
JTC1/SC2/WG2 N2949*Date: 2005-07-12*

From: François Yergeau, Patrick Andries**Re: Comments on N2932****FOR A CORRECT ENCODING OF N'KO****POUR UN CODAGE CORRECT DU N'KO**

RÉSUMÉ INTRODUCTIF

Le Canada avait émis un avis défavorable à l'inclusion de certains caractères n'ko lors du premier vote sur l'amendement 2 à l'ISO/CEI 10646. Les raisons de cet avis avaient été précisées dans un document¹ peu avant la réunion du GT2 à Xiamen (Chine). À l'issue de cette réunion, il avait signalé par écrit² plusieurs manières de convaincre le Canada d'accepter ces caractères. Malheureusement, nous devons rapporter que nous n'avons pas reçu de réponse à nos questions, mais de simples affirmations péremptoires. Ceci est regrettable. Nous attirons l'attention de tous les organismes nationaux sur cette question et le risque réel qu'il y a de coder des caractères qui vont à l'encontre du modèle caractère-glyphe. Nous montrerons ici pourquoi nous pensons que ces caractères sont inutiles ou dangereux. Nous apporterons également le témoignage d'un expert linguiste ayant étudié le n'ko qui partage notre point de vue. Nous avons aussi démontré grâce à la première mise en œuvre que nous connaissions du n'ko Unicode (moins les caractères litigieux) que les propositions canadiennes ne mettent en rien en péril la mise en œuvre du n'ko par d'autres éditeurs. Nous pensons qu'une explication compréhensible par tous les organismes nationaux s'impose : pourquoi ces caractères contestés ne constituent-ils pas une infraction au modèle caractère-glyphe? Nous ne pensons pas qu'il suffit de répéter que certains membres de la communauté d'utilisateurs disent que ces caractères doivent être codés, nous voulons savoir pourquoi et pourquoi ceci n'est ni inutile ni dangereux. En l'absence de ces explications, nous recommandons que le n'ko soit bien codé mais qu'on omette de cette version les caractères litigieux, ils pourraient être codés par la suite si le besoin de leur codage venait à être prouvé.

INTRODUCTORY SUMMARY

Canada voiced an unfavourable vote to the inclusion of some N'ko characters during the first ballot on Amendment 2 of ISO/CEI 10646. The reasons were clarified in a document³ sometime before the WG2 meeting in Xiamen (China). At the conclusion of this meeting, several ways to convince Canada to accept these characters were mentioned in writing⁴. Unfortunately, we have to report that our questions have not been answered, but that we have only received peremptory affirmations. This is regrettable. We would like to draw the attention of all national bodies on this issue and the very real risk that characters may be encoded conflicting with the character-glyph model. We will show here why we believe these

¹ < <http://std.dkuug.dk/jtc1/sc2/wg2/docs/n2911.pdf> >

² < <http://std.dkuug.dk/jtc1/sc2/wg2/docs/N2927.pdf> >

³ < <http://std.dkuug.dk/jtc1/sc2/wg2/docs/n2911.pdf> >

⁴ < <http://std.dkuug.dk/jtc1/sc2/wg2/docs/N2927.pdf> >

characters are not useful or are risky. We will also bring forward the testimony of a leading expert linguist who has studied N'ko and who shares our standpoint. We have also demonstrated, through the first UCS implementation of N'ko (minus the disputable characters) we know of, that the Canadian suggestions do in no way endanger the implementation of N'ko by other implementers. We believe that an answer comprehensible by all national bodies is imperative: why do the disputed characters not violate the character-glyph model? We believe that it is not sufficient to repeat that some user community members say that these characters must be encoded, we want to know why and why this is neither useless nor dangerous. In the absence of these explanations, we recommend that N'ko be encoded but that the disputable characters be omitted in this version, they could be encoded at a later date once the need of their encoding will have been supplied.

TABLE DES MATIÈRES

I. LES CARACTÈRES LITIGIEUX	4
II. LES ARGUMENTS POUR ET CONTRE.....	5
II. a) les anciennes formes.....	5
III. b) les diacritiques	8
III. UNE MISE EN ŒUVRE DU N'KO.....	11
IV. LE SOUTIEN DE L'INSTITUT N'KO.....	12
V. CONCLUSION ET PROPOSITION	13
Annexe A – Prototype de mise en œuvre JUC du n'ko description et résultats	14
Annexe B – Appui pour l'analyse canadienne par Valentin Vydrine (expert linguiste en langues mandingues).....	16

TABLE OF CONTENTS

I. THE DISPUTED CHARACTERS	18
II. PROS AND CONS	18
II. a) The Old Forms.....	19
III. b) The Combining Marks	22
III. A UCS N'KO IMPLEMENTATION	25
IV. SUPPORT FROM THE N'KO INSTITUTE.....	26
V. CONCLUSION AND PROPOSAL	27
Annex A – proof of concept of a UCS N'ko implementation description and output.....	28
Annex B – Support for the Canadian analysis by Valentin Vydrine (expert linguist in Manding languages)	30

I. LES CARACTÈRES LITIGIEUX

Le bloc n'ko ci-dessous comprend certains caractères litigieux.

	0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F
07Cx	0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F
07Dx	0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F
07Ex	0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F
07Fx	0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F

Le premier groupe de caractères litigieux est composé de 3 caractères dont des formes anciennes et modernes sont proposées comme caractères distincts dans le bloc n'ko.

Nom	Forme ancienne		Forme moderne	
	Glyphe	N°	Glyphe	N°
Dja	Ƶ	U+07E8	ɥ	U+07D6
Tcha	ƶ	U+07E9	ɗ	U+07D7
Ra	ɒ	U+07EA	ɓ	U+07D9

Le second groupe de caractères contestables est composé d'un sous-ensemble des diacritiques proposés pour le n'ko, dont la forme et le comportement sont similaires aux diacritiques génériques (communs) du bloc U+03xx, à savoir les suivants :

U+07EB	̄	=	U+0304	̄	?
U+07ED	◌̇	=	U+0307	◌̇	?
U+07EE	◌̂	=	U+0302	◌̂	?
U+07F2	◌̈	=	U+0323	◌̈	?
U+07F3	◌̈̈	=	U+0308	◌̈̈	?

II. LES ARGUMENTS POUR ET CONTRE

Un certain nombre d'arguments ont été avancés pour soutenir ou au contraire déplorer le codage des caractères litigieux, que ce soit dans la proposition initiale⁵, dans les commentaires de vote ou ailleurs. Cette section reprend ces arguments, en les mettant en regard autant que possible.

