UC Berkeley
Proposals from the Script Encoding Initiative

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

Permalink
https://escholarship.org/uc/item/34q5n9gk

Author
Everson, Michael

Publication Date
2015-12-29

Peer reviewed

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	Ŧ	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.)

ߏ߰ߊ߸ߏ߸ߏ߰ߊ *mɔ-ɔ-mɔ* 'everyone'		ߓ߰ߝ߸ߏ߸ߏߝ *bɔ-ɔ-bɔ* 'every outing'
ߞ߸ߏ߸ߞ *ko-o-ko* 'each thing'		ߛ߲ߏ߸ߒ߸ߛ߲ߏ *si-i-si* 'anyone'
ߘߊ߸ߌ߸ߘߊ *da-a-da* 'each price'		ߛߎ߸ߎ߸ߛߎ *su-u-su* 'every night'
ߗ߯ߋ߸ߋ߸ߗ߯ߋ *ce-e-ce* 'each work'		ߕߍ߸ߋ߸ߕߍ *tɛ-ɛ-tɛ* '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 precomposed shaping glyphs for each letter with the DANTAYALAN fused to it. Andrij 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 instace, it is used with ߞ *ka* as ߞ to abbreviate ߞߎߘߍ *kúdɛ* 'kilometre', with ߝ *fa* as ߝ to abbreviate ߝߋߟߋ *fele* 'megametre', with ߜ *gba* as ߜ to abbreviate ߜߝߟߜߝ *gbàlàgbala* 'metre', with ߛ *sa* as ߛ to abbreviate ߛߌߘߐ *sidɔ* 'gram', and with ߕ *ta* as ߕ to abbreviate ߟߎߕ *tóngba* 'litre'. Examples of letters with DANTAYALAN connected to another letter are ߜߊ *gbaw.* 'mm.' and ߜ *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 ߀߀߸ߠߒ߯₣ is *đ15,00* (cf. $15.00), ߉߉₣ is *đ99*. Another form is ߒߟŦ߸ߟ₣ *đ5.ŧ95* (~ $5.¢95 for $5.95), or ߒߟŦ ߀߀߸ߟ₣ *đ5.00 ŧ95* ($5.00 ¢95). (See Fig. 4.)

1

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.
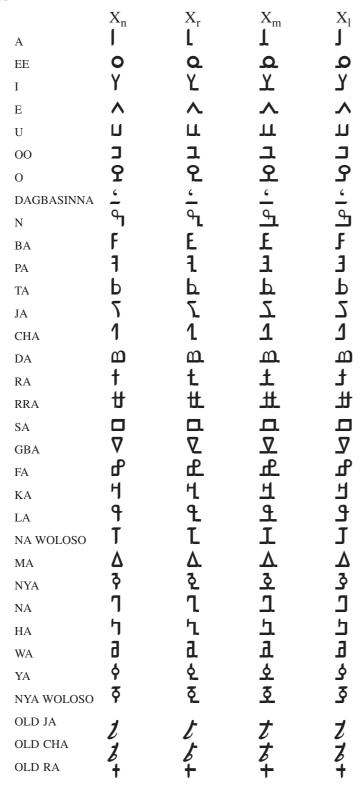
Figures

	X_n	X_r	X_m	X_l
A				
EE				
I				
E				
U				
OO				
O				
DAGBASINNA				
N				
BA				
PA				
TA				
JA				
CHA				
DA				
RA				
RRA				
SA				
GBA				
FA				
KA				
LA				
NA WOLOSO				
MA				
NYA				
NA				
HA				
WA				
YA				
NYA WOLOSO				
OLD JA				
OLD CHA				
OLD RA				

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

	X_n	X_r	X_m	X_l
A				
EE				
I				
E				
U				
OO				
O				
DAGBASINNA				
N				
BA				
PA				
TA				
JA				
CHA				
DA				
RA				
RRA				
SA				
GBA				
FA				
KA				
LA				
NA WOLOSO				
MA				
NYA				
NA				
HA				
WA				
YA				
YA WOLOSO				
OLD JA				
OLD CHA				
OLD RA				

Figure 2. Chart showing the N'Ko repertoire with shaping forms combined with DANTAYALAN.

