

Universal Multiple-Octet Coded Character Set
International Organization for Standardization
Organisation internationale de normalisation
Международная организация по стандартизации

Doc Type: Working Group Document

Title: Proposal to encode four N’Ko characters in the BMP of the UCS
Source: UC Berkeley Script Encoding Initiative (Universal Scripts Project)
Author: Michael Everson
Status: Liaison Contribution
Action: For consideration by JTC1/SC2/WG2 and UTC
Date: 2015-12-19

1. Character additions. This document requests the addition of four new N’Ko characters to the UCS. One of these letters is a punctuation character, one is a combining mark, and two are currency signs. If this proposal is accepted, the following characters will be encoded:

07FB	-	NKO TE-KERENDE
07FD	◌̣	NKO DANTAYALAN
07FE	₪	NKO DOROME SIGN
07FF	₮	NKO TAMAN SIGN

2. TE-KERENDE. The TE-KERENDE $_$ is a character used to link compounds together. It rests on the baseline like U+07F8 $_$ NKO COMMA and like that it functions as a shaping boundary. N’Ko orthography makes use of the usual non-baseline hyphen for word-breaking, and so TE-KERENDE cannot be unified with HYPHEN-MINUS. A line can break after but not before a TE-KERENDE. The usage of the TE-KERENDE is quite common. (See Figure 5.)

ḡḏ_ḡ_ḡḏ <i>mɔ-ɔ-mɔ</i> ‘everyone’	ḡḡ_ḡ_ḡḡ <i>bɔ-ɔ-bɔ</i> ‘every outing’
ḡḡ_ḡ_ḡḡ <i>ko-o-ko</i> ‘each thing’	ḡḡ_ḡ_ḡḡ <i>si-i-si</i> ‘anyone’
ḡḡ_ḡ_ḡḡ <i>da-a-da</i> ‘each price’	ḡḡ_ḡ_ḡḡ <i>su-u-su</i> ‘every night’
ḡḡ_ḡ_ḡḡ <i>ce-e-ce</i> ‘each work’	ḡḡ_ḡ_ḡḡ <i>tɛ-ɛ-tɛ</i> ‘every space’

3. DANTAYALAN. The DANTAYALAN is used with individual letters to abbreviate units of measure. This character is a combining mark which fuses with the base-line of the letter in question. It does not interfere with shaping behaviour; one appropriate font mechanism for handling this might be to use four pre-composed shaping glyphs for each letter with the DANTAYALAN fused to it. Andriy Rovenchak proposed this character in L2/14-126 “Proposal to encode an additional Nko character”. The use of this character is productive, and can appear with many (and in principle all) N’Ko letters. For instance, it is used with \mathfrak{H} *ka* as \mathfrak{H} to abbreviate \mathfrak{H} *kúde* ‘kilometre’, with \mathfrak{P} *fa* as \mathfrak{P} to abbreviate \mathfrak{P} *fele* ‘megametre’, with \mathfrak{V} *gba* as \mathfrak{V} to abbreviate \mathfrak{V} *gbàlāgbala* ‘metre’, with \mathfrak{Q} *sa* as \mathfrak{Q} to abbreviate \mathfrak{Q} *sid* ‘gram’, and with \mathfrak{B} *ta* as \mathfrak{B} to abbreviate \mathfrak{B} *tóngba* ‘litre’. Examples of letters with DANTAYALAN connected to another letter are \mathfrak{B} *gbaw*. ‘mm.’ and \mathfrak{B} *gbach*. ‘cm.’. (See Figures 1, 2, 3.)

4. Currency signs. Two currency signs based on N’Ko letters exist, for the *dorome* and for the *taman*. These precede the figures which run from right to left, so \mathfrak{Q} \mathfrak{Q} \mathfrak{H} \mathfrak{H} is *d15,00* (cf. \$15.00), \mathfrak{P} \mathfrak{P} is *d99*. Another form is \mathfrak{H} \mathfrak{H} \mathfrak{B} \mathfrak{B} \mathfrak{H} \mathfrak{H} *d5.95* (~ \$5.95 for \$5.95), or \mathfrak{H} \mathfrak{H} \mathfrak{Q} \mathfrak{Q} \mathfrak{H} \mathfrak{H} *d5.00.95* (\$5.00 95). (See Fig. 4.)

