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 10 Latin letters for pre-1921 Latvian orthography
Source: German NB
Status: National Body Contribution
Action: For consideration by JTC1/SC2/WG2 and UTC
Date: 2009-03-26, revised 2009-04-30 (font for representative glyphs changed; note regarding properties added)

Additions for pre-1921 Latvian orthography

Ǥ U+A7A0 LATIN CAPITAL LETTER G WITH OBLIQUE STROKE

ǥ U+A7A1 LATIN SMALL LETTER G WITH OBLIQUE STROKE

Ꝁ U+A7A2 LATIN CAPITAL LETTER K WITH OBLIQUE STROKE

ꝁ U+A7A3 LATIN SMALL LETTER K WITH OBLIQUE STROKE

Ꞥ U+A7A4 LATIN CAPITAL LETTER N WITH OBLIQUE STROKE

ꞥ U+A7A5 LATIN SMALL LETTER N WITH OBLIQUE STROKE

Ꞧ U+A7A6 LATIN CAPITAL LETTER R WITH OBLIQUE STROKE

ꞧ U+A7A7 LATIN SMALL LETTER R WITH OBLIQUE STROKE

Ꞩ U+A7A8 LATIN CAPITAL LETTER S WITH OBLIQUE STROKE
· Also used in pre-1950 Lower Sorbian orthography

ꞩ U+A7A9 LATIN SMALL LETTER S WITH OBLIQUE STROKE
→ 1E9C latin small letter long s with diagonal stroke

Informative notes and cross references to be added for existing characters:

U+1E9C LATIN SMALL LETTER LONG S WITH DIAGONAL STROKE
→ A7A9 latin small letter s with oblique stroke
· Medievalist use for several abbreviations
· Used in pre-1921 Latvian orthography
· Used in Blackletter types of pre-1950 Lower Sorbian orthography

All these letters are used in the orthography of Latvia used before 1921 [4] [5] [7], see fig. 1 to 6, 9 to 11. The LATIN SMALL LETTER S WITH OBLIQUE STROKE also occurs in a long s form; this form is unified with the already encoded U+1E9C LATIN SMALL LETTER LONG S WITH DIAGONAL STROKE.

Also, the stroked forms of L/l which are part of the pre-1921 Latvian orthography are proposed to be unified with the already encoded U+0141/0142 LATIN CAPITAL/SMALL LETTER WITH STROKE.

The encoding of the letters proposed here closes a gap regarding Blackletter (Fraktur) types:

According to [4] (which appeared 1880) pp. 227-234, Blackletter was in common use during the 19th century for the following languages: German, Danish, Swedish, Latvian, Finnish, Czech, Slovak. Even as this list may be incomplete, all letters used for the 19th century orthographies of all major European languages are now included in Unicode, with the exception of Latvian. Encoding the 10 letters proposed here would enable to create Blackletter fonts which cover the complete 19th century and early 20th century usage in Europe.

The stroked forms of G/g, K/k, L/l, N/n, R/r (but not the stroked form of S/s/ſ) roughly conform to the modern use of cedilla in the modern Latvian orthography (which is frequently displayed like a comma below for these members, or as a turned comma above in the case of the lowercase g). However, the stroke cannot be simply regarded as a typographic variant of the modern cedilla, as its appearance is

completely different not only in Blackletter (which was the prevalent printing type before 1921 in Latvia), but also in Roman type (see fig. 1, 9 to 11).

Besides the historic Latvian use, the s with oblique stroke also was used in Lower Sorbian orthography until about 1950 (until the current orthography was introduced) [6]. – See fig. 7 and 8.

Regarding the representative glyphs, the stroke is presented in a flat angle, according to the glyphs presented in fig. 1. Therefore, in the names the term "oblique stroke" is used regarding the stroke.

