and animals are looking. This is unlike Egyptian hieroglyphs, which are read into the faces of the glyphs for people and animals. Meroitic hieroglyphs are also written vertically in columns.

Shaping. In Meroitic cursive, the letter for *i* usually connects to a preceding consonant. There is no other connecting behavior.

Punctuation. The Meroitic scripts were among the earliest to use word division—not always consistently—to separate basic sentence elements, such as noun phrases, verb forms, and so on. For this purpose Meroitic hieroglyphs use three vertical dots, represented by U+205D TRICOLON. When Meroitic hieroglyphs are presented in vertical columns, the orientation of the three dots shifts to become three horizontal dots. This can be represented either with U+2026 HORIZONTAL ELLIPSIS, or in more sophisticated rendering, by glyphic rotation of U+205D TRICOLON. Meroitic cursive uses two vertical dots, represented by U+003A COLON.

Symbols. Two ankh-like symbols are used with Meroitic hieroglyphs.

Meroitic Cursive Numbers. Meroitic numbers are found only in Meroitic Cursive. The system consists of numbers one through nine and bases for ranks: tens, hundreds, thousands, ten thousands, and hundred thousands. The numbers for 100 and higher are systematically formed by attaching the numbers for one through nine as a multiplier to the respective base for each rank. There is also a notation for a fractional system based on twelfths, which simply uses one to eleven dots to represent each fraction.

11.6 Anatolian Hieroglyphs

Anatolian Hieroglyphs: U+14400–U+1467F

Anatolian hieroglyphs appeared on personal seals, monumental inscriptions, and other objects in the second and first millennia BCE in present-day Turkey and surrounding areas. The script, known also as Luwian or Luvian hieroglyphs, was used primarily to write the Luwian language.

Structure. Anatolian hieroglyphs contain both syllabic and logographic elements. Words can be represented by logographs alone, by logographs with a phonetic complement, or solely by syllabic values.

Directionality. Anatolian hieroglyphs can be written left-to-right, right-to-left, or boustrophedon, and lines are often divided by horizontal rules. Within a line, characters are grouped vertically, typically from top to bottom, although the characters may be placed out of phonetic or logical order for aesthetic reasons.

The characters in the Anatolian Hieroglyphs block have a strong left-to-right directionality (Bidi_Class = L), because publications typically lay out hieroglyphs from left to right. When Anatolian hieroglyphs are displayed right to left, the glyphs should be mirrored from those shown in the code charts.

Repertoire. The repertoire of characters is broadly based on the sign catalog of Laroche (1960), supplemented by additions from later handbooks. Some signs contained in Laroche are considered variants today, but have been encoded separately to represent the complete history of Anatolian scholarship and discussions about the decipherment.

Character names for variant signs are usually distinguished by an “A”, “B”, or “C” appended to the catalog number of the main sign. For example, U+14600 ANATOLIAN HIEROGLYPH A457A is a variant of U+145FF ANATOLIAN HIEROGLYPH A457.

A few hieroglyphs developed a simplified, cursive shape, based on the more pictorial shape of the signs found on monuments. The simplified forms are encoded separately, and are differentiated in their names.

U+1442B 𔐫 ANATOLIAN HIEROGLYPH A041 (monumental style)
= capere
= syllabic tà

U+1442C 𔐬 ANATOLIAN HIEROGLYPH A041A (cursive style)
= syllabic tà

The script contains a productive grapheme, U+145B1 𔖱 ANATOLIAN HIEROGLYPH A383 RA OR RI, which appears as a part of several other signs, such as U+145B9 𔖹 ANATOLIAN HIEROGLYPH A389. The characters containing this graphic element as part of their form are not decomposable.

Annotations. Latin names are used traditionally to describe characters used logographically and appear as annotations in the names list. Those characters which have a Luwian phonetic value or are logosyllabic are identified in the annotations. When a plus sign appears between two elements in the annotation, the elements are considered a single graphic unit, whereas a period between the two elements indicates the two elements are considered graphically separate.

U+1447E 𔑾 ANATOLIAN HIEROGLYPH A107A
= bos+mi

U+14480 𔒀 ANATOLIAN HIEROGLYPH A107B
= bos.mi

Punctuation. In some texts, word division is indicated by U+145B5 ANATOLIAN HIEROGLYPH A386 or its variant U+145B6 ANATOLIAN HIEROGLYPH A386A. U+145CE ANATOLIAN HIEROGLYPH A410 BEGIN LOGOGRAM MARK and U+145CF ANATOLIAN HIEROGLYPH A410A END LOGOGRAM MARK sometimes occur in text to mark logograms.

The characters U+145F7 ANATOLIAN HIEROGLYPH A450 and U+144EF ANATOLIAN HIEROGLYPH A209 are occasionally used to fill blank spaces, often at the end of a word. Spaces are used in modern renditions of hieroglyphic text.

Numbers. Some of the hieroglyphic signs have been interpreted as having numeric values. These include values for 1–5, 8–10, 12, 100, and 1000. However, all of the Anatolian hieroglyphs have the General_Category = Other_Letter and no specific numeric values for them are assigned in the Unicode Character Database.

Rendering. Just as for Egyptian hieroglyphs, only the basic text elements of the script are encoded. A higher-level protocol is required for the display Anatolian hieroglyphs in a nonlinear layout.