ISO/IEC JTC1/SC2/WG2 N2962
L2/05-194
**2005-08-08**

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 add LATIN SMALL LETTER GLOTTAL STOP to the UCS**
**Source:** **Canada (SCC) and Ireland (NSAI)**
**Status:** **Member Body Contribution**
**Date:** **2005-08-08**

**Request.** This document asks for the disunification of a new *U+2C70 LATIN SMALL LETTER GLOTTAL STOP from the existing U+0294 LATIN LETTER GLOTTAL STOP. It also asks for the deletion of the case-pair relationship between U+0294 and U+0241 LATIN CAPITAL LETTER GLOTTAL STOP, and the addition of the case-pair relationship between U+0241 and the new *U+2C70. (The asterisk is used to show that this character is not yet encoded.)

If this proposal is adopted, the following three characters would exist:

ʔ 0294 LATIN LETTER GLOTTAL STOP
- caseless use in IPA and other phonetic notation (technical notation)
- caseless use in Nootka, Nitinaht, Musqueam, Kootenai, Thompson (Canadian aboriginal orthographies)
- does not have an uppercase equivalent

x 0241 latin capital letter glottal stop
x 2C70 latin small letter glottal stop
x 0C20 modifier letter glottal stop

Ɂ 0241 LATIN CAPITAL LETTER GLOTTAL STOP
- casing use in Chipewyan, Dogrib, Slavey (Canadian aboriginal orthographies)
- uppercase is 2C70 latin small letter glottal stop

x 0294 latin letter glottal stop
x 0C20 modifier letter glottal stop

ⱥ 2C70 LATIN SMALL LETTER GLOTTAL STOP
- casing use in Chipewyan, Dogrib, Slavey (Canadian aboriginal orthographies)
- uppercase is 0241 latin capital letter glottal stop

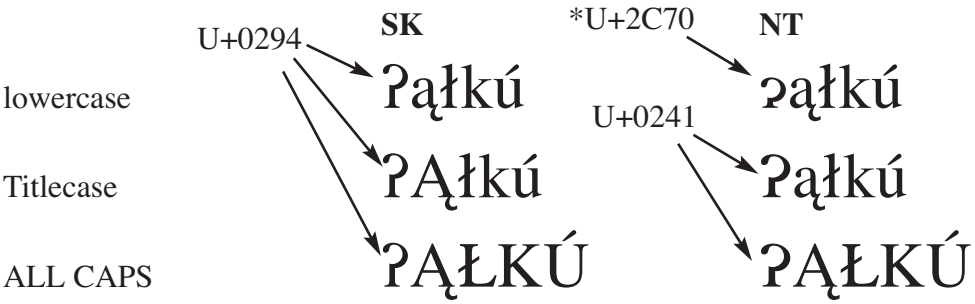
x 0294 latin letter glottal stop
x 0C20 modifier letter glottal stop

with the following properties:

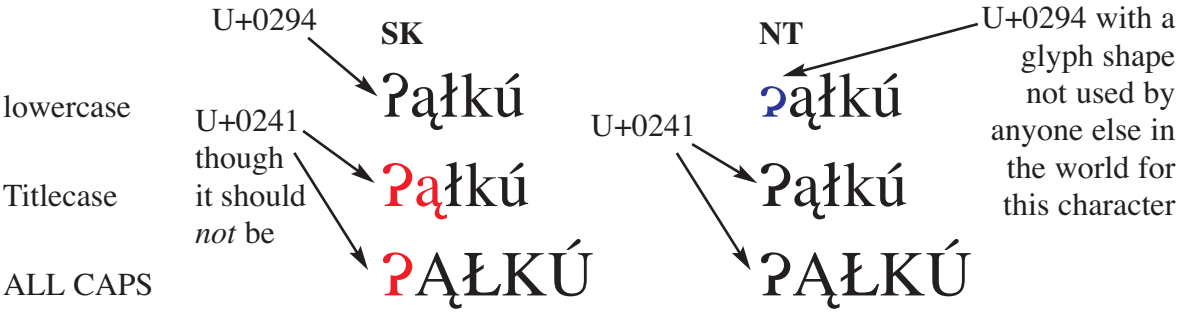
```
0294;LATIN LETTER GLOTTAL STOP;Ll;0;L;;;;;N;;;;;
0241;LATIN CAPITAL LETTER GLOTTAL STOP;Lu;0;L;;;;;N;;;;2C70;
2C70;LATIN SMALL LETTER GLOTTAL STOP;Ll;0;L;;;;;N;;;0241;;0241
```