Properties:
```
A7A0;LATIN CAPITAL LETTER G WITH OBLIQUE STROKE;Lu;0;L;;;;;N;;;;A7A1;
A7A1;LATIN SMALL LETTER G WITH OBLIQUE STROKE;Ll;0;L;;;;;N;;;A7A0;;A7A0
A7A2;LATIN CAPITAL LETTER K WITH OBLIQUE STROKE;Lu;0;L;;;;;N;;;;A7A3;
A7A3;LATIN SMALL LETTER K WITH OBLIQUE STROKE;Ll;0;L;;;;;N;;;A7A2;;A7A2
A7A4;LATIN CAPITAL LETTER N WITH OBLIQUE STROKE;Lu;0;L;;;;;N;;;;A7A5;
A7A5;LATIN SMALL LETTER N WITH OBLIQUE STROKE;Ll;0;L;;;;;N;;;A7A4;;A7A4
A7A6;LATIN CAPITAL LETTER R WITH OBLIQUE STROKE;Lu;0;L;;;;;N;;;;A7A7;
A7A7;LATIN SMALL LETTER R WITH OBLIQUE STROKE;Ll;0;L;;;;;N;;;A7A6;;A7A6
A7A8;LATIN CAPITAL LETTER S WITH OBLIQUE STROKE;Lu;0;L;;;;;N;;;;A7A9;
A7A9;LATIN SMALL LETTER S WITH OBLIQUE STROKE;Ll;0;L;;;;;N;;;A7A8;;A7A8
```

Properties to be changed for already encoded letters: (reference to new uppercase letter)
```
1E9C;LATIN SMALL LETTER LONG S WITH DIAGONAL STROKE;Ll;0;L;;;;;N;;;A7A8;;A7A8
```

Note:

The properties (including the case mapping) for the triplet S/ſ/ẜ of the (proposed) U+A7A8/A7A9 and the already encoded U+1E9C were modeled after the already encoded triplet Ṡ/ṡ/ẛ U+1E60/1E61/1E9B (except that the latter triplet shows a canonical decomposition which the proposed one does not):

```
1E60;LATIN CAPITAL LETTER S WITH DOT ABOVE;Lu;0;L;0053 0307;;;;N;;;;1E61;
1E61;LATIN SMALL LETTER S WITH DOT ABOVE;Ll;0;L;0073 0307;;;;N;;;1E60;;1E60
1E9B;LATIN SMALL LETTER LONG S WITH DOT ABOVE;Ll;0;L;017F 0307;;;;N;;;1E60;;1E60
```

They resemble also the triplet S/s/ſ (except that this employs a compatibility equivalence for the long s; however, adding an equivalent compatibility equivalence for the already encoded U+1E9C to the proposed U+A7A9 was not considered for several reasons, like the use of U+1E9C in medieval context where it has a distinctive meaning):

```
0053;LATIN CAPITAL LETTER S;Lu;0;L;;;;;N;;;;0073;
0073;LATIN SMALL LETTER S;Ll;0;L;;;;;N;;;0053;;0053
017F;LATIN SMALL LETTER LONG S;Ll;0;L;<compat> 0073;;;;N;;;0053;;0053
```

Bibliography:

[1] ALLEN, C. G.: A Manual of European Languages for Librarians.
London & New York 1981, ISBN 0-85935-028-2

[2] BEZZENBERGER, Adalbert: Lettische Dialekt-Studien. Göttingen 1885

[3] BIELENSTEIN, A: Lettische Grammatik. Mitau 1863

[4] FAULMANN, Carl: Das Buch der Schrift.
Wien 1880. Reprint Nördlingen 1985, ISBN 392156851X

[5] HELLWIG, Wilhelm: Der Satz und die Behandlung fremder Sprachen. 5th. ed., Frankfurt/M. 1930

[6] KAULFÜRST, Fabian (Sorbisches Institut / Serbski Institut, Bautzen/Budyšin, Germany):
personal communication 2005/2008/2009.

[7] REICHSDRUCKEREI: Alphabete und Schriftzeichen des Morgen- und des Abendlandes. Berlin 1924.

[8] WISCHMANN, Johann: Der Unteutsche Opitz. Oder Kurtze Anleitung zur Lettischen Dicht⸗Kunst.
(The Not-German [i.e. Latvian] Opitz. Or, Short Guidance to the Latvian Art of Poetry.)
Rīga 1697

[9] Agenda. sa tu evangelisku zerkwu we Pschuskej. Worjejce (Hoyerswerda; Germany), 1898.