5. Unicode Character Properties. Character properties are proposed here.

```

07FB;NKO TE-KERENDE;Po;0;ON;;;N;;;
07FD;NKO DANTAYALAN;Mn;1;R;;;N;;;
07FE;NKO DOROME SIGN;Sc;0;R;;;N;;;
07FF;NKO TAMAN SIGN;Sc;0;R;;;N;;;

```

The Dantalayan is listed as an overlay, because it “fuses” with the baseline, but UTC should check that this is right.

Figures

	X _n	X _r	X _m	X _l
A				
EE	○	○	○	○
I	Y	Y	Y	Y
E	^	^	^	^
U	U	U	U	U
OO	∩	∩	∩	∩
O	⊙	⊙	⊙	⊙
DAGBASINNA	‘	‘	‘	‘
N	𐀀	𐀀	𐀀	𐀀
BA	F	F	F	F
PA	𐀁	𐀁	𐀁	𐀁
TA	b	b	b	b
JA	𐀂	𐀂	𐀂	𐀂
CHA	1	1	1	1
DA	𐀃	𐀃	𐀃	𐀃
RA	†	†	†	†
RRA	𐀄	𐀄	𐀄	𐀄
SA	□	□	□	□
GBA	∇	∇	∇	∇
FA	𐀅	𐀅	𐀅	𐀅
KA	𐀆	𐀆	𐀆	𐀆
LA	𐀇	𐀇	𐀇	𐀇
NA WOLOSO	T	T	T	T
MA	Δ	Δ	Δ	Δ
NYA	𐀈	𐀈	𐀈	𐀈
NA	𐀉	𐀉	𐀉	𐀉
HA	𐀊	𐀊	𐀊	𐀊
WA	𐀋	𐀋	𐀋	𐀋
YA	𐀌	𐀌	𐀌	𐀌
NYA WOLOSO	𐀍	𐀍	𐀍	𐀍
OLD JA	𐀎	𐀎	𐀎	𐀎
OLD CHA	𐀏	𐀏	𐀏	𐀏
OLD RA	+	+	+	+

Figure 1. Chart showing the N’Ko repertoire with its ordinary shaping forms.

	07C	07D	07E	07F
0	0 07C0	♀ 07D0	⊥ 07E0	⋈ 07F0
1	⌈ 07C1	˘ 07D1	Δ 07E1	⋈ 07F1
2	⌈ 07C2	♀ 07D2	♁ 07E2	⋈ 07F2
3	⌈ 07C3	F 07D3	⌈ 07E3	⋈ 07F3
4	⌈ 07C4	⌈ 07D4	⌈ 07E4	˘ 07F4
5	⌈ 07C5	b 07D5	⌈ 07E5	˘ 07F5
6	⌈ 07C6	⌈ 07D6	♁ 07E6	♀ 07F6
7	⌈ 07C7	⌈ 07D7	♁ 07E7	⋈ 07F7
8	⌈ 07C8	⌈ 07D8	⌈ 07E8	⋈ 07F8
9	♀ 07C9	⌈ 07D9	⌈ 07E9	⋈ 07F9
A	⌈ 07CA	⌈ 07DA	⌈ 07EA	⌈ 07FA
B	⊙ 07CB	⊠ 07DB	⋈ 07EB	⌈ 07FB
C	Υ 07CC	▽ 07DC	⋈ 07EC	⌈
D	^ 07CD	♁ 07DD	⋈ 07ED	⋈ 07FD
E	⌈ 07CE	⌈ 07DE	⋈ 07EE	♁ 07FE
F	⌈ 07CF	♀ 07DF	⋈ 07EF	⌈ 07FF

Digits

07C0	0	NKO DIGIT ZERO
07C1	1	NKO DIGIT ONE
07C2	2	NKO DIGIT TWO
07C3	3	NKO DIGIT THREE
07C4	4	NKO DIGIT FOUR
07C5	5	NKO DIGIT FIVE
07C6	6	NKO DIGIT SIX
07C7	7	NKO DIGIT SEVEN
07C8	8	NKO DIGIT EIGHT
07C9	9	NKO DIGIT NINE

