The Unicode® Standard
Version 14.0 – Core Specification

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

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

Includes index.

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

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

QA268.U545 2021

ISBN 978-1-936213-29-0

Published in Mountain View, CA

September 2021

Chapter 19
Africa

This chapter covers the following scripts of Africa:

Ethiopic	***Vai***	***Mende Kikakui***
Osmanya	***Bamum***	***Adlam***
Tifinagh	***Bassa Vah***	***Medefaidrin***
N'Ko		

Ethiopic and Tifinagh are scripts with long histories. Although their roots can be traced back to the original Semitic and North African writing systems, they would not be classified as Middle Eastern scripts today.

The remaining scripts in this chapter have been developed relatively recently. Some of them show roots in Latin and other letterforms. They are all original creative contributions intended specifically to serve the linguistic communities that use them.

Osmanya is an alphabetic script developed in the early 20th century to write the Somali language. N'Ko is a right-to-left alphabetic script devised in 1949 as a writing system for Manden languages in West Africa. Vai is a syllabic script used for the Vai language in Liberia and Sierra Leone; it was developed in the 1830s, but the standard syllabary was published in 1962. Bamum is a syllabary developed between 1896 and 1910, used for writing the Bamum language in western Cameroon. Modern Bassa Vah is an alphabetic script developed early in the 20th century. Mende Kikakui is a right-to-left script used for writing Mende. It was also created in the early 20th century.

Adlam is an alphabetic script used to write Fulani and other African languages. The Fulani are a widespread ethnic group in Africa, and the Fulani language is spoken by more than 40 million people. The script was developed in the late 1980s, and was subsequently widely adopted among Fulani communities, where it is taught in schools.

The Medefaidrin script is used to write the liturgical language Medefaidrin by members of an indigenous Christian church in Nigeria. According to community tradition, the language was revealed to one of the founders of the community in 1927 by divine inspiration. It is presently used for Sunday school lessons and prayers or meditation.

19.1 Ethiopic

Ethiopic: U+1200–U+137F

The Ethiopic syllabary originally evolved for writing the Semitic language Ge'ez. Indeed, the English noun "Ethiopic" simply means "the Ge'ez language." Ge'ez itself is now limited to liturgical usage, but its script has been adopted for modern use in writing several languages of central east Africa, including Amharic, Tigre, and Oromo.

Basic and Extended Ethiopic. The Ethiopic characters encoded here include the basic set that has become established in common usage for writing major languages. As with other productive scripts, the basic Ethiopic forms are sometimes modified to produce an extended range of characters for writing additional languages.

Encoding Principles. The syllables of the Ethiopic script are traditionally presented as a two-dimensional matrix of consonant-vowel combinations. The encoding follows this structure; in particular, the codespace range U+1200..U+1357 is interpreted as a matrix of 43 consonants crossed with 8 vowels, making 344 conceptual syllables. Most of these consonant-vowel syllables are represented by characters in the script, but some of them happen to be unused, accounting for the blank cells in the matrix.

Variant Glyph Forms. A given Ethiopic syllable may be represented by different glyph forms, analogous to the glyph variants of Latin lowercase "a" or "g", which do not coexist in the same font. Thus the particular glyph shown in the code chart for each position in the matrix is merely one representation of that conceptual syllable, and the glyph itself is not the object that is encoded.

Labialized Subseries. A few Ethiopic consonants have labialized ("W") forms that are traditionally allotted their own consonant series in the syllable matrix, although only a subset of the possible vowel forms are realized. Each of these derivative series is encoded immediately after the corresponding main consonant series. Because the standard vowel series includes both "AA" and "WAA", two different cells of the syllable matrix might represent the "consonant + W + AA" syllable. For example:

U+1257 = QH + WAA: potential but unused version of QHWAA

U+125B = QHW + AA: ETHIOPIC SYLLABLE QHWAA

In these cases, where the two conceptual syllables are equivalent, the entry in the labialized subseries is encoded and not the "consonant + WAA" entry in the main syllable series. The six specific cases are enumerated in *Table 19-1*. In three of these cases, the -WAA position in the syllable matrix has been reanalyzed and used for encoding a syllable in -OA for extended Ethiopic.

Table 19-1. Labialized Forms in Ethiopic -WAA

-WAA Form	Encoded as	Not Used	Contrast
QWAA	U+124B ቋ	1247	U+1247 ቇ QOA
QHWAA	U+125B ቻ	1257	
XWAA	U+128B ኋ	1287	U+1287 ኧ XOA
KWAA	U+12B3 ኳ	12AF	U+12AF ኰ KOA
KXWAA	U+12C3 ዃ	12BF	
GWAA	U+1313 ጓ	130F	

Also, *within* the labialized subseries, the sixth vowel ("-E") forms are sometimes considered to be second vowel ("-U") forms. For example:

U+1249 = QW + U: unused version of QWE

U+124D = QW + E: ETHIOPIC SYLLABLE QWE

In these cases, where the two syllables are nearly equivalent, the "-E" entry is encoded and not the "-U" entry. The six specific cases are enumerated in *Table 19-2*.

Table 19-2. Labialized Forms in Ethiopic -WE

"-WE" Form	Encoded as	Not Used
QWE	U+124D ቍ	1249
QHWE	U+125D ቝ	1259
XWE	U+128D ኍ	1289
KWE	U+12B5 ኵ	12B1
KXWE	U+12C5 ዅ	12C1
GWE	U+1315 ጕ	1311

Keyboard Input. Because the Ethiopic script includes more than 300 characters, the units of keyboard input must constitute some smaller set of entities, typically 43+8 codes interpreted as the coordinates of the syllable matrix. Because these keyboard input codes are expected to be transient entities that are resolved into syllabic characters before they enter stored text, keyboard input codes are not specified in this standard.

Syllable Names. The Ethiopic script often has multiple syllables corresponding to the same Latin letter, making it difficult to assign unique Latin names. Therefore the names list makes use of certain devices (such as doubling a Latin letter in the name) merely to create uniqueness; this device has no relation to the phonetics of these syllables in any particular language.

