

Gujarati Script LGR Proposal

Introduction, Current Analysis and Next Steps

Akshat Joshi

NBGP F2F Meeting, Colombo
13th December 2017



Agenda

1

Introduction to
Gujarati Script

2

Classification of
Characters

3

Repertoire Analysis

4

Within & Cross-
Script Variants

5

WLE Rules

6

Current Status and
Next Steps for
Completion

Introduction to Gujarati Script

- ⦿ **Population:** With 54.45 million speakers in India and 65.5 million worldwide, Gujarati is the 26th most spoken native language in the world.
- ⦿ **Geographical area:** Outside the Indian state of Gujarat, Gujarati is extensively spoken in parts of Africa, Madagascar, UK and the USA as well as emigrant communities around the world.
- ⦿ **Languages written in Gujarati script:** Gujarati, and 1 EGIDS 4 language and 11 EGIDS 5 languages.
- ⦿ Gujarati is a variant of Devanāgarī, the main difference being the absence of the shirorekha or the line above the character and also more rounded shapes. Since initially it was used for commercial ends, it has been referred to as śarāphi (banker's) or mahājani (trader's) script.

Classification of Characters: Consonants

Gujarati consonants have an implicit schwa /ə/ included in them. As per traditional classification they are categorized according to their phonetic properties. There are 5 Varga groups (classes) and one non-Varga group. Each Varga, which corresponds to Stops, contains five consonants classified as per their properties. The first four consonants are classified on the basis of Voicing and Aspiration and the last is the corresponding nasal.

Non Varga	ચ	ર	લ	ળ	વ
------------------	---	---	---	---	---

Table 1: Non-Varga consonants

Varga	Unvoiced		Voiced		Nasal
	_Asp	+Asp	_Asp	+Asp	
Velar	ક	ખ	ગ	ઘ	ઙ
Palatal	ચ	છ	જ	ઝ	ઞ
Retroflex	ટ	ઠ	ડ	ઢ	ણ
Dental	ત	થ	દ	ધ	ન
Bi-labial	પ	ફ	બ	ભ	મ

Table 2: Varga classification of consonants

Classification of Characters: Vowels

Separate symbols exist for all Vowels, which are either pronounced independently at the beginning or attached to a consonant. To indicate the latter (other than the implicit one), a Vowel modifier (Matra) is attached to the consonant. Since the consonant has a built in schwa, there are equivalent Matras for all vowels excepting the અ. The correlation is shown as under:

અ	આ	ઇ	ઈ	ઉ	ઊ	ઋ	એ	ઐ	ઓ	ઔ
	ા	િ	ી	ુ	ૂ	ૃ	ે	ૈ	ો	ૌ

In addition to show sounds borrowed from English, Gujarati admits two vowels and their corresponding Matras as in Table 4.1: “Borrowed” Vowels with corresponding Matras as in એટલેન્ટિક /Atlantic/ ઓર /or/

એ	ઓ
ે	ૌ

Repertoire Included - 1

Sr. No	Unicode Code Point	Glyph	Character Name	Unicode General Category (gc)	Indic Syllabic Category	Reference	Y/N
1	0A82	◌̣	GUJARATI SIGN ANUSVARA	Mn	Anusvara (Bindu)	[Omniglot]	
2	0A83	◌̣:	GUJARATI SIGN VISARGA	Mc	Visarga	[Omniglot]	
3	0A85	અ	GUJARATI LETTER A	Lo	Vowel	[Omniglot]	
4	0A86	અા	GUJARATI LETTER AA	Lo	Vowel	[Omniglot]	
5	0A87	ઇ	GUJARATI LETTER I	Lo	Vowel	[Omniglot]	
6	0A88	ઈ	GUJARATI LETTER II	Lo	Vowel	[Omniglot]	
7	0A89	ઉ	GUJARATI LETTER U	Lo	Vowel	[Omniglot]	
8	0A8A	ઉા	GUJARATI LETTER UU	Lo	Vowel	[Omniglot]	
9	0A8B	૨	GUJARATI LETTER VOCALIC R	Lo	Vowel	[Omniglot]	
10	0A8C	૨્	GUJARATI LETTER VOCALIC L	Lo	Vowel	[Omniglot]	

