

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 Latin letters for the Slovene Metelko alphabet in the UCS**

**Author: Karl Pentzlin**

**Status: Individual Contribution**

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

**Date: 2011-05-01**

## 1. Introduction

The Metelko alphabet (Slovene: metelčica) was developed by Franc Serafin Metelko in 1825 and presented in [1]. It was designed to overcome the shortcomings of the then used Slovene orthography (bohoričica) by trying to establish an one-to-one relation for phonemes and letters.

The alphabet was used by some authors in printed works until 1833.

Some of the letters introduced in the basic Latin alphabet are modeled after the Cyrillic letters Ш/ш, Ц/ц, З/з, Ж/ж, and Ч/ч. However, these letters are modified to comply with the overall design of Latin letters. All these small letters have descenders while none of the Cyrillic originals have them. Moreover, the designs of Ш/ш, Ц/ц, and Ж/ж were considerably reworked by used curved forms where the Cyrillic originals only have straight constituents.

## 2. Encoding Considerations

The name and code position of U+A794/A795 LATIN CAPITAL/SMALL LETTER LUNATE EPSILON are coordinated with two other proposals currently being worked on: A proposal for characters used in German dialectology by Michael Everson et al., and an upcoming revision of L2/11-040 "Revised Proposal to encode characters for the English Phonotypic Alphabet".

The code position of the other letters is coordinated with the mentioned upcoming revision of L2/11-040, using the code space in the used blocks directly following the code points used there.

The Metelko alphabet uses a letter pair to denote the schwa sound which resembles a reversed S/s. Due to the principle of encoding Latin letters by shape than by meaning, this letter is unified with the existing U+01A7/U+01A8 LATIN CAPITAL/SMALL LETTER TONE TWO, in spite of the fundamentally different meaning.

Regarding U+A7DA/A7DB LATIN CAPITAL/SMALL LETTER ZE and U+A7DE/A7DF LATIN CAPITAL/SMALL LETTER CHE, the capital letters resemble the Cyrillic form, while the small letters are modified by having a descender. Even if it is advocated that in pure historic alphabets the use of Cyrillic letters is appropriate to denote Latin letters of an identical form, a cross-script special casing is considered problematic. To avoid this, the capital forms are proposed as Latin letters also, yielding unproblematic case pairs.

Regarding U+A7DF LATIN SMALL LETTER CHE, this letter is superficially similar to U+0265 LATIN SMALL LETTER TURNED H. However, in quality serified fonts, the bottom termination is different. While the turned h has a peak there (like the normal h as its top termination; see the representative glyph of U+0265 in the Unicode 6.0 charts), the small Latin CHE has a full serif

at its bottom termination, like a q (and a Cyrillic CHE). Therefore, the small Latin CHE is not unified with the turned h.

The situation is different for U+FB0F LATIN CAPITAL LETTER SLOVENE METELKO TSE. Its lowercase form incidentally is identical to U+019E LATIN SMALL LETTER WITH LONG RIGHT LEG, thus the lowercase forms have to be unified due to the principle of encoding Latin letters by shape than by meaning.

This leaves the U+FB0F LATIN CAPITAL LETTER SLOVENE METELKO TSE as a capital Latin letter with no formal lowercase counterpart.

Therefore it is put into the "Alphabetic presentation Forms" block, due to the principles laid out in L2/11-040 = WG2 N3981 "Revised Proposal to encode characters for the English Phonotypic Alphabet".

### 3. Proposed Characters

Annotations in parentheses address special issues for a character, or reference to figures where such special issues are discussed. (These annotations are not intended to be retained in the character list when copied into the standard.)

#### **Block: Latin Extended-D**

##### **Additions for historical and phonetic use**

- € U+A794 LATIN CAPITAL LETTER LUNATE EPSILON  
→ 0404 cyrillic capital letter ukrainian ie  
→ 2108 scruple  
→ A792 latin capital letter c with bar  
· the serif on the middle bar is sometimes retained or modified in sans-serif fonts  
· denotes the sound [tʃ] in the English Phonotypic Alphabet  
· denotes the sound [e] in the Slovene Metelko alphabet
- € U+A795 LATIN SMALL LETTER LUNATE EPSILON  
→ 03F5 greek lunate epsilon symbol  
· used in German dialectology

