SCRIPT GRAMMAR FOR GUJARATI LANGUAGE

Prepared by

Technology Development for Indian Languages (TDIL) Programme

Department of Information Technology, Government of India

in association with

Centre for Development of Advanced Computing (C-DAC)

Table of Contents

0. INTRODUCTION

The term **script grammar** refers to the behaviour pattern of the writing system of a given language. Languages which have written representations do not use a haphazard manner of storing the information within the system, but use a coherent pattern which is similar to the linguistic grammar of a given language. With the help of specialists (not necessarily linguist) who work in the area of the written representation of the language, the manner in which the shapes of the characters of the language and the representation of the conjunct forms is provided. In other words the Script Grammar deals with the surface structure of the language and tries to provide the best possible "fit" for shapes and their representation. Since this is a highly subjective issue, the shapes provided here are recommendations at the best and conform to the perception of the mandating body/evaluators who consensually arrive at the "best possible fit" which is acceptable to a majority of users. An example from the Devanāgarī script will make the above clear. Although Marathi and Nepali share the same script Devanāgarī, not only do they not share the same character inventory but in addition the representation of certain characters is different. Thus the Nepali /la/ is different from the Marathi /la/ in so far as the placement of the stem is concerned Nepali ल Marathi ल. This ensures that the Script Grammar conforms to the language in question and provides the character shapes acceptable to a given user community. It should be noted that this does not mean monotony. The Marathi and Nepali /la/ can have a variety of forms once the intrinsic structure of the character is determined.

Script Grammar is the term used to define:

- the writing system used to inscribe a given language
- the history of the script and language (wherever available)
- the syllabic structure of the writing system of the language
- the rule ordering of the characters within the syllable (akshar)
- description of the syllabic clusters
- collation order of the characters: lexical / dictionary sorting order

1. OBJECTIVES OF SCRIPT GRAMMAR

The Objectives of the script grammar for each language can be divided into two major parts:

Societal:

- Provide a visual representation of shapes that are deemed to be in conformity with the perception of a given community
- Ensure thereby that this perception is safe-guarded
- Through wide-spread dissemination and creation of appropriate tools ensure that within the given linguistic community, all media tries to adopt the given shape.

Technical:

- Classify the language in terms of its ISO and also whether it belongs to the Abjad, Akshar (Alphasyllabary) class.
- Provide an inventory of the characters pertinent to the language and classify the same in terms of their taxonomy.
- As a corollary determine whether the inventory is in conformity to the Syllable formalism as stipulated in ISCII'91 and subsequently adopted by Unicode.
- Since Brahmi is written from left to right, and since certain characters do not follow the linear L to R order, provide an inventory of displaced catenators i.e. characters such as Mātrās that concatenate to the Consonant
- Propose the best shape representation of the individual characters as well as of the ligatures used within a given script. As a corollary request the expert(s) to identify the largest possible strings of such ligatures.
- Finally provide the collation order pertinent to that script/language, which would be of great utility to high-end NLP as well as to CLDR's in the pertinent language. Thus for example, the collation order for Marathi is different from Hindi although both languages share the same script. Thus in Marathi क्ष, ज्ञ are placed at the end of the consonant inventory i.e. after ह in the sort order. In Hindi क्ष is sorted along with क and ज्ञ with ज

2. END USERS FOR SCRIPT GRAMMAR

The script-grammar specific to a given language can be used by a large number of users.

- Most importantly it can be used by font developers desirous of developing a font which is compliant with the perception of the characters and ligatures of a language by its user community.

- Certain features of the script grammar such as the shapes can also be used for testing OCR and OHWR. Similarly information regarding Ligatures as well as collation order can help in high-end NLP work such as detecting invalid combinations, correct implementation of syllable structure, prediction routines to name a few. Information regarding collation and character sets can be also used for CLDR.

- They allow the font designer to design a font which is in compliance with the norms and standards of that particular script. A major problem which will be dealt with in the template is one of ligatures. The final list of ligatures defined by the script grammar allows the font designer to write specific rules for such glyphs.

- It permits the software developer to design and implement the keyboard and the input mechanism which will meet the requirement of the particular linguistic community.

- The collation or sort order as described in a Script Grammar permits the software developer to write software functions/ routines for sorting data in all applications.

- Script Grammars are equally important for keyboard design, especially when supplemented by frequency data from a corpus.

As can be seen the script grammar has a wide range of use and can be of utility to font developers, Indian language developers and linguists in the area of computation.

3. SCOPE

This script grammar document contains following information about the language and the script used for writing the language.

1. Name of the language and its representation in the 3 letter mnemonic as per ISO 639.3 standard.
2. Script used to inscribe the given language
3. The structure of the script used for writing the language
    - Rule ordering of the characters within the syllable formation is a language
    - Description of the syllabic clusters of the script
    - Collation order of the characters: lexical / dictionary sorting order
    - Compliance of the script with Unicode.

These will be treated within the relevant sections of the script grammar

4. TERMINOLOGY[1]

Abjad: A writing system in which each symbol always or usually stands for a consonant. The long vowels are indicated. However the short vowels are rarely marked and the reader needs to supply these. Example: Urdu written in Perso-Arabic Script is an example of this writing system.

Abugida: also called an alphasyllabary, is a segmental writing system in which consonant–vowel sequences are written as a unit: each unit is based on a consonant letter, and vowel notation is obligatory but secondary[2]

Akshar: see **Abugida**

Allographs: Variants of the representation of a character. Thus ae and æ [U+00E6] in Latin alphabet are allographs.

Allo-Script: The term relates to languages which share a common script. Thus Devanāgarī is used to write 9 official languages. However these languages do not use the same set of characters. Thus Marathi uses the retroflex lla ळ [U+ 0933] which Hindi does not use. Flaps used in Hindi ड़ [U+095C] ढ़ [U+095D] are not used in Marathi. These sub-sets of scripts based on a single "matricial" script are termed as allo-scripts.

Alphabet: A set of letters used in writing a language. Example: The English Alphabet.

Aspirated consonant: A consonant which is pronounced with an extra puff of air coming out at the time of release of the oral obstruction. This has a sound of an extra "h".

Basic alphabet: The minimal set of letters which can be used for uniquely encoding every word of a language. The basic alphabet for English consists of only the upper-case letters A-Z

Catenators: Also termed as Concatenators are characters which are concatenated to another character. In the Brahmi script these are the Mātrās or Vowel modifiers which are adjoined to the consonant and add a vocalic value to the consonant.

Conjunct: The Indic scripts are noted for a large number of consonant conjunct forms that serve as orthographic abbreviations (ligatures) of two or more adjacent letterforms. This abbreviation takes place only in the context of a consonant cluster. Under normal circumstances, a consonant cluster is depicted with a conjunct glyph if such a glyph is available in the current font. In the absence of a conjunct glyph, the one or more dead

[1] As in the case of the BIS Document, in order to make the terminology accessible for all readers, examples have been chosen from English/Latin scripts, wherever possible. Some definitions have been excerpted from the BIS ISCII91 document and suitably modified where necessary.
[2] Wikipedia definition

consonants that form part of the cluster are depicted using half-form glyphs. In the absence of half-form glyphs, the dead consonants are depicted using the nominal consonant forms combined with visible virama signs.[3]

Consonant: A letter representing a speech sound in which the breath is at least partly obstructed.

Diacritic:A mark added to a letter which distinguishes it from the same letter without a mark, usually having a different phonetic value or stress.

Displaced Catenator: (see Catenator) Within the Brahmi script, the writing system is linear and moves from left to right. However in the case of some catenators this rules is not observed and the catenator (wholly or partially) is placed to the right of the consonant to which it relates. The short vowel I /ि/ in Devanāgarī is an example of a displaced catenator.

Display composing: The process of organizing the basic shapes available in a font in order to display (or print) a word.

Display rendition: The process by which a string of characters is displayed (or printed). In this process several consecutive characters may combine with each other on the screen. The sequence of display of the characters may become different.

Eyebrow repha: (See **Eyelash ra)**

Eyelash ra: The eyelash ra is used in Konkani, Gujarati and Marathii. It is treated as different from the र् (repha) by certain linguists. While the former is treated as a flap, the latter is a continuant trill (*cf.*, Kalyan Kale and Anjali Soman. 1986).

Font: A set of symbols used for display or printing of a script in a particular style.

International numerals: The conventional 0 to 9 digits used in English for denoting numbers. these are also known as Indo-Arabic numerals (to differentiate them from the Roman numerals like IX for 9).

