SCRIPT GRAMMAR FOR PUNJABI 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 Hindi 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 Hindi /la/ is different from the Marathi /la/ in so far as the placement of the stem is concerned Hindi 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 Hindi /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. The collation order for Hindi is different from Marathi although both languages share the same script. Thus in Marathi क़, ज़ are placed at the end of the consonant inventory i.e. after ह in the sort order. In Hindi क़ is sorted along with क and ज़ 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 639.1 & 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 Konkani. 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, Hindi 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. 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: PUNJABI/PANJABI[5]
ISO Mnemonics: *pan*

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 Punjabi is written using the Punjabi 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 क्ल or alternatively form an entirely new shape क्ष = क्ष.

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 Punjabi language

None has been proposed by the experts who have mandated the script grammar.

[5] SIL prefers the term Panjabi. Cf. http://www.sil.org/iso639-3/codes.asp?order=639_3&letter=p

6.2. CONFORMITY TO THE SYLLABLE STRUCTURE

Punjabi 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 Punjabi. The full set of Punjabi is given below

(C) Consonants

ਕ	ਖ	ਗ	ਘ	ਙ
ਚ	ਛ	ਜ	ਝ	ਞ
ਟ	ਠ	ਡ	ਢ	ਣ
ਤ	ਥ	ਦ	ਧ	ਨ
ਪ	ਫ	ਬ	ਭ	ਮ
ਯ	ਰ	ਲ	ਲ਼	ਵ
ਸ	ਹ			

Nukta Consonants[6]:
ਖ਼ GURMUKHI LETTER KHA GURMUKHI SIGN NUKTA
ਗ਼ GURMUKHI LETTER GA GURMUKHI SIGN NUKTA
ਜ਼ GURMUKHI LETTER JA GURMUKHI SIGN NUKTA
ਫ਼ GURMUKHI LETTER PHA GURMUKHI SIGN NUKTA
ਸ਼ GURMUKHI LETTER SHA
ਲ਼ GURMUKHI LETTER LA GURMUKHI SIGN NUKTA
ਕ਼ GURMUKHI LETTER KA GURMUKHI SIGN NUKTA

Flapped form: ੜ

[6] Note :- ਸ਼ is a Perso-Arabic sound. ਲ਼ is not a Perso-Arabic sound. According to Duni Chandar ਕ਼ is Perso-Arabic sound in Gurmukhi.

(V) Vowels

ਅ	ਅ	ਇ	ਈ	ਉ	ਊ	ਏ	ਐ	ਓ	ਔ

(M) Mātrās or Vowel Modifiers

◌ੁ	◌ਿ	◌ੀ	◌ੂ	◌ੇ	◌ੈ	◌ੋ	◌ੌ	

(D) Diacritics

NASALS : Punjabi admits two nasals: Bindi and Tippi whose distribution is as under:

Bindi ◌ਂ Tippi ◌ੰ

Bindi and Tippi are allographic variants of the nasal. The rules for placement of Bindi and Tippi are as under:

1. The initial forms of *u, uu* i.e. vowels take Bindi;
2. Matras *u, uu* after a consonant take Tippi;
3. all other short vowels/matras take Tippi;
4. all other long vowels/matras take Bindi.

Addak Bindi ◌ੱ◌ਂ[7]:Not used in Punjabi

Other diacritics viz. Avagraha, Visarga[8] are not used in Punjabi

H (HALANTA/VIRAMA) ◌੍ The *halanta* character is not used when writing Punjabi in Gurmukhi. However, it may occasionally be used in Sanskritised text or in dictionaries for extra phonetic information. When it is used, it represents the suppression of the inherent vowel.

A (ADDAK) ◌ੱ : Addak is used with geminated consonants. Addak is placed on the preceding syllable. Addak cannot be at the beginning of a word.

N (NUKTA)[9] ◌਼ :Used in Punjabi for the following consonants:

ਖ਼ ਗ਼ ਜ਼ ਫ਼ ਸ਼ ਲ਼ ਕ਼

With the exception of Gurmukhi letter sha, the others are combinations of Consonant+Nukta

[7] Although Unicode provides for the combination of Addak+Bindi U+0A01, the experts have deemed that such a combination is not possible.

[8] The visarg symbol (ः U+0A03) is used very occasionally in Gurmukhi. It can either represent an abbreviation (like period is used in English) or it can act like a Sanskrit Visarga where a voiceless h sound is pronounced after the vowel. (cf. Wiki: http://en.wikipedia.org/wiki/Gurmukhi_script)

[9] The nukta is a small dot placed under a character in certain scripts to show that they are flapped or for deriving consonants required for Urdu Loan words: क़,ख़,ग़,ज़,फ़

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**	N,M,D,H
C	**N**	C,M,D,H
-	**V**	D
C, N	**M**	D
C, N,V,M	**D**	-
C, N	**H**	C
C, N,V,M	**A**	Next syllable

C can be preceded by H or no subtype and followed by any one of the following: M,D,H
N can be preceded by C and followed by any one of the following: N,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.
Addak is used exclusively for gemmination of consonants. It has a peculiar property in that it is placed over the preceding syllable, although it is part of the next syllable
Eg. ਕੱਕਾ

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 weak vowel or a mātrā)
The four other subsets viz. Mātrās, Vowel Modifiers, Halanta and Nukta cannot constitute a syllable (akshar) by themselves or in combination among themselves.
1. The Vowel syllable (akshar) is of the following types:

1.1. A pure vowel all by itself: , ਅ/a/ ਆ /ā/ etc.

1.2. A vowel followed by a modifier i.e. either a nasal marker : ਅੰ / ã/

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

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

2.2. A consonant[10] (with or without Nukta) followed by a mātrā i.e. the inherent vowel being substituted by another vowel: ਕਿ /ki:/

2.3. A consonant (with or without Nukta) followed by a modifier:ਕੰ /kə̃/[11]/

[10] For purposes of Simplification, C here will automatically be treated as being also consonant+nukta C+ N
[11] This character represents phonetically the weak implicit vowel, termed as schwa and often shown as /a/ also.