II. a) les anciennes formes

Pour un codage séparé	Contre un codage séparé
<p>Caractères ou glyphes : Ce sont des caractères distincts, ils doivent donc être codés séparément.</p>	<p>Ce sont des variantes de glyphes des mêmes caractères, or l'ISO 10646 code des caractères abstraits, pas des glyphes (modèle caractères-glyphes)</p>
<p>Continuité historique : « L'aspect historique [de] ces caractères offert par le délégué irlandais est vital à l'expression de la continuité du développement de l'écriture. » [traduction de N2898]</p>	<p>L'expression de cette continuité serait mieux servie par l'utilisation d'un seul code pour chaque lettre. Il n'est pas du ressort de l'ISO 10646 d'exprimer le développement des formes au sein d'une même écriture.</p>
<p>Besoin pour anciens écrits : Il faut coder les anciennes formes pour pouvoir publier les premiers écrits de Souleymane Kanté.</p> <p>« il sera nécessaire de publier ces manuscrits d'abord avec les anciennes lettres, ne serait-ce que pour l'authenticité. » [traduction de N2868]</p>	<p>« Ce besoin n'est pas contesté, mais doit être comparé avec le même besoin dans toute autre écriture dont les formes de lettre ont changé. La solution est d'utiliser une police idoine et non pas de coder les anciennes formes. L'écriture latine a connu bien des formes maintenant désuètes comme l'onciale et la gothique (<i>fraktur</i>). Ces formes ne sont et ne seront pas codées. Quand il faut reproduire un document ancien dans sa forme originale [...], une police à cet effet est utilisée, contenant les anciennes formes des lettres. » [traduction de N2911]</p>

⁵ <http://std.dkuug.dk/jtc1/sc2/wg2/docs/n2765.pdf>

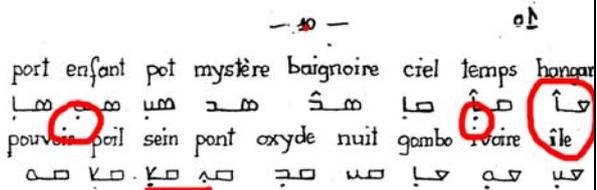
Pour un codage séparé	Contre un codage séparé
<p>Besoin pour discuter des formes: Il faut coder les anciennes formes pour pouvoir en discuter, en contraste avec les nouvelles formes.</p> <p>« Par exemple, le Dr Mamadi Diané, éditeur du saint Coran en n’ko, a dû expliquer dans sa préface l’utilisation des nouveaux caractères. » [traduction de N2898]</p>	<p>Cette situation est la même pour n’importe quelle autre écriture possédant des variantes glyphiques dont on souhaite discuter les différentes formes – par exemple pour parler de la caroline en écriture latine. La solution est la même : utiliser une police idoine.</p> <p>Le professeur Valentin Vydrine, expert dans les langues mandingues et qui a publié plusieurs articles sur le n’ko, est également de cet avis :</p> <p><i>« Oui, je suis de l’opinion que les vieilles lettres ne méritent pas des cases spéciales dans l’UNICODE. Dans ce cas particulier, il s’agit d’un métatexte, tandis que dans un contexte naturel ils ne se retrouvent pratiquement jamais ensemble. »</i> [correspondance privée, cf. Annexe B]</p> <p>Notons enfin que les trois caractères proposés par l’Irlande ne sont pas suffisants pour coder toutes les anciennes formes. Voir le tableau reproduit en annexe A (les résultats).</p>
<p>Texte brut : La distinction entre les anciennes et nouvelles formes est du niveau du texte brut.</p>	<p>Rien ne prouve cette assertion, et des demandes répétées de la soutenir n’ont rien donné. Dans toute autre écriture, la distinction entre des formes historiques et modernes est considérée du ressort du texte enrichi (choix de police).</p> <p>Il est important de noter qu’on ne retrouve jamais les anciennes et nouvelles formes dans un même texte, à l’exception de métatextes discutant précisément de ces formes (et ceci comprend les « preuves » offertes dans N2914⁶). Le Coran et le dictionnaire n’ko n’utilisent que les nouvelles formes.</p>
<p>Wynn : Le cas est similaire à celui du wynn (Pp, U+01BF et U+01F7) qui est distinct du « w ».</p>	<p>Le « w » n’est pas une évolution graphique du wynn. Le wynn a été abandonné en moyen anglais au profit d’un « uu », devenu plus tard « w ».</p>

⁶ <http://std.dkuug.dk/jtc1/sc2/wg2/docs/n2914.pdf>

Pour un codage séparé	Contre un codage séparé
<p>Orthographe : La forme graphique n'est pas seule à avoir changé, l'orthographe a aussi changé.</p>	<p>Ceci n'a rien d'unique, ceci se produit dans toutes autres écritures qui évoluent. [paraphrase de N2911]</p>
<p>Recherche :</p>	<p>« [N]ous voulons aussi attirer l'attention sur un argument qui milite contre ce codage : le fait d'avoir un seul code pour, disons, ANCIEN RA et RA ferait en sorte que plusieurs types de traitement sur le texte, notamment la recherche, les traiteraient de façon identique. Ceci signifie que quelqu'un cherchant un mot dans une base de données ou dans internet n'aurait pas à se souvenir de chercher à la fois l'ancienne et la nouvelle forme (sauf bien sûr en cas de changement orthographique). Il s'agit d'un réel bénéfice du codage unique, qui ne devrait pas être négligé. » [traduction de N2911]</p> <p>Ajoutons ici que la proposition initiale N2765 admet déjà cette équivalence pour le tri : « pour ce qui a trait à l'ordre lexicographique, les ANCIENNES LETTRES DJA, TCHA et RA sont considérées comme des variantes du DJA, TCHA et RA. ». Le codage séparé n'est pas rédhibitoire pour le tri – les mises en œuvre suffisamment bien faites d'algorithmes de tri peuvent être adaptées – mais est beaucoup plus nuisible pour la recherche, surtout dans un contexte global comme celui d'internet.</p>

III. b) les diacritiques

Précisons d'emblée que seuls sont discutés ici les 5 diacritiques litigieux U+07EB, U+07ED, U+07EE, U+07F2 et U+07F3. Le codage des autres diacritiques utilisés en n'ko n'étant pas contesté, il n'en est pas question ici.

Pour un codage séparé	Contre un codage séparé
<p>Directionnalité : « il nous paraît donc préférable de considérer ces caractères comme propres à cette écriture afin d'éviter tout problème de directionnalité... » [N2765]</p>	<p>Cet argument est totalement infondé, les diacritiques adoptent la direction de leur caractère de base dans l'algorithme bidi.</p>
<p>Variation graphique : « Certains diacritiques n'ko possèdent plus d'œils que les diacritiques génériques, ils sont également dessinés un peu plus haut et avec plus de graisse que les diacritiques génériques. » [N2765]</p>	<p>Les diacritiques génériques font déjà l'objet de variations plus grandes que celles observées entre le n'ko et les formes de références du standard. Le tréma/umlaut U+0308, par exemple, peut se trouver à diverses hauteurs, y compris jusqu'à l'intérieur de la lettre, et apparaître comme deux points de graisse variable, comme un petit e suscrit ou même comme un double accent aigu. Les variations du n'ko sont parfaitement conformes à la variabilité déjà observée. L'accent aigu peut passer de pratiquement vertical (en polonais notamment) à pratiquement plat.</p> <p>Les diacritiques n'ko, par exemple dans cet extrait de la main même de Souleymane Kanté, son inventeur, sont parfois indistincts des diacritiques génériques présents à leurs côtés. Cf. ci-dessous. Tous les styles n'ko ne sont pas gras. Il s'agit, bien évidemment, d'un attribut lié à la largeur de trait, à la police choisie.</p> 
<p>Caractères étrangers : Il ne faut pas devoir utiliser en n'ko des caractères étrangers au n'ko.</p> <p>« Afin de ne pas sous-entendre que certains caractères sont secondaires ou ne font pas partie du jeu de caractères n'ko et qu'ils ne</p>	<p>Les diacritiques génériques sont spécifiquement prévus pour être utilisés avec n'importe quelle écriture. Ils ne sont ni propres ni étrangers à aucune écriture. Cf. section 7.7 du standard Unicode</p> <p>(http://www.unicode.org/versions/Unicode4.0)</p>

