

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

**Doc Type: Working Group Document**

**Title: Second revised proposal to encode four historic Latin letters for Sakha (Yakut)**

**Authors: Ilya Yevlampiev, Nurlan Jumagueldinov, Karl Pentzlin**

**Status: Expert Contribution**

**Action: For consideration by JTC1/SC2/WG2 and UTC**

**Date: 2012-04-26**

**Replaces: L2/12-044, WG2 N4213**

*Some remarks regarding distantly similar Cyrillic letters are added at the end of the introduction, addressing a question raised during UTC #130.  
In consequence, two characters were renamed.*

## **1. Introduction**

Sakha (ISO 639: sah, also known as Yakut) is a language spoken by 456.288 people according to the 2002 census [2], mainly in the area which today is the Sakha Republic, a federal subject of Russia located in the north east of Siberia.

Some Cyrillic orthographies devised in the 19<sup>th</sup> century did not yield a larger success in alphabetization. Based on this fact and the feeling that the former orthographies did not reflect the phonemes of the language sufficiently, the linguist Semyon Andreyevich Novgorodov (semen nəʃurʊdɑp, Səmən Noğoruodap, 1892–1924) devised a Latin orthography using a subset of IPA, augmented by four letters for specific diphthongs which are proposed here. This orthography did use lowercase letters only (like IPA).

This orthography was introduced 1917 in the wake of the Russian revolution, and it was the official orthography of the Sakha language until 1927, when the alphabet was changed again.

Details can be found in [3].

Regarding the letter **Ǝ** proposed as U+AB60 LATIN SMALL LETTER IOTIFIED E, it is noted that this letter somewhat resembles the Cyrillic letter **Ѧ** U+0465 CYRILLIC SMALL LETTER IOTIFIED E. In fact, it is possible that Novgorodov was inspired by the "iotified" Cyrillic letters when devising the alphabet. However, the Latin IOTIFIED E resembles a "iotified Latin e", while U+0465 resembled a "iotified Cyrillic e" despite of its name based on its use rather than on its glyphic appearance. The following lines show the Cyrillic letters U+0435, U+0454, U+0465 using some common fonts:

**е € Ǝ** — Arial

**е € Ǝ** — Times New Roman

**е € Ǝ** — Roman Cyrillic (<http://kodeks.uni-bamberg.de/AKSL/Schrift/RomanCyrillicStd.htm> )

This shows that the IOTIFIED E is, if any, a "iotified Latin e", not a "iotified Cyrillic e". Thus, independent of the question whether unification of Latin and Cyrillic letters is appropriate at all, it cannot be unified with **Ѧ**, as **е** and **Ѧ** are distinct Cyrillic letters. Thus, unifying the iotified forms of these letters while the base letters are different in form and function would cause confusion.

## 2. Proposed Characters

### **Block: Latin Extended-E**

#### **Historic letters for Sakha (Yakut)**

*These letters were used 1917–1927 in the official IPA-based Latin orthography of that era.*

Ӑ U+AB60 LATIN SMALL LETTER SAKHA IOTIFIED A  
→ A657 cyrillic small letter iotified a

ӕ U+AB61 LATIN SMALL LETTER IOTIFIED E  
→ 0465 cyrillic small letter iotified e

œ U+AB62 LATIN SMALL LETTER OPEN OE  
→ 0254 latin small letter open o

uo U+AB63 LATIN SMALL LETTER UO

#### **Notes on the character names**

The names of the characters proposed here as Latin small letters "IOTIFIED E", "OPEN OE", and "UO" are derived from their shape. It is not unlikely that such letters are found when still unencoded repertoires of other historical orthographies or of dialectological transcriptions are re-searched (see e.g. some examples of similar letters in recent proposals for German dialectology). Thus, names which describe their obvious composition are proposed here.

As the "LATIN SMALL LETTER SAKHA IOTIFIED A" does not resemble glyphically an "a" (neither a Latin or modern Cyrillic nor a Church Slavonic one) and was devised for Sakha, it is named according to its use in the Sakha orthography which it is part of.