2.4. A consonant (with or without Nukta) followed by a mātrā and a modifier: ਕੂੰ /kũ/

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.*

ਪ੍ਰਬੰਧ (prəbənd), ਸ੍ਵਰਗ (svərəg)

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" (secondary derivative from the adjective कृत्स्न /kṛtsna/ meaning "whole, complete".)

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[12]	CM	CD	CMD
ACC/CHC[13]	ACCM/CHCM	ACCD/CHCD	ACCMD/CHCMD
-	-	-	-
-	-	-	-

A total number of 8 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.

Since the formal structure script grammar of the syllable (akshar) is common to all Brahmi based scripts, it will not be treated in the sample template, but it will form the basis of an exhaustive description of the characters as well as their ligatural representations.

[12] C here will automatically be treated as being also consonant+nukta, C+N to simplify the explanation
[13] The combination CHC is also assumed to represent geminate clusters shown as A

6.3 SCRIPT GRAMMAR PROPER

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

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

6.3.1. The Character Set of Punjabi.

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 Punjabi.
This comprises the following:
6.3.1.1. The Consonant Set
6.3.1.2. The Vowel Set[9]
6.3.1.3. The Mātrā Set[9]
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 Punjabi comprises the following characters:

A basic Consonant inventory arranged as per their Vargas.

	STOPS				NASAL	FRICATIVE
	-voiced -aspirated	-voiced +aspirated	+voiced -aspirated	+voiced +aspirated	+voiced -aspirated	-+/voiced -aspirated
Pharyngal	ਕ਼		ਖ਼	ਗ਼		
Velar	ਕ	ਖ	ਗ	ਘ	ਙ	
Palatal	ਚ	ਛ	ਜ	ਝ	ਞ	ਜ਼
Retroflex	ਟ	ਠ	ਡ	ਢ	ਣ	
Flaps			ੜ			
Dental	ਤ	ਥ	ਦ	ਧ	ਨ	
B-labial	ਪ	ਫ	ਬ	ਭ	ਮ	ਫ਼

Other consonants

ਯ	ਰ	ਲ	ਲ਼	ਵ
ਸ਼	ਸ	ਹ		

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

	STOPS				NASAL	FRICATIVE
	-voiced -aspirated	-voiced +aspirated	+voiced -aspirated	+voiced +aspirated	+voiced -aspirated	-+/voiced -aspirated
Pharyngal	ਕ਼		ਖ਼	ਗ਼		
Velar	ਕ	ਖ	ਗ	ਘ	ਙ	
Palatal	ਚ	ਛ	ਜ	ਝ	ਞ	ਜ਼
Retroflex	ਟ	ਠ	ਡ	ਢ	ਣ	
Flaps			ੜ			
Dental	ਤ	ਥ	ਦ	ਧ	ਨ	
B-labial	ਪ	ਫ	ਬ	ਭ	ਮ	ਫ਼

Other consonants

ਯ	ਰ	ਲ	ਲ਼	ਵ
ਸ਼	ਸ	ਹ		

6.3.1.2. The Vowel Set

The Vowel set of Punjabi is as under:

ਅ	GURMUKHI LETTER A	Mukta
ਆ	GURMUKHI LETTER AA	Kanna
ਇ	GURMUKHI LETTER I	Sihari
ਈ	GURMUKHI LETTER II	Bihari

ੁ	GURMUKHI LETTER U	Onkaṛ
ੂ	GURMUKHI LETTER UU	Dulankaṛ
ੇ	GURMUKHI LETTER EE	Lava
ੈ	GURMUKHI LETTER AI	Dulava
ੋ	GURMUKHI LETTER OO	Hora
ੌ	GURMUKHI LETTER AU	Kanora

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

ਅ	ਆ	ਇ	ਈ	ੁ	ੂ	ੇ	ੈ	ੋ	ੌ

6.3.1.3. The Mātrā Set

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

Mātrā Sign	Mātrā Name	Corresponding vowel	Consonant Shapes formed
ਾ	GURMUKHI VOWEL SIGN AA	ਆ	ਕ+ਾ=ਕਾ
ਿ	GURMUKHI VOWEL SIGN I	ਇ	ਕ+ਿ=ਕਿ
ੀ	GURMUKHI VOWEL SIGN II	ਈ	ਕ+ੀ=ਕੀ
ੁ	GURMUKHI VOWEL SIGN U	ੁ	ਕ+ੁ=ਕੁ
ੂ	GURMUKHI VOWEL SIGN UU	ੂ	ਕ+ੂ=ਕੂ
ੇ	GURMUKHI VOWEL SIGN EE	ਏ	ਕ+ੇ=ਕੇ
ੈ	GURMUKHI VOWEL SIGN AI	ਐ	ਕ+ੈ=ਕੈ
ੋ	GURMUKHI VOWEL SIGN OO	ਓ	ਕ+ੋ=ਕੋ

ੌ	GURMUKHI VOWEL SIGN AU	ਔ	ਕ+ੌ=ਕੌ

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). Punjabi has a single displaced catenator.