Letters

07CA	l	NKO LETTER A
07CB	o	NKO LETTER EE
07CC	Y	NKO LETTER I
07CD	^	NKO LETTER E
07CE	u	NKO LETTER U
07CF	o	NKO LETTER OO
07D0	o	NKO LETTER O
07D1	˘	NKO LETTER DAGBASINNA
07D2	ŋ	NKO LETTER N
07D3	f	NKO LETTER BA
07D4	ɸ	NKO LETTER PA
07D5	b	NKO LETTER TA
07D6	ɣ	NKO LETTER JA
07D7	ɬ	NKO LETTER CHA
07D8	ɓ	NKO LETTER DA
07D9	ɾ	NKO LETTER RA
07DA	ʒ	NKO LETTER RRA
07DB	ɔ	NKO LETTER SA
07DC	ɗ	NKO LETTER GBA
07DD	ɸ	NKO LETTER FA
07DE	ɰ	NKO LETTER KA
07DF	ɱ	NKO LETTER LA
07E0	T	NKO LETTER NA WOLOSO
07E1	Δ	NKO LETTER MA
07E2	ɸ	NKO LETTER NYA
07E3	ɱ	NKO LETTER NA
07E4	ɱ	NKO LETTER HA
07E5	ɱ	NKO LETTER WA
07E6	ɸ	NKO LETTER YA
07E7	ɸ	NKO LETTER NYA WOLOSO

Archaic letters

07E8	ɰ	NKO LETTER JONA JA
07E9	ɰ	NKO LETTER JONA CHA
07EA	ɰ	NKO LETTER JONA RA
		→ 07D9 ɸ nko letter ra

Tone marks

07EB	˘	NKO COMBINING SHORT HIGH TONE
		→ 0304 combining macron
07EC	˘	NKO COMBINING SHORT LOW TONE
		→ 0303 combining tilde
07ED	˘	NKO COMBINING SHORT RISING TONE
		→ 0307 combining dot above
07EE	˘	NKO COMBINING LONG DESCENDING TONE
		→ 0302 combining circumflex accent
07EF	˘	NKO COMBINING LONG HIGH TONE
07F0	˘	NKO COMBINING LONG LOW TONE
07F1	˘	NKO COMBINING LONG RISING TONE
07F2	˘	NKO COMBINING NASALIZATION MARK
		→ 0323 combining dot below

07F3	˘	NKO COMBINING DOUBLE DOT ABOVE
		→ 0308 combining diaeresis
07F4	˘	NKO HIGH TONE APOSTROPHE
		→ 02BC modifier letter apostrophe
07F5	˘	NKO LOW TONE APOSTROPHE
		→ 02BB modifier letter turned comma

Symbol

07F6	2	NKO SYMBOL OO DEENE
------	---	---------------------

Punctuation

07F7	⊗	NKO SYMBOL GBAKURUNEN
07F8	˘	NKO COMMA
07F9	˘	NKO EXCLAMATION MARK

Letter extender

07FA	-	NKO LAJANYALAN
		→ 005F _ low line
		→ 0640 - arabic tatweel

Punctuation

07FB	-	NKO TE-KERENDE
		→ 002D - hyphen-minus

Abbreviation sign

07FD	⊗	NKO DANTAYALAN
		• used to abbreviate units of measure

Currency signs

07FE	ɸ	NKO DOROME SIGN
		→ 07D8 ɓ nko letter da
07FF	ɸ	NKO TAMAN SIGN
		→ 07D5 b nko letter ta

A. Administrative

1. Title

Proposal to encode four N’Ko characters in the BMP of the UCS

2. Requester’s name

Script Encoding Initiative (author: Michael Everson)

3. Requester type (Member body/Liaison/Individual contribution)

Liaison contribution.

4. Submission date

2015-12-29

5. Requester’s reference (if applicable)

6. Choose one of the following:

6a. This is a complete proposal

Yes.

6b. More information will be provided later

No.

B. Technical – General

1. Choose one of the following:

1a. This proposal is for a new script (set of characters)

No.

1b. Proposed name of script

1c. The proposal is for addition of character(s) to an existing block

Yes.

1d. Name of the existing block

Nko.