Latin alphabet: The alphabet used for writing the language of ancient Rome. Also known as the Roman alphabet. The alphabet is used today for writing English and European languages.

Letter: A character representing one or more of the simple or compound sounds used in speech. It can be any of the alphabetic symbols.

Ligature: (see **Conjunct**)

[3] Unicode ver. 6.0 Chapter 9.0 pp 6-7

Nasal consonant: A consonant pronounced with the breath passing through the nose. Example *m n* in English.

Nasalized vowel: A vowel pronounced with the breath passing both through the nose and the mouth. In Indian scripts this is denoted by a Chandrabindu and gives the vowel/vowel modifier over which it placed a nasal value. Example: जाँच

Phonetic alphabet: An alphabet which has direct correspondence between letters and sounds Example: The International Phonetic Alphabet..

Pure consonant: A consonant which does not have any vowel implicitly associated with it.

Rafar: A special case of a ligature constituted by the adjunction of ra followed by a halanta to consonant. The resultant combination places the ra on top of the consonant to which it is adjoined. In case the consonant itself is adjoined to another consonant, the rafar is placed above the consonant e.g.in Devanagari: र्+क= र्क , र्+घ्+य=र्घ्य

Rakar: A special case of a ligature constituted by the adjunction of a consonant followed by a halanta to ra. In a large number of Brahmi scripts the ra is adjoined to the stem of consonant to which it relates. In the case of consonants which have no stem such as the dental retroflexes in Devanāgarī, the rakar is placed below the consonant to which it relates.

Repha: (see **Rafar)**

Roman script: The script based on the ancient Roman alphabet, with the letters A-Z and additional diacritic marks. Used for writing a language which is not usually written in the Roman alphabet.

Script: A distinctive and complete set of characters used for the written form of one or more languages.

Script numerals: The 0 to 9 digits in a script, which have shapes distinct from their international counterparts.

Syllable: A unit of pronunciation uttered without interruption, forming whole or part of a word, and usually having one vowel or diphthong sound optionally surrounded by one or more consonants

Transliteration: Representation of words with the closest corresponding letters in an alphabet of a different language.

Vowel: A letter representing a speech sound made with the vibration of the vocal cords, but without audible obstruction

Vowel sign: A graphic character associated with a letter, to indicate a vowel to be associated with that character (Mātrā in Hindi).

5. PHILOSOPHY AND UNDERLYING PRINCIPLES

The script grammar is based on the following principles:

1. The Grammar aims to depict the surface grammar of the written language: the manner in which characters as well as conjuncts are depicted

2. Where a given script admits many languages, it is pre-suppose that such languages will prescribe different representations for a given shape or conjunct according to the perception of the native users of that language

3. Corollary to the above the result is a script and allo-scripts i.e. a given script shared by many languages is not uniformly deployed across all the languages but is subject to variations and modulations.

4. The term Grammar is used here in a non-normative sense: what is prescribed is in the form of recommendations provided by experts who visualize the shape of the given script in their mother tongue in a specific manner. Subjective variations may occur[4]

5. The Grammar is limited to its synchronic use i.e. the manner in which a given language as of today admits a character set within the script used to write it. It is not diachronic or historical in nature and does not study the evolution of the given script across centuries.

[4] It is recommended that such variations be culled by placing the Grammars of different scripts in public review.

6. SCRIPT GRAMMAR STRUCTURE

The script grammar provided below has the following parts.

Part 6.1. deals with peripheral elements such as the ISO of the language, the writing system used: (Alphasyllabic) Abugida or Abjad.

Part 6.2. treats of the syllabic structure. It verifies whether the character set of the language complies with the ISCII syllabic structure and if not which cases are not compliant.

Part 6.3 is the script grammar proper and describes the character set as well as the conjunct shapes of the given script along with the collation order

6.1. PERIPHERAL ELEMENTS OF THE SCRIPT GRAMMAR

These constitute the elements that are peripheral to the Script Grammar. The main parameters considered are the mnemonic and name of the language (needed for CLDR and also for language tags), the writing system used to inscribe the language and wherever possible a short history of the language.

6.1.1. Name of the language and its representation in the 3 letter mnemonic as per ISO 639.1. & 639.3

Name of the Language: GUJARATI
ISO Mnemonics: *guj*

This refers to a one line description of the language and its mnemonic representation as per the ISO.

6.1.2. Identification of the writing system(s) used to inscribe the given language

Gujarati is written using the Gujarati script. It is an alphasyllabary with the akshar as its core.

This is a one line description of the script used to write the language. However in case the language uses more than one script, all the scripts in question are specified, provided these constitute the official language of the given state.
All scripts derived from Brahmi are Abugidas i.e. syllabary driven systems. The main features of Abugidas are as under:

- The consonant has an implicit vowel built-in which is normally the schwa.
- The inherent vowel can be modified by the addition of other vowels or muted by a diacritic termed as a Virama or Halanta
- Vowels can be handled as full vowels with a vocalic value
- When two or more consonants join together they form ligatures which can be recognized by their shape ક્ત or alternatively form an entirely new shape ક્ + ષ = ક્ષ.

Abugidas/Alphasyllabaries because of their syllabic structure require a special description which is the subject of the discussion in 6.2. below.

6.1.3. Amendments needed in Unicode for Gujarati language

None have been proposed by the *Gujarati Sahitya Parishad* which has mandated the script grammar.

6.2. CONFORMITY TO THE SYLLABLE STRUCTURE

Gujarati language complies with the syllable (akshar) structure described above. It can admit up to 3 consonant clusters.

Alphasyllabaries are determined by the notion of the syllable or the Akshar. The compositional grammar of the syllable determines it well-formedness. This is through a series of formal constraints based on a Backus-Naur Formalism which is given below. The syllable (akshar), first defined in the ISCII document (1991), identifies the following character 'sub-sets' for the purposes of identifying the syllable (akshar). In what follows the syllable analysis will be restricted to Gujarati.

(C) Consonants

ક	ખ	ગ	ઘ	ઙ
ચ	છ	જ	ઝ	ઞ
ટ	ઠ	ડ	ઢ	ણ
ત	થ	દ	ધ	ન
પ	ફ	બ	ભ	મ
ય	ર	લ	વ	શ
ષ	સ	હ	ળ	

(V) Vowels

અ	આ	ઇ	ઈ	ઉ	ઊ	ઋ	એ	એ	ઐ	ઑ	ઓ	ઔ

(M) Mātrās or Vowel Modifiers

ા	િ	ી	ુ	ૂ	ૃ	ૅ	ે	ૈ	ૉ	ો	ૌ

(D) Diacritics

ં: Anuswar	Anuswara, a nasal, denoted by a dot above the letter after which it is to be pronounced. This falls under Nasal category.
ઁ:Chandrabindu	Chandrabindu, a nasal, denoted by a breve with a dot superposed above the letter after which it is to be pronounced. This falls under Nasal category.
ઃ Visarga	Visarga, denoted by two dots placed above the other.

ऽ: Avagraha	For extra length with long vowels as seen in the Sanskrit text સર્વતોઽમવત્

(H):Halanta ્- Halant used in most writing systems to signify the lack of an inherent vowel.

(N)[5] Nukta ़ - is Not used in Gujarati

Each of these sub-types has its restrictions in terms of what can precede or follow it within a syllable (akshar), as shown in the table below:

PRECEDED BY	SUBTYPE	FOLLOWED BY
-, H	**C**	M,D,H
-	**V**	D
C	**M**	D
C,V,M	**D**	-
C	**H**	C

C can be preceded by H or no subtype and followed by any one of the following: M,D,H
V can be preceded by no subtype and followed by D but not by another sub-type.
M can be preceded by C and followed by D.
D can be preceded by C, V, M and followed by no other subtype. It closes the syllable (akshar).
H can be preceded by C alone and followed only by C and no other sub-set.

6.2.1. Syllable (akshar) Types

The formalism defines the syllable (akshar) in terms of both what can constitute a syllable (akshar) and what cannot. A valid syllable (akshar) as per this definition can be of only two types:

1. A vowel syllable (akshar): a full vowel.
2. A consonant syllable (akshar): a full consonant (having a mātrā)

The three other subsets viz. Mātrās, Diacritics, Halanta cannot constitute a syllable (akshar) by themselves or in combination among themselves.