CATENATOR	POSITION	EXAMPLE
ਿ	consonant	ਕਿ

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

In Indian scripts derived from Brahmi the ra consonant when conjoined to a consonant, 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. When it is adjoined to the consonant by means of a halanta it constitutes a rakar. Punjabi has no rafars

However Punjabi admits rakars, which form a kind of "subscript" i.e. the ra is placed below the consonant to which it relates:

ਜਰ ਜ੍ਰ

Examples of words where Rakars are used in Punjabi language are given below:

ਪ੍	ਪ੍ਰਬੰਧ
ਗ੍	ਉਪਗ੍ਰਹਿ

6.3.1.6. Diacritics

These are as under in the case of Punjabi:

NASALS : Punjabi admits two nasals: Bindi and Tippi whose distribution is as under:

Bindi ਂ Tippi ੰ:

Bindi and Tippi are allographic variants of the nasal. The rules for placement of Bindi and Tippi are as under:

1. The initial forms of *u, uu* i.e. vowels take Bindi;
2. Matras *u, uu* after a consonant take Tippi;
3. all other short vowels/matras take Tippi;
4. all other long vowels/matras take Bindi.

Other diacritics viz. Avagraha, Visarga[14] are not used in Punjabi

H (HALANTA/VIRAMA) ◌੍ The *halanta* character is not used when writing Punjabi in Gurmukhi. However, it may occasionally be used in Sanskritised text or in dictionaries for extra phonetic information. When it is used, it represents the suppression of the inherent vowel.

A (ADDAK) ◌ੱ : Addak is used with geminated consonants. Addak is placed on the preceding syllable. Addak cannot be at the beginning of a word.

6.3.1.7. Numerals

The International number set (Latino-Arabic set: 0,1,2,3,4,5,6,7,8,9) is used in official documents in Punjabi.
Following are the numbers used in Punjabi language.
੦ ੧ ੨ ੩ ੪ ੫ ੬ ੭ ੮ ੯

Numeral Shapes	Explanation
੦	Punjabi Digit Zero
੧	Punjabi Digit One
੨	Punjabi Digit Two
੩	Punjabi Digit Three
੪	Punjabi Digit Four
੫	Punjabi Digit Five
੬	Punjabi Digit Six
੭	Punjabi Digit Seven
੮	Punjabi Digit Eight

[14] The visarg symbol (ः U+0A03) is used very occasionally in Gurmukhi. It can either represent an abbreviation (like period is used in English) or it can act like a Sanskrit Visarga where a voiceless h sound is pronounced after the vowel. (cf. Wiki: http://en.wikipedia.org/wiki/Gurmukhi_script)

੯	Punjabi Digit Nine

In addition Punjabi admits the following Fraction Signs[15]:

੧	PUNJABI FRACTION ONE QUARTER
੫	PUNJABI FRACTION ONE HALF
੭	PUNJABI FRACTION THREE QUARTERS
ੴ	PUNJABI FRACTION ONE SIXTEENTH
੯	PUNJABI FRACTION ONE EIGHTH
੩	PUNJABI FRACTION THREE SIXTEENTHS

[15] These are rarely used in every day use

6.3.1.8. Punctuation Markers

Punjabi uses punctuation markers from the Latin set. such as . , ; : “ ‘ ( ) [ ] etc.
Purna and Deergha Virama (full-stop/danda) Devanagari code block: U+0964, U+0965 .
The first is used as a Punctuation marker. The second is used in religious texts as in the example below from the Sri Guru Granth Sahib:

ੴ ਸਤਿ ਨਾਮੁ ਕਰਤਾ ਪੁਰਖੁ ਨਿਰਭਉ ਨਿਰਵੈਰੁ ਅਕਾਲ ਮੂਰਤਿ ਅਜੂਨੀ ਸੈਭੰ ਗੁਰ ਪ੍ਰਸਾਦਿ ॥

ਪੰਨਾ 1, ਸਤਰ 1

A list of punctuations is provided below:

Sr. No.	Name of the marker	Marker Shape
1.	Full Stop or Period	.
2.	Question Mark	?
3.	Comma	,
4.	Exclamation Mark	!
5.	Apostrophe	’
6.	Semi Colon	;
7.	Colon	:
8.	Hyphen	-
9.	Dash	--
10.	Ellipsis mark	...
11.	Oblique	/
12.	Double quotation mark	" "
13.	Single quotation mark	‘ ‘
14.	Cross	XXX
15.	As Above	- - " - -
16.	Round Brackets	( )
17.	Square Brackets	[ ]
18.	Curly Brackets	{ }
19.	Devanagari Purna Virama ਡੰਡੀ	।
20.	Devanagari Deergha Virama	॥

6.3.1.9 Other Symbols

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

YAKASH	**U+0A75**
EK ONKAR ੧ੴ	**U+0A74**
UDATTA	**U+0A51**

₹: Rupee Sign as mandated by Government of India.

Fraction signs are listed in 6.3.1.7 above

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:	ਕ	ਖ	ਗ	ਘ	ਙ	ਚ	ਛ	ਜ	ਝ	ਞ	
Table 2:	ਟ	ਠ	ਡ	ਢ	ਣ	ਤ	ਥ	ਦ	ਧ	ਨ	
Table 3:	ਪ	ਫ	ਬ	ਭ	ਮ	ਯ	ਰ	ਲ	ਲ਼	ਵ	ਸ਼
	ਸ	ਹ									
Table 4:	ਫ਼	ਖ਼	ਗ਼	ਜ਼	ੜ						

Wherever there is an X it implies that the combination does not exist. For the font developer this is an indication that for this particular combination which is not possible in the language but needs to be accommodated in the font table, a simple linear combination be provided.

e.g. Although the combination of ਙ +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: ਙਾ

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. These are with Bindi and Tippi.

