

Proposal to Encode Orthographic Glottal Stops in the UCS

Date: 2004-02-01

Author: Peter Constable, Microsoft

Address: One Microsoft Way
Redmond, WA 98052
USA

Tel: +1 425 722 1867

Email: petercon@microsoft.com

A. Administrative

- | | |
|--------------------------------------|--|
| 1. Title | Proposal to Encode Orthographic Glottal Stops in the UCS |
| 2. Requester's name | SIL International (contact: Jonathan Kew), Peter Constable |
| 3. Requester type | Expert contribution |
| 4. Submission date | 2004-02-01 |
| 5. Requester's reference | |
| 6a. Completion | This is a complete proposal |
| 6b. More information to be provided? | Only as required for clarification. |
-

B. Technical—General

- | | |
|---|-------------------------------------|
| 1a. New Script? Name? | No |
| 1b. Addition of characters to existing block? Name? | Yes — Latin Extended B is suggested |
| 2. Number of characters in proposal | 2 |
| 3. Proposed category | A |
| 4. Proposed level of implementation and rationale | 1 (no combining marks) |
| 5a. Character names included in proposal? | Yes |
| 5b. Character names in accordance with guidelines? | Yes |
| 5c. Character shapes reviewable? | Yes |
| 6a. Who will provide computerized font? | SIL International |

6b. Font currently available?	Yes
6c. Font format?	TrueType
7a. Are references (to other character sets, dictionaries, descriptive texts, etc.) provided?	Yes
7b. Are published examples (such as samples from newspapers, magazines, or other sources) of use of proposed characters attached?	Yes
8. Does the proposal address other aspects of character data processing?	Yes, suggested character properties are included (see section D).

C. Technical—Justification

1. Has this proposal for addition of character(s) been submitted before?	No
2a. Has contact been made to members of the user community?	Yes
2b. With whom?	Linguists working with the user communities, and other representatives of the communities
3. Information on the user community for the proposed characters is included?	Dogrib, Chipewyan, North and South Slavey are Athapaskan languages spoken in northwest Canada.
4. The context of use for the proposed characters	general orthographic usage
5. Are the proposed characters in current use by the user community?	Yes
6a. Must the proposed characters be entirely in the BMP?	Preferably
6b. Rationale?	If possible, should be kept with other Latin characters in the BMP.
7. Should the proposed characters be kept together in a contiguous range?	Yes
8a. Can any of the proposed characters be considered a presentation form of an existing character or character sequence?	No
8b. Rationale for inclusion?	N/A

- 9a. Can any of the proposed characters be considered to be similar (in appearance or function) to an existing character? The character LATIN CAPITAL LETTER GLOTTAL STOP has the same appearance as U+0294 LATIN LETTER GLOTTAL STOP, but has different character properties.
- 9b. Rationale for inclusion? Distinct characters (see the discussion in §E.2).
10. Does the proposal include the use of combining characters and/or use of composite sequences? No.
11. Does the proposal contain characters with any special properties? No.
-

D. Proposed Characters

Two characters are proposed: general category and case mapping properties are as shown:

Glyph	Name	Gen. Cat.	Properties
ʔ	LATIN CAPITAL LETTER GLOTTAL STOP	Lu	lowercase map = LATIN SMALL LETTER GLOTTAL STOP
ʔ	LATIN SMALL LETTER GLOTTAL STOP	Ll	uppercase map = LATIN CAPITAL LETTER GLOTTAL STOP

Other properties for these remaining characters should match those of similar characters, such as U+0061 LATIN SMALL LETTER A.

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
ʔ hdaa	jackfish
se ʔ eè	my jacket
we ʔ òò	beyond it
nà ʔ eeli	he or she is sewing
nì ʔ ò	it arrived
k' ʔ erà	(animals) are roaming

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

ʔ ehgotitsoa	'Small Clear Lake'
ʔ ehgotitsoa	'Small Elbow Lake'
Ts' ʔ ehdaà	'Spruce Point'
Ts' ʔ edaa	'Living Spruce'

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

ʔ hdaak'èatì ʔ etsìììì
ʔ hdaak'èatì ʔ echìììì

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

Chìà tì'i k'e dawheda ts'**ʔ**ò nàhdò hòt'e.

ʔasii 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

The uppercase glottal stop used in Athapaskan languages is visually identical with the existing character U+0294 LATIN LETTER GLOTTAL STOP. It might be supposed, then, that U+0294 could be used orthographically for these languages, and that only one new character is required. The problem, however, is that U+0294 has a general category property of Lm, lowercase letter, and must stand in a case relationship with a case-paired character used in the orthographies of these languages. Two possible solutions involving addition of only one new character exist:

1. 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.
2. Change the case of U+0294 to uppercase (i.e. change the general category property to Lu), and add a new lowercase character.

The first 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. There would be considerable confusion both for users of phonetic transcription and for users of Athapaskan orthographies, and inconsistency in usage would result.

The second of these alternatives is problematic in that changing the case of an existing character can create problems for implementations such as domain-name protocols.

It is considered preferable, then, that two new characters be added, and that the intended purposes of U+0294 and the visually-identical LATIN CAPITAL LETTER GLOTTAL STOP be documented, which can easily be done using annotations in the names list.

F. References

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

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

Whaèhdòè Nàowòè Kò, 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.