#### **Properties:**

```
AB60;LATIN SMALL LETTER SAKHA IOTIFIED A;Ll;0;L;;;;;N;;;;;
AB61;LATIN SMALL LETTER IOTIFIED E;Ll;0;L;;;;;N;;;;;
AB62;LATIN SMALL LETTER OPEN OE;Ll;0;L;;;;;N;;;;;
AB63;LATIN SMALL LETTER UO;Ll;0;L;;;;;N;;;;;
```

## 3. References

- [1] Юшманов, Николай Владимирович: Определитель Языков. Москва/Ленинград 1941
- [2] [http://ru.wikipedia.org/wiki/%D0%AF%D0%BA%D1%83%D1%82%D1%81%D0%BA%D0%B8%D0%B9\\_%D1%8F%D0%B7%D1%8B%D0%BA](http://ru.wikipedia.org/wiki/%D0%AF%D0%BA%D1%83%D1%82%D1%81%D0%BA%D0%B8%D0%B9_%D1%8F%D0%B7%D1%8B%D0%BA) (retrieved 2011-09-10)
- [3] <http://ilin-yakutsk.narod.ru/1999-12/50-2.htm> (retrieved 2011-09-05; in Russian)
- [4] Новгородов С.А. Первые шаги якутской письменности. Статьи и письма. - М.: Наука, 1977.
- [5] saqalw: suruk bicik, Yakutsk 1917
- [6] nәfurwodap, semen (Novgorodov, S. A.); et al., suruk bicik. mōskuba (Moscow) 1923
- [7] Будищев, И. П.; Русско-якутский словарь. Якутск 1926

## 4. Examples and Figures

**Fig. 1:** [1], p. 28/29. This table shows a compilation of special Latin letters used in orthographies of languages of the Soviet Union before the publication of [1] (1941). The Yakut letters proposed here are marked. (The table shows several other letters not contained in Unicode 6.0, which are not subject of this proposal.)

Таблица II  
Латиницы языков СССР

В алфавитный список не включены первичные латиницы, замененные впоследствии другими: *абхазская* (особенности — ə, ħ, ı, j, φ, ш, ђ, ѿ; буквы с точками, как ѓ, ѣ, кружками, как пе, зубчиками, как ј, черточкой, как †); *кабардинская* (особенности — à, è, ù; ç, ı, ş; x; h; сочетания типа fh, хu, qhu); *якутская* (особенности — э, ъ, ѓ; ħ [= Δ], њ, џ, ѡ, ѣ, ф, ѡ, з; двоеточие как знак долготы, напр. а:; прописных букв и знаков препинания нет).

*Ингушская, осетинская и чеченская* латиницы употребляли много сочетаний согласных с -h взамен особых букв кавказских языков. Языки народов Севера пользовались значками: ¯ над гласными (*удэйский, эвенкийский, эвенский*), ¨ под согласными (*ительменский, ненецкий, нивхский, саамский, селькупский*) и ‘ после согласных (*ительменский, нивхский, нымыланский, эскимосский*).