4.	07FD	᷿	NKo DANTAYALAN						
	It.is.used.with.	ᴾ	to become	ᴾ	To.abbreviate			*fele :* [Megametre](#) (Mm)	
	It.is.used.with.		to become		To.abbreviate			*kida*	kg
	It.is.used.with.	□	to become		To.abbreviate			*sido*	g
	It.is.used.with.		to become		To.abbreviate			*sidɔtanden*	dg
	It.is.used.with.		to become		To.abbreviate			*sidɔcɛmɛden*	cg
	It.is.used.with.		to become		To.abbreviate			*sidɔwaaden*	mg
	It.is.used.with.	∇ᵁ	to become		To.abbreviate			*kúdɛ*	km
	It.is.used.with.	∇	to become		To.abbreviate			*gàlàgala*	m
	It.is.used.with.		to become		To.abbreviate			*gàlàtanden*	dm
	It.is.used.with.		to become		To.abbreviate			*gàlàcɛmɛden*	cm
	It.is.used.with.		to become		To.abbreviate			*gàlàwaaden:* [Millimetre](#) (mm)	
	It.is.used.with.		to become		To.abbreviate			*gàlàkoe*	µm
	It.is.used.with.		to become		To.abbreviate			*gàlàɲánta*	nm
	It.is.used.with.	ᴾ∇ᵁ	to become		To.abbreviate			*kúdɛfúdúben*	Km²
	It.is.used.with.	ᴾ∇	to become		To.abbreviate			*gàlàfúdúben*	m²
	It.is.used.with.	ᴾ	to become		To.abbreviate			*gàlàtanfúdúben*	dm²
	It.is.used.with.	ᴾ	to become		To.abbreviate			*gàlàcɛmɛfúdúben*	cm²
	It.is.used.with.	ᴾ	to become		To.abbreviate			*gàlàwááfúdúben*	mm²
	It.is.used.with.	ᴶ∇	to become		To.abbreviate			*gàlàfɛdɛben*	m³
	It.is.used.with.	ᴶ□	to become		To.abbreviate			*sènkan*	
	It.is.used.with.	ᴶ	to become		To.abbreviate			*gàlàfɛdɛben*	dm³
	It.is.used.with.	ᴶ	to become		To.abbreviate			*gàlàtanfɛdɛben*	cm³
	It.is.used.with.		to become		To.abbreviate			*gàlàcɛmɛfɛdɛben*	mm³
	It.is.used.with.	∆∇	to become		To.abbreviate			*gàlàmɔnɔ*	Ml
	It.is.used.with.	□∇	to become		To.abbreviate			*gàlàsɔeyi*	Kl
	It.is.used.with.		to become		To.abbreviate			*gàlàfele*	hl
	It.is.used.with.	Ь	to become		To.abbreviate			*tónga*	l
	It.is.used.with.	ЬЬ	to become		To.abbreviate			*tóngatanden*	dl
	It.is.used.with.		to become		To.abbreviate			*tóngacɛmɛden*	cl
	It.is.used.with.		to become		To.abbreviate			*tóngawaaden*	ml
	It.is.used.with.	Ⴕ	to become		To.abbreviate			*kíríti*	t

Figure 3. Chart showing the DANTAYALAN in use in various N'Ko abbreviations

1.	07FE	ꛯ	NKo DOROME TOOSERE			
	It.is.used.with.	Nombres	$15.00 = 00.ʰ⌐ꛯ	$45.00 = 00.ʰ⌐ꛯ	$75.00 = 00.ʰⱴꛯ	$99.00 = 00.ၣၣꛯ
			$15 = ʰ⌐ꛯ	$45 = ʰ⌐ꛯ	$75 = ʰⱴꛯ	$99 = ၣၣꛯ
2.	07FF	꛰	NKo TAMAN TOOSERE			
	It.is.used.with.	Nombres	¢ 0.15 = ʰ⌐.0꛰	¢ 0.45 = ʰ⌐.0꛰	¢ 0.75 = ʰⱴ.0꛰	¢ 0.99 = ၣၣ.0꛰
			¢ 15 = ʰ⌐꛰	¢ 45 = ʰ⌐꛰	¢ 75 = ʰⱴ꛰	¢ 99 = ၣၣ꛰

Figure 4. Chart showing the DOROME SIGN and TOMAN SIGN in various usage configurations.

Figure 5. Example showing the TE-KERENDE in use.