6.2.1.1. The Vowel syllable (akshar) is of the following types:

[5] The nukta is a small dot placed under a character in Northern scripts to show that they are flapped or for deriving 5 other consonants in the Devanāgarī and Punjabi scripts, required for Urdu क़,ख़,ग़,ज़,फ़

6.2.1.1.1. A pure vowel all by itself: અ /a/ આ /ā/ etc.

6.2.1.1.2. A vowel followed by a modifier i.e. either a nasal marker or a visarga or an avagraha: ઁ /ĩ /, અઃ / āH / આऽ /ā:/

6.2.1.2. The Consonant syllable (akshar) can be of the following types:

6.2.1.2.1. A full consonant (with or without Nukta) i.e. with the inherent vowel : ક : /ka/

6.2.1.2.2. A consonant followed by a mātrā i.e. the inherent vowel being substituted by another vowel: કી /ki:/

6.2.1.2.3. A consonant followed by a modifier: કં /kə̃/, હઃ /haH[6]/

6.2.1.2.4. A consonant followed by a mātrā and a modifier: કું /kũ/, દુઃ /duH/.

6.2.1.2.5. A consonant cluster i.e. a dead or half consonant (Consonant+Halanta) followed by a full consonant followed optionally by a mātrā, a modifier or a combination of both. These result in a ligature or what is often termed as *yuktakshara.*

ત્ક /tka/ ત્કં /tkə̃/, ત્કઃ /tkaH/ ત્કું /tkũ/, ત્કુ /tku/.

The above permutations and combinations result in 7 major syllable (akshar) types. Of these the last type introduces the problem of the number of consonant clusters. ISCII (91, p.23) provides for up to three consonant clusters as the "worst case" i.e. the largest possible string. This is functional for modern Prakrits where the largest consonantal cluster rarely exceeds three consonant. Sanskrit is an exception where in a single word, four consonants can come together: कार्त्स्न्य /kārtsnya/ "wholeness", "entirety".

This means that theoretically the following forms can be postulated:

1. Vowel Set: With the Vowel as the node.
   V VD
2. Consonant set: With the Consonant as the node (an implicit or modified vowel is pre-implied).

Node	Mātrā	Modifier	Mātrā+Modifier
C	CM	CD	CMD
CHC	CHCM	CHCD	CHCMD
CHCHC	CHCHCM	CHCHCD	CHCHCMD
CHCHCHC	CHCHCHCM	CHCHCHCD	CHCHCHCM

A total number of 16 theoretical syllables is therefore possible. It will be seen that the written syllable (akshar) is not very different in structure from the phonetic syllable and that the movement from the written to the spoken levels is made feasible by application of certain rules.

[6] This character represents phonetically the weak implicit vowel, termed as schwa and often shown as /a/ also.

This formal structure of the syllable (akshar) explained above is common to all Brahmi based scripts (with a few variations). It will form the basis of an exhaustive description of the characters as well as their ligatural representations.

6.3 SCRIPT GRAMMAR PROPER

This section lays down in detail the different parameters of the Script Grammar for Gujarati. These are:

6.3.1. The Character Set of Gujarati.
6.3.2. The Consonant mātrā combinations of Gujarati.
6.3.3. The Ligature Set of Gujarati.
6.3.4. Collocation Order of Gujarati

6.3.1. The Character Set of Gujarati.

This section provides detailed information about the characters in the language and the list of the same and also more importantly shows the manner in which the character is to be written. Each subsection comprises therefore two parts: the basic character set and the shape each character should have, as mandated by the experts who have designed the script grammar of Gujarati.

This comprises the following:
6.3.1.1. The Consonant Set
6.3.1.2. The Vowel Set
6.3.1.3. The Mātrā Set
6.3.1.4. Displaced Catenators
6.3.1.5. Shape of the combination of ra (rakar,repha)
6.3.1.6. The Set of Diacritics
6.3.1.7. Numerals
6.3.1.8. Punctuation marks
6.3.1.9. Other symbols
Each of these will be analysed in detail:

6.3.1.1. The Consonant Set

The Consonant set of Gujarati comprises the following characters:

A basic Consonant inventory arranged as per their Vargas.

	-voiced -aspirated	-voiced +aspirated	+voiced -aspirated	+voiced +aspirated	Nasal
Velar	ક	ખ	ગ	ઘ	ઙ
Palatal	ચ	છ	જ	ઝ	ઞ
Retroflex	ટ	ઠ	ડ	ઢ	ણ
Dental	ત	થ	દ	ધ	ન
B-labial	પ	ફ	બ	ભ	મ

Other consonants

ય	ર	લ	ળ	વ
શ	ષ	સ	હ	

Note: Ligatures ક્ષ જ્ઞ ત્ર are not listed in the consonants list as they are ligatures.

The exact shapes as desired by the experts are provided in the table below:

	-voiced -aspirated	-voiced +aspirated	+voiced -aspirated	+voiced +aspirated	Nasal
Velar	ક	ખ	ગ	ઘ	ઙ
Palatal	ચ	છ	જ	ઝ	ઞ
Retroflex	ટ	ઠ	ડ	ઢ	ણ
Dental	ત	થ	દ	ધ	ન
B-labial	પ	ફ	બ	ભ	મ

Other consonants

ય	ર	લ	ળ	વ
શ	ષ	સ	હ	

6.3.1.2. The Vowel Set

The Vowel set of Gujarati is as under:

અ	GUJARATI LETTER A
આ	GUJARATI LETTER AA
ઇ	GUJARATI LETTER I
ઈ	GUJARATI LETTER II
ઉ	GUJARATI LETTER U
ઊ	GUJARATI LETTER UU

ઋ	GUJARATI LETTER VOCALIC R
એ	GUJARATI LETTER E
ઐ	GUJARATI LETTER AI
ઓ	GUJARATI LETTER O
ઔ	GUJARATI LETTER AU
ઍ	GUJARATI LETTER CANDRA E
ઑ	GUJARATI LETTER CANDRA O

As per expert recommendations the character set should be written as under:

અ આ ઇ ઈ ઉ ઊ ઋ એ ઐ ઓ ઔ ઍ ઑ

6.3.1.3. The Mātrā Set

The Mātrā (Vowel Modifier Set) of Gujarati is as under:

Mātrā Names	Mātrās Sign	Where is it used ?	Consonant Shapes formed
1. Gujarati sign AA	◌ા	આ	ક્+આ=કા
2. Gujarati sign I (stands to the left of the consonant)	◌િ	ઇ	ક્+ઇ=કિ
3. Gujarati sign II	◌ી	ઈ	ક્+ઈ=કી
4. Gujarati sign U	◌ુ	ઉ	ક્+ઉ=કુ
5. Gujarati sign UU	◌ૂ	ઊ	ક્+ઊ=કૂ
6. Gujarati sign vocalic R	◌ૃ	ઋ	ક્+ઋ=કૃ
7. Gujarati sign vocalic candra E	◌ૅ	ઍ	ક્+ઍ =કૅ
8. Gujarati sign E	◌ે	એ	ક્+એ=કે
9 Gujarati sign AI	◌ૈ	ઐ	ક્+ઐ=કૈ
10.Gujarati sign candra O	◌ૉ	ઑ	ક્+ઑ=કૉ

11.Gujarati sign O	ો	ઓ	ક્+ઓ=કો
12. Gujarati sign AU	ૌ	ઔ	ક્+ઔ=કૌ

As per expert recommendations the character set should be written as under:

ા, િ, ી, ુ, ૂ, ૃ, ૅ, ે, ૈ, ૉ, ો, ૌ,

6.3.1.4. Displaced Catenators

Under normal circumstances Vowel Modifiers also known as catenators (since they concatenate to the preceding consonant) in Brahmi based scripts are written from left to right in linear order (with the exception of Consonant stacks). However certain modifiers are displaced and are placed to the left of the consonant to which they concatenate. As a general rule in all Devanāgarī script driven languages there is only one displaced catenator: ि

CATENATOR	POSITION	EXAMPLE
ि	To left of character	કિ,તિ,રિ

6.3.1.5. Shape of the combination of ra (rakar,repha)

The र takes a variety of shapes known as rakar and repha (rafar) depending on its position. When conjoined before a consonant by means of the halanta, it changes shape and is placed on top of the consonant or consonant clusters to which it relates. This is called a repha or rafar. Gujarati admits a special repha known as eyelash ra. When it is conjoined after a consonant with the help of a halanta, it appends to the consonant in the shape of a slanting stroke attached to the stem (side rakar) or in the case of consonants which have no stem such as ट, it is appended in the shape of a ^ to the bottom of the character (bottom rakar). Gujarati has the following combinations of ra:

RAFARS

Top rafar: ર્ક ર્ત for ex top rafars will be formed in case of following words.