ä вепс., ижор., ингуш., карел.	ı хакасский	ï чеченский
а адыгейский	ʃ тувинский	ü см. ü, ù
æ ингуш., кетский, осет., селькупский, удэйский	қ кавказские, курдский	ÿ казах., каракалпакский
ë ингуш., осет., чеченский	к, в адыгейский	ÿ таджикский
є курдский, саамский	ç адыг., кабардинский	х ингуш., чеченский
ç [большинство <sup>1</sup> ]	ʎ авар., ительм., хант., эским.	х дагестанские
đ абхаз., адыг., башк., шугн.	ı абаз., авар., адыг., кабард.	х дагест., шугнанский
đ адыг., белуджский	ʎ хавтыйский	х абаз., адыг., кабард.
ə [большинство <sup>2</sup> ]	ц ингуш., тюркские, <sup>4</sup> чеченский	ž ингуш., осет., чеченский
é курдский	ц языки народов Севера	z кавказские
ı абаз., абхаз., адыг., кабард.	ö вепс., ижор., карел., чеченский	z дунганский
đ абазинский, абхазский	о [большинство <sup>5</sup> ]	z [большинство <sup>7</sup> ]
ç адыгейский	р адыгейский	z даргинский, табасаранский
oj [большинство <sup>3</sup> ]	р абаз., абхаз., кабард.	z кавказские, нанайский, саам., удэйск., эвенк., эвенск.
ħ абаз., адыг., дагест., курд., татск.	р дагест. (кроме авар.), курд.	z коми
ħ дагест., татск., шугнанский	р адыгейский	z саамский, шугнанский
ħ абаз., кабардинский	q абаз., абхаз., дагест., кабард.	ç, ç абазинский, абхазский
ħ абаз., адыг., кабардинский	г нивхский	ь [большинство <sup>8</sup> ]
ħ ингуш., кабардинский, кетский, маньс., ненецкий, нивх., осет.	š адыг., ингуш., чеченский	г, ħ адыгейский
ђ адыгейский	ş [большинство <sup>6</sup> ]	т, т абаз., абхаз., кабард.
ı абазинский, абхазский	s дагест., маньс., саам., хант.	ʎ абаз., абхаз., адыгейский
ı таджикский	ғ абазинский, абхазский	т, адыгейский
	т белудж., кавк., курд., сир.	ц, ц абаз., абхаз., кабард.
	ѣ адыг., башк., шугнанский	’ вепс., таджикский, тюркские (реже другие).
	р, р абхазский	

<sup>1</sup> Вепсский, ижорский, иранские, кавказские, карельский, сирийский, тюркские (кроме башк., ногайск., тувин.).  
<sup>2</sup> Иранские (кроме таджик.), кавказские (кроме абаз., авар., кабард.), северные (кроме ительм., маньс., нымыл.), сирийский, тюркские (кроме карач.-балк., киргиз., крым.-тат., кумык., ногайск., ойрот., тувин., узбек., хакас., шор., якут.).  
<sup>3</sup> Иранские, кавказские (кроме адыг.), тюркские (кроме ойрот., туркмен., хакас., шор.).  
<sup>4</sup> Кроме азербайджанского.  
<sup>5</sup> Тюркские (кроме узбек.) и еще: бурятский, калмыцкий, лакский, лезгинский, селькупский, удэйский, хантыйский, шугнанский.  
<sup>6</sup> Вепсский, ижорский, иранские, кавказские (кроме адыг.), карельский, сирийский, тюркские (кроме казах.).  
<sup>7</sup> Вепсский, ижорский, иранские (кроме татск.), кавказские (кроме адыг.), карельский, кетский, саамский, сирийский, тюркские (кроме казах., каракалп.).  
<sup>8</sup> Ижорский, курдский, северные (кроме маньс., нанайск., нивх., удэйск., эвенк.), сирийский, тюркские (кроме узбек., уйгур.).

Fig. 2: Title page and alphabet page from [5], pictures taken from {4}.