	07C	07D	07E	07F
0	0 07C0	੧ 07D0	T 07E0	͡◌ 07F0
1	⅃ 07C1	᷄ 07D1	Δ 07E1	◌̰ 07F1
2	Ⱶ 07C2	॰ 07D2	ȝ 07E2	◌̣ 07F2
3	ⅎ 07C3	Ϝ 07D3	٦ 07E3	◌̈ 07F3
4	⅃ 07C4	Ꭶ 07D4	ᒣ 07E4	᾽ 07F4
5	Ⱶ 07C5	b 07D5	ᗧ 07E5	᾿ 07F5
6	Ⴑ 07C6	Γ 07D6	ϙ 07E6	੨ 07F6
7	Ⅴ 07C7	�१ 07D7	ȝ̃ 07E7	⸕ 07F7
8	ᏼ 07C8	ᗰ 07D8	ℓ 07E8	◌̤ 07F8
9	ᖰ 07C9	† 07D9	ℓ̃ 07E9	◌⋮ 07F9
A	⎥ 07CA	♯ 07DA	† 07EA	◌̱ 07FA
B	⬤ 07CB	☐ 07DB	◌̄ 07EB	◌̱ 07FB
C	Υ 07CC	▽ 07DC	◌̃ 07EC	⧄
D	∧ 07CD	ᑆ 07DD	◌̇ 07ED	◌̚ 07FD
E	⊔ 07CE	ᕭ 07DE	◌̂ 07EE	₤ 07FE
F	⅃ 07CF	ꝯ 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	Ɥ	NKO LETTER A
07CB	o	NKO LETTER EE
07CC	Y	NKO LETTER I
07CD	⋀	NKO LETTER E
07CE	⊔	NKO LETTER U
07CF	コ	NKO LETTER OO
07D0	ọ	NKO LETTER O
07D1	ᵼ	NKO LETTER DAGBASINNA
07D2	ᴎ	NKO LETTER N
07D3	Ϝ	NKO LETTER BA
07D4	ꓱ	NKO LETTER PA
07D5	ხ	NKO LETTER TA
07D6	⁊	NKO LETTER JA
07D7	⅂	NKO LETTER CHA
07D8	ധ	NKO LETTER DA
07D9	†	NKO LETTER RA
07DA	ⱦ	NKO LETTER RRA
07DB	▫	NKO LETTER SA
07DC	▽	NKO LETTER GBA
07DD	ꝑ	NKO LETTER FA
07DE	Ꮋ	NKO LETTER KA
07DF	ꟼ	NKO LETTER LA
07E0	T	NKO LETTER NA WOLOSO
07E1	⋀	NKO LETTER MA
07E2	ꝫ	NKO LETTER NYA
07E3	ꓩ	NKO LETTER NA
07E4	ꓧ	NKO LETTER HA
07E5	ꓷ	NKO LETTER WA
07E6	⟡	NKO LETTER YA
07E7	ꝫ	NKO LETTER NYA WOLOSO

Archaic letters

07E8	⌿	NKO LETTER JONA JA
07E9	⌿	NKO LETTER JONA CHA
07EA	✝	NKO LETTER JONA RA
		→ 07D9 † nko letter ra

Tone marks

07EB	ō	NKO COMBINING SHORT HIGH TONE
		→ 0304 ̄ combining macron
07EC	õ	NKO COMBINING SHORT LOW TONE
		→ 0303 ̃ combining tilde
07ED	ȯ	NKO COMBINING SHORT RISING TONE
		→ 0307 ̇ combining dot above
07EE	ô	NKO COMBINING LONG DESCENDING TONE
		→ 0302 ̂ combining circumflex accent
07EF	ö̌	NKO COMBINING LONG HIGH TONE
07F0	o͢	NKO COMBINING LONG LOW TONE
07F1	ǒ	NKO COMBINING LONG RISING TONE
07F2	ọ	NKO COMBINING NASALIZATION MARK
		→ 0323 ̣ combining dot below

07F3	ö	NKO COMBINING DOUBLE DOT ABOVE
		→ 0308 ̈ combining diaeresis
07F4	’	NKO HIGH TONE APOSTROPHE
		→ 02BC ʼ modifier letter apostrophe
07F5	‘	NKO LOW TONE APOSTROPHE
		→ 02BB ʻ 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 ـ arabic tatweel

Punctuation

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

Abbreviation sign

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

Currency signs

07FE	ꝕ	NKO DOROME SIGN
		→ 07D8 ധ nko letter da
07FF	ƫ	NKO TAMAN SIGN
		→ 07D5 ხ 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?