Pour un codage séparé	Contre un codage séparé
<p>sont empruntés d'autres systèmes » [traduction de N2898]</p>	<p>.0/ch07.pdf): « Les signes diacritiques de ce bloc sont destinés à être utilisés avec n'importe quelle écriture. » [traduction de N2911]</p>
<p>Un seul bloc : Les utilisateurs doivent pouvoir trouver tous les caractères n'ko dans le bloc n'ko.</p> <p>« ...imposer que le système n'ko cherche ses caractères dans d'autres parties d'Unicode ou qu'il surimpose d'autres diacritiques serait une erreur de convenance de proportion similaire à celle qui consista à n'attribuer que deux chiffres plutôt que quatre aux années du calendrier avant l'an 2000. »</p> <p>[traduction de N2898]</p>	<p>Le lot commun de la plupart des écritures est de devoir trouver ses caractères dans plusieurs blocs. Ceci est conforme à la conception du standard, qui ne garantit aucunement la contiguïté de tous les caractères d'une écriture. La section 2.8 du standard Unicode le précise (http://www.unicode.org/versions/Unicode4.0.0/ch02.pdf) :</p> <p>« Les caractères utilisés par un système d'écriture peuvent se retrouver dans plusieurs blocs. Ainsi, trouve-t-on des caractères qui servent de lettres pour des systèmes d'écritures à base latine dans au moins neuf blocs : Latin de base, Supplément latin-1, Latin étendu A, Latin étendu B, les Suppléments phonétiques, Latin étendu additionnel, Lettres modificatives avec chasse et Diacritiques. »</p> <p>[traduction du N2911]</p> <p>En outre, c'est déjà le lot du n'ko d'utiliser des caractères d'autres blocs, comme l'admet d'ailleurs la proposition N2765 elle-même :</p> <p>« Le n'ko utilise une série de signes de ponctuation communs à d'autres écritures. Le n'ko se sert à la fois du POINT D'INTERROGATION ARABE U+061F et du VIRGULE ARABE U+060C, du POINT-VIRGULE ARABE U+061B, et les PARENTHÈSES ORNÉES appariées U+FD3E-U+FD3F (souvent sous des formes différentes de l'arabe), »</p> <p>Les signes de paraphrases pour le n'ko</p>

Pour un codage séparé	Contre un codage séparé
	ont également déjà été introduits dans un autre bloc dès Unicode 4.1 (U+2E1C et U+2E1D).
<p>Origine distincte : il a été soulevé (discussion à Markham) que les diacritiques n'ko ont une origine distincte des diacritiques génériques et que ceci justifie leur codage distinct. [N2927]</p>	<p>« Ceci est pour le moins peu pratique : devoir prouver un lien historique partagé entre les diacritiques d'une nouvelle écriture dont certains diacritiques ressemblent aux diacritiques génériques et ces diacritiques du bloc commun signifierait, s'il fallait l'appliquer de manière systématique à l'avenir, qu'il faudrait s'embarquer dans de longues études qui seraient toujours sujettes à révision ou, pire, qui pourraient se révéler inexactes (faut-il alors changer les affectations ?). Mais dans quel but faudrait-il se lancer dans cette étude laborieuse ? Pour quelle raison technique impérieuse ? Nous pensons qu'il vaut mieux considérer les similarités d'aspect et de comportement des diacritiques, et non leur origine, quand on détermine si les diacritiques du bloc commun sont appropriés. L'usage actuel semble appuyer ce point de vue. En effet, le U+0308 est, à la fois, utilisé pour représenter le <i>dialytika</i> grec, l'<i>umlaut</i> allemand, le tréma français, la dérivée double, le tréma cyrillique, comme un signe numéral hébreu et comme diacritique hébreu, sans qu'une origine commune (ou sémantique commune, par ailleurs) ne soit du tout évidente »</p> <p>[traduction de N2911]</p>
<p>Association de polices : « il nous paraît donc préférable de considérer ces caractères comme particuliers à cette écriture afin d'éviter tout problème [...] d'association de polices. » [N2765]</p>	<p>Cet argument n'a jamais été soutenu par le moindre exemple concret de difficulté d'association de polices qui émanerait de l'utilisation des diacritiques génériques.</p> <p>Une mise en œuvre du rendu n'ko (cf. ci-dessous) démontre aujourd'hui que cette mise en œuvre est tout exactement aussi facile (ou aussi difficile) avec les diacritiques génériques qu'avec des diacritiques codés dans le bloc n'ko.</p>
<p>Hameçonnage : à notre connaissance, aucun argument pour contrer l'augmentation du risque d'hameçonnage (cf. ci-contre) n'a été présenté.</p>	<p>Le codage de nouveaux diacritiques semblables à d'autres déjà codés augmente le risque d'hameçonnage par le biais de la parodie de, par exemple, des noms de</p>

Pour un codage séparé	Contre un codage séparé
<p>Traduction d'un extrait de la liste unicode@unicode.org, message de Michael Everson du 21 février 2005 :</p> <p>>Sinon, je suppose que les diacritiques >propres à une écriture particulière qui >ressemblent et se comportent(*) comme >les diacritiques génériques peuvent >constituer un risque de parodie.</p> <p>[à quoi M. Everson répond]</p> <p>Ce n'est pas une raison pour ne pas coder des diacritiques propres à une écriture.</p> <p>http://www.unicode.org/mail-arch/unicode-ml/y2005-m02/0304.html</p>	<p>domaines internationalisés (IDN). Un nom de domaine (ou tout autre identifiant intéressant un fraudeur) en n'importe quelle écriture utilisant un ou plusieurs des cinq diacritiques litigieux pourrait être parodié en utilisant un des nouveaux diacritiques n'ko. Ce risque peut bien sûr être éliminé en plaçant les cinq diacritiques n'ko sur liste noire, mais ceci interdirait l'utilisation légitime d'identifiants en n'ko contenant ces diacritiques.</p> <p>Il est admis que ce risque existe déjà (il y a plusieurs diacritiques ressemblant à un point suscrit par exemple), mais ce n'est pas une raison de l'augmenter encore dans un domaine qui a beaucoup attiré l'attention ces derniers temps.</p>

Bref, nous pensons qu'aucun des arguments avancés n'autorisent à s'écarter des lignes directrices du codage de l'ISO 10646/Unicode. En outre, il faut rappeler que ces écarts non justifiés ne sont pas sans inconvénient en termes de facilité de recherche textuelle ou de sécurité.