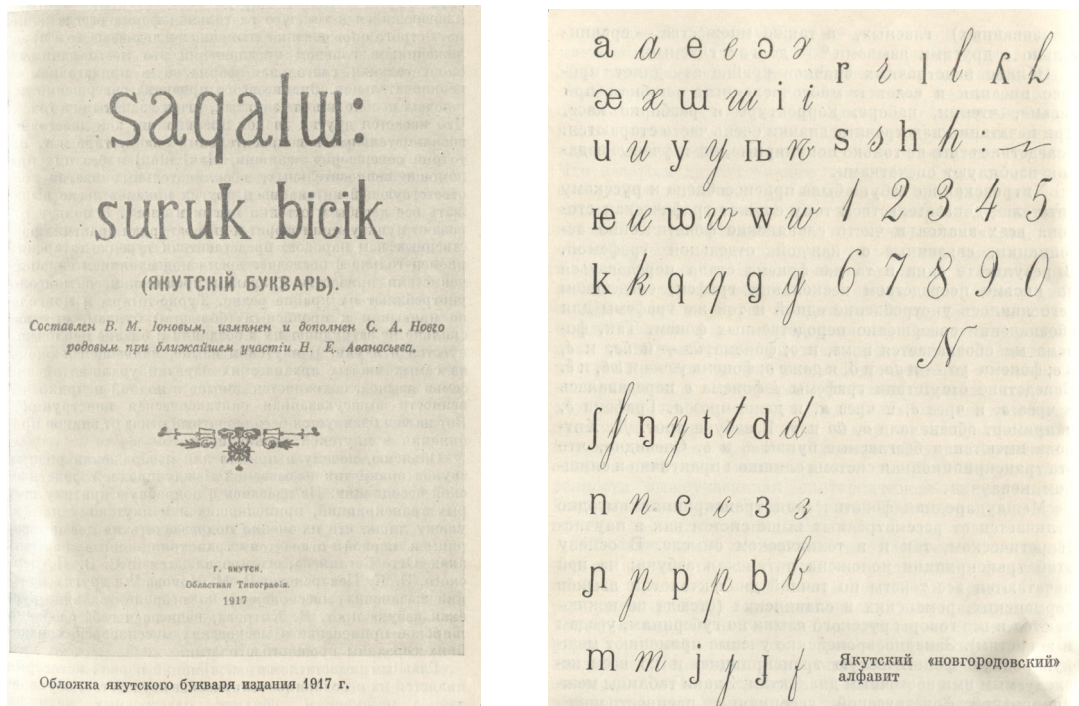
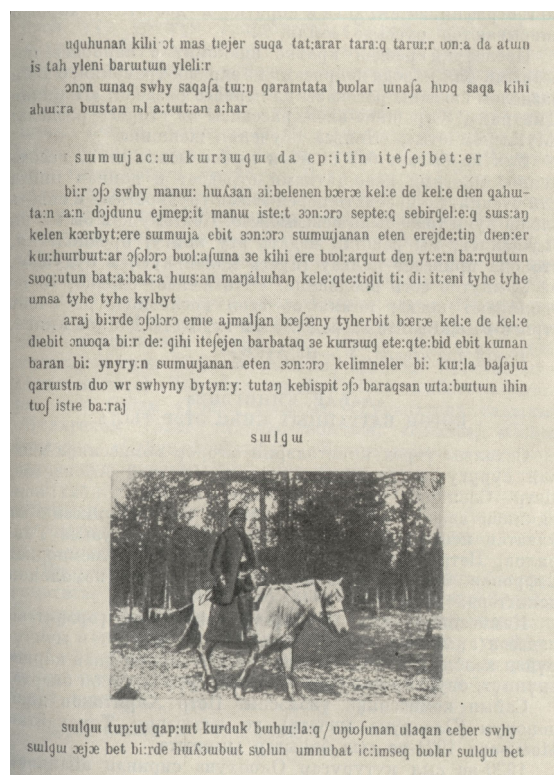


Fig. 3: Title page from {6}.





**Fig. 5:** Sample pages from several books using the Sakha orthography containing the proposed letters..



# ije u:hun atu:lu:r artıla

## J

*ofo ijetin y:ty:n ispetitten xler*



ты 300 bulbut astarun, bulbut bultarun toj eugahuna:tu atu:huitarga atu:luc:llar. atu:hut tus tuspa 30000 30000 qaja bafarar albunna:n twkejde:n ulan bejetin bassun qaqatar bu:lar. atu:hut bassun qaqatıq keriete, beje bassun qaqatar 30000 bulwıq ete. 30000 qaja ije u:ha ije u:hunan qalbh3n artı: 30000, bular talar astarun wllerin, bultarun altarun keppereti:pke ızen du:, k3myske tıje:n ızen du:, 30000 natu:gar atu:l3n tuhanıq etiler. ty kihite bular talar aha wle elbeq—aru:, y:t, qajaq, swhyty:n ete tırite, baha tuja:fa. ty kihite bultu:r bulta alfa da: elbeq—ty bultuttan tabata taja:fa, k3t3r kumatta:qtan kuha, qa:ha, u: bultuttan ejin ejin tıgnarw qaju:la:fun ara:ha... tuq da: k3ma swq... artı: bu manw barutun ty:nen qalbh3n tıpsaran 30000, barar siriger keppereti:pgerge du: kemısterge du: atu:luc: turdafuna, bu baruta ulaqan k3zy:ske turu ete. bu mantan kura 300 ulaqannuk tuhanı etiler.