ધર્મ, ચર્ખો

RAKARS

1. Bottom rakar ટ્ર ડ્ર

2. Side rakar ક્ર ત્ર

Bottom rakar	રાષ્ટ્ર
Side rakar	વ્રત, ચક્ર, પ્રસાદ

6.3.1.6. Diacritics

These are as under in the case of Gujarati:

ં: - Anuswar રંગ

ઁ: -Chandrabindu/Anunasika Rarely used in Gujarati

્: - Halant સંવત્

ઃ - Visarga દુઃખ

ऽ: - Avagaraha: for extra length with long vowels, mainly in Sanskrit texts

e.g. / सर्वतोऽमवत् /

6.3.1.7. Numerals

Following are the numbers used in Gujarati language.
Latino-Arabic set: (0,1,2,3,4,5,6,7,8,9) is used in official documents. But in the Gujarati text, Gujarati numerals should be preferred. They are as follows.

૦ ૧ ૨ ૩ ૪ ૫ ૬ ૭ ૮ ૯

Numeral Shapes	Explanation
૦	Gujarati Digit Zero
૧	Gujarati Digit One
૨	Gujarati Digit Two
૩	Gujarati Digit Three
૪	Gujarati Digit Four
૫	Gujarati Digit Five
૬	Gujarati Digit Six
૭	Gujarati Digit Seven

૮	Gujarati Digit Eight
૯	Gujarati Digit Nine

6.3.1.8. Punctuation Markers

Gujarati uses punctuation markers from the Latin set. such as . , ; : " ' ( ) [ ] etc. English fullstop [.] is used, the use of Purna and Deergha Virama (full-stop/danda) Devanāgarī code block: U+0964, U+0965 ।, ॥ is commonly used in poetry..

A list of punctuations is provided below:

Sr. No.	Name of the marker	Marker Shape
01	Full Stop or Period	.
02	Question Mark	?
03	Exclamation Mark	!
04	Apostrophe	'
05	Semi Colon	;
06	Colon	:
07	Hyphen	-
08	Dash	--
09	Ellipsis mark	...
10	Oblique	/
11	Double quotation mark	" "
12	Single quotation mark	' '
13	Cross	XXX
14	As Above	- - " - -
15	Round Brackets	( )
16	Square Brackets	[ ]
17	Curly Brackets	{ }
19	Devanāgarī Danda	|
20	Devanāgarī Double Danda	||

6.3.1.9 Other Symbols

These are religious, currency markers etc. included in Unicode:

ૐ Om (as written in Gujarati)

₹: Rupee Sign as mandated by Government of India.

Note: The old sign for Rupee in Gujarati ૱ [0AF1] which was ૩ followed by the abbreviation marker has been replaced by ₹ [U+20B9]

6.3.2. Consonant Mātrā Combinations.
These refer to the shapes generated when a Mātrā is adjoined to the Consonant. The layout of these is in the shape of a matrix where the first horizontal row refers to the active consonant and the first vertical column refers to the vowel-modifier.

Due to constraints of space and also for reasons of clarity, for each class a series of 3 tables are provided.

Table 1: ક ખ ગ ઘ ઙ ચ છ જ ઝ ઞ

Table 2: ટ ઠ ડ ઢ ણ ત થ દ ધ ન

Table 3: પ ફ બ ભ મ ય ર લ ળ વ શ
ષ સ હ

Wherever there is an X it implies that the experts have deemed that such a combination is not used in the language. However for the font developer this is an indication that for this particular combination which is not existent in the language but needs to be accommodated in the font table, a simple linear combination be provided.

e.g. Although the combination of ઙ +Mātrā is theoretically not possible it needs to be handled at the font level in the anticipation that a user could type this combination. The font would show the following: ઙા

The classes are as under:
6.3.2.1. refers to a simple concatenation of Consonant and Mātrā combinations.
6.3.2.2. refers to a concatenation of Consonant and Mātrā + Nasal marker combinations.
Other diacritics such as avagraha and visarga have been avoided, since these are linear in nature, are adjoined to the combination and do not in any way modify the structure of the shapes.

6.3.2.1 Consonant and Mātrā combinations.

This set refers to a simple concatenation of Consonant and Mātrā.

Consonant and Mātrā combinations Set 1

	ક	ખ	ગ	ઘ	ઙ	ચ	છ	જ	ઝ	ઞ
◌ા	કા	ખા	ગા	ઘા	X	ચા	છા	જા	ઝા	X
◌િ	કિ	ખિ	ગિ	ઘિ	X	ચિ	છિ	જિ	ઝિ	X
◌ી	કી	ખી	ગી	ઘી	X	ચી	છી	જી	ઝી	X
◌ૃ	કૃ	ખૃ	ગૃ	ઘૃ	X	X	છૃ	જૃ	X	X
◌ુ	કુ	ખુ	ગુ	ઘુ	X	ચુ	છુ	જુ	ઝુ	X
◌ૂ	કૂ	ખૂ	ગૂ	ઘૂ	X	ચૂ	છૂ	જૂ	ઝૂ	X
◌ૈ	કૈ	ખૈ	ગૈ	ઘૈ	X	ચૈ	છૈ	જૈ	ઝૈ	X
◌ો	કો	ખો	ગો	ઘો	X	ચો	છો	જો	ઝો	X
◌ૌ	કૌ	ખૌ	ગૌ	ઘૌ	X	ચૌ	છૌ	જૌ	ઝૌ	X
◌ે	કે	ખે	ગે	ઘે	X	ચે	છે	જે	ઝે	X
◌ૅ	કૅ	X	ગૅ	ઘૅ	X	ચૅ	છૅ	જૅ	ઝૅ	X
◌ૉ	કૉ	X	ગૉ	ઘૉ	X	ચૉ	છૉ	જૉ	ઝૉ	X

Remark 1- ઙ and ઞ are rarely used as the first members of clusters

Consonant and Mātrā combinations Set 2

This set is in continuation of set 1 which shows consonant and Matra combinations.

	ટ	ઠ	ડ	ઢ	ણ	ત	થ	દ	ધ	ન
◌ા	ટા	ઠા	ડા	ઢા	ણા	તા	થા	દા	ધા	ના
◌િ	ટિ	ઠિ	ડિ	ઢિ	ણિ	તિ	થિ	દિ	ધિ	નિ
◌ી	ટી	ઠી	ડી	ઢી	ણી	તી	થી	દી	ધી	ની
◌ૃ	ટૃ	X	X	X	X	તૃ	X	X	ધૃ	નૃ
◌ુ	ટુ	ઠુ	ડુ	ઢુ	ણુ	તુ	થુ	દુ	ધુ	નુ
◌ૂ	ટૂ	ઠૂ	ડૂ	ઢૂ	ણૂ	તૂ	થૂ	દૂ	ધૂ	નૂ

◌ે	ટે	ઠે	ડે	ઢે	ણે	તે	થે	દે	ધે	ને
◌ૈ	ટૈ	ઠૈ	ડૈ	ઢૈ	ણૈ	તૈ	થૈ	દૈ	ધૈ	નૈ
◌ો	ટો	ઠો	ડો	ઢો	ણો	તો	થો	દો	ધો	નો
◌ૌ	ટૌ	ઠૌ	ડૌ	ઢૌ	X	તૌ	થૌ	દૌ	ધૌ	નૌ
◌ૅ	ટૅ	ઠૅ	ડૅ	ઢૅ	X	X	થૅ	X	X	નૅ
◌ૉ	ટૉ	ઠૉ	ડૉ	X	X	તૉ	થૉ	X	X	નૉ

Consonant and Mātrā combinations Set 3

This set is in continuation of set 2 which shows consonant and Matra combinations.