[10] Skohlas-Likkumi. Rīga 1789.

Acknowledgements:

LETTISCH.

Antiqua	Fraktur	Wert	Antiqua	Fraktur	Wert	Antiqua	Fraktur	Wert
A a	𝔄 α	a	J j	ℑ i	y	Pj pj	𝔓j pj	p'
Â â	𝔄̂ ā	ā	K k	𝔎 f	k hart	R r	𝔑 r	r
B b	𝔅 b	b	K k	𝔎 f	k lind	R r	𝔑 r	r'
Bj bj	𝔅j bj	b'	L l	𝔏 l	l	S f	𝔖 i	z
D d	𝔇 d	d	L l	𝔏 l	l'	S f	𝔖 ſ s	s
Ds ds	𝔇s ds	dz	M m	𝔐 m	m	Sch sch	𝔖ch ſch	ž
Dsch dsch	𝔇ch dſch	dž	Mj mj	𝔐j mj	m'	Sch ſch	𝔖ch ſch	š
E e	𝔈 e	e	N n	𝔑 n	n	T t	𝔗 t	t
Ė ė	𝔈 ê	ē	ng	ng	ń	Tsch tsch	𝔗ch tſch	lš
G g	𝔊 g	g hart	N n	𝔑 n	n'	U u	𝔘 u	u
G g	𝔊 g	g lind	ng	ng	ń'	Û û	𝔘 û	ū
H h	ℌ h		O o	𝔒 o	o	W w	𝔚 w	w
I i	ℑ i	i	Ô ô	𝔒̂ ô	ō	Wj wj	𝔚j wj	w'
Î î	ℑ̂ î	ī	P p	𝔓 p	p	Z z	ℨ ʒ	dz

Fig. 1: Scan from [4], p. 231, listing all letters of the Latvian orthography valid at its date of original print (1880), including all letters proposed here (the LATIN SMALL LETTER S WITH OBLIQUE STROKE is found in its round form in the "Fraktur" (Blackletter) column).

LETTISCH

Zeichen alte Orthogr.	neue Orthogr.	Lautwert	Zeichen alte Orthogr.	neue Orthogr.	Lautwert	
α	a	a	n	ņ	n (fast wie nj)	Allgemeines. Die lettische Sprache wird in Kurland, im südlichen Livland und in einem kleinen Teil des Gouvernements Witebsk gesprochen. Zu ihrer Wiedergabe bedient man sich meist der Fraktur. Seit 1921 ist eine neue Rechtschreibung eingeführt, die die Antiqua verwendet. Sie ist eine rein phonetische und bedient sich u. a. der Zeichen č, š und ž für die entsprechenden Zeichen tſch, fch und ſch der alten Orthographie.
â	ā	ā	o	o	o	
b	b	b	ô	ō	ō	
d	d	d	p	p	p	
e	e	e	r	r	r	
ê	ē	ē	r	ŗ	r (fast wie rj)	
f	f	f	ſ	z	ʃ	
g	g	g	ſch	ž	sch (stimmhaft)	Silbentrennung. Diejenigen Buchstaben, die einen Laut bilden, dürfen nicht getrennt werden. Ebenso darf h (das in der alten Orthographie als Dehnungszeichen dient) von dem vorhergehenden Vokal nicht getrennt werden.
g	ģ	g (fast wie gj)	f, ß	s	ß	
h	h	Hilfszeichen	fch	š	sch	
i	i	i	t	t	t	
î	ī	ī	u	u	u	
j	j	j	û	ū	ū	
k	k	k	w	v	w	
k	ķ	k (fast wie kj)	ʒ	c	z (= tß)	Interpunktion. Wie im Deutschen.
l	l	l (dumpf)	dſch	dž	dsch (stimmhaft)	Literatur. Grammatiken von Bielenstein (1863) und Brentano; Wörterbuch von Ulmann u. Brasche. Endselin u. Mühlenbach, Latwenšu gramatika (1907).
l	ļ	l (fast wie lj)	dſ	dz	dſ	
m	m	m	tfch	č	tsch	
n	n	n				