**Difficulties for natural orthographies.** U+0294 ʔ LATIN LETTER GLOTTAL STOP is a letter which has long been used in the International Phonetic Alphabet and other transcription notations. Linguistic transcriptions using this letter inspired a number of natural orthographies for languages in Canada. Some Athapascan communities in the Northwest Territories innovated a *new* bicameral character

*from* U+0294. The two characters U+0241 Ɂ LATIN CAPITAL LETTER GLOTTAL STOP and *U+2C70 ɂ LATIN SMALL LETTER GLOTTAL STOP are distinct from U+0294—and both characters *also* need to be able to co-occur with U+0294 in plain text. The current unification and case-mapping causes unexpected and incorrect behaviour. An example follows here: the Chipewyan language in Saskatchewan uses U+0294 unicamerally. The Chipewyan language in the Northwest Territories uses U+0241 and *U+2C70 bicamerally. Here is the word [ʔãɬkú] 'sometimes' as it *should* appear in regular spelling, in titlecase, and in all caps:




In Saskatchewan, U+0294 is caseless in all three lines; titlecasing applies to the first letter following it, and to all the letters following it when in all caps. In the Northwest Territories, *U+2C70 is lowercase; titlecasing applies to it, and to it and all the letters following it when in all caps. Now here is the word [ʔãɬkú] as it is *currently* specified according to the Unicode Standard, in regular spelling, in titlecase, and in all caps:



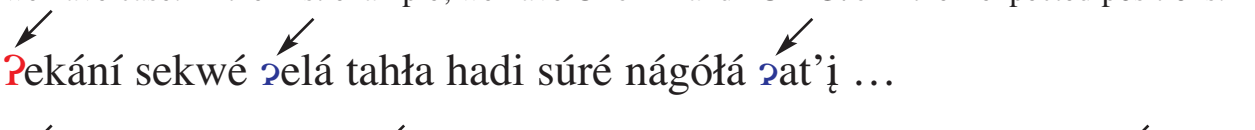

Because U+0294 currently uppercases to U+0241, titlecasing *does not work correctly* for Saskatchewan Chipewyan (or for the other natural orthographies—Nootka, Nitinaht, Musqueam, Kootenai, and Thompson—which also use the traditionally *unicameral* U+0294). The glottal stop is changed to a *different* character (which is not correct) and because the first letter is titlecased, the second letter does not capitalize when titlecasing, although it should. Moreover, special fonts have to be used for Northwest Territories Chipewyan—and Dogrib and North Slavey—in order to give the unified U+0294 the correct shape. That shape is *only* used in the Northwest Territories languages which use the *bicameral* glottal stop.

The current Unicode specification, in giving casing properties to the normally unicameral U+0294, disadvantages the users of Saskatchewan orthographies who will not get the casing behaviour they expect without special software tailoring. Similarly, users of Northwest Territories orthography are disadvantaged in terms of getting the special x-height glyph for their lowercase glottal without resorting to special fonts and language tagging. There is an important corollary to this: when plain text is displayed—without language tagging and control over font selection—legibility for the distinction between U+0294 and U+0241 can be *lost*.

Special Saskatchewan-specific software tailorings and special Northwest-Territories-specific fonts are only necessary because U+0294 and *U+2C70 have been unified. Disunification allows both communities to implement, process, and display their characters simply without special software for

either. It may be the case that such software could be made to solve the problems the unification causes. But who will provide that software, which needs to be supported in all applications on all platforms? Simply adding one character will allow Northwest Territories Athapascan users to get the behaviour and shapes they need, while leaving U+0294 untouched for all of its worldwide users.

**Difficulties for linguistic research.** Linguists and other researchers working with Athapascan in the Northwest Territories are also disadvantaged by the current unification. An example can be found in North Slavey. It would be perfectly natural for a linguist to write about a text in natural orthography and also to give it in phonetic notation. In natural orthography, as normally written and in all caps, we have case: In the first example, we have U+0241 and *U+2C70 in their expected positions:





And in IPA transcription, we have the caseless glottal stop U+0294:


Clearly, in this common linguistic context where all three forms are required, the encoding for the natural orthography and the phonetic transcription are *incompatible* if *U+2C70 and U+0294 are unified—unless the linguist is equipped with specialized fonts and language-tagging software. Of course, linguists do use special fonts which contain their specialized characters, but here the unification requires one of the special characters to have special rules to create a contextualized shape used only in Northwest Territories Athapascan orthography.

