

ISO/IEC JTC 1/SC 2/WG 2
PROPOSAL SUMMARY FORM TO ACCOMPANY SUBMISSIONS
FOR ADDITIONS TO THE REPERTOIRE OF ISO/IEC 10646¹

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: Revised Proposal to Encode Orthographic Glottal Stops in the UCS

2. Requester's name: SIL International (contact: Jonathan Kew), Peter Constable

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

4. Submission date: 2004-06-08

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 B is suggested

2. Number of characters in proposal: 1

3. Proposed category (select one from below - see section 2.2 of P&P document):
A-Contemporary X B.1-Specialized (small collection) _____ 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. Proposed Level of Implementation (1, 2 or 3) (see Annex K in P&P document): 1
Is a rationale provided for the choice? Yes
If Yes, reference: no combining marks in proposal

5. 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

6. Who will provide the appropriate computerized font (ordered preference: True Type, or PostScript format) for publishing the standard? SIL International
If available now, identify source(s) for the font (include address, e-mail, ftp-site, etc.) and indicate the tools used: _____

7. 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

8. 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)?
Yes: suggested character properties included

9. 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.

¹ Form number: N2652-F (Original 1994-10-14; Revised 1995-01, 1995-04, 1996-04, 1996-08, 1999-03, 2001-05, 2001-09, 2003-11)

C. Technical - Justification

1. Has this proposal for addition of character(s) been submitted before?	<u>Yes</u>
If YES explain	<u>Previous version submitted (L2/04-065) for UTC#98, but not discussed in that meeting due to lack of time. Proposal has been revised to respond to anticipated objections.</u>
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.)?	<u>Yes</u>
If YES, with whom?	<u>Linguists working with the user communities, and other representatives of the communities</u>
If YES, available relevant documents:	<u>See information provided below.</u>
3. Information on the user community for the proposed characters (for example: size, demographics, information technology use, or publishing use) is included?	<u>Yes</u>
Reference:	<u>Dogrib, Chipewyan, North and South Slavey are Athapaskan languages spoken in northwest Canada.</u>
4. The context of use for the proposed characters (type of use; common or rare)	<u>common</u>
Reference:	<u>see information provided below</u>
5. Are the proposed characters in current use by the user community?	<u>Yes</u>
If YES, where? Reference:	<u>In communities cited above. See additional information provided below.</u>
6. After giving due considerations to the principles in the P&P document must the proposed characters be entirely in the BMP?	<u>Preferably, yes</u>
If YES, is a rationale provided?	<u>Yes</u>
If YES, reference:	<u>If possible, should be kept with other Latin characters in the BMP.</u>
7. Should the proposed characters be kept together in a contiguous range (rather than being scattered)?	<u>N/A</u>
8. Can any of the proposed characters be considered a presentation form of an existing character or character sequence?	<u>No</u>
If YES, is a rationale for its inclusion provided?	<u>N/A</u>
If YES, reference:	<u>N/A</u>
9. Can any of the proposed characters be encoded using a composed character sequence of either existing characters or other proposed characters?	<u>No</u>
If YES, is a rationale for its inclusion provided?	<u>N/A</u>
If YES, reference:	<u>N/A</u>
10. Can any of the proposed character(s) be considered to be similar (in appearance or function) to an existing character?	<u>Yes</u>
If YES, is a rationale for its inclusion provided?	<u>Yes</u>
If YES, reference:	<u>The character U+0294 LATIN LETTER GLOTTAL STOP has the same appearance in some uses as the proposed character LATIN CAPITAL LETTER GLOTTAL STOP, but is case pair to that character, and the two are given distinct appearances when used together in an orthography.</u>
11. Does the proposal include use of combining characters and/or use of composite sequences?	<u>No</u>
If YES, is a rationale for such use provided?	<u>N/A</u>
If YES, reference:	<u>N/A</u>
Is a list of composite sequences and their corresponding glyph images (graphic symbols) provided?	<u>N/A</u>
If YES, reference:	<u>N/A</u>
12. Does the proposal contain characters with any special properties such as control function or similar semantics?	<u>No</u>
If YES, describe in detail (include attachment if necessary)	<u>N/A</u>
13. Does the proposal contain any Ideographic compatibility character(s)?	<u>No</u>
If YES, is the equivalent corresponding unified ideographic character(s) identified?	<u>N/A</u>
If YES, reference:	<u>N/A</u>

D Proposed Characters

One character is proposed: general category and case mapping properties are as shown:

Glyph	Name	Gen. Cat.	Properties
ʔ	LATIN CAPITAL LETTER GLOTTAL STOP	Lu	lowercase map = U+0294 LATIN LETTER GLOTTAL STOP

Other properties should match those of similar characters, such as U+0041 LATIN CAPITAL LETTER A.

It is further proposed that the case mapping properties of the existing character U+0294 be revised as follows:

uppercase map = LATIN CAPITAL LETTER GLOTTAL STOP
titlecase map = LATIN CAPITAL LETTER GLOTTAL STOP

Also, it is suggested that annotations be added to the new and existing character as follows (assuming *nnnn* as the code point for the new character):

0294 * IPA and other phonetic notation (technical notation)
* Chipewyan, Dogrib, Slavey (orthography)
* different glyphs may be required for orthographic use than for technical notations
x 02C0 modifier letter glottal stop
x *nnnn* latin capital letter glottal stop

nnnn * Chipewyan, Dogrib, Slavey (orthography)
x 0294 latin letter glottal stop

E Other Information

E.1 Rationale

The Chipewyan, Dogrib and Slavey languages are Athabaskan languages spoken in northwest Canada. These languages have phonemic glottal stop and use the glottal stop character orthographically. Moreover, the orthographies of these languages have an orthographic case distinction between upper- and lower-case glottal stop.