Encoding Order and Sorting. The order of the consonants in the encoding is based on the traditional alphabetical order. It may differ from the sort order used for one or another language, if only because in many languages various pairs or triplets of syllables are treated as equivalent in the first sorting pass. For example, an Amharic dictionary may start out with a section headed by *three* H-like syllables:

U+1200 ETHIOPIC SYLLABLE HA

U+1210 ETHIOPIC SYLLABLE HHA

U+1280 ETHIOPIC SYLLABLE XA

Thus the encoding order cannot and does not implement a collation procedure for any particular language using this script.

Diacritical Marks. The Ethiopic script generally makes no use of diacritical marks, but they are sometimes employed for scholarly or didactic purposes. In particular, U+135F ETHIOPIC COMBINING GEMINATION MARK and U+030E COMBINING DOUBLE VERTICAL LINE ABOVE are sometimes used to indicate emphasis or gemination (consonant doubling).

Numbers. Ethiopic digit glyphs are derived from the Greek alphabet, possibly borrowed from Coptic letterforms. In modern use, European digits are often used. The Ethiopic number system does not use a zero, nor is it based on digital-positional notation. A number is denoted as a sequence of powers of 100, each preceded by a coefficient (2 through 99). In each term of the series, the power 100^n is indicated by n HUNDRED characters (merged to a digraph when n = 2). The coefficient is indicated by a *tens* digit and a *ones* digit, either of which is absent if its value is zero.

For example, the number 2,345 is represented by

2,345 = (20 + 3)*100^1 + (40 + 5)*100^0

= 20 3 100 40 5

= TWENTY THREE HUNDRED FORTY FIVE

= 1373 136B 137B 1375 136D 

A language using the Ethiopic script may have a *word* for "thousand," such as Amharic "SHI" (U+123A), and a quantity such as 2,345 may also be written as it is spoken in that language, which in the case of Amharic happens to parallel English:

2,345 = TWO thousand THREE HUNDRED FORTY FIVE

= 136A 123A 136B 137B 1375 136D 

In Ge'ez language manuscripts the conjunction "**ወ**" is frequently used to write numbers as they would be spoken.

For example, the number 2,345 would then be written in a Ge'ez language document as

2,345 = TWENTY and THREE HUNDRED with FORTY and FIVE

= 136A 12C8 136B 137B 12C8 1375 12C8 136D 

Word Separators. The traditional word separator is U+1361 ETHIOPIC WORDSPACE (:). In modern usage, a plain white wordspace (U+0020 SPACE) is becoming common.

Section Mark. One or more s*ection marks* are typically used on a separate line to mark the separation of sections. Commonly, an odd number is used and they are separated by spaces.

Ethiopic Extensions

The Ethiopic script is used for a large number of languages and dialects in Ethiopia and in some instances has been extended significantly beyond the set of characters used for major languages such as Amharic and Tigre. There are four blocks of extensions to the Ethiopic script: Ethiopic Supplement U+1380..U+139F, Ethiopic Extended U+2D80..U+2DDF, Ethiopic Extended-A U+AB00..U+AB2F, and Ethiopic Extended-B U+1E7E0..U+1E7FF. Those extensions cover such languages as Me'en, Blin, and the Gurage languages, which use many additional characters. The Ethiopic Extended-A block, in particular, includes characters for the Gamo-Gofa-Dawro, Basketo, and Gumuz languages. Several other characters for Ethiopic script extensions can be found in the main Ethiopic script block in the range U+1200..U+137F, including combining diacritical marks used for Basketo.

The Ethiopic Extended-B block contains characters for the modern Gurage orthography, which covers the Inor, Mesqan, Sebatbeit, and Soddo languages. Additional characters for this orthography can be found in the Ethiopic Supplement and Ethiopic Extended blocks. Some of the character names in these blocks include the word "SEBATBEIT" because they were originally encoded for the older Sebatbeit orthography. The modern Gurage orthography uses some of these characters for all Gurage languages, including Sebatbeit.

The Ethiopic Supplement block also contains a set of tonal marks. They are used in multiline scored layout. Like other musical (an)notational systems of this type, these tonal marks require a higher-level protocol to enable proper rendering.

19.2 Osmanya

Osmanya: U+10480–U+104AF

The Osmanya script, which in Somali is called 𐒍𐒖𐒇 𐒈𐒝𐒑𐒛𐒐𐒘 *far Soomaali* "Somali writing" or 𐒋𐒘𐒈𐒑𐒛𐒒𐒕𐒖 *Cismaanya*, was devised in 1920–1922 by 𐒋𐒘𐒈𐒑𐒛𐒒 𐒕𐒓𐒈𐒚𐒍 𐒏𐒜𐒒𐒖𐒆𐒕𐒆 (Cismaan Yuusuf Keenadiid) to represent the Somali language. It replaced an attempt by Sheikh Uweys of the Confraternity Qadiriyyah (died 1909) to devise an Arabic-based orthography for Somali. It has, in turn, been replaced by the Latin orthography of Muuse Xaaji Ismaaciil Galaal (1914–1980). In 1961, both the Latin and the Osmanya scripts were adopted for use in Somalia, but in 1969 there was a coup, with one of its stated aims being the resolution of the debate over the country's writing system. A Latin orthography was finally adopted in 1973. Gregersen (1977) states that some 20,000 or more people use Osmanya in private correspondence and bookkeeping, and that several books and a biweekly journal *Horseed* ("*Vanguard*") were published in cyclostyled format.

Structure. Osmanya is an alphabetic script, read from left to right in horizontal lines running from top to bottom. It has 22 consonants and 8 vowels. Unique long vowels are written for U+1049B 𐒛 OSMANYA LETTER AA, U+1049C 𐒜 OSMANYA LETTER EE, and U+1049D 𐒝 OSMANYA LETTER OO; long *uu* and *ii* are written with the consonants U+10493 𐒓 OSMANYA LETTER WAW and U+10495 𐒕 OSMANYA LETTER YA, respectively.