III. UNE MISE EN ŒUVRE DU N'KO

M. Patrick Andries a réalisé une mise en œuvre du rendu du n'ko, dans le but de démontrer la faisabilité d'une part, et de découvrir d'éventuelles difficultés de mise en œuvre d'autre part. Cette mise en œuvre est fondée sur la technologie standard XSL du W3C et transforme un document XML source contenant du texte en n'ko codés en utilisant les caractères du JUC retenus par le Canada en un fichier PDF, en utilisant des polices OpenType produites pour la circonstance. Le fichier PDF peut dès lors être affiché ou imprimé par un lecteur PDF standard. On trouvera en annexe A des détails de mise en œuvre ainsi qu'un exemple de rendu.

Cette mise en œuvre n'est ni très complète ni très raffinée, mais elle vaut démonstration de principe. Elle démontre ce qui suit :

- ☒ Le n'ko tel que proposé par le Canada peut être mis en œuvre de façon satisfaisante sans grande difficulté.
- ☒ Cette mise en œuvre utilise les diacritiques génériques pour les cinq diacritiques litigieux. *Aucune* difficulté supplémentaire de mise en œuvre ni *aucune* limitation de l'apparence ou du positionnement des diacritiques n'en découlent. La mise en œuvre aurait été la même avec des diacritiques dans le bloc n'ko, il ne s'agit au moteur de façonnage n'ko que de remplacer un nombre par un autre dans une table.

La principale difficulté de mise en œuvre est l'algorithme bidi. À défaut d'une mise en œuvre existante ouverte (et utilisable dans le contexte XSL) que nous aurions pu modifier, il a fallu écrire un petit algorithme bidi minimal pour les besoins de la démonstration. Cet algorithme n'est pas complet et robuste, et par conséquent cette mise en œuvre *ne* démontre *pas* si les propriétés bidirectionnelles proposées pour les lettres (AL) et les chiffres n'ko (AN) sont correctes (plutôt que, plus simplement, R). Les différences entre R et AL sont bien trop fines pour notre algorithme grossier. Mais il semble que la classe bidirectionnelle pour les chiffres n'ko proposée par l'Irlande (à savoir AN) soit incorrecte puisque les chiffres n'ko s'écrivent de droite à gauche à l'inverse des chiffres arabes.

IV. LE SOUTIEN DE L'INSTITUT N'KO

Les auteurs de la proposition prétendent que la communauté n'ko soutient la proposition. Ce soutien prend notamment la forme d'un message [N2932⁷] du président de l'Institut N'ko, M. Mamady Doumbouya, établi aux États-Unis, que nous analysons en en reproduisant ici l'essentiel [notre traduction de l'anglais] :

Nous avons examiné la liste des noms de caractère et le tableau de caractères pour le codage du n'ko dans l'ISO/CEI 10646 et nous sommes convaincus que ceux-ci satisfont à tous nos besoins techniques pour le stockage, le traitement et l'échange de textes n'ko.

Il est bien entendu que le n'ko tel que proposé est satisfaisant à ces égards ; qui plus est, la mise en œuvre de démonstration établit que ce codage peut être implémenté. Nos objections ont trait à ce que l'amendement 2 propose de coder plus que ce qui est nécessaire, sans bénéfice pour quiconque et au détriment des principes de codage de l'ISO/CEI 10646, ce qui entraîne des risques ou des désavantages.

Il est intéressant de noter que le même Mamady Doumbouya admet que la position canadienne est tout aussi valable que celle proposée :

« Je n'essaye pas de mettre en doute le bien-fondé de la solution canadienne, mais simplement de dire qu'il ne s'agit pas de la seule solution valable. »
[Mamady Doumbouya, traduction d'une correspondance privée du 18 mars 2005]

Le soutien de la communauté n'ko prend donc un nouvel aspect. Certes, la communauté soutiendrait la proposition par la voix du président de l'Institut N'ko – la seule que nous ayons entendue. Mais ce même président Doumbouya ne rejette pas la position canadienne, qu'il dit valide.

Dans ce même N2932, Doumbouya déclarait :

Nous aimerions réitérer notre exigence de coder les trois lettres archaïques séparément des lettres qui les ont remplacées. Cette distinction en texte brut nous permet de représenter précisément les phases pertinentes de l'évolution du n'ko, à l'instar de caractères déjà codés similaires qui permettent cette distinction pour l'anglais et d'autres langues européennes.

⁷ <http://std.dkuug.dk/jtc1/sc2/wg2/docs/N2932.doc>

Il s'agit ici d'une redite (cf. II. a ci-dessus). Et la réponse est la même : le n'ko devrait utiliser le même moyen que l'anglais et les autres langues européennes pour faire de telles distinctions, à savoir un changement de police.

Nous avons posé plusieurs fois la question de savoir pourquoi le n'ko devrait être traité différemment (messages privés, N2911), sans obtenir le moindre linéament de réponse.

M. Doumbouya poursuit :

Nous aimerions également répéter notre appui pour le codage de diacritiques propres au n'ko pour toutes les marques de ton. Celles-ci diffèrent autant dans leur aspect que leur position des signes « génériques » codés ailleurs dans la norme ; la codification de tous les diacritiques dans le seul bloc n'ko, tel que précisé dans le projet de norme proposé aux voix, facilitera la production de polices.

L'argument sur les formes et positions est une redite (cf II. b ci-dessus pour le contre-argument). Quant à l'argument de difficulté de mise en œuvre, qui n'a jamais été soutenu autre chose que des affirmations gratuites, nous espérons maintenant que la mise en œuvre mentionnée ci-dessus en démontre définitivement la vacuité.

V. CONCLUSION ET PROPOSITION

Pour les anciennes formes des lettres DJA, TCHA et RA, faut-il sacrifier les principes de l'ISO/CEI 10646 (codage de caractères abstraits et modèle caractères-glyphes) pour satisfaire un illusoire besoin d'une distinction au niveau du texte brut ? Notre réponse est non.

Pour les cinq diacritiques litigieux, faut-il augmenter les risques de parodie et d'hameçonnage, sans le moindre bénéfice perceptible pour qui que soit ? Notre réponse est non.

Il n'est toutefois pas question d'empêcher le codage du n'ko. Il est question, au contraire, de le coder *correctement*, de façon pleinement satisfaisante pour la communauté n'ko mais sans nuire à toutes les autres communautés d'utilisateur et d'implémenteurs de l'ISO/CEI 10646 en mettant en péril son architecture ou en augmentant les risques de l'utilisation d'Unicode en réseau.

La prudence et la recherche de consensus exigent de retirer ces caractères litigieux. Le processus de normalisation permettrait de les ajouter plus tard si, d'aventure, un réel besoin s'en faisait sentir. L'inverse est impossible : une fois codés, les caractères restent dans la norme *ad vitam aeternam*.

En résumé, nous proposons de retirer les caractères U+07E8, U+07E9, U+07EA (anciennes formes) et U+07EB, U+07ED, U+07EE, U+07F2 et U+07F3 (diacritiques) et de procéder au codage du reste. Nous savons par la correspondance du président Doumbouya⁸ que cette proposition est valide pour la communauté n'ko. Nous croyons aussi que c'est la meilleure position à adopter pour le reste de la communauté 10646/Unicode.