	પ	ફ	બ	ભ	મ	ય	ર	લ	વ	શ	ષ	સ	હ	ળ	ક્ષ	જ્ઞ
◌ા	પા	ફા	બા	ભા	મા	યા	રા	લા	વા	શા	ષા	સા	હા	ળા	ક્ષા	જ્ઞા
◌િ	પિ	ફિ	બિ	ભિ	મિ	યિ	રિ	લિ	વિ	શિ	ષિ	સિ	હિ	ળિ	ક્ષિ	જ્ઞિ
◌ી	પી	ફી	બી	ભી	મી	યી	રી	લી	વી	શી	ષી	સી	હી	ળી	ક્ષી	જ્ઞી
◌ૃ	પૃ	X	બૃ	ભૃ	મૃ	X	X	લૃ	વૃ	શૃ	X	સૃ	X	X	X	X
◌ુ	પુ	ફુ	બુ	ભુ	મુ	યુ	રુ	લુ	વુ	શુ	ષુ	સુ	હુ	ળુ	ક્ષુ	X
◌ૂ	પૂ	ફૂ	બૂ	ભૂ	મૂ	યૂ	રૂ	લૂ	વૂ	શૂ	ષૂ	સૂ	હૂ	ળૂ	ક્ષૂ	X
◌ે	પે	ફે	બે	ભે	મે	યે	રે	લે	વે	શે	ષે	સે	હે	ળે	ક્ષે	જ્ઞે
◌ૈ	પૈ	ફૈ	બૈ	ભૈ	મૈ	યૈ	રૈ	લૈ	વૈ	શૈ	ષૈ	સૈ	હૈ	ળૈ	ક્ષૈ	X
◌ો	પો	ફો	બો	ભો	મો	યો	રો	લો	વો	શો	ષો	સો	હો	ળો	ક્ષો	જ્ઞો
◌ૌ	પૌ	ફૌ	બૌ	ભૌ	મૌ	યૌ	રૌ	લૌ	વૌ	શૌ	ષૌ	સૌ	હૌ	X	ક્ષૌ	X
◌ૅ	પૅ	ફૅ	બૅ	X	મૅ	X	રૅ	લૅ	વૅ	શૅ	X	સૅ	હૅ	X	X	X
◌ૉ	પૉ	ફૉ	બૉ	X	મૉ	યૉ	રૉ	લૉ	વૉ	શૉ	X	સૉ	હૉ	X	X	X

6.3.2.2 Consonant and Mātrā +Nasal combinations.

This set refers to a Consonant and Mātrā + Nasal marker combinations.

Consonant and Mātrā + Nasal combinations - Set 1

	ક	ખ	ગ	ઘ	ઙ	ચ	છ	જ	ઝ	ઞ
◌ં	કં	ખં	ગં	ઘં	X	ચં	છં	જં	ઝં	X
◌ાં	કાં	ખાં	ગાં	ઘાં	X	ચાં	છાં	જાં	ઝાં	X
◌િં	કિં	ખિં	ગિં	X	X	ચિં	X	જિં	ઝિં	X
◌ીં	કીં	ખીં	ગીં	ઘીં	X	ચીં	છીં	જીં	ઝીં	X
◌ૃં	X	X	X	X	X	X	છૃં	જૃં	X	X
◌ું	કું	ખું	ગું	ઘું	X	ચું	છું	જું	ઝું	X
◌ૂં	કૂં	ખૂં	ગૂં	ઘૂં	X	ચૂં	છૂં	જૂં	ઝૂં	X
◌ેં	કેં	ખેં	ગેં	ઘેં	X	ચેં	છેં	જેં	ઝેં	X
◌ૈં	કૈં	X	X	ઘૈં	X	X	X	X	X	X
◌ોં	કોં	ખોં	ગોં	ઘોં	X	ચોં	છોં	જોં	ઝોં	X
◌ૌં	કૌં	ખૌં	ગૌં	ઘૌં	ઙૌં	ચૌં	છૌં	જૌં	ઝૌં	X
◌ૅં	કૅં	ખૅં	ગૅં	ઘૅં	ઙૅં	ચૅં	છૅં	જૅં	ઝૅં	X
◌ૉં	કૉં	ખૉં	ગૉં	ઘૉં	ઙૉં	ચૉં	છૉં	જૉં	ઝૉં	ખૉં

Consonant and Mātrā +Nasal combinations - Set 2

This set is in continuation of set 1 above which shows combinations of Consonant and Mātrā + Nasal marker

	ટ	ઠ	ડ	ઢ	ણ	ત	થ	દ	ધ	ન
◌ં	ટં	ઠં	ડં	ઢં	ણં	તં	થં	દં	ધં	નં
◌ાં	ટાં	ઠાં	ડાં	ઢાં	ણાં	તાં	થાં	દાં	ધાં	નાં
◌િં	ટિં	ઠિં	ડિં	ઢિં	ણિં	તિં	X	દિં	ધિં	નિં
◌ીં	ટીં	ઠીં	ડીં	ઢીં	ણીં	તીં	થીં	દીં	ધીં	નીં
◌ૃં	X	X	X	X	X	X	X	X	X	X
◌ું	ટું	ઠું	ડું	ઢું	ણુું	તુું	થુું	દુું	ધુું	નુું
◌ૂં	ટૂં	ઠૂં	ડૂં	ઢૂં	ણૂું	તૂું	થૂું	દૂું	ધૂું	નૂું
◌ેં	ટેં	ઠેં	ડેં	ઢેં	ણેં	તેં	થેં	દેં	ધેં	નેં
◌ૈં	ટૈં	ઠૈં	ડૈં	ઢૈં	ણૈં	તૈં	થૈં	દૈં	ધૈં	નૈં
◌ોં	ટોં	ઠોં	ડોં	ઢોં	X	તોં	થોં	દોં	ધોં	નોં
◌ૌં	ટૌં	ઠૌં	ડૌં	ઢૌં	ણૌં	તૌં	થૌં	દૌં	ધૌં	નૌં
◌ૅં	ટૅં	ઠૅં	ડૅં	ઢૅં	ણૅં	તૅં	થૅં	દૅં	ધૅં	નૅં
◌ૉં	ટૉં	ઠૉં	ડૉં	ઢૉં	ણૉં	તૉં	થૉં	દૉં	ધૉં	નૉં

Consonant and Mātrā +Nasal combinations - Set 3

This set is in continuation of set 2 above which shows combinations of Consonant and Mātrā + Nasal marker

	પ	ફ	બ	ભ	મ	ય	ર	લ	વ	શ	ષ	સ	હ	ળ	ક્ષ[7]	જ્ઞ[8]
◌ં	પં	ફં	બં	ભં	મં	યં	રં	લં	વં	શં	ષં	સં	હં	ળં	X	X
◌ાં	પાં	ફાં	બાં	ભાં	માં	યાં	રાં	લાં	વાં	શાં	ષાં	સાં	હાં	ળાં	ક્ષાં	જ્ઞાં
◌િં	પિં	ફિં	બિં	ભિં	મિં	X	રિં	લિં	વિં	શિં	ષિં	સિં	હિં	ળિં	X	X

[7] Inserted by expert although this is not a single consonant but a ligature
[8] Inserted by expert although this is not a single consonant but a ligature

◌ીં	પીં	ફીં	બીં	ભીં	મીં	X	રીં	લીં	વીં	શીં	ષીં	સીં	હીં	ળીં	X	X
◌ૃં	X	X	બૃં	ભૃં	X	X	X	X	વૃં	શૃં	ષૃં	X	X	X	X	X
◌ું	પું	ફું	બું	ભું	મું	યું	રું	લું	વું	શું	ષું	સું	હું	ળું	X	X
◌ૂં	પૂં	ફૂં	બૂં	ભૂં	મૂં	X	રૂં	લૂં	વૂં	શૂં	ષૂં	સૂં	હૂં	ળૂં	X	X
◌ંં	પંં	ફંં	બંં	ભંં	મંં	X	રંં	લંં	વંં	શંં	ષંં	સંં	હંં	ળંં	X	X
◌ૈં	પૈં	ફૈં	બૈં	X	મૈં	X	રૈં	લૈં	વૈં	શૈં	ષૈં	સૈં	હૈં	ળૈં	X	X
◌ોં	પોં	ફોં	બોં	ભોં	મોં	X	રોં	લોં	વોં	શોં	ષોં	સોં	હોં	ળોં	X	X
◌ૌં	પૌં	ફૌં	બૌં	ભૌં	મૌં	X	રૌં	લૌં	વૌં	શૌં	ષૌં	સૌં	હૌં	ળૌં	X	X
◌ંં	પંં	ફંં	બંં	ભંં	મંં	યંં	રંં	લંં	વંં	શંં	ષંં	સંં	હંં	ળંં	X	X
◌ૉં	પૉં	ફૉં	બૉં	ભૉં	મૉં	યૉં	રૉં	લૉં	વૉં	શૉં	ષૉં	સૉં	હૉં	ળૉં	X	X

Consonant and Mātrā + Nasal combinations: With Chandrabindu
Since Chandrabindu is rarely used in Gujarati, the experts have deemed the same as invalid

6.3.3. The Ligature Set of Gujarati.

Gujarati has a large set of ligatural forms. These are combinations of Consonant+Halanta+Consonant (CHC) or CHCHC or even rarer CHCHCHC. The CHC combinations which are the most frequent are arranged in the shape of a matrix: the abscissa or horizontal axis refers to the Consonant which constitutes the ligature and the ordinate or vertical axis shows the consonant which forms the ligature and which is followed by a halanta.