Ordering. Alphabetical ordering is based on the order of the Arabic alphabet, as specified by Osman Abdihalim Yuusuf Osman Keenadiid. This ordering is similar to the ordering given in Diringer (1996).

Character Names and Glyphs. The character names used in the Unicode Standard are as given by Osman. The glyphs shown in the code charts are taken from *Afkeenna iyo fartysa* ("Our language and its handwriting") 1971.

19.3 Tifinagh

Tifinagh: U+2D30–U+2D7F

The Tifinagh script is used by approximately 20 million people who speak varieties of languages commonly called Berber or Amazigh. The three main varieties in Morocco are known as Tarifite, Tamazighe, and Tachelhite. In Morocco, more than 40% of the population speaks Berber. The Berber language, written in the Tifinagh script, is currently taught to approximately 300,000 pupils in 10,000 schools—mostly primary schools—in Morocco. Three Moroccan universities offer Berber courses in the Tifinagh script leading to a Master's degree.

Tifinagh is an alphabetic writing system. It uses spaces to separate words and makes use of Western punctuation.

History. The earliest variety of the Berber alphabet is Libyan. Two forms exist: a Western form and an Eastern form. The Western variety was used along the Mediterranean coast from Kabylia to Morocco and most probably to the Canary Islands. The Eastern variety, Old Tifinagh, is also called Libyan-Berber or Old Tuareg. It contains signs not found in the Libyan variety and was used to transcribe Old Tuareg. The word *tifinagh* is a feminine plural noun whose singular would be *tafniqt*; it means "the Phoenician (letters)."

Neo-Tifinagh refers to the writing systems that were developed to represent the Maghreb Berber dialects. A number of variants of Neo-Tifinagh exist, the first of which was proposed in the 1960s by the Académie Berbère. That variant has spread in Morocco and Algeria, especially in Kabylia. Other Neo-Tifinagh systems are nearly identical to the Académie Berbère system. The encoding in the Tifinagh block is based on the Neo-Tifinagh systems.

Source Standards. The encoding consists of four Tifinagh character subsets: the basic set of the Institut Royal de la Culture Amazighe (IRCAM), the extended IRCAM set, other Neo-Tifinagh letters in use, and modern Tuareg letters. The first subset represents the set of characters chosen by IRCAM to unify the orthography of the different Moroccan modern-day Berber dialects while using the historical Tifinagh script.

Ordering. The letters are arranged according to the order specified by IRCAM. Other Neo-Tifinagh and Tuareg letters are interspersed according to their pronunciation.

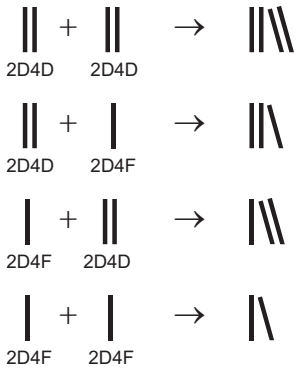
Directionality. Historically, Berber texts did not have a fixed direction. Early inscriptions were written horizontally from left to right, from right to left, vertically (bottom to top, top to bottom); boustrophedon directionality was also known. Modern-day Berber script is most frequently written in horizontal lines from left to right; therefore the bidirectional class for Tifinagh letters is specified as strong left to right. Displaying Berber texts in other directions can be accomplished by the use of directional overrides or by the use of higher-level protocols.

Diacritical Marks. Modern Tifinagh variants tend to use combining diacritical marks to complement the Tifinagh block. The Hawad notation, for example, uses diacritical marks from the Combining Diacritical Marks block (U+0300–U+036F). These marks are used to

represent vowels and foreign consonants. In this notation, <U+2D35, U+0307> represents "a", <U+2D49, U+0309> represents a long "i" /i:/, and <U+2D31, U+0302> represents a "p". Some long vowels are represented using two diacritical marks above. A long "e" /e:/ is thus written <U+2D49, U+0307, U+0304>. These marks are displayed side by side above their base letter in the order in which they are encoded, instead of being stacked.

Yal and Yan. While the neo-Tifinagh glyph for U+2D4D TIFINAGH LETTER YAL in Morocco is typically rendered with two bars linked by a small slanted stroke ⵍ, traditional texts from all areas, and modern-day materials from areas outside Morocco often represent *yal* with two vertical strokes ⵍⵍ. However, the two vertical bar shape can cause visual ambiguity in words with consonant clusters, because *yal* may be mistaken for two instances of U+2D4F TIFINAGH LETTER YAN, whose glyph is a single vertical stroke ⵏ. Individual font designers, local traditions, and national preferences employ various means to prevent confusion, including varying the spacing between the bars, and slanting or lowering the bars. *Figure 19-1* shows examples that illustrate contextual shaping by slanting the bars of *yal* and *yan*.

Figure 19-1. Tifinagh Contextual Shaping

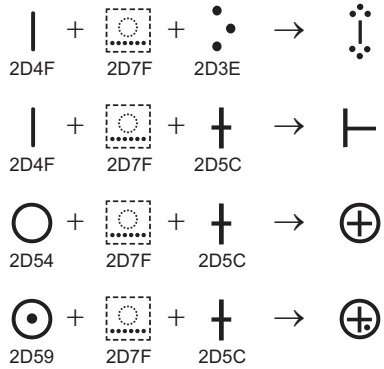


Bi-Consonants. Bi-consonants are additional letterforms used in the Tifinagh script, particularly for Tuareg, to represent a consonant cluster—a sequence of two consonants without an intervening vowel. These bi-consonants, sometimes also referred to as bigraphs, are not directly encoded as single characters in the Unicode Standard. Instead, they are represented as a sequence of the two consonant letters, separated either by U+200D ZERO WIDTH JOINER or by U+2D7F TIFINAGH CONSONANT JOINER.