The unification means that unusual glyph-shaping behaviour would have to be applied to U+0294 *only* for Athapascan of the Northwest Territories. We believe that this is unreasonable, both in terms of the burden of font development for the user community, and because U+0294 has been in use by linguists for a century *without* this kind of size variation. The Northwest Territory Athapascan communities have innovated a new bicameral Latin letter, and the disunification requested here will permit them to use it, alongside the unicameral U+0294, without having to resort to any special language tagging or language-specific fonts.

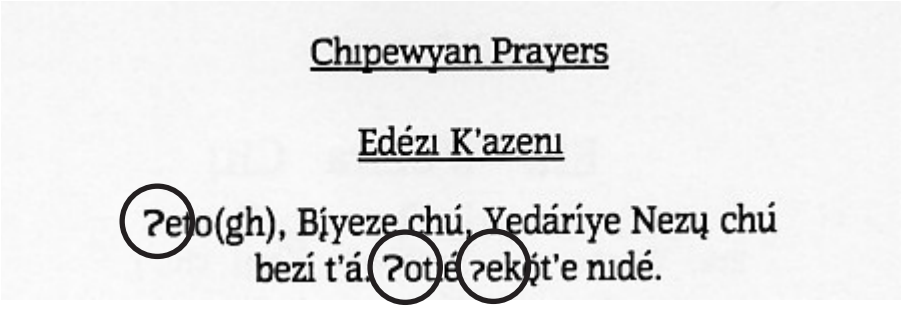
**Reasons for urgent disunification.** A number of new lowercase characters were proposed by the US Member Body to be added to FPDAM2 of ISO/IEC 10646:2004, in order to ensure case-folding stability. It is, apparently, important to the UTC that these characters be added before the publication of Unicode 5.0. We also understand that after the publication of Unicode 5.0, no new lowercase letter pairings will be added to existing uppercase letters which at the time of publication had no lowercase partner. In addition to the lowercase letters proposed by the US Member Body on its ballot comment on FPDAM2, this one character *U+2C70, used in Aboriginal Canadian orthographies in distinction to U+0294, *must* also be added as a matter of urgency, because the current mapping of U+0241 is to the *wrong* lowercase character.

U+0294 LATIN LETTER GLOTTAL STOP is used in a number of natural orthographies for Canadian languages. Nootka (Nuu-chah-nulth), Nitinaht (Diitiidʔatx̣), Musqueam (Hən̓q̓əmin̓əm̓), Kootenai (Ktunaxa), and Thompson (Nłeʔkepmxcin) all use this character, with its normal, tall glyph shape, and *without* casing behaviour for it.

The implications for multilingual data-processing should be clear. A Canadian government database, for instance, containing the names of persons from the different communities, cannot represent the names correctly. Even if language-tagging were able to specify an x-height glyph for U+0294 used in Dogrib, it still remains that casing behaviour has been *introduced* into the non-Northwest Territories languages where that behaviour is *not* part of their orthographies. And as noted above, the unification of U+2C70 with U+0241 causes needless difficulty for researchers who wish to use Athapascan Northwest Territories orthography side by side with IPA phonetic transcription.

Another area in which the addition of casing behaviour to U+0294 is that of case-sensitive searching. Users outside of the Northwest Territories will not expect to find U+0241 in their data, because their glottal stop does not case. But since the link between U+0294 and U+0241 is specified by the Unicode properties , data could be transformed by a casing operation, and users might fail to find words in a search which they would have been able to find before U+0241 was added to the standard with its link to U+0294.