Repertoire Included - 2

Sr. No.	Unicode Code Point	Glyph	Character Name	Unicode General Category (gc)	Indic Syllabic Category	Reference	Y/N
11	0A8D	એ	GUJARATI VOWEL CANDRA E	Lo	Vowel	[Omniglot]	
12	0A8F	એ	GUJARATI LETTER E	Lo	Vowel	[Omniglot]	
13	0A90	ઐ	GUJARATI LETTER AI	Lo	Vowel	[Omniglot]	
14	0A91	ઓ	GUJARATI VOWEL CANDRA O	Lo	Vowel	[Omniglot]	
15	0A93	ઓ	GUJARATI LETTER O	Lo	Vowel	[Omniglot]	
16	0A94	ઔ	GUJARATI LETTER AU	Lo	Vowel	[Omniglot]	
17	0A95	ક	GUJARATI LETTER KA	Lo	Consonant	[Omniglot]	
18	0A96	ખ	GUJARATI LETTER KHA	Lo	Consonant	[Omniglot]	
19	0A97	ગ	GUJARATI LETTER GA	Lo	Consonant	[Omniglot]	
20	0A98	ઘ	GUJARATI LETTER GHA	Lo	Consonant	[Omniglot]	

Repertoire Included - 3

Sr. No	Unicode Code Point	Glyph	Character Name	Unicode General Category (gc)	Indic Syllabic Category	Reference	Y/N
21	0A99	ઠ	GUJARATI LETTER NGA	Lo	Consonant	[Omniglot]	
22	0A9A	ઘ	GUJARATI LETTER CA	Lo	Consonant	[Omniglot]	
23	0A9B	ઙ	GUJARATI LETTER CHA	Lo	Consonant	[Omniglot]	
24	0A9C	ઞ	GUJARATI LETTER JA	Lo	Consonant	[Omniglot]	
25	0A9D	ઝ	GUJARATI LETTER JHA	Lo	Consonant	[Omniglot]	
26	0A9E	ઞ	GUJARATI LETTER NYA	Lo	Consonant	[Omniglot]	
27	0A9F	ટ	GUJARATI LETTER TTA	Lo	Consonant	[Omniglot]	
28	0AA0	ઠ	GUJARATI LETTER TTHA	Lo	Consonant	[Omniglot]	
29	0AA1	ડ	GUJARATI LETTER DDA	Lo	Consonant	[Omniglot]	
30	0AA2	ઢ	GUJARATI LETTER DDHA	Lo	Consonant	[Omniglot]	

Repertoire Included - 4

Sr. No.	Unicode Code Point	Glyph	Character Name	Unicode General Category (gc)	Indic Syllabic Category	Reference	Y/N
31.	0AA3	ણ	GUJARATI LETTER NNA	Lo	Consonant	[Omniglot]	
32.	0AA4	ત	GUJARATI LETTER TA	Lo	Consonant	[Omniglot]	
33.	0AA5	થ	GUJARATI LETTER THA	Lo	Consonant	[Omniglot]	
34.	0AA6	દ	GUJARATI LETTER DA	Lo	Consonant	[Omniglot]	
35.	0AA7	ધ	GUJARATI LETTER DHA	Lo	Consonant	[Omniglot]	
36.	0AA8	ન	GUJARATI LETTER NA	Lo	Consonant	[Omniglot]	
37.	0AAA	પ	GUJARATI LETTER PA	Lo	Consonant	[Omniglot]	
38.	0AAB	ફ	GUJARATI LETTER PHA	Lo	Consonant	[Omniglot]	
39.	0AAC	બ	GUJARATI LETTER BA	Lo	Consonant	[Omniglot]	
40.	0AAD	ભ	GUJARATI LETTER BHA	Lo	Consonant	[Omniglot]	