When a bi-consonant is considered obligatory in text, it is represented by the two consonant letters, with U+2D7F TIFINAGH CONSONANT JOINER inserted between them. This use of U+2D7F is comparable in function to the use of U+0652 ARABIC SUKUN to indicate the absence of a vowel after a consonant, when Tuareg is written in the Arabic script. However, instead of appearing as a visible mark in the text, U+2D7F TIFINAGH CONSONANT JOINER indicates the presence of a bi-consonant, which should then be rendered with a preformed

glyph for the sequence. Examples of common Tifinagh bi-consonants and their representation are shown in *Figure 19-2*.

Figure 19-2. Tifinagh Consonant Joiner and Bi-consonants



If a rendering system cannot display obligatory bi-consonants with the correct, fully-formed bi-consonant glyphs, a fallback rendering should be used which displays the TIFINAGH CONSONANT JOINER visibly, so that the correct textual distinctions are maintained, even if they cannot be properly displayed.

When a bi-consonant is considered merely an optional, ligated form of two consonant letters, the bi-consonant can be represented by the two consonant letters, with U+200D ZERO WIDTH JOINER inserted between them, as a hint that the ligated form is preferred. If a rendering system cannot display the optional, ligated form, the fallback display should simply be the sequence of consonants, with no visible display of the ZWJ.

Bi-consonants often have regional glyph variants, so fonts may need to be designed differently for different regional uses of the Tifinagh script.

19.4 N'Ko

N'Ko: U+07C0–U+07FF

N'Ko is a literary dialect used by the Manden (or Manding) people, who live primarily in West Africa. The script was devised by Solomana Kante in 1949 as a writing system for the Manden languages. The Manden language group is known as *Mandenkan*, where the suffix *-kan* means "language of." In addition to the substantial number of Mandens, some non-Mandens speak *Mandenkan* as a second language. There are an estimated 20 million Mandenkan speakers.

The major dialects of the Manden language are Bamanan, Jula, Maninka, and Mandinka. There are a number of other related dialects. When Mandens from different subgroups talk to each other, it is common practice for them to switch—consciously or subconsciously—from their own dialect to the conventional, literary dialect commonly known as *Kangbe*, "the clear language," also known as N'Ko. This dialect switching can occur in conversations between the Bamanan of Mali, the Maninka of Guinea, the Jula of the Ivory Coast, and the Mandinka of Gambia or Senegal, for example. Although there are great similarities between their dialects, speakers sometimes find it necessary to switch to *Kangbe* (N'Ko) by using a common word or phrase, similar to the accommodations Danes, Swedes, and Norwegians sometimes make when speaking to one another. For example, the word for "name" in Bamanan is *togo*, while it is *tooh* in Maninka. Speakers of both dialects will write it as ߞ̂ , although each may pronounce it differently.

Character Names and Block Name. Although the traditional name of the N'Ko language and script includes an apostrophe, apostrophes are disallowed in Unicode character and block names. Because of this, the formal block name is "NKo" and the script portion of the Unicode character names is "NKO".

Structure. The N'Ko script is written from right to left. It is phonetic in nature (one symbol, one sound). N'Ko has seven vowels, each of which can bear one of seven diacritical marks that modify the tone of the vowel as well as an optional diacritical mark that indicates nasalization. N'Ko has 19 consonants and two "abstract" consonants, U+07E0 NKO LETTER NA WOLOSO and U+07E7 NKO LETTER NYA WOLOSO, which indicate original consonants mutated by a preceding nasal, either word-internally or across word boundaries. Some consonants can bear one of three diacritical marks to transcribe foreign sounds or to transliterate foreign letters.

U+07D2 NKO LETTER N is considered neither a vowel nor a consonant; it indicates a syllabic alveolar or velar nasal. It can bear a diacritical mark, but cannot bear the nasal diacritic. The letter U+07D1 NKO LETTER DAGBASINNA has a special function in N'Ko orthography. The standard spelling rule is that when two successive syllables have the same vowel, the vowel is written only after the second of the two syllables. For example, ߓ߬ <ba, la, oo> is pronounced [bolo], but in a foreign syllable to be pronounced [blo], the *dagbasinna* is inserted for ߓ᳽ <ba, dagbasinna, la, oo> to show that a consonant cluster is intended.

Diacritical Marks. N'Ko diacritical marks are script-specific, despite superficial resemblances to other diacritical marks encoded for more general use. Some N'Ko diacritics have a wider range of glyph representation than the generic marks do, and are typically drawn rather higher and bolder than the generic marks.

Two of the tone diacritics, when applied to consonants, indicate specific sounds from other languages—in particular, Arabic or French language sounds. U+07F3 NKO COMBINING DOUBLE DOT ABOVE is also used as a diacritic to represent sounds from other languages. The combinations used are as shown in *Table 19-3*.

Table 19-3. N'Ko Diacritic Usage

Character	Applied To	Represents
U+07EB NKO COMBINING SHORT HIGH TONE	SA	[s] or Arabic ص SAD
	GBA	[ɣ] or Arabic غ GHAIN
	KA	[q] or Arabic ق QAF
U+07ED NKO COMBINING SHORT RISING TONE	BA	[ɓ]
	TA	[ŧ] or Arabic ط TAH
	JA	[z] or Arabic ز ZAIN
	CHA	[ð] or Arabic ذ THAL and also French [ʒ]
	DA	[đ] or Arabic ض ZAD
	RA	French [ʀ]
	SA	[ʃ] or Arabic ش SHEEN
	GBA	[g]
	FA	[v]
	KA	[x] or Arabic خ KHAH
	LA	[lˤ]
	MA	[ɱ]
	NYA	[ŋ]
	HA	[ħ] or Arabic ح HAH
	YA	[j]
U+07F3 NKO COMBINING DOUBLE DOT ABOVE	A	[ʕa] or Arabic عَ AIN + A
	EE	French [ə]
	U	French [y]
	JA	[ʒ] or Arabic ظ ZAH
	DA	[ɖ]
	SA	[θ] or Arabic ث THEH
	GBA	[kp]

Table 19-4 shows the use of the tone diacritics when applied to vowels.