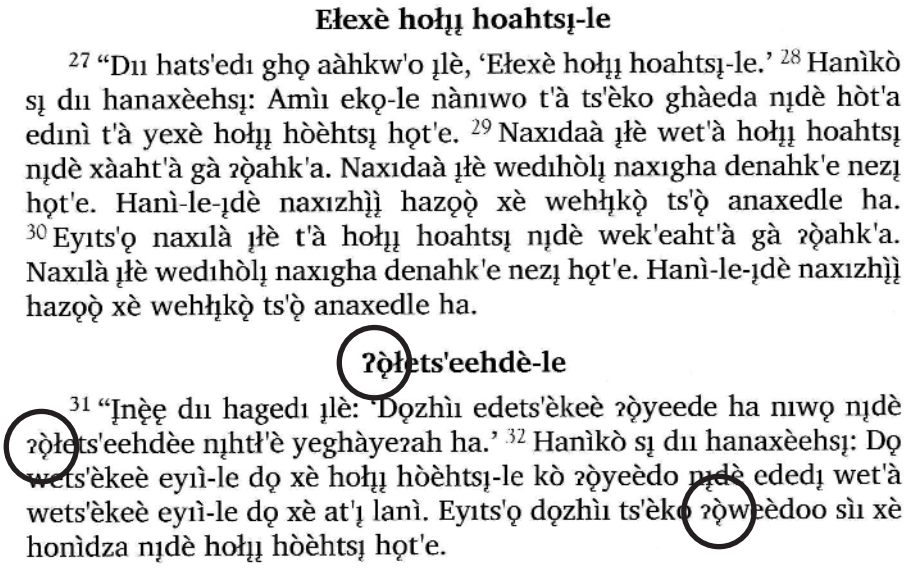
Adding U+2C70 would have little or no impact on existing Athapascan data. Most of the Northwest Territories material is probably encoded using SIL 8-bit fonts and fonts with PUA characters at present anyway, and it was SIL which originally requested both U+0241 and U+2C70.


Chipewyan Prayers

Edézı K'azenı

ʔeɬo(gh), Bı̨yeze chú, Yedáríye Nezų chú
bezí t'á. ʔotı̨é ʔekǫ́t'e nıdé.

**Figure 1.** Sample from Yellowknives Dene First Nation, 2000, showing Northwest Territories Chipewyan, where U+0241 LATIN CAPITAL LETTER GLOTTAL STOP and *U+0241 LATIN SMALL LETTER GLOTTAL STOP are clearly distinguished, with U+0241 in use in titlecasing.