⁸ Correspondance également reçue par le modérateur du forum Unicode-Afrique, Don Osborn.

Annexe A – Prototype de mise en œuvre JUC du n’ko description et résultats

La source utilise les caractères du JUC proposés pour le n’ko

Les fichiers source de notre prototype utilisent les caractères du JUC proposés pour le n’ko à l’exception des caractères litigieux. En voici un extrait. Afin que les caractères soient bien visibles ci-dessous, nous les avons affichés ci-dessous sous la forme d’appels de caractère numérique (à savoir &#xxxx;).

```
<?xml version="1.0" encoding="UTF-8"?>
<nko:document xmlns:nko="http://hapax.qc.ca/nko"
xmlns="http://www.w3.org/1999/xhtml">
  <h1>Alphabet de la langue N&#x2019;KO</h1>
  <h1>&#xa0;&#x07D2;&#x07DE;&#x07CF;
&#x07DB;&#x07D3;&#x07CD;&#x07DB;&#x07CE;&#x0323;&#xa0;&#x07DB;&#x07CC;&#x07F0;&#x07D9;&#x07CA;&#x07EC;&#x07DF;&#x07CA;&#x0323;&#x07C7;</h3>
  <p>Le N&#x2019;KO comme l'arabe s'écrit de droite à gauche. Mais
comme le latin s'écrit avec deux sortes de lettres (voyelles et consonnes),
c'est ainsi que le n&#x2019;ko est indépendant car il n'est ni occidental
ni oriental.</p>
  <p>Donc il se base sur une neutralité positive.</p>
  <h3>Les sept voyelles =
&#x07DB;&#x07CC;&#x07F0;&#x07D9;&#x07CA;&#x07EC;&#x07DF;&#x07CA;&#x0323;&#x07C7;</h3>
  <table align="center" style="font-size:16pt;" width="280"
border="1">
    <tr>
      <td width="40" align="center">&#x07CA;</td>
      <td width="40" align="center">&#x07CB;</td>
      <td width="40" align="center">&#x07CC;</td>
      <td width="40" align="center">&#x07CD;</td>
      <td width="40" align="center">&#x07CE;</td>
      <td width="40" align="center">&#x07CF;</td>
      <td width="40" align="center">&#x07D0;</td>
    </tr>
```

Une version complète est disponible ici : <http://hapax.qc.ca/m%E9thodepratiqnenko.xml>

Le prototype utilise également deux polices OpenType (une pour les variantes modernes, l’autre pour les variantes anciennes) qui contiennent les formes contextualisées des caractères n’ko.

Annexe B – Appui pour l'analyse canadienne par Valentin Vydrine (expert linguiste en langues mandingues⁹)

Sujet:Re: Vieilles formes n'ko

Date:Samedi 29 jan 2005 19h26'22'' +0300 (MSK)

De:vydrine

Répondre à:[vydrine](#)

Pour:[patrick.andries](#)

Références:<41F691A6.000003.05202@camay.yandex.ru>

<41F69362.1030103@xcential.com> <41FA6597.40100@xcential.com>

Cher M. Andries,

Merci de m'avoir envoyé les documents en PDF. Malheureusement, le n2846 ne s'est pas ouvert, mais les deux autres sont bien lisibles.

En fait, l'argumentation du Canada (N2911.pdf) me semble raisonnable; on pourrait résoudre le problème des caractères anciens par la création des polices spéciales utilisant les mêmes codes que les nouvelles variantes des caractères. Il m'est difficile d'imaginer la situation où on aura besoin de se servir simultanément des nouvelles et des anciennes variantes dans le même texte. Et s'il s'agit des textes différents, le problème sera résolu par les polices différentes.

Je suis également d'accord avec l'argumentation canadienne concernant les signes diacritiques. A mon avis, là où ces signes existent pour les autres systèmes graphiques, il faut en profiter, plutôt que créer leurs doubles pour le n'ko.

Veuillez agréer, Monsieur Andries, l'expression de mes salutations les plus respectueuses,

Valentin Vydrine

⁹ Brève bibliographie du professeur Valentin F. Vydrine :

- ❖ « Souleymane Kanté, un philosophe-innovateur traditionnaliste maninka vu à travers ses écrits en nko », *Mande Studies* n° 3, 2001, pp. 99–131.
- ❖ Esquisse contrastive du Kgoro, (Manding), Rudiger Koppe Verlag, 2001.
- ❖ Dictionnaire mandingue-russe (« manden-rousskii slovar »): maninka, bamana (russe), Izd-vo « D. Boulanine», 1999.
- ❖ Sur le dictionnaire Nko, Mandenkan, 1996, Paris

----- Message original -----

Date: Mercredi 2 Fév 2005 21h15'56'' +0300 (MSK)

De: vydrine

Répondre à: vydrine

Pour: patrick.andries

Cher M. Andries,

Cette fois-ci, le fichier PDF s'est ouvert. J'essayerai donc de répondre a vos questions.

>1) dans l'extrait du Coran les anciennes formes semblent être entourées de >parenthèses (même de nouvelles lettres le sont dans le titre) que >représentent ces parenthèses (des parenthèses ornées arabes, des sortes de >guillemets, autre chose ?) Pourquoi ces anciennes lettres sont-elles si >souvent entourées de parenthèses?

Tout d'abord, le texte que vous m'avez envoyé n'est pas en relation directe avec le Coran. Il est intitulé [Préface et il traite des innovations dans l'orthographe n'ko. Souleymane Kanté explique quels anciens caractères il remplace par des nouveaux graphèmes (qui sont donnés entre les parenthèses). Bien évidemment, dans ce texte les deux types de caractères apparaissent ensemble. A mon avis, pour les rares cas comme celui-ci, on pourrait utiliser une police différente, plutôt que réserver des cases pour les anciens graphèmes.

>2) dans l'extrait du livre de Souleymane Kanté, savez-vous ce qu'il est >dit dans l'encadré où il semble qu'une rangée parle des anciennes formes >et la suivante des nouvelles ? Kanté décrit-il simplement le passage des >anciennes formes aux nouvelles en recommandant les nouvelles ?

Oui, exactement. Comme il était le créateur et le suprême autorité, il ne recommandait pas, plutôt il imposait les nouvelles lettres.

>3) dans l'extrait du Coran un seul mot l'avant-dernière ligne (cf. pièce >jointe) semble utiliser une ancienne forme, est-ce un lapsus calami, >l'auteur ayant oublié la nouvelle forme, une fantaisie ?

Voici ce que dit ce passage: « Quand la nasalisation transforme (l) en (n), il faut l'écrire comme ça : (signe de « N assimilé ») sanni, senni, jonni. Avant on les écrivait comme ça : sanni jonni (les deux derniers mots, encadrés en rouge, sont donnés en ancienne orthographe, ce qui est dans la logique du texte).

>Selon vous il n'existe donc pas de raison de conserver cette vieille >lettre ici ? (Je le pense aussi.) La preuve serait évidemment qu'ailleurs >dans ce Coran (le connaissez-vous ?) le même mot ne serait pas écrit avec les >nouvelles formes.