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

	ਕ	ਖ	ਗ	ਘ	ਙ	ਚ	ਛ	ਜ	ਝ	ਞ
ਾ	ਕਾ	ਖਾ	ਗਾ	ਘਾ	ਙਾ	ਚਾ	ਛਾ	ਜਾ	ਝਾ	ਞਾ
ਿ	ਕਿ	ਖਿ	ਗਿ	ਘਿ	ਙਿ	ਚਿ	ਛਿ	ਜਿ	ਝਿ	ਞਿ
ੀ	ਕੀ	ਖੀ	ਗੀ	ਘੀ	ਙੀ	ਚੀ	ਛੀ	ਜੀ	ਝੀ	ਞੀ
ੁ	ਕੁ	ਖੁ	ਗੁ	ਘੁ	ਙੁ	ਚੁ	ਛੁ	ਜੁ	ਝੁ	ਞੁ
ੂ	ਕੂ	ਖੂ	ਗੂ	ਘੂ	ਙੂ	ਚੂ	ਛੂ	ਜੂ	ਝੂ	ਞੂ
ੇ	ਕੇ	ਖੇ	ਗੇ	ਘੇ	ਙੇ	ਚੇ	ਛੇ	ਜੇ	ਝੇ	ਞੇ
ੈ	ਕੈ	ਖੈ	ਗੈ	ਘੈ	ਙੈ	ਚੈ	ਛੈ	ਜੈ	ਝੈ	ਞੈ
ੋ	ਕੋ	ਖੋ	ਗੋ	ਘੋ	ਙੋ	ਚੋ	ਛੋ	ਜੋ	ਝੋ	ਞੋ
ੌ	ਕੌ	ਖੌ	ਗੌ	ਘੌ	ਙੌ	ਚੌ	ਛੌ	ਜੌ	ਝੌ	ਞੌ

Remark 1- ਙ and ਞ are rarely used only as the first members of clusters and mostly as alternatives of Anuswara

Consonant and Mātrā combinations Set 2

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

	ਟ	ਠ	ਡ	ਢ	ਣ	ਤ	ਥ	ਦ	ਧ	ਨ
ਾ	ਟਾ	ਠਾ	ਡਾ	ਢਾ	ਣਾ	ਤਾ	ਥਾ	ਦਾ	ਧਾ	ਨਾ
ਿ	ਟਿ	ਠਿ	ਡਿ	ਢਿ	ਣਿ	ਤਿ	ਥਿ	ਦਿ	ਧਿ	ਨਿ
ੀ	ਟੀ	ਠੀ	ਡੀ	ਢੀ	ਣੀ	ਤੀ	ਥੀ	ਦੀ	ਧੀ	ਨੀ
ੁ	ਟੁ	ਠੁ	ਡੁ	ਢੁ	ਣੁ	ਤੁ	ਥੁ	ਦੁ	ਧੁ	ਨੁ
ੂ	ਟੂ	ਠੂ	ਡੂ	ਢੂ	ਣੂ	ਤੂ	ਥੂ	ਦੂ	ਧੂ	ਨੂ
ੇ	ਟੇ	ਠੇ	ਡੇ	ਢੇ	ਣੇ	ਤੇ	ਥੇ	ਦੇ	ਧੇ	ਨੇ
ੈ	ਟੈ	ਠੈ	ਡੈ	ਢੈ	ਣੈ	ਤੈ	ਥੈ	ਦੈ	ਧੈ	ਨੈ
ੋ	ਟੋ	ਠੋ	ਡੋ	ਢੋ	ਣੋ	ਤੋ	ਥੋ	ਦੋ	ਧੋ	ਨੋ
ੌ	ਟੌ	ਠੌ	ਡੌ	ਢੌ	ਣੌ	ਤੌ	ਥੌ	ਦੌ	ਧੌ	ਨੌ

Consonant and Mātrā combinations Set 3

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

	ਪ	ਫ	ਬ	ਭ	ਮ	ਯ	ਰ	ਲ	ਲ਼	ਵ	ਸ਼	ਸ	ਹ
◌ਾ	ਪਾ	ਫਾ	ਬਾ	ਭਾ	ਮਾ	ਯਾ	ਰਾ	ਲਾ	ਲ਼ਾ	ਵਾ	ਸ਼ਾ	ਸਾ	ਹਾ
◌ਿ	ਪਿ	ਫਿ	ਬਿ	ਭਿ	ਮਿ	ਯਿ	ਰਿ	ਲਿ	ਲ਼ਿ	ਵਿ	ਸ਼ਿ	ਸਿ	ਹਿ
◌ੀ	ਪੀ	ਫੀ	ਬੀ	ਭੀ	ਮੀ	ਯੀ	ਰੀ	ਲੀ	ਲ਼ੀ	ਵੀ	ਸ਼ੀ	ਸੀ	ਹੀ
◌ੁ	ਪੁ	ਫੁ	ਬੁ	ਭੁ	ਮੁ	ਯੁ	ਰੁ	ਲੁ	ਲ਼ੁ	ਵੁ	ਸ਼ੁ	ਸੁ	ਹੁ
◌ੂ	ਪੂ	ਫੂ	ਬੂ	ਭੂ	ਮੂ	ਯੂ	ਰੂ	ਲੂ	ਲ਼ੂ	ਵੂ	ਸ਼ੂ	ਸੂ	ਹੂ
◌ੇ	ਪੇ	ਫੇ	ਬੇ	ਭੇ	ਮੇ	ਯੇ	ਰੇ	ਲੇ	ਲ਼ੇ	ਵੇ	ਸ਼ੇ	ਸੇ	ਹੇ
◌ੈ	ਪੈ	ਫੈ	ਬੈ	ਭੈ	ਮੈ	ਯੈ	ਰੈ	ਲੈ	ਲ਼ੈ	ਵੈ	ਸ਼ੈ	ਸੈ	ਹੈ
◌ੋ	ਪੋ	ਫੋ	ਬੋ	ਭੋ	ਮੋ	ਯੋ	ਰੋ	ਲੋ	ਲ਼ੋ	ਵੋ	ਸ਼ੋ	ਸੋ	ਹੋ
◌ੌ	ਪੌ	ਫੌ	ਬੌ	ਭੌ	ਮੌ	ਯੌ	ਰੌ	ਲੌ	ਲ਼ੌ	ਵੌ	ਸ਼ੌ	ਸੌ	ਹੌ