Repertoire Included - 5

Sr. No.	Unicode Code Point	Glyph	Character Name	Unicode General Category (gc)	Indic Syllabic Category	Reference	Y/N
41.	0AAE	મ	GUJARATI LETTER MA	Lo	Consonant	[Omniglot]	
42.	0AAF	ય	GUJARATI LETTER YA	Lo	Consonant	[Omniglot]	
43.	0AB0	ર	GUJARATI LETTER RA	Lo	Consonant	[Omniglot]	
44.	0AB2	લ	GUJARATI LETTER LA	Lo	Consonant	[Omniglot]	
45.	0AB3	ળ	GUJARATI LETTER LLA	Lo	Consonant	[Omniglot]	
46.	0AB5	વ	GUJARATI LETTER VA	Lo	Consonant	[Omniglot]	
47.	0AB6	શ	GUJARATI LETTER SHA	Lo	Consonant	[Omniglot]	
48.	0AB7	ષ	GUJARATI LETTER SSA	Lo	Consonant	[Omniglot]	
49.	0AB8	સ	GUJARATI LETTER SA	Lo	Consonant	[Omniglot]	
50.	0AB9	હ	GUJARATI LETTER HA	Lo	Consonant	[Omniglot]	

Repertoire Included - 6

Sr. No.	Unicode Code Point	Glyph	Character Name	Unicode General Category (gc)	Indic Syllabic Category	Reference	Y/N
51.	0ABC	◌̣	GUJARATI SIGN NUKTA	Mn	Nukta	[Omniglot]	
52.	0ABE	◌̇	GUJARATI VOWEL SIGN AA	Mc	Matra	[Omniglot]	
53.	0ABF	◌̈	GUJARATI VOWEL SIGN I	Mc	Matra	[Omniglot]	
54.	0AC0	◌̉	GUJARATI VOWEL SIGN II	Mc	Matra	[Omniglot]	
55.	0AC1	◌̊	GUJARATI VOWEL SIGN U	Mn	Matra	[Omniglot]	
56.	0AC2	◌̋	GUJARATI VOWEL SIGN UU	Mn	Matra	[Omniglot]	
57.	0AC3	◌̌	GUJARATI VOWEL SIGN VOCALIC R	Mn	Matra	[Omniglot]	
58.	0AC4	◌̍	GUJARATI VOWEL SIGN VOCALIC RR	Mn	Matra	[Omniglot]	
59.	0AC5	◌̎	GUJARATI VOWEL SIGN CANDRA E	Mn	Matra	[Omniglot]	
60.	0AC7	◌̏	GUJARATI VOWEL SIGN E	Mn	Matra	[Omniglot]	

Repertoire Included - 7

Sr. No.	Unicode Code Point	Glyph	Character Name	Unicode General Category (gc)	Indic Syllabic Category	Reference	Y/N
61.	0AC8	ૐ	GUJARATI VOWEL SIGN AI	Mn	Matra	[Omniglot]	
62.	0AC9	ૐ	GUJARATI VOWEL SIGN CANDRA O	Mc	Matra	[Omniglot]	
63.	0ACB	ૐ	GUJARATI VOWEL SIGN O	Mc	Matra	[Omniglot]	
64.	0ACC	ૐ	GUJARATI VOWEL SIGN AU	Mc	Matra	[Omniglot]	
65.	0ACD	્	GUJARATI SIGN VIRAMA	Mn	Halanta / Virama	[Omniglot]	

MSR-2 Excluded

- Code Points excluded in MSR-2 because they are either not in common use or used for special purpose only (historic, religious, etc.):

Code Point	Glyph	Type	Notes
0ABD	॥	GUJARATI SIGN AVAGRAHA	• obsolete (Sanskrit)
0AD0	ॐ	GUJARATI OM	• religious use • symbol
0AE0	ꣳ	GUJARATI LETTER VOCALIC RR	• obsolete (Sanskrit)
0AE1	ꣴ	GUJARATI LETTER VOCALIC LL	• obsolete (Sanskrit)
0AE2	ꣵ	GUJARATI VOWEL SIGN VOCALIC L	• obsolete (Sanskrit)
0AE3	ꣶ	GUJARATI VOWEL SIGN VOCALIC LL	• obsolete (Sanskrit)

Repertoire Excluded

- Code Points excluded in MSR-2 because they are either not in common use or used for special purpose only (historic, religious, etc.):

Code Point	Glyph	Type	Notes
0A81	◌̣	GUJARATI SIGN CANDRABINDU	• obsolete

Within Script Variants for Gujarati

No cases of within script variants.