qamu:nafa qalbh3n 30000 30000  
30000 swq bulwıqta:q.

## bu:kubalar.

a <sub>a</sub>	e <sub>e</sub>	3 <sub>3</sub>	3e <sub>3e</sub>
ı <sub>ı</sub>	ıe <sub>ıe</sub>	w <sub>w</sub>	wı <sub>wı</sub>
w <sub>w</sub>	i <sub>i</sub>	u <sub>u</sub>	y <sub>y</sub>
k <sub>k</sub>	q <sub>q</sub>	g <sub>g</sub>	ı <sub>ı</sub>
ı <sub>ı</sub>	t <sub>t</sub>	d <sub>d</sub>	n <sub>n</sub>
c <sub>c</sub>	3 <sub>3</sub>	ı <sub>ı</sub>	p <sub>p</sub>
b <sub>b</sub>	j <sub>j</sub>	ı <sub>ı</sub>	m <sub>m</sub>
r <sub>r</sub>	l <sub>l</sub>	ı <sub>ı</sub>	s <sub>s</sub>
h <sub>h</sub>	:	-	



Fig. 6: Two sample pages from [7].

	<b>A.</b>
Автономия	ap̄tan̄m̄ūja
Август	at̄ur̄z̄aq̄ ūja
Агент	eḡient
Алмаз	al̄ama:s̄ ta:s̄
Амбар	am̄pa:r̄
Амбулатория	amb̄ula:t̄ur̄ja
Америка	em̄er̄ike
Амнистия	em̄ni:st̄ije
Анастасия	na:sta:es̄t̄ija
Андрей	ænd̄ryej̄, ænd̄ry:s̄
Анна	a:na, a:n̄us, a:n̄uska
Англия	e:nḡilije
Алексей	æl̄eks̄ej̄
Александр	sa:nd̄ra, æl̄eks̄:ænd̄rye
Александра	æl̄eks̄:ænn̄ær̄e
Аркан	əfu:r̄
Апрель	mu:s̄ u:st̄ar
Апелляция	ep̄pele:ss̄ije
Арестант	q̄a:j̄u:tl̄a:q̄
Арестовать	q̄a:j̄u:jā uḡw̄q̄q̄a, q̄a:j̄ȳq̄q̄a
Армия	ass̄u:n̄
Армия	a:r̄m̄ūja
Аэродром	k̄æt̄ær̄ a:l̄ tyher̄ t̄əqt̄ur̄ sire
Аэроплан	k̄æt̄ær̄ a:l̄
	<b>Б.</b>
Бабушка	ebe
Баба	z̄aq̄tar̄