Consonant and Mātrā combinations Set 4

This set is in continuation of set 3 which shows consonant and Matra combinations.

	ਛ	ਖ਼	ਗ਼	ਜ਼	ੜ
◌	ਛਾ	ਖ਼ਾ	ਗ਼ਾ	ਜ਼ਾ	ੜਾ
◌ਿ	ਛਿ	ਖ਼ਿ	ਗ਼ਿ	ਜ਼ਿ	ੜਿ
◌ੀ	ਛੀ	ਖ਼ੀ	ਗ਼ੀ	ਜ਼ੀ	ੜੀ
◌ੁ	ਛੁ	ਖ਼ੁ	ਗ਼ੁ	ਜ਼ੁ	ੜੁ
◌ੂ	ਛੂ	ਖ਼ੂ	ਗ਼ੂ	ਜ਼ੂ	ੜੂ
◌ੇ	ਛੇ	ਖ਼ੇ	ਗ਼ੇ	ਜ਼ੇ	ੜੇ
◌ੈ	ਛੈ	ਖ਼ੈ	ਗ਼ੈ	ਜ਼ੈ	ੜੈ
◌ੋ	ਛੋ	ਖ਼ੋ	ਗ਼ੋ	ਜ਼ੋ	ੜੋ
◌ੌ	ਛੌ	ਖ਼ੌ	ਗ਼ੌ	ਜ਼ੌ	ੜੌ

6.3.2.2 Consonant and Mātrā +Nasal combinations.

This set refers to a Consonant and Mātrā + Nasal marker combinations. Since The use of Tippi and Bindi is rule governed, column 1 shows both tippi and bindi. Combinations marked in red have been deemed by experts as “dead combinations” i.e. not pertinent in Punjabi language.

Consonant and Mātrā + Nasal combinations - Set 1

	ਕ	ਖ	ਗ	ਘ	ਙ	ਚ	ਛ	ਜ	ਝ	ਞ
◌ਂ / ◌ੰ	ਕੰ	ਖੰ	ਗੰ	ਘੰ	ਙੰ	ਚੰ	ਛੰ	ਜੰ	ਝੰ	ਞੰ
◌ਾਂ / ◌ੰ	ਕਾਂ	ਖਾਂ	ਗਾਂ	ਘਾਂ	ਙਾਂ	ਚਾਂ	ਛਾਂ	ਜਾਂ	ਝਾਂ	ਞਾਂ
◌ਿਂ / ◌ੰ	ਕਿਂ	ਖਿਂ	ਗਿਂ	ਘਿਂ	ਙਿਂ	ਚਿਂ	ਛਿਂ	ਜਿਂ	ਝਿਂ	ਞਿਂ
◌ੀਂ / ◌ੰ	ਕੀਂ	ਖੀਂ	ਗੀਂ	ਘੀਂ	ਙੀਂ	ਚੀਂ	ਛੀਂ	ਜੀਂ	ਝੀਂ	ਞੀਂ
◌ੁਂ / ◌ੰ	ਕੁਂ	ਖੁਂ	ਗੁਂ	ਘੁਂ	ਙੁਂ	ਚੁਂ	ਛੁਂ	ਜੁਂ	ਝੁਂ	ਞੁਂ
◌ੂਂ / ◌ੰ	ਕੂਂ	ਖੂਂ	ਗੂਂ	ਘੂਂ	ਙੂਂ	ਚੂਂ	ਛੂਂ	ਜੂਂ	ਝੂਂ	ਞੂਂ
◌ੇਂ / ◌ੰ	ਕੇਂ	ਖੇਂ	ਗੇਂ	ਘੇਂ	ਙੇਂ	ਚੇਂ	ਛੇਂ	ਜੇਂ	ਝੇਂ	ਞੇਂ
◌ੈਂ / ◌ੰ	ਕੈਂ	ਖੈਂ	ਗੈਂ	ਘੈਂ	ਙੈਂ	ਚੈਂ	ਛੈਂ	ਜੈਂ	ਝੈਂ	ਞੈਂ
◌ੋਂ / ◌ੰ	ਕੋਂ	ਖੋਂ	ਗੋਂ	ਘੋਂ	ਙੋਂ	ਚੋਂ	ਛੋਂ	ਜੋਂ	ਝੋਂ	ਞੋਂ
◌ੌਂ / ◌ੰ	ਕੌਂ	ਖੌਂ	ਗੌਂ	ਘੌਂ	ਙੌਂ	ਚੌਂ	ਛੌਂ	ਜੌਂ	ਝੌਂ	ਞੌਂ

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. Since The use of Tippi and Bindi is rule governed, column 1 shows both tippi and bindi. Combinations marked in red have been deemed by experts as "dead combinations" i.e. not pertinent in Punjabi language.