The character **ʔ**, called 'glottal' or 'glottal stop', represents a sound like what we hear in the middle of the English expression "oh-oh". In Dogrib this sound is an ordinary consonant. It is found in many words of all types:

ʔoo	spruce boughs
ʔihdaa	jackfish
seʔeè	my jacket
wəʔò	beyond it
nāʔeeli	he or she is sewing
nīʔo	it arrived
k'eʔa	(animals) are roaming

Figure 1. Small glottal stop; from Whaèhdòò Nàowòò Kò (2002), p. 173.

ʔeehgòtìsoa	'Small Clear Lake'
ʔehgòtìsoa	'Small Elbow Lake'
Ts'ʔehdaà	'Spruce Point'
Ts'iedaa	'Living Spruce'

Figure 2. Capital (blue highlight) and small (red highlight) glottal stops; from Whaèhdòò Nàowòò Kò (2002), p. 82.

ʔihdaak'èatiʔetsìììì
ʔihdaak'èatiʔechìììì

Figure 3. Capital (blue highlight) and small (red highlight) glottal stops; from Whaèhdòò Nàowòò Kò (2002), p. 90.

Chìà tʔi k'e dawheda ts'ʔo nàhdò hòt'e.

ʔasiì wìzì whenehtà nì le.

Figure 4. Capital (blue highlight) and small (red highlight) glottal stops; from Koyina (1983).



Tatsòga dekò ts'ò naàhtla kò, k'òòhdzò
 Ekw'ò-ehstì ts'ì k'e dawheda. Ekw'ò-ehstì
 wheze xè hadì, "ʔasì ehʔì."

Figure 5. Capital (blue highlight) and small (red highlight) glottal stops; from O'Rouche (1987).

E.2 Relationship to U+0294 LATIN LETTER GLOTTAL STOP

In attested usage, the uppercase glottal stop used in Athapaskan languages has the same appearance that is generally found for existing uses of the character U+0294 LATIN LETTER GLOTTAL STOP, in particular, with how that character is used in phonetic transcription. There is potential for confusion between different characters based on confusion of shapes. There are three potential solutions to address orthographic needs:

1. Change the case of U+0294 to uppercase (i.e. change the general category property to Lu), and add a new lowercase character.
2. Use U+0294 as the orthographic lowercase, and add a new character for the capital; different glyphs would be used for U+0294 in orthographic usage and phonetic transcription (the current proposal).
3. Add two new characters for orthographic use, LATIN CAPITAL LETTER GLOTTAL STOP and LATIN SMALL LETTER GLOTTAL STOP; the existing character would be used for phonetic transcription, but not for orthographic uses.

The first of these alternatives is problematic in that changing the case of an existing character can create serious problems for implementations such as domain-name protocols. It is herewith rejected as a possibility.

The second of these alternatives has the problem that the new character would have a cap-height glyph, which is what is used as the representative glyph for U+0294. The ambiguity of

one glyph being used for two characters while also one of those characters having a distinct glyph (though not shown as the representative glyph) could be confusing both for users of phonetic transcription and for users of Athapaskan orthographies, with inconsistency in usage resulting.

The third is problematic in that a second lowercase character is added (U+0294 has a general category of Ll and so is considered to be lowercase) in order to provide a glyph differentiation, and so can be seen as a violation of the character/glyph model. There is also potential for confusion among users between the two lowercase characters, which could result in inconsistency in usage.

There is, therefore, no perfect solution. In hindsight, it might have been preferable to have created the existing character U+0294 as an uppercase character, but there was no way to anticipate future needs (orthographic usage may have come to light in which the representative glyph for U+0294 was, in fact, used for the lowercase, with a wider counterpart used for uppercase), and that decision could not be reversed now as it could have seriously detrimental effects for some existing implementations. A choice must be made, then, between the two other alternatives.

Since both have some potential for confusion among users while only one can be construed as a violation of a UCS design principle, it is considered preferable to choose the alternative that does not violate that principle: that one new character be added, LATIN CAPITAL LETTER GLOTTAL STOP, and that the different glyphs that may be required for U+0294 be documented, which can easily be done using annotations in the names list.

F References

Koyina, Laiza. 1983. *Dq weda goòle xè Teèt'o si. (The Blind Man and the Loon.)* Yellowknife, NWT, Canada: Northwest Territories Department of Education.

O'Rouche, Robert (ed.) 1987. *Edàni nqge done gok'èjdi (How the Fox helped the People).* Yellowknife, NWT, Canada: Northwest Territories Department of Education.

Whaèhdqè Nàowò Kè, Dogrib Treaty 11 Council. 2002. "Dogrib knowledge on placenames, caribou and habitat: Final report." Yellowknife, NWT, Canada: West Kitikmeot Slave Study Society.

Available online at:

http://www.wkss.nt.ca/HTML/08_ProjectsReports/PDF/DogribPlaceCaribouHabitat2002.pdf.