84	
Большие шали есть?	ul̄aq̄an̄ sa:l̄ bu:la:t̄ ba:r̄ d̄w̄?
Покажите одеяла?	sūw̄f̄annā k̄ærd̄ær̄?
Шапка есть?	ber̄gehē ba:r̄ d̄w̄?
Картуз есть?	q̄ətt̄u:s̄ ba:r̄ d̄w̄?
Покажите шарфы?	sa:r̄tā k̄ærd̄ær̄ ere?
Пояс есть?	ku:r̄ ba:r̄ d̄w̄?
Машинные нитки есть?	mass̄u:nā s̄abā ba:r̄ d̄w̄?
Серые нитки есть?	d̄wn̄ sap̄ ba:r̄ d̄w̄?
Полушубки есть?	k̄ul̄gas̄ ba:ra:n̄ s̄on̄ ba:r̄ d̄w̄?
Что стоит?	s̄ȳn̄atā t̄ə:h̄ænȳj̄? (tuḡuj̄?)
Дайте кирпич чаю?	bir̄ kip̄pi:cēē ca:j̄dā a:jal̄?
Дайте три фунта чаю байхового	ys̄ mu:n̄tā ət̄ ca:j̄dā a:jal̄?
Дайте два фунта сахара?	ik̄kī mu:n̄tā sa:q̄ardā a:jal̄?
Свесьте карамели	su:l̄a:q̄ q̄am̄p̄it̄tā ūja:
Дайте мюшасе	lam̄pas̄n̄j̄ q̄am̄p̄it̄tā a:jal̄
Рубашечные пуговицы есть?	ar̄ba:fū timefē ba:r̄ d̄w̄?
Пуговицы для пальто есть?	s̄on̄ timefē ba:r̄ d̄w̄?
Какие пуговицы нужны?	q̄aj̄daq̄ timeq̄ na:da?
Что стоит дюжина?	z̄u:h̄un̄atā q̄askā tur̄ar̄?
Сапоги есть?	s̄app̄ūkū ba:r̄ d̄w̄?
Катанки есть?	ka:t̄ūn̄kā ba:r̄ d̄w̄?
Войлок есть?	būx̄z̄əq̄ ba:r̄ d̄w̄?
Почем цена?	s̄ȳn̄atā t̄ə:h̄ænȳj̄?
Покажите швейные машины	i:st̄ener̄ mass̄u:n̄atā k̄ærd̄ær̄ ere?
Хорошая машина?	ȳeȳgej̄ mass̄u:nā d̄w̄?
Что стоит?	s̄ȳn̄atā t̄ə:h̄ænȳj̄?
Машинные иголки есть?	mass̄u:nā iḡ̄netē ba:r̄ d̄w̄?
Отпустите мне один фунт пороку и пять фунтов дроби	m̄eḡ̄ē bir̄ mu:n̄tā būra:fū ku:tt̄ā bīes̄ mu:n̄tā dara:m̄u:n̄atā ūja: ere
Все это что стоит?	bū bar̄ūtā t̄ə:h̄æf̄ē tur̄ar̄?

**PROPOSAL SUMMARY FORM TO ACCOMPANY SUBMISSIONS  
FOR ADDITIONS TO THE REPERTOIRE OF ISO/IEC 10646<sup>1</sup>**

Please fill all the sections A, B and C below.

Please read Principles and Procedures Document (P & P) from <http://www.dkuug.dk/JTC1/SC2/WG2/docs/principles.html> for guidelines and details before filling this form.

Please ensure you are using the latest Form from <http://www.dkuug.dk/JTC1/SC2/WG2/docs/summaryform.html>.

See also <http://www.dkuug.dk/JTC1/SC2/WG2/docs/roadmaps.html> for latest Roadmaps.

**A. Administrative**

1. Title:	<b>Second revised proposal to encode four historic Latin letters for Sakha (Yakut)</b>	
2. Requester's name:	<i>Ilya Yevlampiev, Nurlan Jumagueldinov, Karl Pentzlin</i>	
3. Requester type (Member body/Liaison/Individual contribution):	<i>Individual contribution</i>	
4. Submission date:	<i>2012-04-26</i>	
5. Requester's reference (if applicable):		
6. Choose one of the following:		
This is a complete proposal:	<input type="checkbox"/> Yes	
(or) More information will be provided later:	<input type="checkbox"/>	