	ਟ	ਠ	ਡ	ਢ	ਲ	ਤ	ਥ	ਦ	ਧ	ਨ
◌ਂ / ◌ੰ	ਟੰ	ਠੰ	ਡੰ	ਢੰ	ਲੰ	ਤੰ	ਥੰ	ਦੰ	ਧੰ	ਨੰ
◌ਾਂ / ◌ੰ	ਟਾਂ	ਠਾਂ	ਡਾਂ	ਢਾਂ	ਲਾਂ	ਤਾਂ	ਥਾਂ	ਦਾਂ	ਧਾਂ	ਨਾਂ
◌ਿਂ / ◌ੰ	ਟਿੰ	ਠਿੰ	ਡਿੰ	ਢਿੰ	ਲਿੰ	ਤਿੰ	ਥਿੰ	ਦਿੰ	ਧਿੰ	ਨਿੰ
◌ੀਂ / ◌ੰ	ਟੀਂ	ਠੀਂ	ਡੀਂ	ਢੀਂ	ਲੀਂ	ਤੀਂ	ਥੀਂ	ਦੀਂ	ਧੀਂ	ਨੀਂ
◌ੁਂ / ◌ੰ	ਟੁਂ	ਠੁਂ	ਡੁਂ	ਢੁਂ	ਲੁਂ	ਤੁਂ	ਥੁਂ	ਦੁਂ	ਧੁਂ	ਨੁਂ
◌ੂਂ / ◌ੰ	ਟੂਂ	ਠੂਂ	ਡੂਂ	ਢੂਂ	ਲੂਂ	ਤੂਂ	ਥੂਂ	ਦੂਂ	ਧੂਂ	ਨੂਂ
◌ੇਂ / ◌ੰ	ਟੇਂ	ਠੇਂ	ਡੇਂ	ਢੇਂ	ਲੇਂ	ਤੇਂ	ਥੇਂ	ਦੇਂ	ਧੇਂ	ਨੇਂ
◌ੈਂ / ◌ੰ	ਟੈਂ	ਠੈਂ	ਡੈਂ	ਢੈਂ	ਲੈਂ	ਤੈਂ	ਥੈਂ	ਦੈਂ	ਧੈਂ	ਨੈਂ
◌ੋਂ / ◌ੰ	ਟੋਂ	ਠੋਂ	ਡੋਂ	ਢੋਂ	ਲੋਂ	ਤੋਂ	ਥੋਂ	ਦੋਂ	ਧੋਂ	ਨੋਂ
◌ੌਂ / ◌ੰ	ਟੌਂ	ਠੌਂ	ਡੌਂ	ਢੌਂ	ਲੌਂ	ਤੌਂ	ਥੌਂ	ਦੌਂ	ਧੌਂ	ਨੌਂ

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. Since The use of Tippi and Bindi is rule governed, column 1 shows both tippi and bindi. Combinations marked in red have been deemed by experts as "dead combinations" i.e. not pertinent in Punjabi language.

	ਪ	ਫ	ਬ	ਭ	ਮ	ਯ	ਰ	ਲ	ਲ਼	ਵ	ਸ਼	ਸ	ਹ
◌ਂ / ◌ੰ	ਪੰ	ਫੰ	ਬੰ	ਭੰ	ਮੰ	ਯੰ	ਰੰ	ਲੰ	ਲ਼ੰ	ਵੰ	ਸ਼ੰ	ਸੰ	ਹੰ
◌ਾਂ / ◌ੰ	ਪਾਂ	ਫਾਂ	ਬਾਂ	ਭਾਂ	ਮਾਂ	ਯਾਂ	ਰਾਂ	ਲਾਂ	ਲ਼ਾਂ	ਵਾਂ	ਸ਼ਾਂ	ਸਾਂ	ਹਾਂ
◌ਿਂ / ◌ੰ	ਪਿੰ	ਫਿੰ	ਬਿੰ	ਭਿੰ	ਮਿੰ	ਯਿੰ	ਰਿੰ	ਲਿੰ	ਲ਼ਿੰ	ਵਿੰ	ਸ਼ਿੰ	ਸਿੰ	ਹਿੰ
◌ੀਂ / ◌ੰ	ਪੀਂ	ਫੀਂ	ਬੀਂ	ਭੀਂ	ਮੀਂ	ਯੀਂ	ਰੀਂ	ਲੀਂ	ਲ਼ੀਂ	ਵੀਂ	ਸ਼ੀਂ	ਸੀਂ	ਹੀਂ
◌ੁਂ / ◌ੰ	ਪੁੰ	ਫੁੰ	ਬੁੰ	ਭੁੰ	ਮੁੰ	ਯੁੰ	ਰੁੰ	ਲੁੰ	ਲ਼ੁੰ	ਵੁੰ	ਸ਼ੁੰ	ਸੁੰ	ਹੁੰ
◌ੂਂ / ◌ੰ	ਪੂੰ	ਫੂੰ	ਬੂੰ	ਭੂੰ	ਮੂੰ	ਯੂੰ	ਰੂੰ	ਲੂੰ	ਲ਼ੂੰ	ਵੂੰ	ਸ਼ੂੰ	ਸੂੰ	ਹੂੰ
◌ੇਂ / ◌ੰ	ਪੇਂ	ਫੇਂ	ਬੇਂ	ਭੇਂ	ਮੇਂ	ਯੇਂ	ਰੇਂ	ਲੇਂ	ਲ਼ੇਂ	ਵੇਂ	ਸ਼ੇਂ	ਸੇਂ	ਹੇਂ
◌ੈਂ / ◌ੰ	ਪੈਂ	ਫੈਂ	ਬੈਂ	ਭੈਂ	ਮੈਂ	ਯੈਂ	ਰੈਂ	ਲੈਂ	ਲ਼ੈਂ	ਵੈਂ	ਸ਼ੈਂ	ਸੈਂ	ਹੈਂ
◌ੋਂ / ◌ੰ	ਪੋਂ	ਫੋਂ	ਬੋਂ	ਭੋਂ	ਮੋਂ	ਯੋਂ	ਰੋਂ	ਲੋਂ	ਲ਼ੋਂ	ਵੋਂ	ਸ਼ੋਂ	ਸੋਂ	ਹੋਂ
◌ੌਂ / ◌ੰ	ਪੌਂ	ਫੌਂ	ਬੌਂ	ਭੌਂ	ਮੌਂ	ਯੌਂ	ਰੌਂ	ਲੌਂ	ਲ਼ੌਂ	ਵੌਂ	ਸ਼ੌਂ	ਸੌਂ	ਹੌਂ