Cross Script Variants for Gujarati and <script>

- List of code points which are variants because Gujarati and <script> scripts are closely related to each other:

Code Point 1 + Glyph 1	Code Point 2 + Glyph 2	Type (blocked)

Whole Label Evaluation Rules (in plain English) - 1

- ⊙ C → Consonant
- ⊙ M → Matra
- ⊙ V → Vowel
- ⊙ D → Anusvara (Bindu)
- ⊙ X → Visarga
- ⊙ H → Halanta / Virama
- ⊙ N → Nukta

Whole Label Evaluation Rules (in plain English)

- ⦿ Rule 1: N must be preceded only by either of specific set of Cs
 - § (U+0A15)
 - ㄸ (U+0A16)
 - ㅊ (U+0A17)
 - ㅌ (U+0A1C)
 - ㅍ (U+0A2B)

- ⦿ Rule 2: H must be preceded by C or CN

- ⦿ Rule 3: X must be preceded by either of V, C, N or M

- ⦿ Rule 4: D must be preceded by either of V, C, N or M

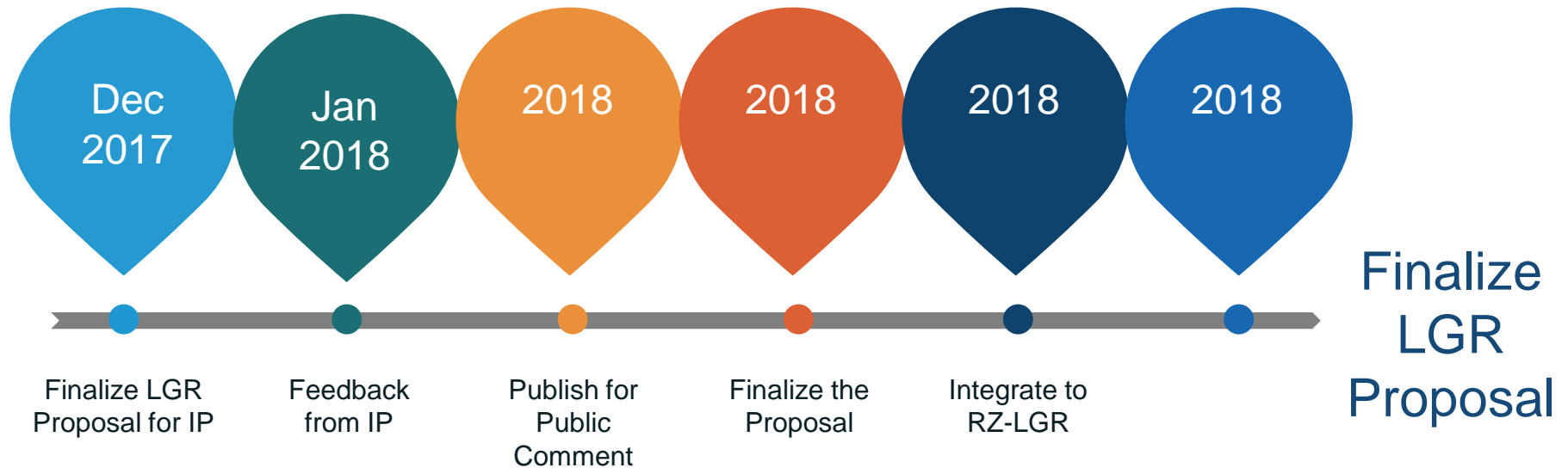
- ⦿ Rule 5: M must be preceded either by C or CN

- ⦿ Rule 6: V Can **NOT** be preceded by H

Current Status

Section in Google Doc	Status (To be initiated / In Progress / Completed / Reviewed)	Additional Notes
Description of Script and use	Completed	
Languages using Script	Completed	
Classification of characters	Completed	
Repertoire Analysis	Completed	
Within-script Variants	In Progress	
Cross-script Variants	In Progress	Devanagari
WLE Rules	Completed	
References		

Plan Till Completion



To Summarize

The LGR proposal for Gujarati script is progressing as most of the components have been completed, particularly just the variants part is still being worked upon. Ideally, we are expecting a final proposal by end of December for feedback from the Integration Panel.

Thank You and Questions

Contributors:

Name	Address	Email
Dr. Raimond Doctor		raymond.doctor@gmail.com
Akshat S. Joshi		akshatj@cdac.in

Thank you !