Oui, je suis de l'opinion que les vieilles lettres ne méritent pas des cases spéciales dans l'UNICODE. Dans ce cas particulier, il s'agit d'un métatexte, tandis que dans un contexte naturel ils ne se retrouvent pratiquement jamais ensemble.

Amicalement,
Valentin Vydrine

I. THE DISPUTED CHARACTERS

The N’Ko block below contains disputable characters.

	0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F
07Cx	0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F
07Dx	0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F
07Ex	0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F
07Fx	0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F

The first disputable group is made up of 3 characters whose old and new forms are proposed as separate characters in the N’ko block.

Name	Old Form		Modern Form	
	Glyph	Code point	Glyph	Code point
Ja	Ƶ	U+07E8	Ɂ	U+07D6
Cha	ƶ	U+07E9	ɂ	U+07D7
Ra	ƴ	U+07EA	Ƀ	U+07D9

The second group of disputable characters is composed of a subset of the combining marks proposed for N’Ko whose form and behaviour is similar to the generic (and common) combining marks block (U+03xx). We specifically refer to the characters shown below:

U+07EB	◌̄	=	U+0304	◌̄ ?
U+07ED	◌̇	=	U+0307	◌̇ ?
U+07EE	◌̂	=	U+0302	◌̂ ?
U+07F2	◌̈́	=	U+0323	◌̈́ ?
U+07F3	◌̈́	=	U+0308	◌̈́ ?

II. PROS AND CONS

A certain number of arguments for or against the encoding of the disputable characters have been made, whether in the initial proposal¹⁰, in the subsequent ballot comments or elsewhere. This section lists these arguments and attempts to show them side-by-side.

¹⁰ <http://std.dkuug.dk/jtc1/sc2/wg2/docs/n2765.pdf>

II. a) The Old Forms

For a separate encoding	Against a separate encoding
Characters or glyphs: These are distinct characters, they must be encoded as separate characters.	These are glyph variants of same characters whereas ISO 10646 encodes only abstract characters, not glyphs (character-glyph model)
Historical continuity: “The historical aspect these characters offered by the delegate from Ireland is very vital to the expression of the continuity of the script development.” [N2898]	The expression of this continuity would be better served by assigning a single code to each letter. Expressing the development of the forms characters have adopted within a given script does not fall within ISO 10646’s scope.
Need for older writings: It is necessary to encode the old forms in order to be able to publish Solomana Kante’s early writings. “it will be necessary to publish these manuscripts first with the old letters, if only for writing authenticity.” [N2898]	“This necessity is not contested, but must be put in perspective with the same necessity facing any other script where letter shapes have changed. The solution is to use an appropriate font, not to encode the old forms. The Latin script has known many now obsolete forms such as Uncial and Fraktur. These forms are not and will not be encoded. When an old document needs to be reproduced in its original form [...] a font for this purpose is used, showing the old shapes for the letters.” [N2911]

For a separate encoding	Against a separate encoding
<p>Need to discuss forms: It is necessary to encode the old forms in order to be able to discuss them, in contrast with the new ones.</p> <p>“As an example, Dr. Mamadi Baba Diane publisher of the holly Qur’an in N’Ko, had to explain in his preface how to use the new characters.” [N2898]</p>	<p>This is the case for any script with glyph variants which would need to be discussed, for instance to discuss the Caroline forms in the Latin script. The solution is identical: use an appropriate font.</p> <p>Mr. Valentin Vydrine, expert linguist in the Manden languages and who has published several articles on N’ko, also shares this opinion:</p> <p>“Yes, I am of the opinion that the old letters do not warrant special codes in UNICODE. In this case, we are dealing with a metatext while in a natural context these will in practice never occur together.”</p> <p>[translation of private correspondence, see annex B]</p> <p>Finally, note that the three characters proposed by Ireland are not sufficient to encode all the old Forms. See the table on top of Annex A (in the results parts).</p>
<p>Plain text: The distinction between the old and new forms is is a plain text distinction.</p>	<p>Nothing proves this assertion, and repeated requests to support it have been vain. In any other script, the distinction between historic and modern forms is considered to be rich text (a choice of font).</p> <p>It is important to note that the old and new forms are never found together, except in metatexts discussing precisely these forms (and this includes the “proofs” provided in N2914¹¹). The N’Ko Koran and the N’Ko dictionary only use the new forms.</p>
<p>Wynn: This case is similar to the <i>wynn</i> case (ƿƿ, U+01BF and U+01F7) which is separately encoded from “w”.</p>	<p>“W” is not a graphical evolution of <i>wynn</i>. <i>Wynn</i> was abandoned in Middle English and replaced by a “uu”, which later became “w”.</p>
<p>Spelling: The graphical form was not the only thing that changed, the spelling of words also changed.</p>	<p>There is nothing singular in this, this happens in all writing systems as they evolve. [paraphrase of N2911]</p>
<p>Searching:</p>	<p>“[W]e also wish to bring attention to an argument that militates against this encoding:</p>

¹¹ <http://std.dkuug.dk/jtc1/sc2/wg2/docs/n2914.pdf>

For a separate encoding	Against a separate encoding
	<p>having a single code for, say, OLD RA and RA means that many kinds of text processing, most importantly search, will treat them identically. This means that someone searching for a word in a database or on the Internet does not have to remember to search it in both old and new forms (except of course in the case of orthographic change). This is a real benefit of not encoding the old shapes, which should not be underestimated.” [N2911]</p> <p>Let us add here that the initial proposal, N2765, already acknowledges this equivalence for sorting: “(though in ordering words, the OLD LETTERs JA, CHA, and RA are treated as variants of JA, CHA, and RA”. Separate coding is not crippling for sorting – well-made enough implementations can be adapted – but is much more troublesome for search, especially in a global context such as the internet.</p>

III. b) The Combining Marks

Let us first state that only the 5 disputable combining marks: U+07EB, U+07ED, U+07EE, U+07F2 and U+07F3 are in scope here. The encoding of the other combining marks used in N’ko is not challenged; we will therefore not discuss them hereafter.

For a separate encoding	Against a separate encoding
<p>Directionality: “it has been considered best to treat these as script-specific characters to avoid problems with directionality...” [N2765]</p>	<p>This argument is totally unfounded; combining marks adopt the directionality of their base character in the bidi algorithm.</p>
<p>Graphical variation: “Some N’Ko diacritics have a wider range of glyph representation than the generic marks do, and, for example, are typically drawn rather higher and bolder than the generic marks.” [N2765]</p>	<p>Generic diacritics already have wider variations in their representation as those observed between N’ko and the reference forms found in the standard. The diæresis-umlaut, for instance, can be found at different heights, even within the base letter, and can appear as two dots of varying boldness, even as a small “e” on top or as a double acute accent. The variations observed in N’ko are perfectly compatible with the variability already observed. The acute accent can be rendered from practically vertical (in Polish for instance) to practically flat.</p> <p>The N’ko combining marks, for example in the extract below written by Souleymane Kanté, the script’s inventor, are sometimes indistinguishable from the generic combining marks found (in French) side by side to them. Not all N’ko styles are bold. It is a matter, quite obviously, of the thickness of the stroke used, of the font used.</p> 
<p>Foreign characters: Characters foreign to N’ko must not be necessary to encode N’Ko texts.</p> <p>“Not to imply that some of the characters are secondary characters or not part of N’Ko character set but merely borrowed from other systems.” [N2898]</p>	<p>The 03xx diacritics are common, meant for use by all scripts and not reserved for any particular script. The Unicode standard is crystal-clear on this (section 7.7, http://www.unicode.org/versions/Unicode4.0.0/ch07.pdf):</p> <p>“The combining diacritical marks in this</p>