Consonant and Mātrā +Nasal combinations - Set 4

This set is in continuation of set 3 above which shows combinations of Consonant and Mātrā + Nasal marker. Since The use of Tippi and Bindi is rule governed, column 1 shows both tippi and bindi. Combinations marked in red have been deemed by experts as "dead combinations" i.e. not pertinent in Punjabi language.

	ਖ਼	ਗ਼	ਜ਼	ਫ਼	ਲ਼
◌ਂ / ◌ੰ	ਖ਼ੰ	ਗ਼ੰ	ਜ਼ੰ	ਫ਼ੰ	ਲ਼ੰ
◌ਾਂ / ◌ੰ	ਖ਼ਾਂ	ਗ਼ਾਂ	ਜ਼ਾਂ	ਫ਼ਾਂ	ਲ਼ਾਂ
◌ਿਂ / ◌ੰ	ਖ਼ਿੰ	ਗ਼ਿੰ	ਜ਼ਿੰ	ਫ਼ਿੰ	ਲ਼ਿੰ
◌ੀਂ / ◌ੰ	ਖ਼ੀਂ	ਗ਼ੀਂ	ਜ਼ੀਂ	ਫ਼ੀਂ	ਲ਼ੀਂ
◌ੁਂ / ◌ੰ	ਖ਼ੁਂ	ਗ਼ੁਂ	ਜ਼ੁਂ	ਫ਼ੁਂ	ਲ਼ੁਂ
◌ੂਂ / ◌ੰ	ਖ਼ੂਂ	ਗ਼ੂਂ	ਜ਼ੂਂ	ਫ਼ੂਂ	ਲ਼ੂਂ
◌ੇਂ / ◌ੰ	ਖ਼ੇਂ	ਗ਼ੇਂ	ਜ਼ੇਂ	ਫ਼ੇਂ	ਲ਼ੇਂ
◌ੈਂ / ◌ੰ	ਖ਼ੈਂ	ਗ਼ੈਂ	ਜ਼ੈਂ	ਫ਼ੈਂ	ਲ਼ੈਂ
◌ੋਂ / ◌ੰ	ਖ਼ੋਂ	ਗ਼ੋਂ	ਜ਼ੋਂ	ਫ਼ੋਂ	ਲ਼ੋਂ
◌ੌਂ / ◌ੰ	ਖ਼ੌਂ	ਗ਼ੌਂ	ਜ਼ੌਂ	ਫ਼ੌਂ	ਲ਼ੌਂ

6.3.3. The Ligature Set of Punjabi.

Punjabi clusters are based on the particularity of Punjabi script. When a "dead consonant" is joined to a "live consonant", two possibilities occur:
1. Gemination: This occurs only when both consonants are similar, in which case the result is "addak+Consonant i.e. a geminated consonant. Addak cannot occur at the beginning of a word and in intraword position, the addak spans across syllables i.e. the addak is placed above the character of the preceding syllable:
e.g. ੱਕ ਛ: Invalid ਕੱਕ: Valid

Since there is no change in shape per se, these are not represented in the table
2. True Ligatures: Apart from the Geminated consonants, PUNJABI only the following ligatures. These are regularly formed with three letters only: ਰ(r), ਵ(v) and ਹ(h).

a. ਰ(r) and ਵ(v) are used to make consonant clusters and behave similarly.

Examples:
ਕਰ /k*ər*/ *vs.* ਕ੍ਰਮਵਾਰ /k*rəmvar*/ ਸਵੇਰ /*səvēr*/ vs. ਸ੍ਵਰਗ /***svə****rəg*/

b. ਹ(h) is a tone marker and raises the tone. Subjoined ਹ(h): behaves the same way as the regular ਹ(h) in non-word-initial positions. The regular ਹ(h) is pronounced at the beginning of words but not in other positions, where it instead raises the tone. The difference in usage is that the regular ਹ is used after vowels and the subscript version when there is no vowel, and is attached to consonants.