76

Fig. 2: Scan from [7], p. 76, showing the small letterforms and documenting the date until the old Latvian orthography was valid (starting in the sixth line of the right column, translating "Since 1921, a new orthography is introduced ...") – *Scan and reference provided by Andreas Stötzner, Signographisches Institut*

Schriftzeichen		Lautwert	Schriftzeichen		Lautwert
Alte Recht-schreibung	Neue Recht-schreibung		Alte Recht-schreibung	Neue Recht-schreibung	
α	a	a	n	n	nʲ
â	ā	ā (lang)	(v)	(o)	o
b	b	b	ô	ō	ō (lang)
ð	d	d	p	p	p
ẹ	e	e	r	r	r
ê	ē	ē (lang)	ɛ	ŗ	rʲ
	f	f	f	z	sanftes s
g	g	g	ſch	ž	stimmhaftes sch (wie franz. j)
g	ǧ	gʲ	ſ ß	s	ß
h	h	(nur Hilfs-zeichen)	ſch	š	sch
i	i	i	t	t	t
î	ī	ī (lang)	u	u	u
j	j	j	û	ū	ū (lang)
k	k	k	w	v	w
ꝁ	ķ	kʲ	ȥ	c	z, tß
l	l	l	ðſch	dž	dsch (stimmhaft)
ł	ļ	lʲ	ðſ	dz	dſ (sanft)
m	m	m	tſch	č	tsch
n	n	n			

Mehrere unserer Laute kommen im Lettischen nicht vor, wie ch ö ü,
andere werden umschrieben, z. B. q = kw, x = kſ, y = i, ä = ɛ. Auch
h wird als selbständiger Buchstabe nicht gebraucht, sondern nur beim
ſch und als Dehnzeichen bei Selbstlauten, c nur in der neuen Recht-
schreibung und als Hilfsbuchstabe bei ſch.

Akzent- und Sonderbuchstaben   2. Die Selbstlaute tragen in der
alten Rechtschreibung bei langer
Aussprache den Zirkumflex: â ê î ô û, mit Trema können ü und ï vor-
kommen, und außerdem gibt es noch mehrere gestrichene Buchstaben:
g ł ꝁ n ſ, welche auch als Versalien gebraucht werden: G Ꝁ L N R S.

Fig. 3: Scan from [5], p. 186, showing the small letterforms in the table and the capital forms in the text
(last line). – *Scan and reference provided by Andreas Stötzner, Signographisches Institut*
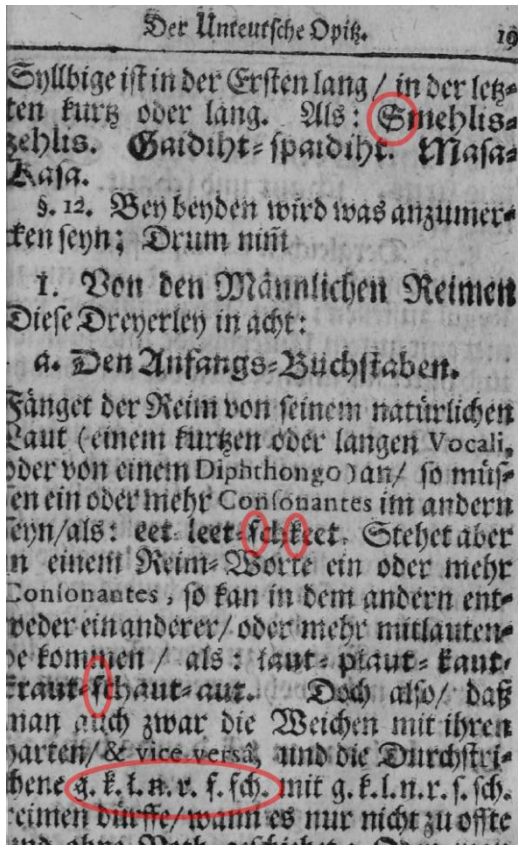
Fig. 4: Scan from [8], p.19, showing all lower case forms of the proposed letters.
The Text is in German, discussing rhyming rules for Latvian. The sentence containing the letter series translates: *"You also may rhyme the soft [letters] with their hard [counterparts], and the slashed g,k,l,n,r,s,sch with g,k,l,n,r,s,sch, if only this happens not too often and without necessity."*
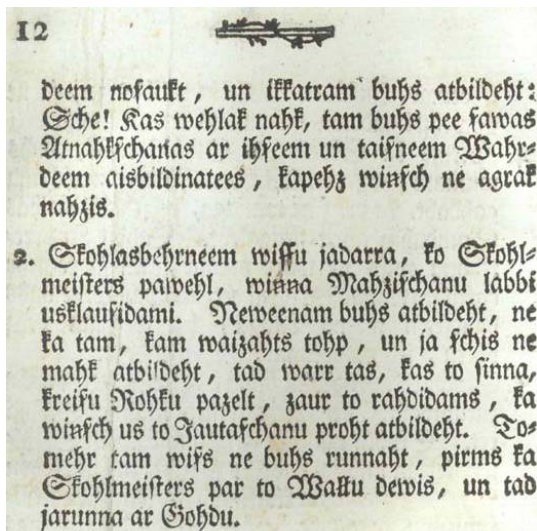– *Scan and reference provided by Prof. Stefan Kessler, Universität Greifswald*