For a separate encoding	Against a separate encoding
	<p>block are intended for general use with any script.”</p> <p>[N2911]</p>
<p>A single block: Users must be able to find all N’ko character in a single block.</p> <p>“...to require the N’ko system to hunt for it characters in the other parts of the Unicode or to superimpose other diacritics will be for N’ko users an error of expediency of proportion similar to the one that assigned two positions instead of four to the calendar year prior to the year 2000.” [N2898]</p>	<p>“Having characters in various blocks is already the common situation for most if not all scripts! Here is another excerpt from the Unicode standard (section 2.8, (http://www.unicode.org/versions/Unicode4.0_0/ch02.pdf):”</p> <p>“Characters used in a single writing system may be found in several different blocks. For example, characters used for letters for Latin-based writing systems are found in at least nine different blocks: Basic Latin, Latin-1 Supplement, Latin Extended-A, Latin Extended-B, IPA Extensions, Phonetic Extensions, Latin Extended Additional, Spacing Modifier Letters, and Combining Diacritical Marks.”</p> <p>[N2911]</p> <p>Furthermore, N’ko already needs characters from other blocks, as acknowledged by the N2765 proposal itself:</p> <p>“N’ko uses a number of punctuation marks in common with other scripts. The ARABIC QUESTION MARK U+061F and ARABIC COMMA U+060C, ARABIC SEMICOLON U+061B, and the paired ORNATE PARENTHESES U+FD3E-U+FD3F are used (often with different shapes than are used in Arabic).</p> <p>The paraphrases signs (used by N’ko) have also already been introduced as soon as Unicode 4.1 in another block (U+2E1C and U+2E1D).</p>
<p>Distinct origin: It was argued (Markham discussions) that the N’ko diacritics are of a distinct origin from the generic ones and that</p>	<p>“This is at best impractical: having to prove a common historical link between a new script’s generic-looking diacritics and the</p>

For a separate encoding	Against a separate encoding
<p>this would justify their encoding [N2927]</p>	<p>ones in the common block would mean, if applied systematically in the future, that one would have to embark on a protracted study which could always be subject to review or worse, proven wrong (do we change the assignments then?). But to what avail would one need to do this laborious study? For which compelling technical reason? We believe it is best to consider the similarities in shape and behaviour of diacritics, not origin, when determining whether those from the common block are appropriate. Current practice seems to support this view, with U+0308 being used for the Greek dialytika, the German umlaut, the French tréma, the double derivative, the Cyrillic diaeresis, as a mark for numerals in Hebrew and as a Syriac diacritic, without any common origin (or meaning, for that matter) being obvious at all.” [N2911]</p>
<p>Font-binding: “it has been considered best to treat these as script-specific characters to avoid problems with [...] font-binding. » [N2765]</p>	<p>This argument has never been supported by any concrete example showing how font binding would be made harder by the use of generic combining marks.</p> <p>A rendering implementation of UCS N’ko (see Annex A) now demonstrates that implementing with generic combining marks is just as easy (or just as difficult) as using combining marks found in the proposed N’ko block.</p>
<p>Phishing: to our best knowledge, no argument has been opposed against the risk of phishing (see opposite).</p> <p>Extract from the unicode@unicode.org list, message from Michael Everson on the 21st of February 2005 :</p> <ul style="list-style-type: none"> >Otherwise, I guess script-specific >diacritics that look and behave(*) like >generic diacritics may well embody some >spoofing potential. <p>[to which Mr Everson responded]</p> <p>That's not a reason not to encode script-</p>	<p>The encoding of diacritics similar to others already encoded increases unduly the risk of phishing through spoofing, for instance, of international domain names (IDN). A domain name (or any other identifier of interest to a hacker) in any script using any or several of the five disputable combining marks could be spoofed in using the new N’ko diacritics. This risk can of course be alleviated by blacklisting the 5 disputable N’ko diacritics, but this would mean that they could not be used as legitimate characters in N’ko identifiers.</p> <p>It is known that this risk already exists (several diacritics look like a subscript dot for</p>

For a separate encoding	Against a separate encoding
specific diacritics. http://www.unicode.org/mail-arch/unicode-ml/y2005-m02/0304.html	instance), but this is not a reason to compound the risk in an area which has generated a lot of publicity recently.

In summary, we believe that none of the arguments put forward authorize us to depart from the ISO 10646/Unicode encoding guidelines. Moreover, making an unjustified exception in this case is not exempt of disadvantages, whether in the area of text searching or of security.

III. A UCS N'KO IMPLEMENTATION

Mr. Patrick Andries has implemented a UCS N'ko prototype using Canada's suggestions in order, on the one hand, to demonstrate its feasibility and, on the other hand, to discover potential implementation issues. This implementation uses the standard W3C XSL technology, it transforms a XML source document containing N'ko texts encoded in UCS as proposed by Canada into a PDF file. It uses two sets of OpenType fonts (one for the old typefaces, one for the new ones). The PDF files can then be displayed or printed by a standard PDF viewer. Annex A provides more details on this implementation, as well as a sample rendered page.

This implementation is neither complete nor very refined, but it does serve as proof of concept. It demonstrates the following:

- ☒ N'ko as proposed by Canada may be implemented in a satisfactory fashion without much difficulty.
- ☒ This implementation uses generic diacritics for all five disputable diacritics. *No* additional implementation complexity and *no* limitation in the shape or position of the diacritics arose from this choice. The implementation would have been similar had we used diacritics encoded in the N'ko proposed block, the N'ko shaping engine simply has to replace a number in a table by another.

The main implementation difficulty is related to the bidi algorithm. In the absence of an existing open implementation (usable in our XSL context) which we could have modified, a minimal bidi algorithm had to be implemented for this proof of concept. This algorithm is neither complete nor robust and hence this implementation *does not* demonstrate whether the bidirectional properties proposed for the N'ko letters (AL) and the N'ko digits (AN) are adequate (rather than, more simply, R). The differences between AL and R are far too subtle for our crude bidi algorithm. But it seems that the bidirectional class for the digits proposed by Ireland (AN) is incorrect since N'ko numbers are written from right to left, contrarily to Arabic.

IV. SUPPORT FROM THE N'KO INSTITUTE

The authors of the proposal purport that N'ko user community supports the original proposal. This support takes the form, among others, of a message [N2932¹²] from the president of the N'ko Institute, Mr. Mamady Doumbouya, established in the United States. We will analyze this message while reproducing here the essential passages in italics.

We have reviewed the list of character names and the code table for the encoding of N'ko in ISO/IEC 10646, and we are satisfied that it meets all of our technical requirements for storage, processing, and interchange of N'ko text.