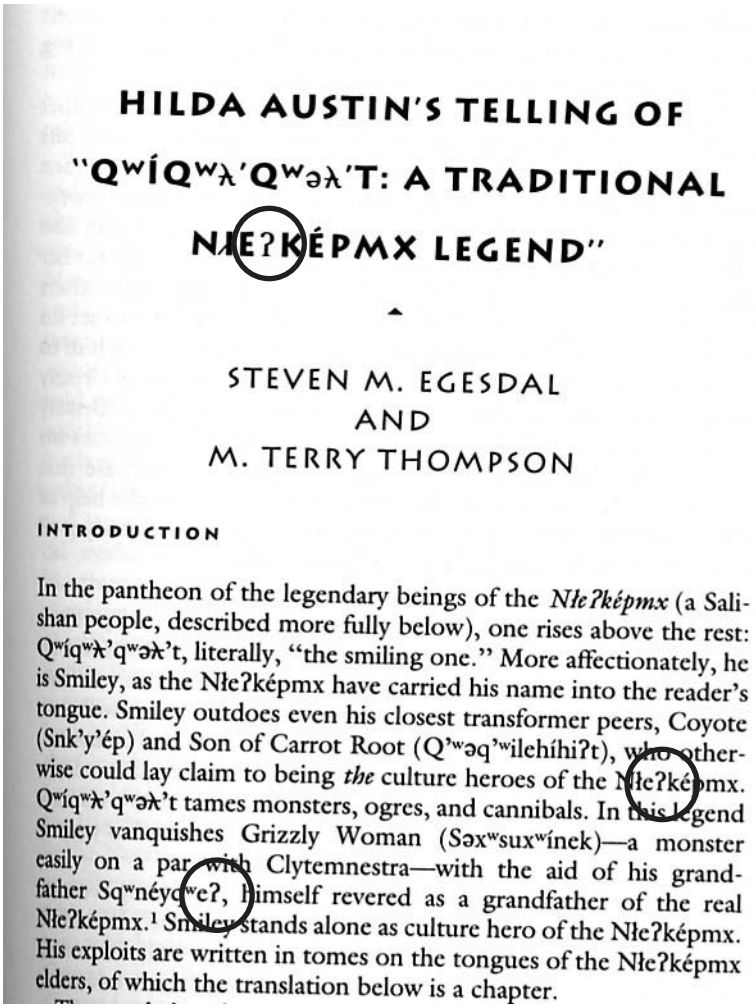

**Ełexè hoɬı̨ hoahtsı̨-le**

27 "Dıı hats'edı ghǫ aàhkw'o ı̨lè, 'Ełexè hoɬı̨ hoahtsı̨-le.' 28 Hanìkò sı̨ dıı hanaxèehsı̨: Amìı ekǫ-le nànıwo t'à ts'èko ghàeda nı̨dè hòt'a edınì t'à yexè hoɬı̨ hòèhtsı̨ hǫt'e. 29 Naxıdaà ı̨ɬè wet'à hoɬı̨ hoahtsı̨ nı̨dè xàaht'à gà ʔǫ̀ahk'a. Naxıdaà ı̨ɬè wedıhòlı̨ naxıgha denahk'e nezı̨ hǫt'e. Hanì-le-ı̨dè naxızhı̨̀ı̨ hazǫǫ̀ xè wehɬı̨kǫ̀ ts'ǫ̀ anaxedle ha. 30 Eyıts'ǫ naxılà ı̨ɬè t'à hoɬı̨ hoahtsı̨ nı̨dè wek'eaht'à gà ʔǫ̀ahk'a. Naxılà ı̨ɬè wedıhòlı̨ naxıgha denahk'e nezı̨ hǫt'e. Hanì-le-ı̨dè naxızhı̨̀ı̨ hazǫǫ̀ xè wehɬı̨kǫ̀ ts'ǫ̀ anaxedle ha.

**ʔǫ̀ɬets'eehdè-le**

31 "Ine̜e̜ dıı hagedı ı̨lè: 'Dǫzhìı edets'èkeè ʔǫ̀yeede ha nıwǫ nı̨dè ʔǫ̀ɬets'eehdèe nı̨htɬ'è yeghàyeʔah ha.' 32 Hanìkò sı̨ dıı hanaxèehsı̨: Dǫ wets'èkeè eyìı-le dǫ xè hoɬı̨ hòèhtsı̨-le kò ʔǫ̀yeèdo nı̨dè ededı̨ wet'à wets'èkeè eyìı-le dǫ xè at'ı̨ lanì. Eyıts'ǫ dǫzhìı ts'èko ʔǫ̀weèdoo sìı xè honìdza nı̨dè hoɬı̨ hòèhtsı̨ hǫt'e.

**Figure 2.** Sample from the Dogrib Translation Committee, 2003, showing part of the gospel of Matthew in Dogrib, where U+0241 LATIN CAPITAL LETTER GLOTTAL STOP and *U+0241 LATIN SMALL LETTER GLOTTAL STOP are clearly distinguished, with U+0241 in use in titlecasing.



HILDA AUSTIN'S TELLING OF "QʷÍQʷƛ'Qʷəƛ'T: A TRADITIONAL NŁE?KÉPMX LEGEND"

STEVEN M. EGESDAL
AND
M. TERRY THOMPSON

INTRODUCTION

In the pantheon of the legendary beings of the *Nłe?képmx* (a Salishan people, described more fully below), one rises above the rest: Qʷíqʷƛ'qʷəƛ't, literally, "the smiling one." More affectionately, he is Smiley, as the Nłe?képmx have carried his name into the reader's tongue. Smiley outdoes even his closest transformer peers, Coyote (Snk'y'ép) and Son of Carrot Root (Q'ʷəq'ʷilehíhi?t), who otherwise could lay claim to being *the* culture heroes of the Nłe?képmx. Qʷíqʷƛ'qʷəƛ't tames monsters, ogres, and cannibals. In this legend Smiley vanquishes Grizzly Woman (Səxʷsuxʷínek)—a monster easily on a par with Clytemnestra—with the aid of his grandfather Sqʷnéyqʷe?, himself revered as a grandfather of the real Nłe?képmx.[1] Smiley stands alone as culture hero of the Nłe?képmx. His exploits are written in tomes on the tongues of the Nłe?képmx elders, of which the translation below is a chapter.

**Figure 3.** Sample from Swann, 1994, in the Thompson (Nłe?kepmxcin) language, where U+0294 LATIN LETTER GLOTTAL STOP is used, without casing distinction, in ordinary text and in all caps.

**Bibliography.**


Dogrib Translation Committee. 2003. *Nǫ̀htsı̨ Nıı̨htł'è: Zezì wegǫ̀hłı̨ tł'axǫǫ̀ı̨*. Toronto: Canadian Bible Society. ISBN 0-88834-111-3.

Swann, Brian ed. 1994. *Coming to Light: Contemporary Translations of the Native Literature of North America*. Toronto: Random House.

Yellowknives Dene First Nation. 2000. *Dëne Yati ʔerehtł'ís / Dǫne Yati Enıı̨htł'è*. Yellowknife: Yellowknives Dene First Nation.

## A. Administrative

**1. Title**

Proposal to add LATIN SMALL LETTER GLOTTAL STOP to the UCS.

**2. Requester's name**

Canada (SCC) and Ireland (NSAI).