As in 6.3.2. the ligature sets are divided into the following

6.3.3.1 CHC (in a matrix)

6.3.3.2 CHCHC

6.3.3.3.CHCHCHC

6.3.3.1. CHC (combination of two consonanats)

These ligatures are presented as in the earlier case of Consonant+Mātrā combinations in three sets. A lot of "slots" have an X marked, showing that the experts have deemed that such a ligature is not possible in the language. However in these cases, the font developer is to assume that the ligature is linear in nature.

The following set shows a combination of two consonants. To know how particular combinations forms, select one consonant from the first column and second from first row. For eg. Combination of consonant "ક" and "ક" is ligature "ક્ક".

CHC(combination of two consonants) - Set 1										
	ક	ખ	ગ	ઘ	ઙ	ચ	છ	જ	ઝ	ઞ
ક્	ક્ક	ક્ખ	X	X	X	ક્ય	ક્છ	X	ક્ઝ	X
ખ્	X	ખ્ખ	X	X	X	X	X	ખ્જ	X	X
ગ્	X	X	ગ્ગ	ગ્ઘ	X	X	X	ગ્જ	ગ્ઝ	X
ઘ્	X	X	X	ઘ્ઘ	X	X	X	X	X	X
ઙ્	X	X	X	X	X	X	X	X	X	X
ચ્	X	X	X	X	X	ચ્ચ	ચ્છ	X	X	X
છ્	X	X	X	X	X	X	છ્છ	X	X	X
જ્	X	X	X	X	X	X	X	જ્જ	જ્ઝ	X
ઝ્	X	X	X	X	X	X	X	X	ઝ્ઝ	X
ઞ્	X	X	X	X	X	X	X	X	X	X
ટ્	ટ્ક	X	X	X	X	ટ્ચ	X	X	X	X

ઠ્	X	X	X	X	X	X	X	X	X	X
ડ્	X	X	ડ્ગ	X	X	X	X	ડ્જ	ડ્ઝ	X
ઢ્	X	X	X	X	X	X	X	X	X	X
ણ્	X	X	X	X	X	X	X	X	X	X
ત્	ત્ક	ત્ખ	X	X	X	X	X	X	ત્ઝ	X
થ્	X	X	X	X	X	X	X	X	X	X
દ્	X	X	દ્ગ	દ્ઘ	X	X	X	X	X	X
ધ્	X	X	X	X	X	X	X	X	X	X
ન્	ન્ક	X	ન્ગ	X	X	ન્ય	X	ન્જ	ન્ઝ	X
પ્	પ્ક	X	X	X	X	પ્ય	X	X	X	X
ફ્	X	X	X	X	X	X	X	ફ્જ	X	X
બ્	X	X	X	X	X	X	X	બ્જ	બ્ઝ	X
ભ્	X	X	X	X	X	X	X	X	X	X
મ્	X	X	X	X	X	X	X	X	X	X
ય્	X	X	X	X	X	X	X	X	X	X
લ્	લ્ક	X	લ્ગ	X	X	લ્ય	X	લ્જ	X	X
ળ્	X	X	X	X	X	X	X	X	X	X
વ્	X	X	X	X	X	X	X	વ્જ	વ્ઝ	X
શ્	શ્ક	શ્ખ	X	X	X	શ્ય	X	X	X	X
ષ્	ષ્ક	ષ્ખ	ષ્ગ	X	X	X	X	X	X	X
સ્	સ્ક	સ્ખ	X	X	X	સ્ય	X	સ્જ	X	X
હ્	X	X	X	X	X	X	X	X	X	X

CHC Set 2:

The following set shows a combination of two consonants. To know how particular combinations forms, select one consonant from the first column and second from first row. For eg. Combination of consonant "ક્" and "ટ" is ligature "ક્ટ".

CHC(combination of two consonants) - Set 2										
	ટ	ઠ	ડ	ઢ	ણ	ત	થ	દ	ધ	ન
ક્	ક્ટ	X	X	X	X	ક્ત	ક્થ	X	X	ક્ન
ખ્	X	X	X	X	X	ખ્ત	X	X	X	X
ગ્	X	X	X	X	ગ્ણ	X	X	ગ્દ	ગ્ધ	ગ્ન
ઘ્	X	X	X	X	X	X	X	X	X	ઘ્ન
ઙ્	X	X	X	X	X	X	X	X	X	X
ચ્	X	X	X	X	X	ચ્ત	X	X	X	X
છ્	X	X	X	X	X	X	X	X	X	X
જ્	X	X	X	X	X	X	X	X	X	જ્ન
ઝ્	X	X	ઝ્ડ	X	X	X	X	ઝ્ધ	X	X
ઞ્	X	X	X	X	X	X	X	X	X	X
ટ્	ટ્ટ	ટ્ઠ	ટ્ડ	X	X	X	X	X	X	X
ઠ્	X	ઠ્ઠ	X	X	X	X	X	X	X	X
ડ્	X	X	ડ્ડ	ડ્ઢ	X	X	X	ડ્ધ	X	X
ઢ્	X	X	X	ઢ્ઢ	X	X	X	X	X	X
ણ્	X	X	X	X	X	X	X	X	X	X
ત્	X	X	X	X	X	ત્ત	ત્થ	X	X	ત્ન
થ્	X	X	X	X	X	X	થ્થ	X	X	X
દ્	X	X	X	X	X	X	X	દ્દ	દ્ધ	X
ધ્	X	X	X	X	X	X	X	X	ધ્ધ	ધ્ન
ન્	ન્ટ	X	ન્ડ	X	X	ન્ત	ન્થ	ન્દ	ન્ધ	ન્ન
પ્	પ્ટ	X	X	X	X	પ્ત	X	X	X	પ્ન

ફ્	ફ્ટ	X	X	X	X	ફ્ત	ફ્થ	X	X	X
બ્	X	X	બ્ડ	X	X	X	X	બ્દ	બ્ધ	બ્ન
ભ્	X	X	X	X	X	X	X	X	X	X
મ્	X	X	X	X	મ્ગ	મ્ત	X	X	X	મ્ન
ય્	X	X	X	X	X	X	X	X	X	X
લ્	લ્ટ	X	લ્ડ	X	X	લ્ત	લ્થ	લ્દ	લ્ધ	X
ળ્	X	X	X	X	X	X	X	X	X	X
વ્	X	X	વ્ડ	X	X	X	X	X	X	X
શ્	X	X	X	X	X	શ્ત	X	X	X	શ્ન
ષ્	ષ્ટ	ષ્ઠ	X	X	ષ્ગ	X	X	X	X	X
સ્	સ્ટ	X	સ્ડ	X	X	સ્ત	સ્થ	સ્દ	X	સ્ન
હ્	X	X	X	X	X	X	X	X	X	હ્ન

CHC SET 3:

The following set shows a combination of two consonants. To know how particular combinations forms, select one consonant from the first column and second from first row. For eg. Combination of consonant "ફ્" and "પ" is the ligature "ફ્પ".