Table 19-4. N'Ko Tone Diacritics on Vowels

Character		Tone	Applied To
U+07EB	NKO COMBINING SHORT HIGH TONE	high	short vowel
U+07EC	NKO COMBINING SHORT LOW TONE	low	short vowel
U+07ED	NKO COMBINING SHORT RISING TONE	rising-falling	short vowel
U+07EE	NKO COMBINING LONG DESCENDING TONE	descending	long vowel
U+07EF	NKO COMBINING LONG HIGH TONE	high	long vowel
U+07F0	NKO COMBINING LONG LOW TONE	long low	long vowel
U+07F1	NKO COMBINING LONG RISING TONE	rising	long vowel

When applied to a vowel, U+07F2 NKO COMBINING NASALIZATION MARK indicates the nasalization of that vowel. In the text stream, this mark is applied before any of the tone marks because combining marks below precede combining marks above in canonical order.

Digits. N'Ko uses decimal digits specific to the script. These digits have strong right-to-left directionality. Numbers are stored in text in logical order with most significant digit first; when displayed, numerals are then laid out in right-to-left order, with the most significant digit at the rightmost side, as illustrated for the numeral 144 in *Figure 19-3*. This situation differs from how numerals are handled in Hebrew and Arabic, where numerals are laid out in left-to-right order, even though the overall text direction is right to left.

Ordinal Numbers. Diacritical marks are also used to mark ordinal numbers. The first ordinal is indicated by applying U+07ED NKO COMBINING SHORT RISING TONE (a dot above) to U+07C1 NKO DIGIT ONE. All other ordinal numbers are indicated by applying U+07F2 NKO COMBINING NASALIZATION MARK (an oval dot below) to the last digit in any sequence of digits composing the number. Thus the nasalization mark under the digit two would indicate the ordinal value 2nd, while the nasalization mark under the final digit four in the numeral 144 would indicate the ordinal value 144th, as shown in *Figure 19-3*.

Figure 19-3. Examples of N'Ko Ordinals

Punctuation. N'Ko uses a number of punctuation marks in common with other scripts. U+061F ARABIC QUESTION MARK, U+060C ARABIC COMMA, U+061B ARABIC SEMICOLON, and the paired U+FD3E ORNATE LEFT PARENTHESIS and U+FD3F ORNATE RIGHT PARENTHESIS are used, often with different shapes than are used in Arabic. A script-specific U+07F8 NKO COMMA and U+07F9 NKO EXCLAMATION MARK are encoded. The NKO COMMA

differs in shape from the ARABIC COMMA, and the two are sometimes used distinctively in the same N'Ko text.

The character U+07F6 NKO SYMBOL OO DENNEN is used as an addition to phrases to indicate remote future placement of the topic under discussion. The decorative U+07F7 ♧ NKO SYMBOL GBAKURUNEN represents the three stones that hold a cooking pot over the fire and is used to end major sections of text.

The two tonal apostrophes, U+07F4 NKO HIGH TONE APOSTROPHE and U+07F5 NKO LOW TONE APOSTROPHE, are used to show the elision of a vowel while preserving the tonal information of the syllable. Their glyph representations can vary in height relative to the baseline. N'Ko also uses a set of paired punctuation, U+2E1C LEFT LOW PARAPHRASE BRACKET and U+2E1D RIGHT LOW PARAPHRASE BRACKET, to indicate indirect quotations.

Ordering. The order of N'Ko characters in the code charts reflects the traditional ordering of N'Ko. However, in collation, the three archaic letters U+07E8 NKO LETTER JONA JA, U+07E9 NKO LETTER JONA CHA, and U+07EA NKO LETTER JONA RA should be weighted as variants of U+07D6 NKO LETTER JA, U+07D7 NKO LETTER CHA, and U+07D9 NKO LETTER RA, respectively.

Rendering. N'Ko letters have shaping behavior similar to that of Arabic. Each letter can take one of four possible forms, as shown in *Table 19-5*.

Table 19-5. N'Ko Letter Shaping

Character	X_n	X_r	X_m	X_l
A	ߊ	ߊ	ߊ	ߊ
EE	ߋ	ߋ	ߋ	ߋ
I	ߌ	ߌ	ߌ	ߌ
E	ߍ	ߍ	ߍ	ߍ
U	ߎ	ߎ	ߎ	ߎ
OO	ߏ	ߏ	ߏ	ߏ
O	ߐ	ߐ	ߐ	ߐ
DAGBASINNA	◌ߑ	◌ߑ	◌ߑ	◌ߑ
N	ߒ	ߒ	◌ߒ	◌ߒ
BA	ߓ	ߓ	ߓ	ߓ
PA	ߔ	ߔ	ߔ	ߔ
TA	ߕ	ߕ	ߕ	ߕ

Table 19-5. N'Ko Letter Shaping (Continued)

Character	X_n	X_r	X_m	X_l
JA	ߖ	ߖ	ߖ	ߖ
CHA	ߗ	ߗ	ߗ	ߗ
DA	ߘ	ߘ	ߘ	ߘ
RA	ߙ	ߙ	ߙ	ߙ
RRA	ߚ	ߚ	ߚ	ߚ
SA	ߛ	ߛ	ߛ	ߛ
GBA	ߜ	ߜ	ߜ	ߜ
FA	ߝ	ߝ	ߝ	ߝ
KA	ߞ	ߞ	ߞ	ߞ
LA	ߟ	ߟ	ߟ	ߟ
NA WOLOSO	ߠ	ߠ	ߠ	ߠ
MA	ߡ	ߡ	ߡ	ߡ
NYA	ߢ	ߢ	ߢ	ߢ
NA	ߣ	ߣ	ߣ	ߣ
HA	ߤ	ߤ	ߤ	ߤ
WA	ߥ	ߥ	ߥ	ߥ
YA	ߦ	ߦ	ߦ	ߦ
NYA WOLOSO	ߧ	ߧ	ߧ	ߧ
JONA JA	ߨ	ߨ	ߨ	ߨ
JONA CHA	ߩ	ߩ	ߩ	ߩ
JONA RA	ߪ	ߪ	ߪ	ߪ

A noncursive style of N'Ko writing exists where no joining line is used between the letters in a word. This is a font convention, not a dynamic style like bold or italic, both of which are also valid dynamic styles for N'Ko. Noncursive fonts are mostly used as display fonts for the titles of books and articles. U+07FA NKO LAJANYALAN is sometimes used like U+0640 ARABIC TATWEEL to justify lines, although Latin-style justification where space is increased tends to be more common.