All other ligatural forms are theoretically possible but do NOT exist in Punjabi language. In what follows therefore ligatures formed by combinations of ਰ(r), ਵ(v) and ਹ(h) with the consonants will be provided:

	ਰ	ਵ	ਹ
ਕ੍	ਕ੍ਰ	ਕ੍ਵ	ਹ੍ਹ
ਖ੍	ਖ੍ਰ	ਖ੍ਵ	ਖ੍ਹ
ਗ੍	ਗ੍ਰ	ਗ੍ਵ	ਗ੍ਹ
ਘ੍	ਘ੍ਰ	ਘ੍ਵ	ਘ੍ਹ
ਙ੍	ਙ੍ਰ	ਙ੍ਵ	ਙ੍ਹ
ਚ੍	ਚ੍ਰ	ਚ੍ਵ	ਚ੍ਹ
ਛ੍	ਛ੍ਰ	ਛ੍ਵ	ਛ੍ਹ
ਜ੍	ਜ੍ਰ	ਜ੍ਵ	ਜ੍ਹ

ਝ਼	ਝ਼	ਝ਼	ਝ਼
ਫ਼	ਫ਼	ਫ਼	ਫ਼
ਟ਼	ਟ਼	ਟ਼	ਟ਼
ਠ਼	ਠ਼	ਠ਼	ਠ਼
ਡ਼	ਡ਼	ਡ਼	ਡ਼
ਢ਼	ਢ਼	ਢ਼	ਢ਼
ਲ਼	ਲ਼	ਲ਼	ਲ਼
ੜ	ੜ	ੜ	ੜ
ਥ਼	ਥ਼	ਥ਼	ਥ਼
ਦ਼	ਦ਼	ਦ਼	ਦ਼
ਧ਼	ਧ਼	ਧ਼	ਧ਼
ਨ਼	ਨ਼	ਨ਼	ਨ਼
ਪ਼	ਪ਼	ਪ਼	ਪ਼
ਢ਼	ਢ਼	ਢ਼	ਢ਼
ਬ਼	ਬ਼	ਬ਼	ਬ਼
ਭ਼	ਭ਼	ਭ਼	ਭ਼
ਮ਼	ਮ਼	ਮ਼	ਮ਼
ਯ਼	ਯ਼	ਯ਼	ਯ਼
ਲ਼	ਲ਼	ਲ਼	ਲ਼
ਲ਼	ਲ਼	ਲ਼	ਲ਼
ਵ਼	ਵ਼	ਵ਼	ਵ਼
ਸ਼	ਸ਼	ਸ਼	ਸ਼
ਸ਼	ਸ਼	ਸ਼	ਸ਼
ਹ਼	ਹ਼	ਹ਼	ਹ਼

ਗ਼ਾ	ਗ਼ਾ	ਗ਼ਾ	ਗ਼ਾ
ਜ਼ਾ	ਜ਼ਾ	ਜ਼ਾ	ਜ਼ਾ
ਡ਼ਾ	ਡ਼ਾ	ਢ਼ਾ	ਫ਼ਾ
ਰ਼ਾ	ਲ਼ਾ	ਵ਼ਾ	ਸ਼ਾ

As per the experts the combinations marked in red are theoretical combinations and not used in Punjabi.

6.3.4 The Collation Order of Punjabi.

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. 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 Punjabi 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[16].

◌ਂ	◌ਁ	◌ਃ	ਅ	ਆ	ਇ	ਈ	ਉ	ਊ	ਏ
ਐ	ਓ	ਔ	ਕ	ਖ	ਖ਼	ਗ	ਗ਼	ਘ	ਙ
ਚ	ਛ	ਜ	ਜ਼	ਝ	ਞ	ਟ	ਠ	ਡ	ੜ
ਢ	ਣ	ਤ	ਥ	ਦ	ਧ	ਨ	ਪ	ਫ	ਫ਼
ਬ	ਭ	ਮ	ਯ	ਰ	ਲ	ਲ਼	ਵ	ਸ਼	ਸ
ਹ	◌਼	◌ਾ	◌ਿ	◌ੀ	◌ੁ	◌ੂ	◌ੇ	◌ੈ	◌ੋ
◌ੌ									

[16] The following are not included in the sort order: Addak Bindi,Ek Onkar, Udaat , Iri, Ura, Yakash. Visarga is included although it is rare.

7. REFERENCES

Bedi Tarlochan Singh, Gurmukhi Lipi da Janam te Vikas. Patiala: Punjabi University, 1999.

Diringer, David. 1996 (1947). The alphabet: a key to the history of mankind. New Delhi: Munshiram Manoharlal. ISBN 81-215-0780-0

Faulmann, Carl. 1990 (1880). Das Buch der Schrift. Frankfurt am Main: Eichborn. ISBN 3-8218-1720-8

Gebrauch mit besonderer Berücksichtigung des Buchgewerbes. Unter Mitwerkung von Fachgelehrten zusammengestellt in der Reichsdruckerei. Berlin: Reichsdruckerei.

Gurbaksh (G.B.) Singh. Gurmukhi Lipi da Janam te Vikas. Chandigarh: Punjab University, 1950.

Haarmann, Harald. 1990. Die Universalgeschichte der Schrift. Frankfurt: Campus. ISBN 3-593-34346-0

ISCII'91

Ishar Singh Tangh, Gurmukhi Lipi da Vigyamulak Adhiyan. Patiala: Jodh Singh Karamjit Singh.

Kala Singh Bedi, Lipi da Vikas. Patiala: Punjabi University, 1995.

Kartar Singh Dakha. Gurmukhi te Hindi da Takra. 1948.

Piara Singh Padam, Gurmukhi Lipi da Itihas. Patiala: Kalgidhar Kalam Foundation Kalam Mandir, 1953.

Reichsdruckerei. 1924. Alphabete und Schriftzeichen des Morgen- und Abendlandes, zum allgemeinen

Sohan Singh Galautra. Punjab dian Lipian.

Taylor, Isaac. 1883. The alphabet: an account of the origin and development of letters. Vol. 1: Semitic alphabets; Vol. 2: Aryan alphabets. London: Kegan Paul.

Unicode 5.2 and upwards. http://www.unicode.org

8. ANNEXURES

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

Annexure 2: Unicode Table of Gurmukhi[17]

Link: http://unicode.org/charts/PDF/U0A00.pdf

[17] 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