CHC(combination of two consonants) - **Set 3**													
	પ	ફ	બ	ભ	મ	ય	ર	લ	વ	શ	ષ	સ	હ
ક્	ક્પ	ક્ફ	X	X	ક્મ	ક્ય	ક્ર	ક્લ	ક્વ	ક્શ	X	ક્સ	X
ખ્	X	X	X	X	ખ્મ	ખ્ય	ખ્ર	X	ખ્વ	ખ્શ	X	ખ્સ	X
ગ્	ગ્પ	X	ગ્બ	ગ્ભ	ગ્મ	ગ્ય	ગ્ર	ગ્લ	ગ્વ	X	X	ગ્સ	X
ઘ્	X	X	X	X	X	ઘ્ય	ઘ્ર	X	ઘ્વ	X	X	X	X
ઙ્	X	X	X	X	ઙ્મ	X	X	X	X	X	X	X	X
ચ્	X	X	X	X	X	ચ્ય	ચ્ર	X	X	X	X	X	X
છ્	X	X	X	X	X	છ્ય	છ્ર	X	છ્વ	X	X	X	X
જ્	X	X	X	X	જ્મ	જ્ય	જ્ર	X	જ્વ	X	X	X	X

ઝ્	X	X	ઝબ	X	ઝમ	ઝય	ઝ	X	X	X	X	X	X
ઞ્	X	X	X	X	X	X	X	X	X	X	X	X	X
ટ્	ટ્પ	X	X	X	ટ્મ	ટ્ય	ટ્ર	X	ટ્વ	ટ્શ	X	ટ્સ	X
ઠ્	X	X	X	X	X	ઠ્ય	X	X	X	X	X	X	X
ડ્	X	X	X	ડ્ભ	ડ્મ	ડ્ય	ડ્ર	X	ડ્વ	X	X	ડ્સ	ડ્હ
ઢ્	X	X	X	X	X	ઢ્ય	ઢ્ર	X	X	X	X	X	X
ણ્	X	X	X	ણ્ભ	ણ્મ	ણ્ય	X	X	X	X	X	X	X
ત્	ત્પ	ત્ફ	ત્બ	X	ત્મ	ત્ય	ત્ર	ત્લ	ત્વ	ત્શ	X	ત્સ	X
થ્	X	X	X	X	X	થ્ય	થ્ર	X	X	X	X	X	X
દ્	દ્પ	X	દ્બ	દ્ભ	દ્મ	દ્ય	દ્ર	X	દ્વ	X	X	X	દ્હ
ધ્	X	X	X	X	ધ્મ	ધ્ય	ધ્ર	X	ધ્વ	X	X	X	X
ન્	X	ન્ફ	X	X	ન્મ	ન્ય	X	ન્લ	ન્વ	ન્શ	X	ન્સ	ન્હ
પ્	પ્પ	પ્ફ	X	X	X	પ્ય	પ્ર	પ્લ	X	પ્શ	X	પ્સ	પ્હ
ફ્	X	X	X	X	X	ફ્ય	ફ્ર	ફ્લ	X	X	X	X	X
બ્	X	X	બ્બ	બ્ભ	X	બ્ય	બ્ર	બ્લ	બ્વ	X	X	બ્સ	બ્હ
ભ્	X	X	X	ભ્ભ	X	ભ્ય	ભ્ર	X	ભ્વ	X	X	X	X
મ્	મ્પ	મ્ફ	મ્બ	મ્લ	મ્મ	મ્ય	મ્ર	મ્લ	X	મ્શ	X	મ્સ	મ્હ
ય્	X	X	X	X	X	ય્ય	X	X	X	X	X	X	X
લ્	લ્પ	લ્ફ	લ્બ	લ્લ	લ્મ	લ્ય	X	લ્લ	લ્વ	લ્શ	X	લ્સ	લ્હ
ળ્	X	X	X	X	X	ળ્ય	X	X	X	X	X	X	X
વ્	X	X	X	X	X	વ્ય	વ્ર	X	વ્વ	X	X	X	વ્હ
શ્	શ્પ	શ્ફ	શ્બ	શ્ભ	શ્મ	શ્ય	શ્ર	શ્લ	શ્વ	શ્શ	X	X	શ્હ
ષ્	ષ્પ	ષ્ફ	X	X	ષ્મ	ષ્ય	X	X	ષ્વ	X	ષ્ષ	X	X
સ્	સ્પ	સ્ફ	સ્બ	X	સ્મ	સ્ય	સ્ર	સ્લ	સ્વ	X	X	સ્સ	X
હ્	X	X	X	X	હ્મ	હ્ય	હ્ર	હ્મ/ હ્મ	હ્વ	X	X	હ્સ	હ્હ

6.3.3.2 CHCHC (combination of three consonanats)

These are not as frequent as the CHC combinations. Only the major are listed below. With a few exceptions these are mainly linear in nature

ક્ષ્ણ	ક્ષ્મ	ક્ષ્ય	ચ્છ્વ	જ્ઞ્ય	જ્જ્વ
ત્ક્ર	ત્ક્ષ	ત્ત્વ	ત્પ્ર	ત્પ્લ	ત્મ્ય
ત્ર્ય	ત્સ્ન	ત્સ્વ	દ્ધ્વ	ન્દ્ર	ન્દ્ર
ન્મ્ય	ચ્છ્	ર્ચ્ય	ર્ઙ્ગ	ર્ણ્ય	ર્ત્ત
ર્દ્ર	ર્ધ્ય	ર્ધ્વ	ર્વ્ય	ર્ષ્ય	ર્હ્ય
લ્પ્ય	ષ્ક્ર	ષ્ટ્ર	ષ્ઠ્ય	ષ્પ્ર	સ્ક્ર
સ્ટ્ર	સ્ત્ર	સ્થ્ય			

6.3.3.3.CHCHCHC (Combination of four Consonanats)

This cluster is rare in a majority of languages and the experts have deemed that it is not found in Gujarati

6.3.4 The Collation Order of Gujarati.

Collation is one of the most important features of a script grammar. It determines the order in which a given culture indexes its characters. This is best seen in a dictionary sort where for easy search words are sorted and arranged in a specific order. Within a given script, each allo-script may have a different sort-order. Thus in Devanagari the conjunct glyph क्ष is sorted along with क, since the first letter of that conjunct is क and on a similar principle ज्ञ is sorted along with ज. Different scripts admit different sort orders and for all high-end NLP applications, sort is a crucial feature to ensure that the applications index data as per the cultural perception of that community. In quite a few States, sort order is clearly defined by the statutory bodies of that state and hence it is crucial that such sort order be ascertained and introduced in the script grammar.

In the case of Gujarati the following is the traditional sort order as determined by the experts. The order as given below is pertinent to sorting by a computer program and is compliant with CLDR as laid down by Unicode and W3C.

ઁ ં ઃ અ આ ઇ ઈ ઉ ઊ ઋ

ઍ એ ઐ ઑ ઓ ઔ ક ખ ગ ઘ ઙ

ચ છ જ ઝ ઞ ટ ઠ ડ ઢ ણ

ત થ દ ધ ન પ ફ બ ભ મ

ય ર લ ળ વ શ ષ સ હ ઽ

ા િ ી ુ ૂ ૃ ૅ ે ૈ ૉ ો ૌ

In Tabular format:

ઁ	ં	ઃ	અ	આ	ઇ	ઈ	ઉ	ઊ	ઋ		
ઍ	એ	ઐ	ઑ	ઓ	ઔ	ક	ખ	ગ	ઘ	ઙ	
ચ	છ	જ	ઝ	ઞ	ટ	ઠ	ડ	ઢ	ણ		
ત	થ	દ	ધ	ન	પ	ફ	બ	ભ	મ		
ય	ર	લ	ળ	વ	શ	ષ	સ	હ	ઽ		
ા	િ	ી	ુ	ૂ	ૃ	ૅ	ે	ૈ	ૉ	ો	ૌ