Fig. 5: Scan from [10], p.12, showing some of the proposed letters.
Retrieved from http://valoda.ailab.lv/korpuss/senie/faksimili/18gs/SL1789/0012a.htm, 2008-10-28.

Note: This example shows ligatures of two long s with stroke (line starting with "2.") and of two l with stroke (second last line), employing a common stroke between the two letters. Such ligatures are common in 18th century prints but we have found no such ligatures in 19th century prints. We regard such ligatures as presentation forms, like other common Blackletter ligatures like "ch".
– *Reference provided by Gunars Lucans*

Fig. 6: Scan from [1], p.595, showing a sample text of Latvian (set in Fraktur), showing some of the proposed letters.

Fig. 7: Scan from a Lower Sorbian text of 1905.
As this text is set in Blackletter, it shows the s with oblique stroke in its long form only.
– Scan provided by Fabian Kaulfürst

Fig. 8: Scan from [9], providing a Lower Sorbian example from 1898, showing also the upper case of the S with oblique stroke. *– Scan and reference provided by Fabian Kaulfürst*

45   Tŏ dĭnu un nâkuschu nakti wìnsch palizis pĭ tâs pĭrmás
      méitas; ŭtrá dĭná gâsch pĭ ŭtras méitas, kuŗa ſĕlta istabá
      dſíwási, un tŏ sweizinásch. Tâ atbĭldési tápat ka pĭrmá:
      „labdĭn, wez' Plêscha! Kas tewi te nesa? Mäns wîrs ir stipris,
      tewi nŭkóus". Plêscha atbĭldésch: „na, na! es jow nâzu
50 tew nakaĭá".

Fig. 9: Scan from [2] (p.48), providing a Latvian example showing some of the proposed letters,
      printed in Antiqua type.

Weens ģēger's jā' us jakti
zaur[1]) äglu meſchīnu,
tur sateek wienu meitīnu.
„Ak meitīna, tu smuk' un daila,
kur tawi doemi stäwe"?
„Man' doemi stā' pee tewis
us ägle meſchīna".
Kur doama saulīt' lazoeti,
tur tumscha nakts atroad,
kur doama siewiņ' prezät,
tur oetris isprezäje.

Fig. 10: Another scan from [2] (p.48), providing a Latvian example showing some of the proposed
      letters, printed in Antiqua type.

8. î — í.

brîdenêt, schelten	br̦ídis, Frist
grìwa, Mündung	grída, Diele
îsts, eigentlich, ächt	íss, kurz
ķîlis, Keil	ķílas, Plur. tant. Pfand
krîts, Kreide	krítu, ich falle

Fig. 11: Scan from [3] (a Latvian grammar in German), showing Latvian examples in Antiqua type.

