Proposal to encode *Gujarati Sign Triple Nukta*

Vinodh Rajan
vrs3@st-andrews.ac.uk

Introduction

Avestan is the liturgical language of the Zoroastrian religion. With the conquest of Persia by the Arabs, a part of the Zoroastrian community sought refuge in the Gujarat region of India circa 8th century CE. The migrant community consequently carried the Zoroastrian texts with them into the Indian subcontinent.

This migrant Persian Zoroastrian community was subsequently Indianized across several generations adopting Gujarati as their native language, becoming what would be known as the "Parsi" community of India. However, scholars and priests preserved the Avestan language and the script for liturgical purposes.

Transliteration of the Avestan Script

When Avestan texts were printed for the consumption of the lay community, it was necessary to transliterate the Avestan texts into the native script of the lay Parsi Zoroastrians i.e. Gujarati. In order to preserve the Avestan specific letters in the native Indic script, special characters had to be invented for transliteration.

Extending the native Gujarati phonemes with the Nukta represented most of the unique Avestan consonants that weren't in the Gujarati script.

၆	TTE	ત	TA + NUKTA
૩	NGE	ઁઘ	ANUSVARA+GHA+NUKTA
∫	ZE	જ઼	JA+NUKTA
ﲲ	KHVE	ખ઼	KHA+NUKTA

(Particularly, the position of Nukta in TA must be noted. This is merely a stylistic choice and may be considered by any Gujarati font that is to be specifically used for Avestan transliteration)

Avestan Letter ZHA

Of particular importance is the Avestan letter ZHE 𐬲, which represents the *voiced palato-alveolar fricative /ʒ/*. The closest letter in the Gujarati script for this phoneme is JA જ. However, JA had already been extended by the Nukta to represent the Avestan ZE ʃ (*voiced dental sibilant*). Therefore, a new nukta character was invented for transliterating the Avestan ZHE.

Gujarati Sign Triple Nukta

The new Nukta character had three dots as opposed to the native Gujarati Nukta which had a single dot. It was probably inspired from the Perso-arabic letter JEH ژ used in Urdu orthography for the same phoneme in the Hindustani language.

<p align="center">𐬲   ZHE   જ⃨   JA + TRIPLE NUKTA</p>

It is proposed that this new Nukta character be named as "GUJARATI SIGN TRIPLE NUKTA" and encoded in the Gujarati range of the UCS.

Character to be encoded

OACE   ◌⃛   GUJARATI SIGN TRIPLE NUKTA

Character Properties

```
0ACE;GUJARATI SIGN TRIPLE NUKTA;Mn;7;NSM;;;;;N;;;;;
```

Annotation

The character is to be annotated as "*used in transliteration of Avestan*"

Encoding *Gujarati Letter ZHA* Vs. *Gujarati Sign Triple Nukta*

DEVANAGARI LETTER ZHA ॹ was encoded in the UCS as an atomic code point at U+0979 to transliterate the Avestan ZHE. This was based on K.E. Kanga's Avestan publications in Devanagari script. A parallel solution in Gujarati would have been the encoding of the entire composite character ઝ as GUJARATI LETTER ZHA.

However, it would be more appropriate to consider ઝ as a composite character rather than an atomic one. This would be inline with other characters used for Avestan transliteration such as ઝ, પ, ધ & વ, which are all nukta-based composite characters. Additionally, it also enhances the script's ability to express additional phonemes using the new diacritic *Triple Nukta*.

Therefore, it is highly recommended to encode the diacritic rather than the composite character in the Gujarati block.
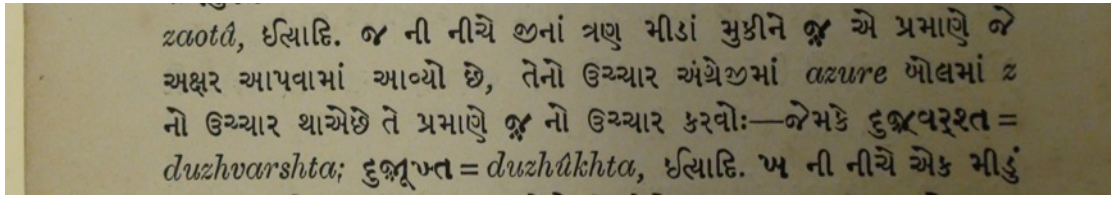
Retrospectively, a better decision would have had been the encoding of a Devanagari *Triple Nukta* diacritic rather than the complete atomic character ZHA in the Devanagari block.
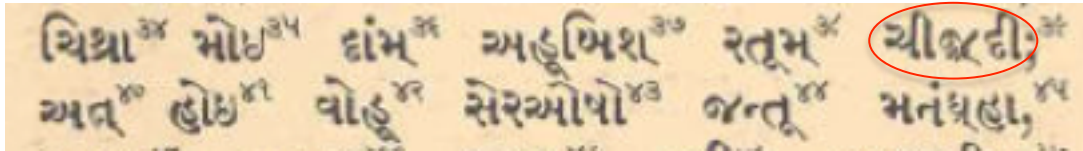
Usage Samples