ઌ ૠ are used only for Sanskrit Loans

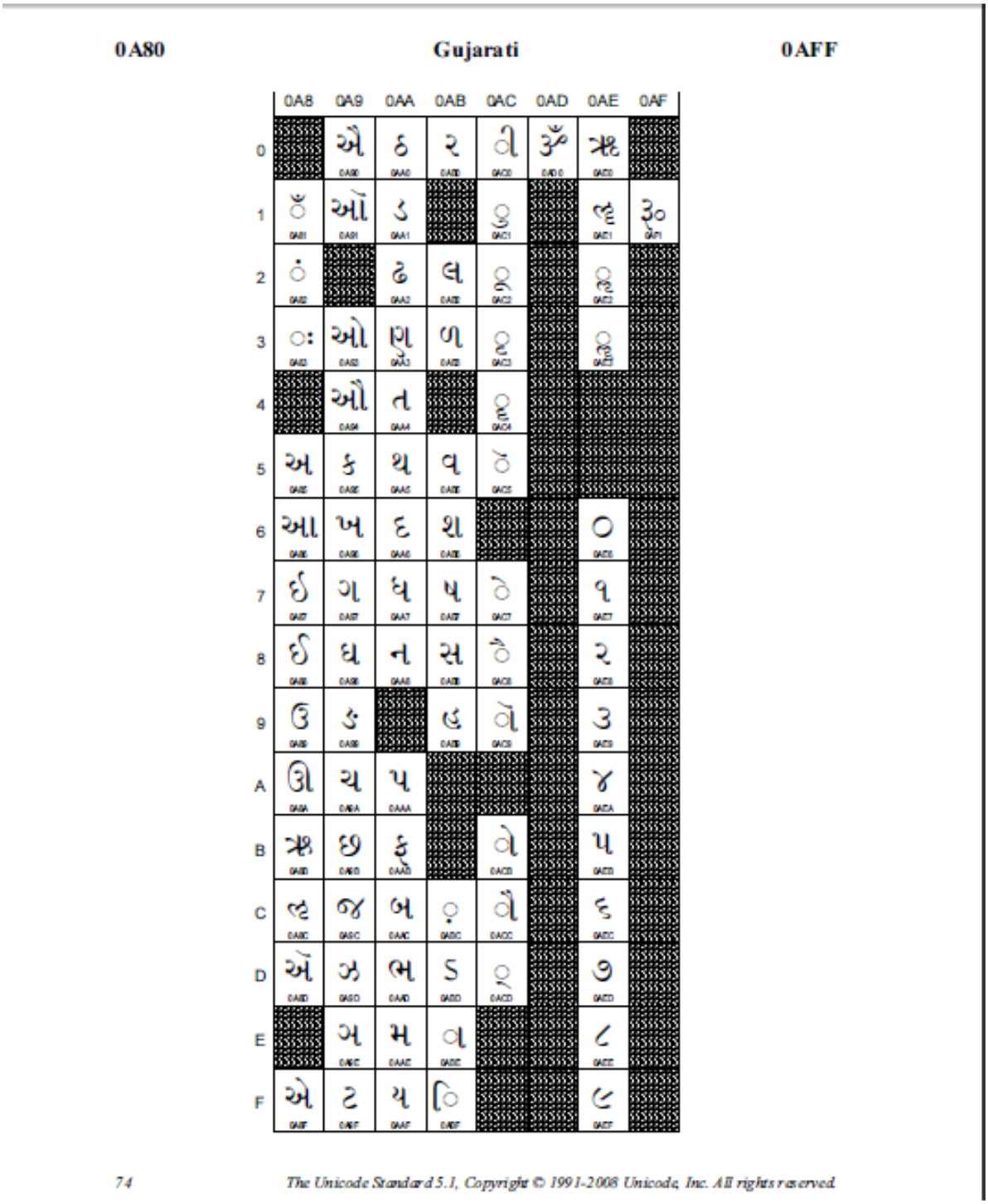
7. REFERENCES

1. http://www.unicode.org
2. ISCII'91

8. ANNEXURES

Annexure 1: Names of experts who have contributed to the script grammar

Annexure 2: Unicode Table of Gujarati[9]



[9] The Unicode chart provided is for version 5.1 since the Script Grammar was prepared at that time. No considerable change in the script grammar can be seen in the updated versions of Unicode, with the possible addition of the Rupee Sign U+02B9

Based on ISCII 1988

Various signs

0A81 ઁ GUJARATI SIGN CANDRABINDU
0A82 ં GUJARATI SIGN ANUSVARA
0A83 ઃ GUJARATI SIGN VISARGA

Independent vowels

0A85 અ GUJARATI LETTER A
0A86 આ GUJARATI LETTER AA
0A87 ઇ GUJARATI LETTER I
0A88 ઈ GUJARATI LETTER II
0A89 ઉ GUJARATI LETTER U
0A8A ઊ GUJARATI LETTER UU
0A8B ઋ GUJARATI LETTER VOCALIC R
0A8C ઌ GUJARATI LETTER VOCALIC L
• used with Sanskrit text
0A8D ઍ GUJARATI VOWEL CANDRA E
0A8E <reserved>
0A8F એ GUJARATI LETTER E
0A90 ઐ GUJARATI LETTER AI
0A91 ઑ GUJARATI VOWEL CANDRA O
0A92 <reserved>
0A93 ઓ GUJARATI LETTER O
0A94 ઔ GUJARATI LETTER AU

Consonants

0A95 ક GUJARATI LETTER KA
0A96 ખ GUJARATI LETTER KHA
0A97 ગ GUJARATI LETTER GA
0A98 ઘ GUJARATI LETTER GHA
0A99 ઙ GUJARATI LETTER NGA
0A9A ચ GUJARATI LETTER CA
0A9B છ GUJARATI LETTER CHA
0A9C જ GUJARATI LETTER JA
0A9D ઝ GUJARATI LETTER JHA
0A9E ઞ GUJARATI LETTER NYA
0A9F ટ GUJARATI LETTER TTA
0AA0 ઠ GUJARATI LETTER TTHA
0AA1 ડ GUJARATI LETTER DDA
0AA2 ઢ GUJARATI LETTER DDHA
0AA3 ણ GUJARATI LETTER NNA
0AA4 ત GUJARATI LETTER TA
0AA5 થ GUJARATI LETTER THA
0AA6 દ GUJARATI LETTER DA
0AA7 ધ GUJARATI LETTER DHA
0AA8 ન GUJARATI LETTER NA
0AA9 <reserved>
0AAA પ GUJARATI LETTER PA
0AAB ફ GUJARATI LETTER PHA
0AAC બ GUJARATI LETTER BA
0AAD ભ GUJARATI LETTER BHA
0AAE મ GUJARATI LETTER MA
0AAF ય GUJARATI LETTER YA
0AB0 ર GUJARATI LETTER RA
0AB1 <reserved>
0AB2 લ GUJARATI LETTER LA
0AB3 ળ GUJARATI LETTER LLA
0AB4 <reserved>
0AB5 વ GUJARATI LETTER VA
0AB6 શ GUJARATI LETTER SHA
0AB7 ષ GUJARATI LETTER SSA
0AB8 સ GUJARATI LETTER SA
0AB9 હ GUJARATI LETTER HA

Various signs

0ABC ઼ GUJARATI SIGN NUKTA
• for extending the alphabet to new letters
0ABD ઽ GUJARATI SIGN AVAGRAHA

Dependent vowel signs

0ABE ા GUJARATI VOWEL SIGN AA
0ABF િ GUJARATI VOWEL SIGN I
• stands to the left of the consonant
0AC0 ી GUJARATI VOWEL SIGN II
0AC1 ુ GUJARATI VOWEL SIGN U
0AC2 ૂ GUJARATI VOWEL SIGN UU
0AC3 ૃ GUJARATI VOWEL SIGN VOCALIC R
0AC4 ૄ GUJARATI VOWEL SIGN VOCALIC RR
0AC5 ૅ GUJARATI VOWEL SIGN CANDRA E
0AC6 <reserved>
0AC7 ે GUJARATI VOWEL SIGN E
0AC8 ૈ GUJARATI VOWEL SIGN AI
0AC9 ૉ GUJARATI VOWEL SIGN CANDRA O
0ACA <reserved>
0ACB ો GUJARATI VOWEL SIGN O
0ACC ૌ GUJARATI VOWEL SIGN AU

Various signs

0ACD ્ GUJARATI SIGN VIRAMA
0ACE <reserved>
0ACF <reserved>
0AD0 ૐ GUJARATI OM

Additional vowels for Sanskrit

0AE0 ૠ GUJARATI LETTER VOCALIC RR
0AE1 ૡ GUJARATI LETTER VOCALIC LL
0AE2 ૢ GUJARATI VOWEL SIGN VOCALIC L
0AE3 ૣ GUJARATI VOWEL SIGN VOCALIC LL

Reserved

For viram punctuation, use the generic Indic 0964 and 0965.

0AE4 <reserved>
→0964 । devanagari danda
0AE5 <reserved>
→0965 ॥ devanagari double danda

Digits

0AE6 ૦ GUJARATI DIGIT ZERO
0AE7 ૧ GUJARATI DIGIT ONE
0AE8 ૨ GUJARATI DIGIT TWO
0AE9 ૩ GUJARATI DIGIT THREE
0AEA ૪ GUJARATI DIGIT FOUR
0AEB ૫ GUJARATI DIGIT FIVE
0AEC ૬ GUJARATI DIGIT SIX
0AED ૭ GUJARATI DIGIT SEVEN
0AEE ૮ GUJARATI DIGIT EIGHT
0AEF ૯ GUJARATI DIGIT NINE

Currency sign

0AF1 ૱ GUJARATI RUPEE SIGN