ISO/IEC JTC 1/SC 2/WG 2

<table>
<tr><td colspan="2">

PROPOSAL SUMMARY FORM TO ACCOMPANY SUBMISSIONS
FOR ADDITIONS TO THE REPERTOIRE OF ISO/IEC 10646[1]
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*.**

</td></tr>
</table>

A. Administrative

1. **Title:**		*Proposal to encode 10 Latin letters for pre-1921 Latvian orthography*
2. Requester's name:		*German NB*
3. Requester type (Member body/Liaison/Individual contribution):		*Member Body Contribution*
4. Submission date:		*2009-03-26, revised 2009-04-30*
5. Requester's reference (if applicable):		
6. Choose one of the following:		
This is a complete proposal:		*Yes*
(or) More information will be provided later:		

B. Technical – General

1. Choose one of the following:	
a. This proposal is for a new script (set of characters):	*No*
Proposed name of script:	
b. The proposal is for addition of character(s) to an existing block:	*Yes*
Name of the existing block:	*Latin Extended-D*
2. Number of characters in proposal:	*10*

3. Proposed category (select one from below - see section 2.2 of P&P document):

A-Contemporary ____	B.1-Specialized (small collection) __X__	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 ____

4. Is a repertoire including character names provided?	*Yes*
a. If YES, are the names in accordance with the "character naming guidelines" in Annex L of P&P document?	*Yes*
b. Are the character shapes attached in a legible form suitable for review?	*Yes*
5. Who will provide the appropriate computerized font (ordered preference: True Type, or PostScript format) for publishing the standard?	*Andreas Stötzner (Signographisches Institut, Pegau/Sa., Germany)*
If available now, identify source(s) for the font (include address, e-mail, ftp-site, etc.) and indicate the tools used:	*http://www.pentzlin.com/AndronLettischglyphen.zip (containing the font as .TTF file)*
6. References:	
a. Are references (to other character sets, dictionaries, descriptive texts etc.) provided?	*Yes*
b. Are published examples of use (such as samples from newspapers, magazines, or other sources) of proposed characters attached?	*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)?	*No*

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 http://www.unicode.org/Public/UNIDATA/UCD.html and associated Unicode Technical Reports for information needed for consideration by the Unicode Technical Committee for inclusion in the Unicode Standard.

C. Technical - Justification

1. Has this proposal for addition of character(s) been submitted before?	*No*
If YES explain *(This proposal is a rework of L2/08-405 which was not decided at UTC #117 of Nov. 2008)*	
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.)?	*Yes*
If YES, with whom? *see references in the text*	
If YES, available relevant documents: *see references in the text*	
3. Information on the user community for the proposed characters (for example: size, demographics, information technology use, or publishing use) is included?	*Yes*
Reference: *see text*	
4. The context of use for the proposed characters (type of use; common or rare)	*see text*
Reference: *see text*	
5. Are the proposed characters in current use by the user community?	*Yes*
If YES, where? Reference: *see text*	
6. After giving due considerations to the principles in the P&P document must the proposed characters be entirely in the BMP?	*Yes*
If YES, is a rationale provided?	*Yes*
If YES, reference: *Keeping in line with other Latin characters*	
7. Should the proposed characters be kept together in a contiguous range (rather than being scattered)?	*Preferred*
8. Can any of the proposed characters be considered a presentation form of an existing character or character sequence?	*No*
If YES, is a rationale for its inclusion provided?	
If YES, reference:	
9. Can any of the proposed characters be encoded using a composed character sequence of either existing characters or other proposed characters?	*No*
If YES, is a rationale for its inclusion provided?	
If YES, reference:	
10. Can any of the proposed character(s) be considered to be similar (in appearance or function) to an existing character?	*No*
If YES, is a rationale for its inclusion provided?	
If YES, reference:	
11. Does the proposal include use of combining characters and/or use of composite sequences?	*No*
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?	*n/a*
If YES, reference:	
12. Does the proposal contain characters with any special properties such as control function or similar semantics?	*No*
If YES, describe in detail (include attachment if necessary)	
13. Does the proposal contain any Ideographic compatibility character(s)?	*No*
If YES, is the equivalent corresponding unified ideographic character(s) identified?	
If YES, reference:	