From K.E. Kanga (1936):
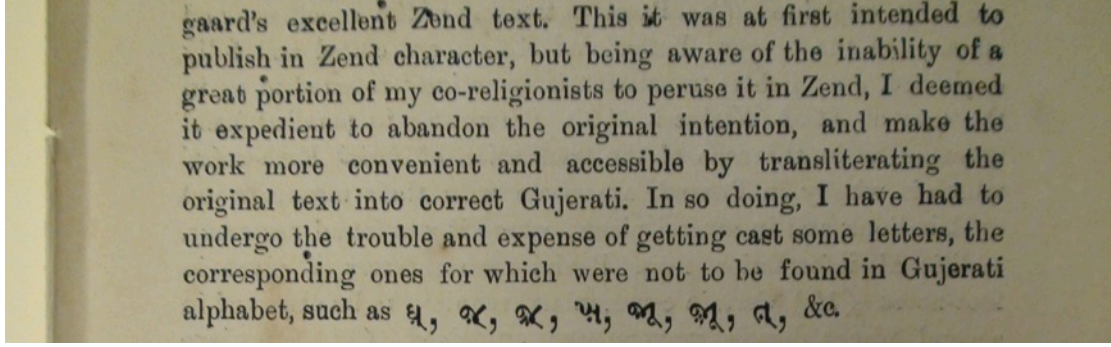
Pg 27 (Preface)

Pg 32 (Preface)

(ઋૂ ૐૂ are formed instead of ૐૂ ૐૂ with the *Vowel Sign UU* probably to avoid overlapping of the vowel sign with the Nuktas in ૐ & ૐ. A font may either implement this or adjust the placement of the vowel signs to avoid overlapping with the Nukta signs.)

Pg 5 (Contents)

ﺣﺠﺴﺮﺪ is transliterated into Gujarati as ચીૐદી

Reference

1. Kanga, K.E. (1936). *Khordeh Avestâ*. Navroji Postonji Kavasji Kanga, Bombay.
   http://www.avesta.org/kanga/Khorda_Avesta_Kanga_Gujarati.pdf
2. Skjærvø, P.O. (2003). *An Introduction to Young Avestan.*

<div style="border:1px solid black; text-align:center">

ISO/IEC JTC 1/SC 2/WG 2
PROPOSAL SUMMARY FORM

</div>

A. Administrative

1. Title:
Proposal to encode Gujarati Sign Triple Nukta
2. Requester's name:
Vinodh Rajan
3. Requester type (Member body/Liaison/Individual contribution):
Individual Contribution
4. Submission date:
15th April 2013
5. Requester's reference (if applicable):
6. (Choose one of the following:)
This is a complete Proposal

B. Technical - General

1. (Choose one of the following:)
a. This proposal is for a new script (set of characters):
No
b. The proposal is for addition of character(s) to an existing block:
Yes. Gujarati block
2. Number of characters in proposal:
One
3. Proposed category
Category B1 (Specialized Small)
4. Is a repertoire including character names provided?:
Yes
a. If YES, are the names in accordance with the 'character naming guidelines' in Annex K of ISO/IEC 10646-1?
Yes
b. Are the character shapes attached in a reviewable form?
Yes
6. Who will provide the appropriate computerized font (ordered preference: True Type, PostScript or 96x96
   bit-mapped format) for publishing the standard?
 Vinodh Rajan
 If available now, identify source(s) for the font (include address, e-mail, ftp-site, etc.) and indicate the tools
 used:
Vinodh Rajan, Fontforge
7. References:
a. Are references (to other character sets, dictionaries, descriptive texts etc.) provided?
Yes
b. Are published examples (such as samples from newspapers, magazines, or other sources) of use 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. It also discusses the transliteration of Avestan script to Gujarati script.

C. Technical - Justification

1. 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?
Joseph H Peterson, **Webmaster of http://www.avesta.org. The proposal has also been circulated in the Zoroastrian mailing lists by him**
If YES, available relevant documents?
The communication was through emails
2. Information on the user community for the proposed characters (for example: size, demographics, information technology use, or publishing use) is included?
Parsi Zoroastrians to print their religious Avestan texts in transliterated Gujarati script
3. The context of use for the proposed characters (type of use; common or rare)
Rare
4. Are the proposed characters in current use by the user community?
Yes
If YES, where?
Avestan texts printed in Gujarati script
5. After giving due considerations to the principles in N 1352 must the proposed characters be entirely in the BMP?
Yes
If YES, is a rationale provided?
It is an addition to the Gujarati block encoded in the BMP
If YES, reference:
6. Should the proposed characters be kept together in a contiguous range (rather than being scattered)?
N/A. A single character is proposed
7. Can any of the proposed characters be considered a presentation form of an existing character or character sequence?
No
8. Can any of the proposed character(s) be considered to be similar (in appearance or function) to an existing character?
Yes. There are several triangular dot characters in Unicode such as SYRIAC THREE DOTS ABOVE, HEBREW ACCENT SEGOL etc.
If YES, is a rationale for its inclusion provided?
It is a combining sign specific to the Gujarati script and also has a different function.
If YES, reference:
9. Does the proposal include use of combining characters and/or use of composite sequences (see clause 4.11 and 4.13 in ISO/IEC 10646-1)?
Yes
If YES, is a rationale for such use provided?
Yes
If YES, reference:
Last section of the proposal discusses this
Is a list of composite sequences and their corresponding glyph images (graphic symbols) provided?
Yes. The composite glyph with GUJARATI LETTER JA is provided. The placement of the sign follows the similar position for other consonants.
If YES, reference:
Inline in the proposal
10. Does the proposal contain characters with any special properties such as control function or similar semantics?
No