##### **Additions for the Slovene Metelko alphabet**

- Ĥ U+A7CE LATIN CAPITAL LETTER H WITH NOTCH
- ĥ U+A7CF LATIN SMALL LETTER H WITH NOTCH
- Ł U+A7D0 LATIN CAPITAL LETTER L WITH PIGTAIL AT BOTTOM  
· resembles the digraph "Lj" in modern Slovene orthography
- ł U+A7D1 LATIN SMALL LETTER L WITH PIGTAIL AT BOTTOM  
→ 027A latin small letter turned r with long leg
- Ń U+A7D2 LATIN CAPITAL LETTER N WITH INWARDS HOOK  
· resembles the digraph "Nj" in modern Slovene orthography
- ń U+A7D3 LATIN SMALL LETTER N WITH INWARDS HOOK  
· also used in modern Lithuanian dialectology

(NOTE: This letter is identical with U+AB83 as proposed in L2/11-135  
" Proposal to add characters used in Lithuanian dialectology to the UCS".)

⓪ U+A7D4 LATIN CAPITAL LETTER VERTICALLY BARRED O

Ⓛ U+A7D5 LATIN SMALL LETTER VERTICALLY BARRED O

Ⓜ U+A7D6 LATIN CAPITAL LETTER ROUND SHA  
→ 0428 cyrillic capital letter sha  
→ 05E9 hebrew letter shin  
→ 0633 arabic letter seen

Ⓨ U+A7D7 LATIN SMALL LETTER ROUND SHA

Ⓢ U+A7D8 LATIN CAPITAL LETTER ROUND SHCHA  
→ 042A cyrillic capital letter shcha

Ⓣ U+A7D9 LATIN SMALL LETTER ROUND SHCHA

Ⓤ U+A7DA LATIN CAPITAL LETTER ZE  
→ 021C latin capital letter yogh  
→ 0417 cyrillic capital letter ze

Ⓥ U+A7DB LATIN SMALL LETTER ZE  
→ 0292 latin small letter ezh

Ⓦ U+A7DC LATIN CAPITAL LETTER ROUND ZHE  
→ 0416 cyrillic capital letter zhe

Ⓧ U+A7DD LATIN SMALL LETTER ROUND ZHE

Ⓨ U+A7DE LATIN CAPITAL LETTER CHE  
→ 0427 cyrillic capital letter che  
→ A792 latin capital letter turned h

Ⓩ U+A7DF LATIN SMALL LETTER CHE  
→ 0265 latin small letter turned h

### ***Block: Alphabetic Presentation Forms***

#### **Addition for the Slovene Metelko alphabet**

ŋ U+FB0F LATIN CAPITAL LETTER SLOVENE METELKO TSE  
· lowercase is 019E latin small letter n with long right leg  
→ 0220 latin capital letter n with long right leg

### **3.1 Annotations to be added for already encoded characters**

ž	U+01A8	LATIN SMALL LETTER TONE TWO
	<i>after:</i>	· Zhuang (old orthography)
	<i>ADD:</i>	· used as schwa in the Slovene Metelko alphabet
ŋ	U+0220	LATIN SMALL LETTER N WITH RIGHT LONG LEG
	<i>after:</i>	· Lakota
	<i>ADD:</i>	· used as small tse in the Slovene Metelko alphabet
	<i>ADD:</i>	→ FBOF latin capital letter slovene metelko tse