2. Number of characters in proposal

4.

3. Proposed category (A-Contemporary; B.1-Specialized (small collection); B.2-Specialized (large collection); C-Major extinct; D-Attested extinct; E-Minor extinct; F-Archaic Hieroglyphic or Ideographic; G-Obscure or questionable usage symbols)

Category A.

4a. Is a repertoire including character names provided?

Yes.

4b. If YES, are the names in accordance with the “character naming guidelines” in Annex L of P&P document?

Yes.

4c. Are the character shapes attached in a legible form suitable for review?

Yes.

5a. Who will provide the appropriate computerized font (ordered preference: True Type, or PostScript format) for publishing the standard?

Michael Everson.

5b. If available now, identify source(s) for the font (include address, e-mail, ftp-site, etc.) and indicate the tools used:

Michael Everson. Fontlab.

6a. Are references (to other character sets, dictionaries, descriptive texts etc.) provided?

No.

6b. Are published examples of use (such as samples from newspapers, magazines, or other sources) of proposed characters attached?

Yes.

7. Does the proposal address other aspects of character data processing (if applicable) such as input, presentation, sorting, searching, indexing, transliteration etc. (if yes please enclose information)?

Yes.

8. Submitters are invited to provide any additional information about Properties of the proposed Character(s) or Script that will assist in correct understanding of and correct linguistic processing of the proposed character(s) or script. Examples of such properties are: Casing information, Numeric information, Currency information, Display behaviour information such as line breaks, widths etc., Combining behaviour, Spacing behaviour, Directional behaviour, Default Collation behaviour, relevance in Mark Up contexts, Compatibility equivalence and other Unicode normalization related information. See the Unicode standard at <http://www.unicode.org> for such information on other scripts. Also see Unicode Character Database <http://www.unicode.org/Public/UNIDATA/UnicodeCharacterDatabase.html> and associated Unicode Technical Reports for information needed for consideration by the Unicode Technical Committee for inclusion in the Unicode Standard.

See above.

C. Technical – Justification

1. Has this proposal for addition of character(s) been submitted before? If YES, explain.

No.

2a. Has contact been made to members of the user community (for example: National Body, user groups of the script or characters, other experts, etc.)?

Yes.

2b. If YES, with whom?

Mamady Doumbouya, Baba Mamadi Diane.

2c. If YES, available relevant documents

3. Information on the user community for the proposed characters (for example: size, demographics, information technology use, or publishing use) is included?

Users of N’Ko.

4a. The context of use for the proposed characters (type of use; common or rare)

Common.

4b. Reference

5a. Are the proposed characters in current use by the user community?

Yes.

5b. If YES, where?

The Manden-speaking community.

6a. After giving due considerations to the principles in the P&P document must the proposed characters be entirely in the BMP?

Yes.

6b. If YES, is a rationale provided?

Yes.

6c. If YES, reference

Keep with other N’Ko characters.

7. Should the proposed characters be kept together in a contiguous range (rather than being scattered)?

No.

8a. Can any of the proposed characters be considered a presentation form of an existing character or character sequence?

No.

8b. If YES, is a rationale for its inclusion provided?

8c. If YES, reference

9a. Can any of the proposed characters be encoded using a composed character sequence of either existing characters or other proposed characters?

No.

9b. If YES, is a rationale for its inclusion provided?

9c. If YES, reference

10a. Can any of the proposed character(s) be considered to be similar (in appearance or function) to an existing character?

No.

10b. If YES, is a rationale for its inclusion provided?

10c. If YES, reference

11a. Does the proposal include use of combining characters and/or use of composite sequences?

Yes.

11b. If YES, is a rationale for such use provided?

Yes.

11c. If YES, reference

One of the characters is a combining character.

11d. Is a list of composite sequences and their corresponding glyph images (graphic symbols) provided?

Yes.

11e. If YES, reference

See Figures 1, 2, 4.

12a. Does the proposal contain characters with any special properties such as control function or similar semantics?

No.

12b. If YES, describe in detail (include attachment if necessary)

13a. Does the proposal contain any Ideographic compatibility character(s)?

No.

13b. If YES, is the equivalent corresponding unified ideographic character(s) identified?