19.5 Vai

Vai: U+A500–U+A63F

The Vai script is used for the Vai language, spoken in coastal areas of western Liberia and eastern Sierra Leone. It was developed in the early 1830s primarily by Mọmọlu Duwalu Bukẹlẹ of Jondu, Liberia, who later stated that the inspiration had come to him in a dream. He may have also been aware of, and influenced by, other scripts including Latin, Arabic, and possibly Cherokee, or he may have phoneticized and regularized an earlier pictographic script. In the years afterward, the Vai built an educational infrastructure that enabled the script to flourish; by the late 1800s European traders reported that most Vai were literate in the script. Although there were standardization efforts in 1899 and again at a 1962 conference at the University of Liberia, nowadays the script is learned informally and there is no means to ensure adherence to a standardized version; most Vai literates know only a subset of the standardized characters. The script is primarily used for correspondence and record-keeping, mainly among merchants and traders. Literacy in Vai coexists with literacy in English and Arabic.

Sources. The primary sources for the Vai characters in Unicode are the 1962 Vai Standard Syllabary, modern primers and texts which use the Standard Syllabary (including a few glyph modifications reflecting modern preferences), the 1911 additions of Momolu Massaquoi, and the characters found in *The Book of Ndole*, the longest surviving text from the early period of Vai script usage.

Basic Structure. Vai is a syllabic script written left to right. The Vai language has seven oral vowels [e i a o u ɔ ɛ], five of which also occur in nasal form [ĩ ã ũ ɔ̃ ɛ̃]. The standard syllabary includes standalone vowel characters for the oral vowels and three of the nasal ones, characters for most of the consonant-vowel combinations formed from each of thirty consonants or consonant clusters, and a character for the final velar nasal consonant [ŋ].

The writing system has a *moraic* structure: the weight (or duration) of a syllable determines the number of characters used to write it (as with Japanese kana). A short syllable is written with any single character in the range U+A500..U+A60B. Long syllables are written with two characters, and involve a long vowel, a diphthong, or a syllable ending with U+A60B VAI SYLLABLE NG. Note that the only closed syllables in Vai—that is, those that end with a consonant—are those ending with VAI SYLLABLE NG. The long vowel is generally written using either an additional standalone vowel to double the vowel sound of the preceding character, or using U+A60C VAI SYLLABLE LENGTHENER, while the diphthong is generally written using an additional standalone vowel. In some cases, the second character for a long vowel or diphthong may be written using characters such as U+A54C VAI SYLLABLE HA or U+A54E VAI SYLLABLE WA instead of standalone vowels.

Historic Syllables. In *The Book of Ndole* more than one character may be used to represent a pronounced syllable; they have been separately encoded.

Logograms. The oldest Vai texts used an additional set of symbols called "logograms," representing complete syllables with an associated meaning or range of meanings; these sym-

bols may be remnants from a precursor pictographic script. At least two of these symbols are still used: U+A618 VAI SYMBOL FAA represents the word meaning "die, kill" and is used alongside a person's date of death (the glyph is said to represent a wilting tree); U+A613 VAI SYMBOL FEENG represents the word meaning "thing."

Digits. In the 1920s ten decimal digits were devised for Vai; these digits were "Vai-style" glyph variants of European digits. They never became popular with Vai people, but are encoded in the standard for historical purposes. Modern literature uses European digits.

Punctuation. Vai makes use of European punctuation, although a small number of script-specific punctuation marks commonly occur. U+A60D VAI COMMA rests on or slightly below the baseline; U+A60E VAI FULL STOP rests on the baseline and can be doubled for use as an exclamation mark. U+A60F VAI QUESTION MARK also rests on the baseline; it is rarely used. Some modern primers prefer these Vai punctuation marks; some prefer the European equivalents. Some Vai writers mark the end of a sentence by using U+A502 VAI SYLLABLE HEE instead of punctuation.

Segmentation. Vai is written without spaces between words. Line breaking opportunities can occur between most characters except that line breaks should not occur before U+A60B VAI SYLLABLE NG used as a syllable final, or before U+A60C VAI SYLLABLE LENGTHENER (which is always a syllable final). Line breaks also should not occur before one of the "h-" characters (U+A502, U+A526, U+A54C, U+A573, U+A597, U+A5BD, U+A5E4) when it is used to extend the vowel of the preceding character (that is, when it is a syllable final), and line breaks should not occur before the punctuation characters U+A60D VAI COMMA, U+A60E VAI FULL STOP, and U+A60F VAI QUESTION MARK.

Ordering. There is no evidence of traditional conventions on ordering apart from the order of listings found in syllabary charts. The syllables in the Vai block are arranged in the order recommended by a panel of Vai script experts. Logograms should be sorted by their phonetic values.

19.6 Bamum

Bamum: U+A6A0–U+A6FF

The Bamum script is used for the Bamum language, spoken primarily in western Cameroon. It was developed between 1896 and 1910, mostly by King Ibrahim Njoya of the Bamum Kingdom. Apparently inspired by a dream and by awareness of other writing, his original idea for the script was to collect and provide approximately 500 logographic symbols (denoting objects and actions) to serve more as a memory aid than as a representation of language.

Using the rebus principle, the script was rapidly simplified through six stages, known as Stage A, Stage B, and so on, into a syllabary known as *A-ka-u-ku*, consisting of 80 syllable characters or letters. These letters are used with two combining diacritics and six punctuation marks. The repertoire in this block covers the *A-ka-u-ku* syllabary, or Phase G form, which remains in modern use.