### 3.2 Properties:

A794;LATIN CAPITAL LETTER LUNATE EPSILON;Lu;0;L;;;;;N;;;A795;  
A795;LATIN SMALL LETTER LUNATE EPSILON;Ll;0;L;;;;;N;;;A794;;A794  
A7CE;LATIN CAPITAL LETTER H WITH NOTCH;Lu;0;L;;;;;N;;;A7CF;  
A7CF;LATIN SMALL LETTER H WITH NOTCH;Ll;0;L;;;;;N;;;A7CE;;A7CE  
A7D0;LATIN CAPITAL LETTER L WITH PIGTAIL AT BOTTOM;Lu;0;L;;;;;N;;;A7D1;  
A7D1;LATIN SMALL LETTER L WITH PIGTAIL AT BOTTOM;Ll;0;L;;;;;N;;;A7D0;;A7D0  
A7D2;LATIN CAPITAL LETTER N WITH INWARDS HOOK;Lu;0;L;;;;;N;;;A7D3;  
A7D3;LATIN SMALL LETTER N WITH INWARDS HOOK;Ll;0;L;;;;;N;;;A7D2;;A7D2  
A7D4;LATIN CAPITAL LETTER VERTICALLY BARRED O;Lu;0;L;;;;;N;;;A7D5;  
A7D5;LATIN SMALL LETTER VERTICALLY BARRED O;Ll;0;L;;;;;N;;;A7D4;;A7D4  
A7D6;LATIN CAPITAL LETTER ROUND SHA;Lu;0;L;;;;;N;;;A7D7;  
A7D7;LATIN SMALL LETTER ROUND SHA;Ll;0;L;;;;;N;;;A7D6;;A7D6  
A7D8;LATIN CAPITAL LETTER ROUND SHCHA;Lu;0;L;;;;;N;;;A7D9;  
A7D9;LATIN SMALL LETTER ROUND SHCHA;Ll;0;L;;;;;N;;;A7D8;;A7D8  
A7DA;LATIN CAPITAL LETTER ZE;Lu;0;L;;;;;N;;;A7DB;  
A7DB;LATIN SMALL LETTER ZE;Ll;0;L;;;;;N;;;A7DA;;A7DA  
A7DC;LATIN CAPITAL LETTER ROUND ZHE;Lu;0;L;;;;;N;;;A7DD;  
A7DD;LATIN SMALL LETTER ROUND ZHE;Ll;0;L;;;;;N;;;A7DC;;A7DC  
A7DE;LATIN CAPITAL LETTER CHE;Lu;0;L;;;;;N;;;A7DF;  
A7DF;LATIN SMALL LETTER CHE;Ll;0;L;;;;;N;;;A7ED;;A7DE  
FBOF;LATIN CAPITAL LETTER SLOVENE METELKO TSE;Lu;0;L;;;;;N;;;;;

### 4. References

- [1] Metelko, Franz Seraph: Lehrgebäude der slowenischen Sprache im Königreiche Illyrien und in den benachbarten Provinzen. Laibach (now Ljubljana) 1825.

## 5. Examples and Figures

**Fig. 1:** From [1], unnumbered page before page 1.  
 Showing the Metelko alphabet (the title translates "Slovene script for printing and handwriting").



Fig. 2: *ibid.*, p. 1:

Showing the Metelko alphabet in relation to contemporary Latin alphabets (referred as "Neu-Krain.", i.e. "new Slovene").

**E i n l e i t u n g.**

---

**Krainische und aus andern Alphabeten gleichbedeutende  
Schriftzeichen:**

**Neu-Krain. Alt-Krain. Kroatische, Deutsche, Französi. Italienische.**

A a	a	a	a	a	a
B b	b	b	b	b	b
D d	d	d	d	d	d
E e	e	e	e	e	e aperto
Ɔ c	č	— (?)	č	é	e chiuso(?)
F f	f	f	f	f	f
G g	g	g	g	g: gant	g: gara
H h	h	— h	h	— h	—
H h	h	h	h	—	—
I i	i	i	i	i	i
Š š	š, à, e	— (?)	— e	e: que	—
J j	j	j	ř	i: mien	j
K k	k	k	ř	c: car	ch
L l	l	l	l	l	l
L l	lj	ly	—	il: ail	gl: gli
M m	m	m	m	m	m
N n	n	n	n	n	n
N n	nj	ny	—	gn	gn
O o	o	o	o.	o	o chiuso
Ŏ ǒ	ò	—	—	oi (?)	o aperto (?)
P p	p	p	p	p	p
R r	r	r	r	r	r
S s	f	sz	š	s: son	s: sono
U u	fh	ss, sh	š	che	sce
У у	fhzh	sch	š	—	—
3 3	s	z	f: lesen	z	s: rosa
Х х	sh	s	—	j	—
T t	t	t	t	t	t
Ů ů	u	u	u	ou	u
V v	v	v	v	v	v
Ŧ ŧ	z	cz	š	—	z: zio
Ч ч	zh	ch	ř	—	c: ciò

1

Fig. 3: *ibid.*, p. 281: *Showing some text samples.*

## **E 3 O P O V E B A S N 2.**

### **1. Maba ine vol.**

Maba je videla vola na travneke, ine æelf nemu enaka bita. Zaune tedej svojo gërbasto kóæo napiheváte, ter vpráwa svoje tovarúñje: Ale nésem takó velíka kakor vol? Nése ne, ji pravijo. Le we bol se napéna, ter vpráwa: Sem zdé vender vèlekemu volu enaka? Spet ji odgovore, de ne. Pa ne néha; ze vso moújó se bol ine bol napéna, ine razpote.