**3. Requester type (Member body/Liaison/Individual contribution)**

Member Body contribution.

**4. Submission date**

2005-08-08

**5. Requester's reference (if applicable)**

**6. Choose one of the following:**

**6a. This is a complete proposal**

Yes.

**6b. More information will be provided later**

No.

## B. Technical – General

**1. Choose one of the following:**

**1a. This proposal is for a new script (set of characters)**

No.

**Proposed name of script**

**1b. The proposal is for addition of character(s) to an existing block**

Yes.

**1b. Name of the existing block**

Latin Extended-C.

**2. Number of characters in proposal**

1

**3. Proposed category (see section II, Character Categories)**

Category A.

**4a. Proposed Level of Implementation (1, 2 or 3) (see clause 14, ISO/IEC 10646-1: 2000)**

Level 1.

**4b. Is a rationale provided for the choice?**

Yes.

**4c. If YES, reference**

Spacing letter.

**5a. Is a repertoire including character names provided?**

Yes.

**5b. If YES, are the names in accordance with the naming guidelines in Annex L of ISO/IEC 10646-1: 2000?**

Yes.

**5c. Are the character shapes attached in a legible form suitable for review?**

Yes.

**6a. Who will provide the appropriate computerized font (ordered preference: True Type, or PostScript format) for publishing the standard?**

Michael Everson. TrueType.

**6b. If available now, identify source(s) for the font (include address, e-mail, ftp-site, etc.) and indicate the tools used:**

Michael Everson. Fontographer.

**7a. Are references (to other character sets, dictionaries, descriptive texts etc.) provided?**

No.

**7b. Are published examples of use (such as samples from newspapers, magazines, or other sources) of proposed characters attached?**

No, but see N2789.

**8. Does the proposal address other aspects of character data processing (if applicable) such as input, presentation, sorting, searching, indexing, transliteration etc. (if yes please enclose information)?**

Yes, casing behaviour is addressed.

**9. 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.**

Functions like other Latin letters.

## C. Technical – Justification

**1. Has this proposal for addition of character(s) been submitted before? If YES, explain.**

Not to WG2.

**2a. Has contact been made to members of the user community (for example: National Body, user groups of the script or characters, other experts, etc.)?**

Yes.

**2b. If YES, with whom?**

Chris Harvey (languagegeek.com) is in contact with these communities.

**2c. If YES, available relevant documents**

**3. Information on the user community for the proposed characters (for example: size, demographics, information technology use, or publishing use) is included?**

No.

**4a. The context of use for the proposed characters (type of use; common or rare)**

This character is used as an orthographic character in Athapascan languages of the Northwest Territories.

**4b. Reference**

**5a. Are the proposed characters in current use by the user community?**

Yes.

**5b. If YES, where?**

In Aboriginal communities in Canada.

**6a. After giving due considerations to the principles in Principles and Procedures document (a WG 2 standing document) must the proposed characters be entirely in the BMP?**

Yes.

**6b. If YES, is a rationale provided?**

Yes.

**6c. If YES, reference**

Keep with other Latin characters.

**7. Should the proposed characters be kept together in a contiguous range (rather than being scattered)?**

N/A.

**8a. Can any of the proposed characters be considered a presentation form of an existing character or character sequence?**

No.

**8b. If YES, is a rationale for its inclusion provided?**

**8c. If YES, reference**

**9a. Can any of the proposed characters be encoded using a composed character sequence of either existing characters or other proposed characters?**

Yes.

**9b. If YES, is a rationale for its inclusion provided?**

Yes.

**9c. If YES, reference**

See above. Both this character and U+2041 are derived from U+2094.

**10a. Can any of the proposed character(s) be considered to be similar (in appearance or function) to an existing character?**

No.

**10b. If YES, is a rationale for its inclusion provided?**

**10c. If YES, reference**

**11a. Does the proposal include use of combining characters and/or use of composite sequences (see clauses 4.12 and 4.14 in ISO/IEC 10646-1: 2000)?**

No.

**11b. If YES, is a rationale for such use provided?**

**11c. If YES, reference**

**12a. Is a list of composite sequences and their corresponding glyph images (graphic symbols) provided?**

No.

**12b. If YES, reference**

**13a. Does the proposal contain characters with any special properties such as control function or similar semantics?**

No.

**13b. If YES, describe in detail (include attachment if necessary)**

**14a. Does the proposal contain any Ideographic compatibility character(s)?**

No.

**14b. If YES, is the equivalent corresponding unified ideographic character(s) identified?**

**14c. If YES, reference**