Structure. Modern Bamum is written left-to-right. One interesting feature is that sometimes more letters than necessary are used to write a given syllable. For example, the word *lam* "wedding" is written using the sequence of syllabic characters, *la* + *a* + *m*. This feature is known as pleonastic syllable representation.

Diacritical Marks. U+A6F0 BAMUM COMBINING MARK KOQNDON may be applied to any of the 80 letters. It usually functions to glottalize the final vowel of a syllable. U+A6F1 BAMUM COMBINING MARK TUKWENTIS is only known to be used with 13 letters—usually to truncate a full syllable to its final consonant.

Punctuation. U+A6F2 BAMUM NJAEMLI was a character used in the original set of logographic symbols to introduce proper names or to change the meaning of a word. The shape of the glyph for *njaemli* has changed, but the character is still in use. The other punctuation marks correspond in function to the similarly-named punctuation marks used in European typography.

Digits. The last ten letters in the syllabary are also used to represent digits. Historically, the last of these was used for 10, but its meaning was changed to represent zero when decimal-based mathematics was introduced.

Bamum Supplement: U+16800–U+16A3F

The Bamum Supplement block contains archaic characters no longer used in the modern Bamum orthography. These historical characters are analogous in some ways to the medievalist characters encoded for the Latin script. Most Bamum writers do not use them, but they are used by specialist linguists and historians.

The main source for the repertoire of Bamum extensions is an analysis in Dugast and Jeffreys 1950. The Bamum script was developed in six phases, labeled with letters. Phase A is the earliest form of the script. Phase G is the modern script encoded in the main Bamum

block. The Bamum Supplement block covers distinct characters from the earlier phases which are no longer part of the modern Bamum script.

The character names in this block include a reference to the last phase in which they appear. So, for example, U+16867 BAMUM LETTER PHASE-B PIT was last used during Phase B, while U+168EE BAMUM LETTER PHASE-C PIN continued in use and is attested through Phase C.

Traditional Bamum texts using these historical characters do not use punctuation or digits. Numerical values for digits are written out as words instead.

19.7 Bassa Vah

Bassa Vah: U+16AD0–U+16AFF

The Bassa Vah script is used for the tonal Bassa language of Liberia, which is distinct from the Basa language of Nigeria and the Basaa (sometimes also spelled Bassa) language of Cameroon. Its modern usage and perhaps form are due primarily to Dr. Thomas Flo Lewis in the early 1900s. According to Bassa tradition, an earlier ideographic script was simplified around 1800 by a man named Dirah, and then remained in use primarily among Bassa brought to the Americas as slaves (as was Dirah). While studying abroad in the United States, Lewis learned a version of that script from Dirah's son Jenni and possibly others of Bassa origin in the Americas, and may have made further improvements. The script may also have been influenced by Vai. Lewis actively published about the script; he also arranged for the production of a typesetting machine and primers for Vah, and on his return to Liberia promoted education in the script.

Structure. Modern Bassa Vah is a simple alphabetic script, written from left to right, consisting of 23 consonants, 7 vowels, and 5 tone marks. Except for discussions about the alphabet itself, the vowel letters are always written with tone marks; these marks are placed in a central open area of each vowel glyph. The tone marks are encoded as combining characters.

Punctuation and Digits. Bassa Vah uses a script-specific full stop resembling a plus sign, as well as the European comma, full stop, and quotation marks. It also uses the European digits 0–9.

19.8 Mende Kikakui

Mende Kikakui: U+1E800–U+1E8DF

The Mende Kikakui script is used for the Mende language of Sierra Leone. It is named Kikakui after the sound of its first three characters. Kikakui was popular for correspondence and record-keeping. However, during the 1940s it was largely supplanted by a Latin-based orthography promoted by the British-established Protectorate Literacy Bureau.

An early version of 42 characters was created around 1917 by the Islamic scholar Mohamed Turay, likely influenced both by the Vai syllabary and by Arabic. It was further developed over the next few years by his student and son-in-law Kisimi Kamara who added over 150 more syllabic characters, actively promoted the script, and is generally credited as its primary inventor. The repertoire is based on Tuchscherer 1996. Annotations in the names list provide occasional references to the syllabaries of Amara Mansaray, a prominent script practitioner, and David Dalby (Dalby 1967). The annotations note where Mansaray or Dalby vary from Tuchscherer.

Structure. The Mende Kikakui script has 185 consonant and vowel (CV) syllabic characters and 12 vowels. No script-specific punctuation is known.

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

Numbers. Although both European digits and Arabic digits have been used with Mende Kikakui, it also has its own unique non-decimal number system. This system uses the following characters:

- A set of digits one through nine
- A set of multiplier subscripts for powers of ten from 10 through 1,000,000, encoded as combining marks
- A special subscript for teens, also encoded as a combining mark

Number units in the range 11 through 19 are represented as a digit plus the teens mark. Numbers such as 20, 300, or 5,000 are represented as a digit plus the appropriate multiplier mark. Complete numbers are written as a right-to-left sequence of number units, largest unit first (displayed on the right), whose values are added to produce the numeric value, as shown in the examples in *Table 19-6*.