Nápeh ve nesreçe perprave. Kdor sam sebe práv pozná, se néma ÷emú napiheváte.

### **2. Gãd ine pila.**

Gãd pride ponóça ve kovátneho, ine hoçe vse oródje razgriste. Narprej se sprave nad nákov (náкло). Ko pa neç ne oprave, ga zapustí, ine gre nad pilo: tukej, méne, bo ló-æe kej opravil. Pa tude lete se zastón perzadéva. Pila se mu posmehúje rekóu: Bedák, kakó hoçew mene se svójime zobní razgriste, ker jáz æelézo premágam, in nákov, kteremu ti neç ne morew, lahko ve prah zdrobím?

Kdor ve svojo moúj zaupa, družeh pa ne pozná, se opéte.

### **3. Volk ine jagne.**

Volk gre ke studénju pit, ine zagleda jagne, kateró veliko: niæe od nega pije. Naglo se volk zadere nad nim rekóu: Zakaj me vodó kaliw? Kakó æe to, lube volk, odgovorí ubogo jagne, jáz pijem spode, ti pa zgore; sej ne teçe voda ve brég! Volk se razserdí, ter prave: Zdé je ravno pol léta, kar se me ti opravlálo. Kakó? odgovorí jagne, takrat me we né belo na svéte. Je bil pa tvoj oçe, ale tvoja mate, zavpije volk, ine razterga nedolæno jagne.

Mouñéjimu vséle obvelá, ne prepiraj se æe nim.

Fig. 4: Cesar in Opat: Valentin Stanič's adaptation of the poem "Der Kaiser und der Abt" by Gottfried August Bürger (1828; Zgodovina slovenskega slovstva II, Lino Legiša).

CESAR IN OPAT. 1)  
(13 BÜRGERJA).

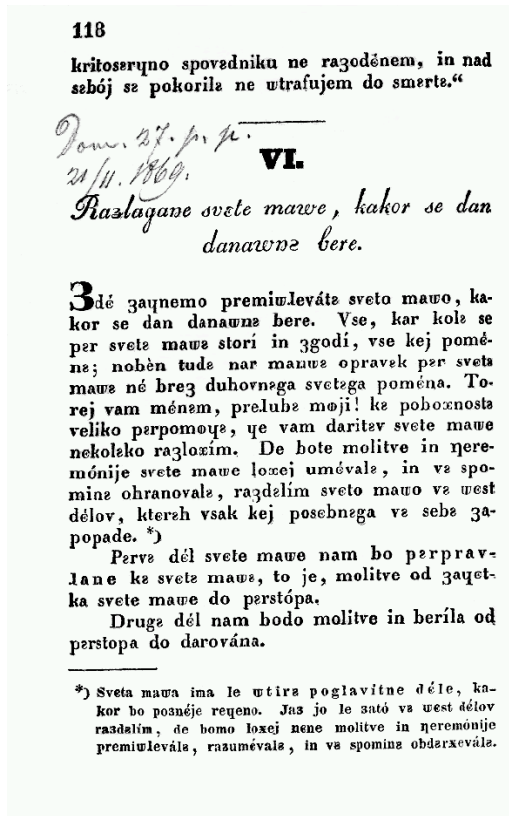
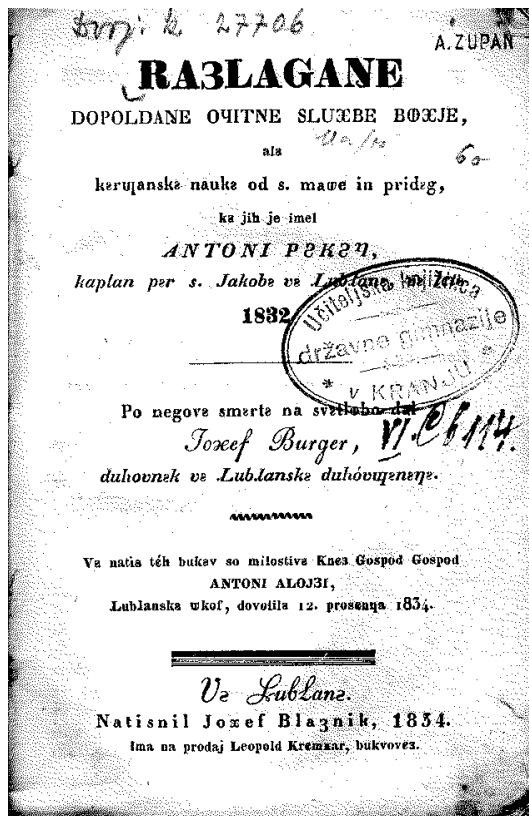
**P**ravilno povedal bom nove postave:  
Enkrat je bil cesar prebrisane glave;  
Opat je tud' bil, mox mnogoum lakrāt,  
Al ukoda, oyar je bal zvit bil k' opat.