**B. Technical – General**

1. Choose one of the following:		
a. This proposal is for a new script (set of characters):	<input type="checkbox"/> No	
Proposed name of script:		
b. The proposal is for addition of character(s) to an existing block:	<input checked="" type="checkbox"/> Yes	
Name of the existing block:	<i>Latin Extended-E</i>	
2. Number of characters in proposal:	<i>4</i>	
3. Proposed category (select one from below - see section 2.2 of P&P document):		
A-Contemporary	<input type="checkbox"/>	B.1-Specialized (small collection) <input checked="" type="checkbox"/>
C-Major extinct	<input type="checkbox"/>	D-Attested extinct <input type="checkbox"/>
F-Archaic Hieroglyphic or Ideographic	<input type="checkbox"/>	G-Obscure or questionable usage symbols <input type="checkbox"/>
B.2-Specialized (large collection)	<input type="checkbox"/>	E-Minor extinct <input type="checkbox"/>
4. Is a repertoire including character names provided?	<input type="checkbox"/> Yes	
a. If YES, are the names in accordance with the "character naming guidelines" in Annex L of P&P document?	<input type="checkbox"/> Yes	
b. Are the character shapes attached in a legible form suitable for review?	<input type="checkbox"/> Yes	
5. Fonts related:		
a. Who will provide the appropriate computerized font to the Project Editor of 10646 for publishing the standard?	<i>Karl Pentzlin; font available at <a href="http://www.pentzlin.com/sakhaletters1.ttf">http://www.pentzlin.com/sakhaletters1.ttf</a></i>	
b. Identify the party granting a license for use of the font by the editors (include address, e-mail, ftp-site, etc.):	<i>Released into the Public Domain by the author</i>	
6. References:		
a. Are references (to other character sets, dictionaries, descriptive texts etc.) provided?	<input type="checkbox"/> Yes	
b. Are published examples of use (such as samples from newspapers, magazines, or other sources) of proposed characters attached?	<input type="checkbox"/> Yes	
7. Special encoding issues:		
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)?	<input type="checkbox"/> n/a	

**8. Additional Information:**

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/reports/tr44/> ) and associated Unicode Technical Reports for information needed for consideration by the Unicode Technical Committee for inclusion in the Unicode Standard.

<sup>1</sup> Form number: N3902-F (Original 1994-10-14; Revised 1995-01, 1995-04, 1996-04, 1996-08, 1999-03, 2001-05, 2001-09, 2003-11, 2005-01, 2005-09, 2005-10, 2007-03, 2008-05, 2009-11, 2011-03)



**C. Technical - Justification**

1. Has this proposal for addition of character(s) been submitted before? If YES explain		No
2. Has contact been made to members of the user community (for example: National Body, user groups of the script or characters, other experts, etc.)? If YES, with whom? If YES, available relevant documents:	<i>Linguists, Librarians</i> <i>se text</i>	Yes
3. Information on the user community for the proposed characters (for example: size, demographics, information technology use, or publishing use) is included? Reference:	<i>see text</i>	Yes
4. The context of use for the proposed characters (type of use; common or rare) Reference:	<i>see text</i>	Historically common
5. Are the proposed characters in current use by the user community? If YES, where? Reference:	<i>see text</i>	Historical and scientific
6. After giving due considerations to the principles in the P&P document must the proposed characters be entirely in the BMP? If YES, is a rationale provided? If YES, reference:	<i>To keep them in line with similar characters</i>	Yes Yes
7. Should the proposed characters be kept together in a contiguous range (rather than being scattered)?		Yes
8. Can any of the proposed characters be considered a presentation form of an existing character or character sequence? If YES, is a rationale for its inclusion provided? If YES, reference:		No
9. Can any of the proposed characters be encoded using a composed character sequence of either existing characters or other proposed characters? If YES, is a rationale for its inclusion provided? If YES, reference:		No
10. Can any of the proposed character(s) be considered to be similar (in appearance or function) to an existing character? If YES, is a rationale for its inclusion provided? If YES, reference:		No
11. Does the proposal include use of combining characters and/or use of composite sequences? If YES, is a rationale for such use provided? If YES, reference: Is a list of composite sequences and their corresponding glyph images (graphic symbols) provided? If YES, reference:		No
12. Does the proposal contain characters with any special properties such as control function or similar semantics? If YES, describe in detail (include attachment if necessary)		No
13. Does the proposal contain any Ideographic compatibility characters? If YES, are the equivalent corresponding unified ideographic characters identified? If YES, reference:		No