Table 19-6. Number Formation in Mende Kikakui

Value	Character Sequence	Display
10	1E8C7 MENDE KIKAKUI DIGIT ONE 1E8D1 MENDE KIKAKUI COMBINING NUMBER TENS	𞣇𞣑
14	1E8CA MENDE KIKAKUI DIGIT FOUR 1E8D0 MENDE KIKAKUI COMBINING NUMBER TEENS	𞣊𞣐
27	1E8C8 MENDE KIKAKUI DIGIT TWO 1E8D1 MENDE KIKAKUI COMBINING NUMBER TENS 1E8CD MENDE KIKAKUI DIGIT SEVEN	𞣈𞣑𞣍
206	1E8C8 MENDE KIKAKUI DIGIT TWO 1E8D2 MENDE KIKAKUI COMBINING NUMBER HUNDREDS 1E8CC MENDE KIKAKUI DIGIT SIX	𞣈𞣒𞣌
417	1E8CA MENDE KIKAKUI DIGIT FOUR 1E8D2 MENDE KIKAKUI COMBINING NUMBER HUNDREDS 1E8CD MENDE KIKAKUI DIGIT SEVEN 1E8D0 MENDE KIKAKUI COMBINING NUMBER TEENS	𞣊𞣒𞣍𞣐
784	1E8CD MENDE KIKAKUI DIGIT SEVEN 1E8D2 MENDE KIKAKUI COMBINING NUMBER HUNDREDS 1E8CE MENDE KIKAKUI DIGIT EIGHT 1E8D1 MENDE KIKAKUI COMBINING NUMBER TENS 1E8CA MENDE KIKAKUI DIGIT FOUR	𞣍𞣒𞣎𞣑𞣊

19.9 Adlam

Adlam: U+1E900–U+1E95F

Adlam is a script used to write Fulani and other African languages. The Fulani are a large, historically nomadic tribe of Africa numbering more than 45 million and spread across Senegambia (Senegal) to the banks of the Nile and the Red Sea. Depending on the language, they are called by different names, including Fulani, Fula, Peul, Pul, Fut, Fellata, Tekruri, Toucouleur, Peulh, Wasolonka, and Kourte.

The Fulani are today a widespread ethnic group in Africa, and the Fulani language is spoken by more than 40 million.

During the late 1980s, brothers Ibrahima and Abdoulaye Barry devised this alphabetic script to represent the Fulani language. After several years of development it was widely adopted among Fulani communities and is currently taught at schools in Guinea, Nigeria, Liberia and other nearby countries. The name *Adlam* is derived from the first four letters of the alphabet (A, D, L, M), standing for *Alkule Dandaydhe Leñol Mulugol* ("the alphabet that protects the peoples from vanishing").

Structure. Adlam is a casing script with right-to-left directionality. Its letters can be written separately or can be cursively joined in the same way that Arabic and N'Ko are. Joining is optional, not obligatory.

Diacritical Marks. A range of diacritical marks is used. The lengthener U+1E944 ADLAM ALIF LENGTHENER is used only on the letters U+1E900 ADLAM CAPITAL LETTER ALIF and U+1E922 ADLAM SMALL LETTER ALIF. The lengthener U+1E945 ADLAM VOWEL LENGTHENER is used with other vowels. The U+1E946 ADLAM GEMINATION MARK marks long consonants. These diacritical marks are typically high with capital letters, and high with small letters with ascenders, but low with other small letters.

The diacritical mark U+1E947 ADLAM HAMZA is used atop a consonant when a glottal stop occurs between it and the following vowel. The hamza has high and low variants. The mark U+1E948 ADLAM CONSONANT MODIFIER is used to indicate foreign sounds, primarily in Arabic transcription.

The U+1E94A ADLAM NUKTA is used to indicate both native and borrowed sounds. When vowels are lengthened, however, the nukta is drawn below the vowels to indicate the change. When drawn above a letter, the nukta is called *hoortobbhere* ("dot above") in Fulani; when drawn below, it is called *lestobbhere* ("dot below").

This varied rendering of the Adlam nukta is similar to the behavior of some accents in Latin typography, for which the rendering often depends on the availability of fonts, cultural preferences, or the geographical area. A Latin example is the preference in Latvian and in Romanian for a comma below diacritic shape for some letters, while a cedilla shape is preferred for the same letters in Turkish and in Marshallese.

Line Breaking. Letters have the same line breaking behavior as N'Ko.

Numbers. Adlam uses ten digits with a right-to-left directionality like the digits in N'Ko.

Punctuation. Adlam uses European punctuation and the U+061F ARABIC QUESTION MARK.

Cursive Joining. Cursive joining is used in some contexts. In a cursive context, all letters are dual-joining with a base form, a left-joining form, a dual-joining form, and a right-joining form. Diacritics do not break cursive connections.

Digits and punctuation do not participate in shaping. In a cursive context, U+0640 ARABIC TATWEEL can be used for elongation.

19.10 Medefaidrin

Medefaidrin: U+16E40–U+16E9F

The Medefaidrin script is used to write the liturgical language Medefaidrin by members of an indigenous Christian church, Oberi Okaime ("Church freely given"), which was active in the Nigerian province of Calabar in the 1930s near the Western bank of Cross River. The main spoken language for this group is Ibibio-Efik, which belongs to the Atlantic family of the Niger-Congo languages.

The Medefaidrin script shows the strong influence of English orthography with the use of capital and small letters, and a special sign for the pronoun "I", which has both an upper and lowercase form (U+16E44 MEDEFAIDRIN CAPITAL LETTER ATIU and U+16E64 MEDEFAIDRIN SMALL LETTER ATIU). The community tradition is that this spirit language was revealed to Bishop Ukpong, one of the founders of the community, in 1927 by divine inspiration. The secretary of the group, Jakeld Udofia, transcribed the language to writing. Presently, the religious community counts about 4,000 members. The Medefaidrin language is used for teaching Sunday school lessons and for saying prayers or meditation on the scriptures.

Structure. Medefaidrin is written left to right. There is a close relationship between the phonological analysis and the writing system: the letters are pronounced mostly as written.

Ordering. The order of Medefaidrin characters in the code charts reflects the traditional ordering of Medefaidrin found in instruction materials.

Punctuation, Digits, and Other Marks. Medefaidrin uses a vigesimal (base-20) number system that requires twenty digits. Script-specific punctuation marks are U+16E97 MEDEFAIDRIN COMMA, U+16E98 MEDEFAIDRIN FULL STOP, and U+16E9A MEDEFAIDRIN EXCLAMATION OH. Another unique mark is a symbol for the conjunction "or," represented by the Medefaidrin *aiva*, U+16E99 MEDEFAIDRIN SYMBOL AIVA.