Best' cesar prestal je vrečine in mraza,  
Spal malo, pa bil skož' veselga obraza;  
Je vede in qren krh komaj imel,  
Uo veokrat je xejo in lakot terpel.

Opatu je mogla bit' druga postrēga,  
Bogato napiten na mehko se vlega,  
Na tolstem obraze se linja bleuje;  
Obsegle trēbuha mu nēso trije.

1) To besede, kišero je Bürger iz stare pripovēste: „King John and the Abbot of Canterbury“ vzel, je se lada V. St. \* K. prestevail imo va lēto 1828 vs Tersta na svatlobe dal.

Fig. 5: Title page and a sample page of a book using the Metelko alphabet (retrieved 2011-02-06 from <http://www.arts.ualberta.ca/~ljubljan/metelcica.html> )





**ISO/IEC JTC 1/SC 2/WG 2  
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:	<i>Proposal to encode Latin letters for the Slovene Metelko alphabet in the UCS</i>
2. Requester's name:	<i>Karl Pentzlin</i>
3. Requester type (Member body/Liaison/Individual contribution):	<i>Individual Contribution</i>
4. Submission date:	<i>2011-05-01</i>
5. Requester's reference (if applicable):	
6. Choose one of the following:	
This is a complete proposal:	<input type="checkbox"/> <i>Yes</i>
(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"/> <i>No</i>	
Proposed name of script:		
b. The proposal is for addition of character(s) to an existing block:	<input type="checkbox"/> <i>Yes</i>	
Name of the existing block:	<i>Latin Extended D; Alphabetic Presentation Forms</i>	
2. Number of characters in proposal:	<input type="checkbox"/> <i>20</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"/>	B.2-Specialized (large collection) <input type="checkbox"/>
C-Major extinct <input type="checkbox"/>	D-Attested extinct <input type="checkbox"/>	E-Minor extinct <input type="checkbox"/>
F-Archaic Hieroglyphic or Ideographic <input type="checkbox"/>	G-Obscure or questionable usage symbols <input type="checkbox"/>	
4. Is a repertoire including character names provided?	<input type="checkbox"/> <i>Yes</i>	
a. If YES, are the names in accordance with the "character naming guidelines" in Annex L of P&P document?	<input type="checkbox"/> <i>Yes</i>	
b. Are the character shapes attached in a legible form suitable for review?	<input type="checkbox"/> <i>Yes</i>	
5. Fonts related:		
a. Who will provide the appropriate computerized font to the Project Editor of 10646 for publishing the standard?	<i>Peter Weiss (Inštitut za slovenski jezik Frana Ramovša ZRC SAZU, Dialektološka sekcija), Ljubljana</i>	
b. Identify the party granting a license for use of the font by the editors (include address, e-mail, ftp-site, etc.):		
6. References:		
a. Are references (to other character sets, dictionaries, descriptive texts etc.) provided?	<input type="checkbox"/> <i>Yes</i>	
b. Are published examples of use (such as samples from newspapers, magazines, or other sources) of proposed characters attached?	<input type="checkbox"/> <i>Yes</i>	
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"/> <i>No</i>	

**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.

<sup>1</sup> Form number: N3702-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)

**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:	Yes <i>Peter Weiss, Ljubljana, Slovenia</i>
3. Information on the user community for the proposed characters (for example: size, demographics, information technology use, or publishing use) is included? Reference:	Yes <i>see text</i>
4. The context of use for the proposed characters (type of use; common or rare) Reference:	Specialized <i>see text</i>
5. Are the proposed characters in current use by the user community? 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? If YES, is a rationale provided? If YES, reference:	Yes Yes <i>To keep them in line with similar characters</i>
7. Should the proposed characters be kept together in a contiguous range (rather than being scattered)?	
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:	Yes Yes <i>See text</i>
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 character(s)? If YES, is the equivalent corresponding unified ideographic character(s) identified? If YES, reference:	No