It is obvious that N'ko as proposed is satisfactory in this regard; moreover our sample implementation demonstrates that it is a workable model. Our objections deal with the fact that the amendment proposes to encode more than is necessary, without any benefit to any user and to the detriment of the ISO/IEC 10646 encoding principles, leading to risks or disadvantages.

It is interesting to note that even Mamady Doumbouya recognizes that the Canadian position is as valid as the one that was proposed in the Amendment:

"I am not trying to question the validity of the Canadian solution, but simply to say that it is not the only valid position." [Mamady Doumbouya, private correspondence dated 18th March of 2005]

The support of the N'ko community therefore takes a different shape. While it is true that the user community could support the proposal through the voice of the president of the N'ko Institute – the sole voice we have heard –, president Doumbouya does not reject the Canadian position which he calls valid.

We would like to reiterate our requirement for the three archaic letters to be encoded as characters distinct from the letters which replaced them. Making this distinction in plain text enables us to accurately represent relevant phases of N'Ko development, just as similar distinct characters enable such a distinction to be made for English and other European languages.

This is but repetition (cf. II. a here above). And the answer remains the same: N'ko should use the same means as English and other European languages to make such distinctions, namely use a different font.

We have asked several times why N'ko should be treated differently (private messages, N2911), without getting ever the slightest outline of an answer.

We would also like to repeat our support for the unique encoding of script-specific diacritics for all of the N'ko tone marks. These differ in form and position from the "generic" marks encoded elsewhere in the standard, and it will facilitate font development and implementation for these to be encoded as they have been proposed in the draft under ballot.

¹² <http://std.dkuug.dk/jtc1/sc2/wg2/docs/N2932.doc>

The argument on the shape and position is but a repetition (cf II. b here above for the counter-argument). As far as the complexity of implementation argument is concerned, this has never been anything but unsupported assertions and we hope that the implementation mentioned above definitively puts it to rest.

V. CONCLUSION AND PROPOSAL

For the old forms of the letters JA, CHA and RA, must the ISO/CEI 10646 principles (encoding of abstract characters, character glyph model) be sacrificed in order to satisfy an illusionary need of a plain text distinction? Our answer is no.

Is it worth increasing the risk of spoofing and phishing for the five disputable diacritics, without any perceptible benefit to anyone? Our answer is no.

However, we do not want in any way to rule out the encoding of N'ko. We believe we should, on the contrary, code it *correctly*, in a way fully satisfactory for the N'ko user community but without impacting all the other user and implementer communities of ISO/IEC 10646 by endangering its architecture or increasing the risks linked to its networked usage.

Prudence and consensus building require that the disputable characters be removed at this stage. The standardization process allows for their later addition later, were a real need to be found. The opposite is impossible: once encoded, characters remain in the standard *ad vitam æternam*.

In summary, we propose that the characters U+07E8, U+07E9, U+07EA (old letter forms) and U+07EB, U+07ED, U+07EE, U+07F2 et U+07F3 (diacritics) be withdrawn and to proceed with the encoding of the other N'ko characters. We know from the correspondence of president Doumbouya¹³ that this proposal is valid for the N'ko community. We believe this to be the best solution for the rest of the 10646/Unicode community.

¹³ Also sent to the moderator of the Unicode-Afrique list, Don Osborn

Annex B – Support for the Canadian analysis by Valentin Vydrine (expert linguist in Manding languages¹⁴)

[Our translation from French]

Subject:Re: N'ko Old Forms

Date:Saturday 29 Jan 2005 19:26:22 +0300 (MSK)

From:vydrine

Reply to:[vydrine](#)

To:[patrick.andries](#)

References:<41F691A6.000003.05202@camay.yandex.ru>

<41F69362.1030103@xcential.com> <41FA6597.40100@xcential.com>

Dear Mr. Andries,

Thank you for having sent me the documents as PDF. Unfortunately, N2846 could not be opened, but the two other ones were readable.

In fact, Canada's arguments (N2911.pdf) seem to me to be very reasonable; the problem with the old characters could be solved by the creation of special font using the same codes as the new variants of the characters. It is difficult for me to fathom a case when one would need to use old and new variants simultaneously in a same text. And in the case of different texts, the problem is solved by different fonts.

I also agree with Canada's argument relative to the combining diacritical marks. In my opinion, where these marks exist in other graphical systems, one must make use of it, rather than create duplicates for N'ko.

Yours respectfully,

Valentin Vydrine

¹⁴ Brief bibliography of professor Valentin F. Vydrine :

- ❖ Souleymane Kanté, a Maninka traditionalist philosopher-innovator seen through his writings in N'ko ("Souleymane Kanté, un philosophe-innovateur traditionnaliste maninka vu à travers ses écrits en nko"), *Mande Studies* no. 3, 2001, pp. 99–131. (In French)
- ❖ Contrastive sketch of Kagoro ("Esquisse contrastive du Kagoro (Manding)"), Rudiger Koppe Verlag, 2001. (in French)
- ❖ Manden-Russian Dictionary ("Manden-rousskii slovar"): Maninka, Bamana, Izd-vo "D. Boulanine", 1999. (in Russian)
- ❖ On the N'ko Dictionary ("Sur le dictionnaire Nko"), (in French) Mandenkan, 1996, Paris. (in French)

----- Original message -----

Date:Mercredi 2 Fév 2005 21h15'56'' +0300 (MSK)

From:vydrine

Reply to:vydrine

To:patrick.andries

Dear Mr. Andries,

This time, I was able to open the PDF file. I will thus try to answer your questions.

>1) in the Koran extract the old forms seem to be surrounded by parentheses
>(even new letters are surrounded in this title), what do these parentheses
>mean (are they ornate Arabic parentheses, some quite of quotation marks,
>something else? Why are these old letters so often surrounded by
>parentheses?

First of all, the text that you have sent me is not directly related to the Koran. Its title is "Preface" and it deals with the innovations in the N'ko orthography. Souleymane Kanté explains which old characters he replaces by new graphemes (who are given in parentheses). Obviously, in this text, the two types of characters appear together. In my opinion, in rare cases such as this one, a different font could be used, rather than reserving special code points for the old graphemes.

>2) In the extract of Souleymane Kanté's book, do you know what is said in
>the framed table where it seems a row speaks about the old forms and the
>next speaks about the new ones. Does Kanté simply describe who to pass
>from the old forms to the new ones, while recommending the new ones?

Yes, exactly. But as he was the creator and the supreme authority, he was not recommending, but rather imposing the new letters.

>3) In the extract of the Koran, a single word on the second last line (see
>attachment) seems to use the old form, is it a slip of the pen, the writer
>having forgotten to write the new form, a whimsical usage?

Here is what this passage says: "When through a nasalition the (l) is transformed into an (n), one must write it like this: ("assimilated N" sign) sannì, senni, jonnu. Before they were written like this: sannì jonni (the last two words, circled in red, are given in the old spelling, which respects the text's logic).

>According to you, there is therefore no reason to keep the old letter here?
>(I also think it.) The proof would be that elsewhere in this Koran (do you
>know it?) the same word would not be written with the new forms.

Yes, I am of the opinion that the old letters do not warrant special codes in UNICODE. In this case, we are dealing with a metatext while in a natural context these will in practice never occur together.

Yours friendly,
Valentin Vydrine